Structure-based design of anti-cancer drugs: The use of biophysical techniques for screening and characterization of novel inhibitors of the initiation factor eIF4E.

A thesis submitted for the degree of Doctor of Philosophy

by Christopher John Brown

This thesis is submitted in part fulfilment for the degree of Doctor of Philosophy at the University of Edinburgh. Unless otherwise stated the work is original and has not been previously submitted, in whole or in part, for any degree at this, or any other university.

Abstract

The eukaryotic translation initiation factor eIF4E is deregulated in many human cancers, and its over-expression in cells leads to malignant transformation. Oncogenic properties of eIF4E are directly linked to its ability to bind 7-methyl guanosine triphosphate (the cap structure) of the 5'mRNA, and initiate translation by recruitment of the eIF4F complex and ribosomal apparatus. Antisense mediated reduction of eIF4E in transformed cell lines reverted their transformed and tumourigenic phenotypes. Over-expression of eIF4E leads to enhanced translation of mRNAs with 5'UTR regions containing large amounts of secondary structure that encode proto-oncogenes, such as c-myc, cyclin D1, orthinine decarboxylase. The cap structure found at the 5' end of mRNA therefore presents itself as a target for small molecule inhibitor development targeting eIF4E.

In this work high resolution crystals were produced of full length human eIF4E complexed to m7GTP and the 4E-BP1 motif peptide. The interactions of eIF4E with the cap structure were also analysed by mass spectrometry, revealing a requirement of eIF4E for a guaninie monophosphates derivative with a positive delocalised charge on the m7G ring. The mass spectrometry technique, developed for the analysis of eIF4E's interaction with the cap-structure, was then applied to screening small molecule cap analogues. Potential inhibitors for screening were either isolated using virtual screening techniques or synthesised to produce a series of cap mimicking molecules containing a positive delocalised charge. Mass spectrometry identified a series of N7 substituted GMP derivatives that bound to eIF4E in the gas-phase. These compounds were then used in co-crystallisation trials with full length human eIF4E complexed with a 4E-BP1 motif peptide. The co-

crystal structures of eIF4E with N7 benzyl derivatives revealed a flipping of the tryptophan 102 to accommodate the bulky N7 group and the expulsion of two structured waters. It also showed that if a para-fluoro group is located on the benzyl modification, then interactions also occur with a structured water and an arginine, which explains its increased binding to eIF4E. The flipping of the tryptophan reveals the inherent flexibility in the cap-binding site. The structural information, revealing that the cap-binding site of eIF4E undergoes a conformational change in binding N7 derivatives of GMP with large bulky groups, provides us with valuable insight that can be used in future drug design efforts. The mass spectrometry assay coupled with a clear structure activity relationship, developed on the basis of various "cap-like" ligands studied in the work, gives an excellent starting point for the development of cap-analogue mimics for anti-cancer therapeutics.

Acknowledgments

During my PhD studies of the last several years I am indebted to Swann building, level 3 for their support and their unwavering good company. Special mentions, however, must be made to the following people for thoughtful discussion and their tolerance. A thank you to Professor Malcolm Walkinshaw for giving me the opportunity to pursue PhD studies and for pointing me in the right directions, Dr. David Millband for introducing me to the wonders of the wet lab, and Dr. Jacqui Dornan and Dr. Martin Wear for advice on numerous subjects. Other people who have in particular helped me with certain areas are Dr Iain McNae, who showed me the dark corners of crystallography, and Dr. Perdita Barran and Hannah Florance for introducing me to mass spectrometry and its application in the study of protein complexes. A thank you must also go to Dr Paul Taylor for his pertinent advice on all matters relating to LIDAEUS, computing and crystallography. Last but not forgotten are Dr C Mcinnes, Dr. G Kontipidis and Professor PM Fischer, from Cyclacel, who gave regular advice and suggestions on my work. The BBSRC and Cyclacel also kindly provided me with monetary support for my PhD studies. And finally and most importantly a thank you to my family for their undying support during my PhD and for all the hard work they have put in for the last 23 years before that.

Abbreviation list.

4E-BP1, 4E-binding protein 1,

4E-BP2, 4E-binding protein 2,

4E-BP3, 4E-binding protein 3,

7bzGMP, 7 benzyl guanosine monophosphate;

7etGMP, 7 ethyl guanosine monophosphate;

7*p*flbzGMP, 7 para-flouro-benzyl guanosine monophosphates;

ACD, available chemicals databse;

ADME, Absorption, Distribution, Metabolism and Excretion;

API, atmospheric pressure ionization;

BCL-2/xL, B cell leukemia 2/xl ;

CAD, collisional activated dissociation;

CBC, cap binding complex;

CBP20, cap binding protein 20;

CBP80, cap binding protein 80;

CCP4, Collaborative Computational Project Number 4;

CD, circular dichroisim;

CMP, cytidine 2' monophosphates;

CTP, cytidine triphosphate;

DCM, dichloromethane;

DEAD, Asp-Glu-Ala-Asp;

DMSO, dimethyl sulphoxide;

DTT, dithioerythritol;

eIF2, elongation factor 2;

eIF3, elongation factor 3;

eIF4, elongation factor 4 (group of);

eIF4A, elongation factor 4A;

eIF4B, elongation factor 4B;

eIF4E, elongation factor 4E;

eIF4F, elongation factor 4F;

eIF4G, elongation factor 4G;

eIF4H, elongation factor 4H;

ERK, extracellular signal-regulated kinase;

ESI, electrospray ionization;

ESRF, The European Synchrotron Radiation Facility;

FRAP/mTOR, FKBP12 and Rapamycin-AssociatedProtein/mammalian Target Of Rapamycin ;

HA, hydrogen bond acceptor;

HD, hydrogen bond donor;

HDM2, human double minute 2;

HNSCC, hean and neck squamous cell carcinomas;

HR, high resolution.

HTS, high throughput screening;

IGF2, insulin-like growth factor 2;

IPTG, Isopropyl β-D-1-thiogalactopyranoside;

IRES, internal ribosome entry site;

ITC, isothermal calorimetry;

m1Ade, 1 methyl adenine;

m1Cyt, 1 methyl cytosine;

m3Ade, 3 methyl adenine;

m3Cyt, 3 methyl cytosine;

m7G, 7 methyl guanosine;

m7GDP, 7 methyl guanosine diphosphate;

m7GMP, 7 methyl guanosine monophosphate;

m7Gppp, 7 methyl guanosine triphosphate;

m7GpppA, 7 methyl guanosine triphosphate adenine;

m7GpppG, 7 methyl guanosine triphosphate guanine;

m7GTP, 7 methyl guanosine triphosphate;

m7Gua, 7 methyl guanosine;
MALDI, Matrix assisted laser-desorption ionization;
MAP, mitogen-activated protein;
MBP, maltose binding protein;
MDM2, mouse double minute 2;
Met-tRNAi, methionine transfer RNA initiator;
MMP-9, matrix metalloprotease 9;
MS, mass spectrometry;
NCS, non crystallographic symmetry;
NMDA, N-methyl-D-aspartate receptor;
ODC, othinine decarboxylase ;
MAPK, mitogen-activated protein kinase;
PAB1P, poly adenine binding 1 protein;
PABP, poly adenine binding protein;
p-Cl-Bz, para-chloro-benzyl;
PDB, protein database;
PDGF, platelet-derived growth factor;
PEG, polyethylene glycol;
PTEN, phosphatase and tensin homologue;
RMS, root mean standard deviation, ;
RNP, ribonucleotide protein;
SAR, stucture activity relationshop;
SBP, spermine binding protein;
SEC, size exclusion chromatography;
SPR, surface plasmon resonance;
SRS, synchrotron radiation source.
TK, thymidine kinase;
TOF, time-of-flight;
UTR, un-translated region;
VEGF, vascular endothelial growth factor;

Figure list.

Chapter 1: eIF4E a potential target for the treatment of cancer.

Chapter 2: Expression, purification and crystallisation of eIF4E.

Chapter 3: Crystallographic and biophysical analysis of full length human eIF4E complexed with m7GTP and the 4E-BP1 motif peptide.

Chapter 4: Probing the interactions of eIF4E cap-analogue binding by mass spectrometry.

Chapter 5: Identifying initial drug leads for eIF4E and development of an overall structure-based drug design strategy.

Chapter 6: Crystallisation of eIF4E with N7 substituted cap-analogues and insights into structure based drug design.

Chapter 7: Materials and methods.

Contents List

Chapter 7: Materials and methods.
7.0 Materials and methods.

Chapter 1: eIF4E a potential therapeutic target for the treatment of cancer.

1.0 Initiation factor 4E (eIF4E).

Control of mRNA translation plays a critical role in cell growth, proliferation and differentiation. In eukaryotes, most mRNAs are translated in a cap-dependent manner. The cap structure m7GpppN (where N is any nucleotide) is found at the 5' terminus of all cellular eukaryotic mRNAs (except those in organelles)[1]. The 7-methylguanine is linked by a 5'-5' tri-phosphate bridge to the first transcribed residue (see figure 1). This group known as the 'cap', functions in splicing, polyadenylation, nuclear export, stability and recognition of the mRNA for translation[2]. A key player in the regulation of translation is the mRNA 5' cap-binding protein eIF4E, which is the rate-limiting member of the eIF4F complex[3]. This complex consists of eIF4E, an ATP-dependant helicase, and a large scaffolding protein, eIF4G, which acts as a docking site for other proteins[4]. eIF4E specifically interacts with the cap structure[1].

Cap-structure attached via a 5'-5' phosphate link

Rest of mRNA chain (5'-3')
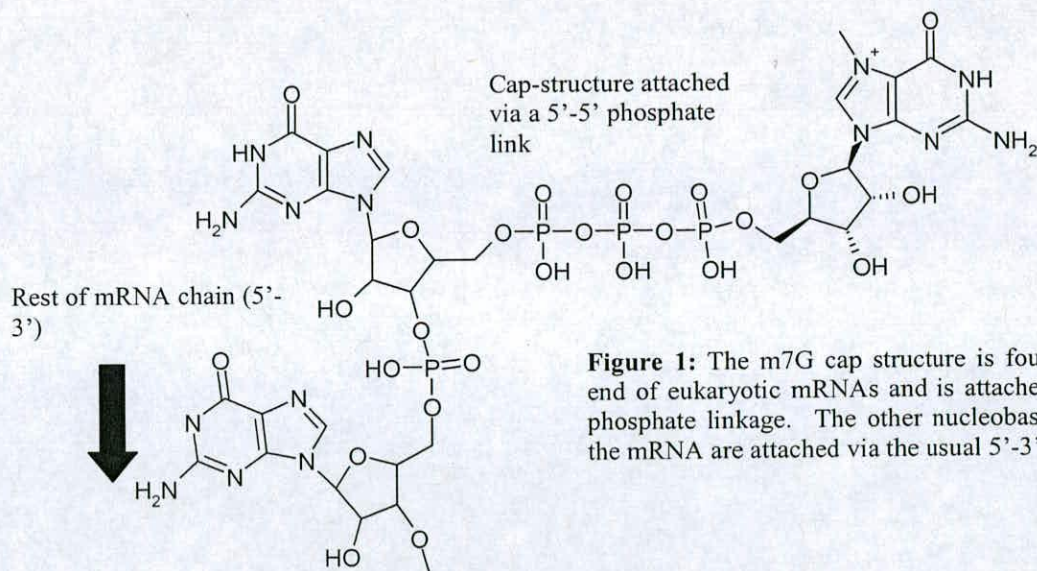
Figure 1: The m7G cap structure is found at the 5' end of eukaryotic mRNAs and is attached by a 5'-5' phosphate linkage. The other nucleobases that form the mRNA are attached via the usual 5'-3' linkage.

1

An alternative mechanism of translation initiation is cap- and eIF4E independent and requires an internal RNA structure termed an Internal Ribosome Entry Site (IRES) to which the 40S subunit directly binds. This mode was first identified in picornaviruses[5,6], but subsequent studies revealed the presence of IRES-dependent cellular translation in some mRNA transcripts during mitosis and apoptosis[1,7-9].
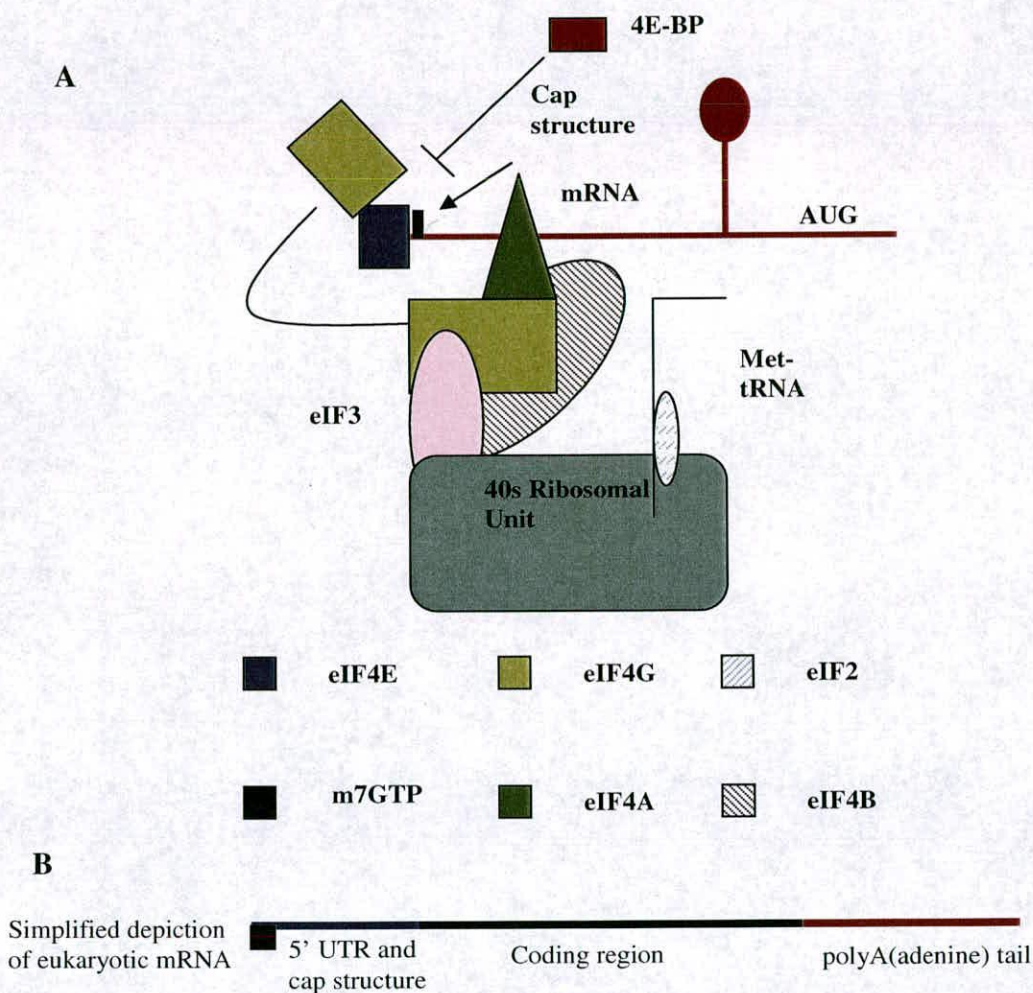
Figure 2: (A) Diagrammatic representation of protein interactions in the 48S preinitiation complex. Interactions between components of the initiation complex are illustrated. Notably, eIF4E is the only factor that specifically recognises the cap structure of mRNA, and so has a crucial role in recruitment of the 40S subunit via its interactions with eIF4G, and indirectly via eIF3, eIF4A and eIF4B. PABP (not shown) interacts with eIF4G and the polyA tail of eukaryotic mRNA (as shown in 2B) and causes the mRNA to circularise.

2

Protein synthesis is usually the most expensive process in the cell and translation rates are tightly regulated[10,11]. In mammals, regulation mainly operates at a level of initiation, rather than elongation or termination. The first step in the initiation process is the binding of the small ribosomal subunit (40s) to the mRNA. Since this is the rate-limiting step in initiation, it is often the target of regulation[11,12]. In prokaryotes, base pairing between rRNA and the 'Shine-Delgarno' sequence preceding the initiation codon of each open reading frame plays a dominant role in the ribosome-binding step[13,14]. In eukaryotes the binding step is much more complex. The model (as shown in figure 2) that best represents this binding step shows that the multi-subunit eukaryotic initiation factor 3 (eIF3), which is associated with the 40S ribosomal unit, also binds to the mRNA associated eIF4G protein and thereby links the ribosome to the mRNA[13,14].

The initiation process consists of three steps: (1) formation of the 43S complex, composed of a 40S ribosomal subunit and the initiation factors eIF-2, eIF-3, Met-tRNA and mRNA, (2) formation of the 48S complex containing mRNA, which is mediated by the eIF4 group of factors; and (3) joining of the 60S subunit to form the complete 80S complex[13,14]. In most circumstances, the second step is rate limiting and hence, a major point of regulation. This step also discriminates between the mRNAs that are to be recruited by the ribosome and remain in the untranslated pool of mRNAs[12]. This process is mediated by the eIF4 group of factors, for which eIF4E is rate limiting[3]. Evidence also showing that the role of eIF4E is rate limiting was obtained experimentally with the application of antisense RNA technology. Antisense RNA-

3

mediated reduction in eIF4E expression suppressed protein synthesis rates and the spectrum of expressed proteins in direct proportion to the level of eIF4E[15].

In most cases, eIF4G does not only bind directly to mRNA, it also associates with the cap structure at the 5' end and the poly(A) tail at the 3' end[16]. Initiation factor eIF4G binds the cap and the poly(A) tail indirectly through its interaction with the cap-binding protein eIF4E and the poly(A) binding protein Pab1p, respectively. The simultaneous interaction of these two proteins apart from placing eIF4G on the mRNA also leads to the circularisation of the mRNA[17].

Translation initiation and cell viability depends upon eIF4G binding to eIF4E at the 5' end of the mRNA. Initiation factor eIF4E binds the methylated guanosine cap of mRNA with sub-micromolar ability, and binds non-methylated precursors at least five-fold less efficiently[18]. Thus, the eIF4G-eIF4E complex binds to the 5' end of mRNA because eIF4E specifically recognises the cap. The association of eIF4G with eIF4E increases eIF4E affinity for the cap ten-fold[19] This allosteric regulation provides a selective advantage for the complex over apo eIF4E in binding to the mRNA. The eIF4E cap interaction in mammals is regulated by the phosphorylation of eIF4E (at Ser 209) via the eIF4G –associated Mnk1 kinase[20].

The phosphorylation state of eIF4E is, in general correlated with the translation rate and growth status of the cell[21]. Stimulation of cell growth activates the ras-ERK signaling pathway, which leads to the direct phosphorylation of eIF4E at Ser 209.

4

Phosphorylation at Ser 209 by the MAP kinases integration kinases (MNK1/2) is a controversial issue as to whether it increases the cap-affinity of eIF4E elevating translation rates or not[21]. It has been demonstrated that the mitogens-stimulated pathway acting through ERKs, and the stress-activated pathway acting through the p38 MAPKs, appear to converge on MNK1 (as shown in figure 3)[22].

Mammalian 4E-BP1, 4E-BP2 and 4E-BP3 and yeast p20 inhibit cap-dependent protein synthesis by binding to eIF4E without affecting cap recognition. Experimental work using point mutants showed that human eIF4G1 and the 4E-BP have shared but non-identical binding sites on human and yeast eIF4E[23-25]. eIF4E inhibition is caused by the blocking of eIF4G binding by the 4E-BPs. Sequence of the 4E-BPs and the eIF4Gs suggest that these two protein families have converged on the same eIF4E binding strategy that employs a *Y-X-X-X-X-L-ϕ* eIF4E recognition motif (where *X* is variable and ϕ is *L, M* or *F*) [26].

Under normal cellular conditions, eIF4E is bound by the inhibitory 4E-BPs, which sequester eIF4E from interaction with eIF4G, the scaffolding protein of the eIF4E translation initiation complex. Upon mitogenic stimulation, the 4E-BPs become phosphorylated at multiple sites as a consequence of both the ras-ERK and PI3 kinase/AKT signalling pathways[27] (see figure 3). These phosphorylation events dislodge the 4E-BPs from eIF4E, thereby freeing eIF4E and enabling eIF4E to deliver mRNAs effectively to the eIF4F complex. 4E-BP1 phosphorylation is also dependent

upon the FKBP-rapamycin associated protein/mammalian target of rapamycin (FRAP/mTOR) kinase (as shown in figure 3). Once bound to the scaffolding protein eIF4G, eIF4E may become phosphorylated at S209 by the kinase MNK. This phosphorylation event may play a key regulatory role for eIF4E binding however it is unclear whether this event stimulates or represses normal cellular translation[21].
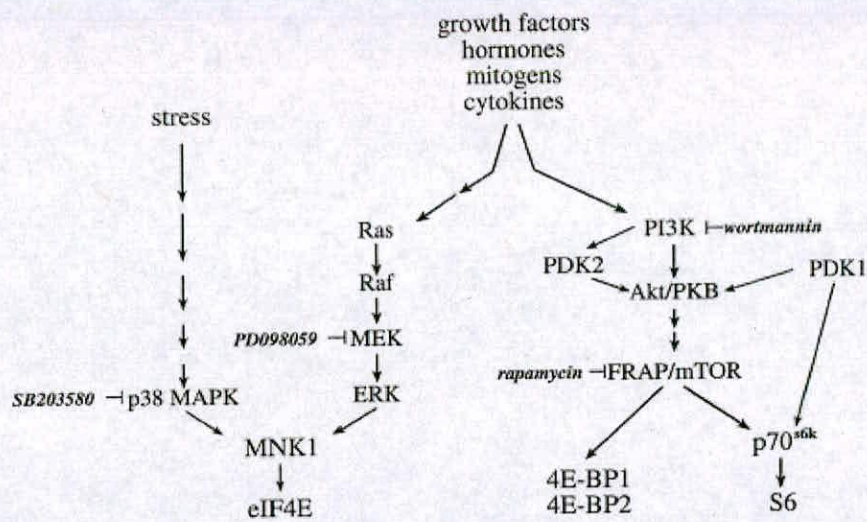
Figure 3: The intracellular signalling pathways impinging upon the translation initiation factors eIF4E and 4E-BPs. The Ras pathway, leading to eIF4E phosphorylation, and the PI3K pathway, leading to 4E-BP and p70S6K phosphorylation, are depicted. Also shown are the targets of several inhibitors used in studying these pathways.

1.1 Disregulation of protein synthesis is a step in malignant progression.

It is well established that defective processes of cellular homeostasis, like functions involved in DNA repair and genomic stability, result in a predisposition toward malignant transformation[28]. Disregulation of protein synthesis is also a major step in malignant progression. A decade ago, studies aimed at understanding the regulation and

function of eIF4E revealed that a moderate over-expression of this translation initiation factor results in dramatic phenotype changes, including rapid proliferation, suppression of apoptosis and malignant transformation[29]. Conversely, the tumorigenic properties of cancer cells can be strongly inhibited by anti-sense RNA-mediated reduction of eIF4E expression or by over expression of the inhibitory eIF4E binding proteins, the 4E-BPs[30]. eIF4E expression was subsequently found to be elevated in carcinomas of the breast, head and neck, bladder, cervical, lung and prostate, when compared with normal tissues and benign lesions[29]. These data firmly implicate elevated eIF4E expression in malignant progression.

Establishing a greater protein synthesis output may be critical for cancer cells to sustain rapid rates of cell division. However, analysis of cells transformed by eIF4E revealed that the synthesis of only select protein were greatly enhanced, while synthesis of most were only minimally increased[31]. These studies indicate that eIF4E specifically stimulates translation of several oncogenic transcripts that, under normal cellular conditions, are translationally repressed. The concerted upregulation of these key malignancy related proteins (e.g. vascular endothelial growth factor (VEGF), cyclin D1, matrix metalloprotease 9 (MMP-9)) may enable tumour cells to override normal growth constraints, stimulate angiogenesis, invade surrounding tissue and establish the autocrine stimulatory growth and survival signals necessary for the successful establishment of mesatases[29]

1.1.1 Protein synthesis and mRNA discrimination.

Protein synthesis plays a fundamental role in nearly every aspect of metabolism. It also constitutes a critical step in the control of gene expression[10] as mRNAs must be recruited for translation. The synthesis of each protein ultimately depends on the relative abundance of its mRNA and its capacity of that mRNA to interact with components of the translation initiation machinery. This property of the translation initiation process establishes a hierarchy among the different mRNAs to be translated because the initiation capacity of protein synthesis is limiting. In eukaryotes, the flow of information from genes to proteins is too slow to accommodate for rapid changes in the environment. Eukaryotes compensate for this problem by maintaining a pool of mRNAs that are not immediately utilised. Examples could include transcripts that encode growth factors that can be rapidly produced under conditions that require the cells to re-enter cell division rapidly[12].

In metozoans, mRNAs vary over a 100-fold range in their translational efficiencies[32]. Also their transcription is dependent on the particular growth conditions of the cell. A theoretical treatment of mRNA competition[10,29,33] hypothesise that the spectrum of translated mRNAs changes with the overall rate of protein synthesis. Weak mRNAs are out competed by strong mRNAs when the rate of translation initiation is reduced, as in quiescent cells[34], whereas weak and strong mRNAs are translated when protein synthesis becomes disregulated. A disproportionate number of mRNAs that would be characterised as weak are those encoding oncoproteins and growth and survival factors, these have 5' UTRs that contain secondary structure elements. Several of these genes

also share the property of being cell-cycle regulated and, in turn, their protein products affect cell-cycle progression[29,34].

1.1.2 Paradigm of strong and weak mRNAs.

By specifically binding the 5'-terminal cap containing end of mRNA, eIF4E, following its release from 4E-BPs, can deliver the mRNA to the eIF4F complex. The eIF4F complex then scans 5'-3' unwinding secondary structure in the 5' untranslated region (5'-UTR) of the mRNA to reveal the initiation codon, trigger ribosome binding and facilitate translation of the mRNA[11]. The abundance of the active eIF4F complex is normally limiting due to the limited expression of eIF4E as well as its binding to the inhibitory 4E-BPs. Therefore, cellular mRNAs must compete with each other for access to the eIF4F complex and for subsequent delivery to the ribosomes. The short, unstructured 5'-UTRs of most cellular mRNAs (e.g. house keeping genes) enable the eIF4E complex to scan readily for the translation initiation codon. These mRNAs are translated efficiently, even when the active eIF4F complex is limiting, and are termed strong mRNAs (see figure 6). By contrast, the lengthy, G + C rich, highly structured 5'-UTRs, typical of growth factor and proto-oncogene mRNAs (termed weak mRNAs) hinder efficient scanning and start codon recognition (see figure 6). This leads to the mRNAs being inefficiently translated, which is further attenuated when the active eIF4F complex is limiting[1,10,13,29,35-37].

Translation of strong mRNAs (see figure 4) will quickly reach a maximum in the presence of low levels of free eIF4E (i.e. active eIF4F complex), even in quiescent cells.

9

Weak mRNAs are poorly translated when free eIF4E is limiting. With increased free eIF4E (i.e. increased levels of active eIF4F), the translation of these weak mRNAs (e.g. growth regulators) is disproportionately enhanced. The levels of free eIF4E can be increased substantially by one or all of the following mechanisms: (1) increased eIF4E expression; (2) decreased 4E-BP expression; or (3) increased phosphorylation of 4E-BP resulting from hyperactivation of intracellular signalling pathways. Cancer cells routinely show activation of the signalling pathways that phosphorylate 4E-BPs and may also reduce expression of the 4E-BPs[38]. Cancer cells also frequently display increased eIF4E expression as well[29].
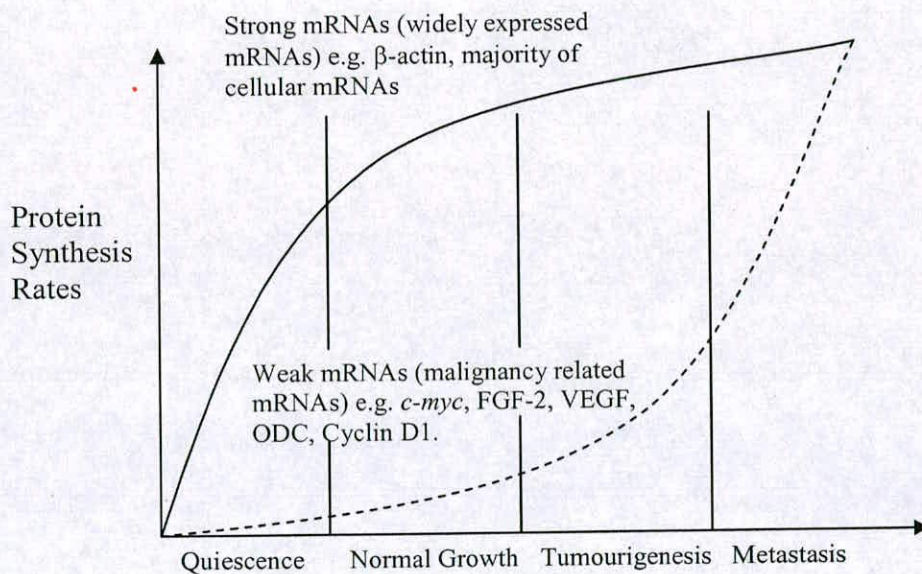
Figure 4: mRNA competition for translation initiation[10,29,33,39]. The products of weak mRNAs become selectively increased with increased eIF4E concentration, which results in higher amounts of eIF4F complex. Disproportionate increased translation of these weak mRNAs would increase with malignant progression in relation to increased free eIF4E. Figure derived from De Benedetti and Graff[12].

Consistent high levels of free eIF4E would selectively and disproportionately enhance the translation of weak mRNAs. As these weak mRNAs almost universally encode growth regulatory proteins, the concerted upregulation of these gene products may profoundly influence the biology of these cells, enabling tumourigenesis and ultimately metastatic progression. There is strong experimental evidence that now supports the model that increasing levels of free eIF4E preferentially enhances the translation of the weak mRNAs driving transformation and malignancy[29].

1.2 Role of eIF4E in malignancy.

Over-expression of eIF4E resulted in cellular transformation and tumourigenesis in the immortalised but nontransformed cell lines, NIH3T3, CREF or MM3MG[39-41]. eIF4E also acts as a potent enhancer of transformation in cooperation with v-myc or E1A[42]. It was then discovered that eIF4E expression is elevated in many malignant cell lines[43,44]. In CREF cells, eIF4E over-expression not only induced transformation and tumour formation but also induced metastatic capacity. Further evidence of a role for eIF4E in malignancy has been provided by studies wherein eIF4E expression was suppressed by antisense RNA expression or eIF4E function was blocked by the expression of 4E-BPs. Ectopic expression of 4E-BP1 reverted the transformed and tumourigenic phenotype of cells transformed by eIF4E or v-src[45]. Expression of antisense RNA to eIF4E in HeLa cells suppressed proliferation and altered cellular morphology[15]. Antisense RNA mediated reduction of eIF4E by ~ 60% in ras-transformed CREF cells suppressed soft-agar colonisation as well as tumour formation and growth[30]. Also the ability of these cells to invade surrounding normal tissues and to metastasise was profoundly

suppressed[46]. Similarly, in human breast and head and neck cancer cell lines, antisense-RNA-mediated reduction of eIF4E suppressed tumour formation and angiogenesis[47-49].

Translation of model mRNAs engineered to have excessive secondary structure in their 5'-UTRs (i.e. model weak mRNAs, see figure 6) were specifically stimulated in cells over-expressing eIF4E[31,50] Further work has shown that eIF4E controls the translation of endogenous mRNAs whose protein products are well recognised to contribute to malignancy (see figure 5). These mRNAs encode proto-oncoproteins such as c-myc, cyclin D1 and ornithine decarboxylase (ODC), angiogenesis factors such as FGF-2 and VEGF and degradative enzymes that enable tumour invasiveness such as MMP-9[12].
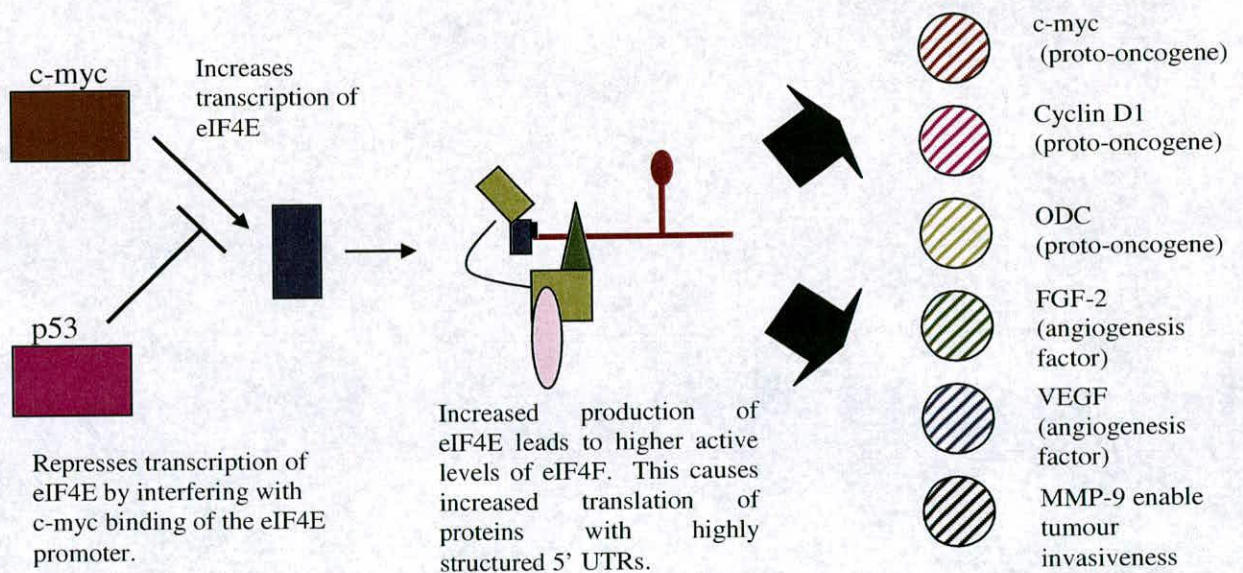
Figure 5: Transciption of eIF4E is controlled by c-myc and leads to increased levels of active eIF4E complex. This causes increased translation of mRNAs with highly structured 5' UTRs, whose protein products are well recognised to contribute to malignancy. These mRNAs encode proto-oncoproteins such as c-myc, cyclin D1 and ornithine decarboxylase (ODC), angiogenesis factors such as FGF-2 and VEGF and degradative enzymes that enable tumour invasiveness such as MMP-9[12]. p53 has recently been shown to transcriptionally repress eIF4E by binding to c-myc and inhibiting its interaction with the eIF4E promoter[51].

1.2.1 *c-myc* and p53.

In contrast to eIF4E, the p53 tumour suppressor plays a pivotal role in preventing tumour occurrence and development. Although p53 may exert its tumour suppressing action through multiple mechanisms, primarily it functions as a transcription factor activating or inactivating the transcription of its target genes[52]. Genes activated by p53 include those that mediate cell cycle arrest, DNA repair and apoptosis, whereas many pro-proliferative and anti-apoptotic genes are suppressed by p53.

The eIF4E promoter contains two canonical CACGTG motifs known to be the c-*myc*-binding site[53,54]. One problematic feature of c-*myc* is its relatively weak activity in reporter gene assays[55]. Importantly, myc could transactivate the eIF4E promoter in standard reporter assays but, more importantly, dominant negative forms of c-*myc* down-regulated eIF4E promoter activity[53]. These formal connections between myc's transactivating activity and a gene involved in protein synthesis revealed an important connection between c-*myc* and growth control. A connection between *myc* and translation initiation makes particular sense because translation initiation elements have particular effects on translation of cell cycle regulators[56,57], and how cell growth might regulate cell division.

Zhu et al[51] showed that p53 negatively regulates eIF4E which is positively regulated by c-myc. p53 interacts directly with c-myc and this interaction significantly inhibits the

binding of c-myc to the eIF4E promoter and c-myc induced eIF4E promoter activity. eIF4E controls the translation of various malignancy-associated mRNAs which are involved in pathological process, the co-regulation of eIF4E by p53 and c-myc may represent one of the mechanisms for p53 and c-myc mediated cancer formation and progression. Zhu et al[51] also showed that the inhibitory effect of p53 on eIF4E promoter activity was abrogated by the p53 inhibitory oncogene MDM2[58].

The translation of the c-myc mRNA is normally repressed in B cells unless stimulated by mitogens or growth factors[59]. Deletion mutations in the c-myc 5'-UTR, which are present in several Burkitt lymphomas (resulting from translocation of the c-myc gene), result in enhanced translation of the c-myc mRNA[60-64]. C-myc is implicated in several biological processes such as cell growth, proliferation and apoptosis[65,66]. C-myc is expressed ubiquitously during embryogenesis and in post-developmental tissues with a high proliferative capacity. C-myc also inhibits terminal differentiation of most cell types and sensitises cells to growth factor withdrawal induced apoptosis[67,68]. Activated oncogenic c-myc plays a critical role in the progression of Burkitt's Lymphona, its expression is elevated or deregulated in a wide range of other human cancers and is associated with aggressive tumours with poor prognosis[65,66].

14

1.2.2 Ornithine decarboxylase – ODC.

ODC mRNA contains a GC-rich 5'UTR (see figure 6), rendering it poorly translated[69,70]. Over-expression of eIF4E in cells leads to a 30-fold increase in ODC protein levels, as well, depletion of eIF4E using anti-sense RNA suppressed ODC mRNA translation in the eIF4E over-expressing cells[71,72]. Polyamines are ubiquitous cellular components that are involved in normal and neoplastic growth. Polyamine biosynthesis is tightly regulated in mammalian cells by the activities of two key decarboxylases, ornithine- and S-adenosylmethionine decarboxylase. Polyamines are essential for growth, differentiation, survival and mediate cellular transformation[73,74]. Over-expression of ODC transforms NIH 3T3 cells[75,76]. Transgenic expression of polyamine biosynthetic enzymes, such as ODC and S-adenosylmethionine decarboxylase in mice, causes skin cancer[77]. Also, both intracellular polyamine concentrations and ODC activity are increased in colorectal cancer tissue and in premalignant polyps[78]. The polyamine pathway is also implicated in carcinogenesis and tumour progression of breast cancer[79]. Thus, the polyamine synthetic pathway is considered to be an anti-cancer target[80,81]. Since eIF4E directly regulates ODC mRNA translation, targeting both eIF4E and ODC may serve as an important target for anti-cancer therapy.

1.2.3 Cyclin D1.

eIF4E expression leads to an increase in cyclin D1 protein[82,83]. eIF4E also enhances the transport of the cyclin D1 mRNA from the nucleus[83-87]. The progression of mammalian cells through G1 phase of the cell cycle is governed by the D-type cyclins (D1, D2, D3). These proteins are induced at the beginning of the G1 phase and associate with

serine/threonine cyclin dependent kinases to form activated holoenzymes. Cyclin D1 expression is increased in several human cancers (colorectal and mammary adenocarcinomas and pancreatic tumours)[88-91], as well as in several cancer cell lines[92]. Both cyclin D1 and polyamines are required for entry into S phase and their increased expression has been linked to transformation[93].

1.2.4 VEGF and FGF.

eIF4E over-expression results in a dramatic increase in the secretion of vascular endothelial growth factor (VEGF) without affecting mRNA levels[94]. VEGF mRNA in eIF4E transfected cells is associated with heavy polysomes indicating that increased VEGF expression is achieved through its translational upregulation[94]. Similarly, fibroblast growth factor-2 (FGF-2) mRNA is also loaded onto heavy polysomes in cells expressing eIF4E, leading to increased secretion of FGF-2[50]. The formation of new blood vessels from pre-existing ones is a well-controlled process[95,96]. Currently, several VEGF inhibitors which inhibit VEGF and VEGF receptor production or VEGF-VEGFR interactions are undergoing clinical testing in several malignancies[97].

Increased eIF4E expression in cells regulates the translation of several other proteins involved in autocrine growth stimulation (PDGF, IGF2), cell survival (BCL-2, BCL-xL), invasion (MMP-9) and communication with the extracellular environment (NMDA)[29,93]. Thus, the eIF4E mediated translational modulation of these malignancy associated mRNAs plays a pivotal role in both tumour formation and metastasis.

16

| β-actin | ODC | VEGF |
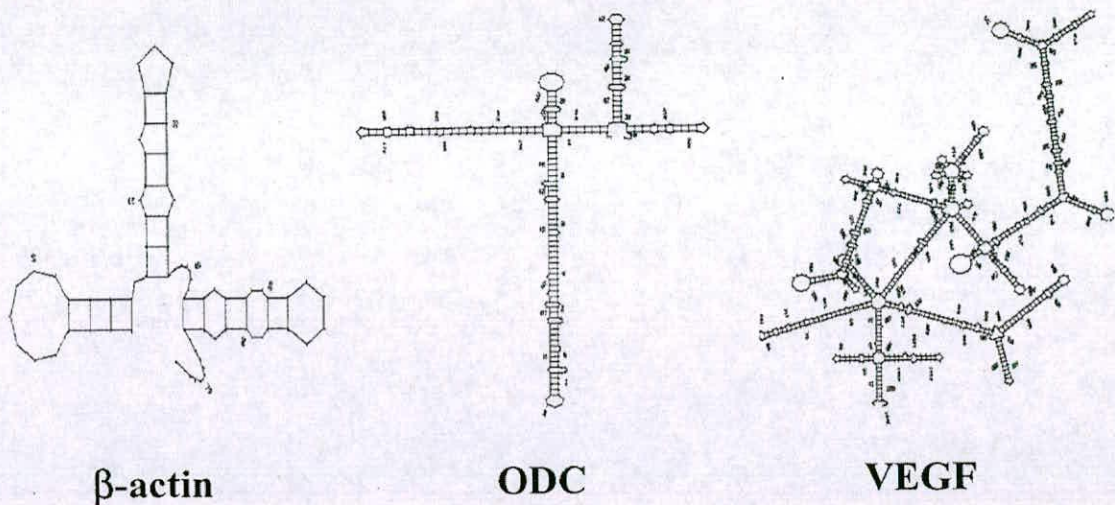
Figure 6: 5' UTR structures of competitive and non-competitive mRNAs[37]. The 5'-untranslated regions from a strong mRNA (*β*-actin) and two weak mRNAs (Vascular endothelial growth factor, VEGF, and ornithine decarboxylase, ODC) are shown. These were generated using the Zucker m-fold program[37]. Note that the non-competitive mRNAs have highly structured, lengthy 5' UTRs, which encumber efficient scanning and recognition of the translation start codon. These structures must be resolved by the helicase activity of the eIF-4F complex to enable translation.

1.3 Cap-independent translation.

Translation of FGF-2, VEGF and c-myc mRNAs that are found to be enhanced in eIF4E over-expressing cells can also be driven by an internal ribosome entry site, or IRES[9,98,99]. Their translation may occur even in the presence of limiting eIF4E (i.e. cap independent) under certain conditions. For example, translation of c-myc mRNA can occur at the M-phase of the cell cycle where cap-dependent translation is inhibited[100]. Another study has shown that the mechanism of c-myc translation is more consistent with shunting[63] a form of initiation that requires cap-proximal entry of ribosomes and eIF4E. It is currently unclear what contribution IRES-dependent translation has or may have on the increased expression of these potent growth regulatory molecules in cancer cells.

1.4 Elevated expression of eIF4E is common in human carcinomas.

High levels of eIF4E expression are found in several cancers; breast cancers, head and neck cancers, bladder cancers, prostatic adenocarcinomas, lung cancers, cervival cancers and lymphomas.

Breast cancer: Elevated eIF4E expression was initially researched in breast cancer. The basal level of eIF4E in normal breast tissue is low, making the abnormally elevated eIF4E expression in carcinomas very evident. eIF4E is elevated three- to 30 fold above normal breast and benign lesions and is a ubiquitous feature of breast carcinoma's[101,102]. eIF4E levels are also a strong, independent prognostic indicator of recurrence and death from breast cancer [103,104].

Head and Neck cancers: Head and neck squamous cell carcinomas (HNSCC) have a high local occurrence rate due in part to incomplete tumour resection. The high lethality of this type of cancer is not due to metastasis, but due to local occurrence in most cases. Complete resection of the tumour is critically important for survival. Incomplete excision will almost invariably lead to rapid reoccurrence and death from tumour burden. eIF4E expression is low in benign lesions, such as polyps and leukoplakia, but is clearly elevated in HNSCC. The presence of as little as 5% of eIF4E positive cells present at the surgical margins was sufficient to predict local reoccurrence within two thirds of the patients with the reoccurrence occurring in two years[48,105].

Lung cancer: eIF4E expression in bronchial adenocarcinomas (but not squamous cell carcinomas) is three to eight-fold higher than in the normal lung[106]. In an independent study of adenocarcinomas of the peripheral lung, eIF4E was found to be elelvated by up to seven fold relative to normal lung tissue and was specifically associated with histological grade and invasiveness of the tumour[107]. These studies indicate enhanced eIF4E expression in the development and progression of lung cancers.

1.5 Role of eIF4E in metastatic progression.

The metastatic phenotype is extremely complex requiring the metastasising tumour cells to hurdle numerous divergent challenges. To establish a successful metastatic colony, a tumour cell must induce angiogenesis, which enables continuous tumour growth and provides access to the vasculature. The tumour cell must also dislodge from the primary tumour site, break down and penetrate surrounding normal tissue, enter the blood or the lymphatic system, survive with in that circulation and exit at a site distal to the primary tumour site. Challenged by a different growth factor milieu in the tissue of the metastatic site, a metastasising tumour cell must be able to attract a new blood supply and to generate the necessary growth and survival factors to support and establish a successful metastatic colony. This shift towards growth factor autonomy is essential for the successful establishment of metastases. Indeed, although many tumour cells are able to break from the primary tumour, enter the circulation and colonise at a distal site, only a very few will be able to establish a metastatic colony successfully[108].

Enhanced eIF4E function can stimulate autocrine growth factor signalling. eIF4E expression stimulated expression of activin, a member of the tumour growth factor β family, which in turn stimulated mesodermal development in Xenopus oocytes[109]. Over-expression of eIF4E in NIH3T3 cells induced transformation, in part, by establishing an autocrine growth stimulatory loop that triggered signalling through the ras-MAP kinase pathway[110]. In CREF cells over-expression of eIF4E triggered signalling through the ERK pathway[37]. These reports indicate that enhanced eIF4E function can play a role in the acquisition of growth factor autonomy – a critical phenotype necessary for the successful establishment of metastases[29]. Enhanced eIFE function has also been linked to metastasis in both experimental tumour systems and in human disease.

Reduction of eIF4E by antisense RNA has been shown to suppress the ability of ras-transformed CREF cells to metastasise to the lungs in an experimental metastasis assay. The few metastases that did form regained normal levels of eIF4E expression, indicating that the successful establishment of metastases selected for cells with greater levels of eIF4E expression[46]. CREF cells transformed by the ectopic over-expression of eIF4E were capable of forming both spontaneous and experimental metastases[37]. The metastatic colonies formed routinely showed even higher eIF4E expression levels than that of the parental, transfected cells[37]. Tumour progression to metastases involves selection for increased eIF4E expression.

Inhibiting eIF4E function by blocking the signalling cascades that release eIF4E from 4E-BPs also profoundly suppresses metastasis. The macrolide antibiotic rapamycin, which suppresses mTOR activity and thereby prevents the the release of 4E-BPs from eIF4E (figure 1), suppresses the capacity of murine colon tumour cells to grow as experimental liver metastases[111]. Rapamycin treatment also prevents pulmonary metastasis of human renal cell carcinoma cells[112].

Transcriptional profiling of metastatic human solid tumours revealed a molecular signature for metastasis. Expression of nine genes were routinely decreased across a diverse range of solid tumour metastases, whereas eight genes were found to be consistently upregulated. eIF4E was identified as one of this set of genes. Enhanced expression of eIF4E serves as part of a molecular signature for human solid tumour metastases[113]. These reports support the role that eIF4E may play a role in metastatic progression and further supports eIF4E as a target for anti-cancer therapy.

1.6 Exploiting eIF4E as therapeutic target.

Metastatic progression is driven by the concerted actions of many diverse protein-protein interactions that govern cell cycle progression and growth factor autonomy (ODC, c-myc, cyclin D1), angiogenesis (VEGF, FGF-2), survival BCL-2 and invasion (MMP-9, CD44v6). Though expression of these key genes is specifically regulated by distinct stimuli, translation of every one of these proteins is suppressed under cellular conditions wherein the activity of the eIF4F translation initiation complex is low. The activity of this complex is regulated by the availability of eIF4E. Both human and

experimental tumours commonly show a marked increase in the ratio of eIF4E:4E-BP; a consequence of eIF4E overexpression, diminished 4E-BP expression and/or enhanced intracellular signalling through the PI3 kinase and ras-ERK pathways. Such enhanced eIF4E availability facilitates formation of the eIF4F complex and selectively upregulates the translation of malignancy related mRNAs. Inhibiting the function of eIF4E should diminish the assembly and function of the eIF4F complex and selectively diminish the translation of these malignancy related mRNAs[12,29].

Currently several approaches are being used to exploit eIF4E as a therapeutic target including antisense RNA[114], standard small molecule therapy[111,112] and, excitingly gene therapy[29]. Antisense RNA-mediated reduction of eIF4E in both epithelial and fibroblast models reduced translation of key malignancy related molecules (ODC, VEGF and FGF-2) in concert with diminished tumour growth, invasion and metastasis. Whether anti-sense RNA strategies prove useful for therapy remains to be fully explored. However a number of anti-sense oligonucleotides have now entered clinical trials for cancer[12,29].

Rapamycin and its derivatives inhibit the kinase mTOR and thus prevent phosphorylation of 4E-BP and the subsequent release of 4E-BP from eIF4E. Consequently, rapamycin treatment prohibits eIF4E from engaging and activating the eIF4F complex. Treatment of murine colon carcinoma cells or human renal cell carcinoma cells with rapamycin suppresses metastasis[111,112]. Rapamycin treatment may be particularly effective in tumours lacking PTEN function, such as glioblastomas and

prostate cancers[115,116]. Early phase clinical trials with the rapamycin derivative, CCI-779, have indicated that this approach may be a promising therapy, even as a single agent, for cancer treatment[117,118].

The increased function of eIF4E in tumour cells has provided a potential avenue for a gene therapy approach. Metastatic murine breast cancer cells were injected into the tail vein of mice to facilitate the formation of pulmonary metastases. Mice were subsequently injected with plasmids expressing thymidine kinase (TK) to facilitate the cytotoxic conversion of ganciclovir. Mice injected with plasmids encoding TK linked to the translationally repressive 5'-UTR showed a 90% reduction in pulmonary metastases. Importantly, these mice showed no evidence for systemic toxicity as the translationally suppressed TK was exclusively expressed only in the tumour tissues, where eIF4E activity was highest. In contrast, the mice injected with the uncontrolled TK plasmid showed significant toxicity attributable to the expression of the TK protein in all tissues and, subsequently, the systemic cytotoxic conversion of ganciclovir[119].

1.7 Structure based drug design rationale.

There are three main ways of interfering with eIF4E action utilising a structure based drug design rationale (see figure 7): (1) to develop cap-analogs with high affinity and specificity for eIF4E, (2) to develop peptide mimics to interfere with the eIF4G and eIF4E interaction and (3) to develop kinase inhibitors to interfere with specific points within the signal transduction pathway responsible for regulation of eIF4E and 4E-BP activity. Cap analogues are a promising lead, with the major advantage of structures

23

being available[120-123] and the mechanism of binding being reasonably understood[18,124] making this an ideal candidate for virtual screening (discussed in chapter 3 and 5). Signal transduction pathways are not a prime candidate as the best defined kinase in this class has already been identified as the target of rapamycin. Rapamycin is a known anti-proliferative, having been developed as an immunosuppressant. It has also been tested and developed for uses in oncology[111,112]. There are several other kinases in the pathway that are potential targets, but developing understanding of them suggests they would be difficult to use to target eIF4E activation. The number of proteins that interact with eIF4E make it a promising target for the utility of peptides as drug development templates (chapter 3).

To develop cap analogues

To develop peptide mimetics.

To develop kinase inhibitors to interfere with the signal transduction pathway regulating eIF4E and 4E-BP.
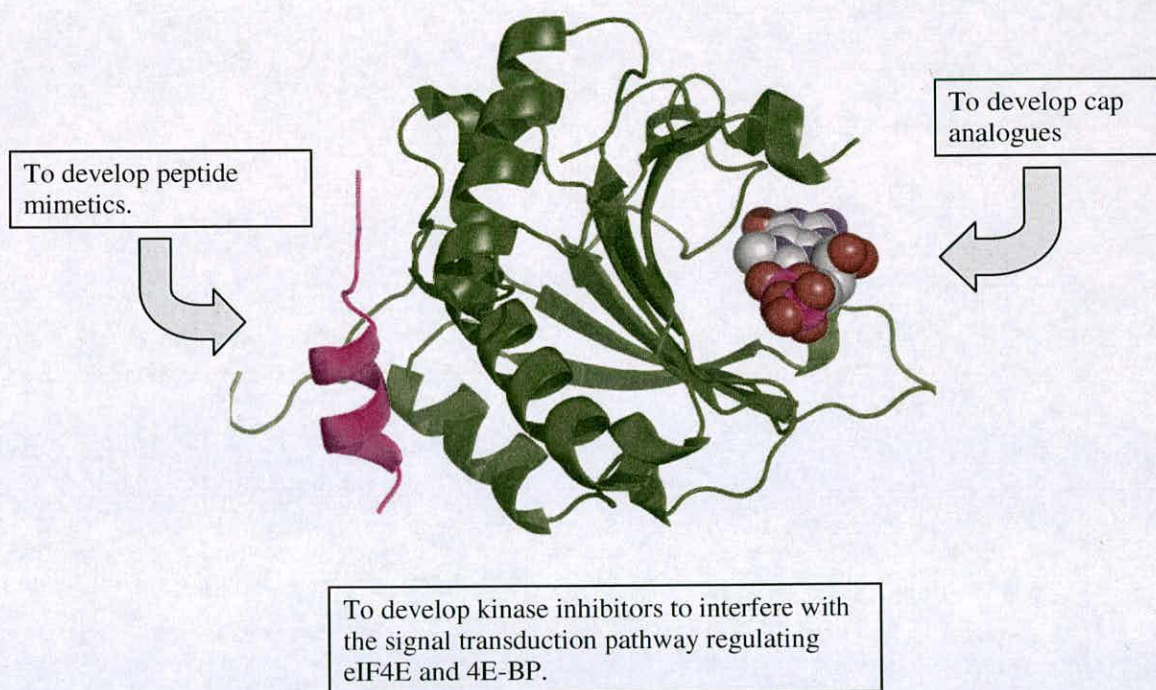
Figure 7: eIF4E can be targeted with 3 distinct structure based drug design strategies: (1) to develop cap-analogs with high affinity and specificity for eIF4G, (2) to develop peptide mimics to interfere with the eIF4G and eIF4E interaction and (3) to develop kinase inhibitors to interfere with specific points within the signal transduction pathway responsible for regulation of eIF4E and 4E-BP activity.

It is the aim of this work to develop cap-mimetics for the mRNA cap structure. To achieve this high resolution crystals of eIF4E are needed for soaking and crystallisation experiments and a primary screen to identify potential hits must be developed. To facilitate this a source of purified protein is also required. As well as these main goals the interactions of the cap-analogue must also be studied to identify the main determinants required for binding as these are of great importance to structure based drug design efforts.

Reference List

1. Gingras,A.C., Raught,B. & Sonenberg,N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913-963 (1999).

2. Lewis,J.D. & Izaurralde,E. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247**, 461-469 (1997).

3. Hiremath,L.S., Webb,N.R. & Rhoads,R.E. Immunological detection of the messenger RNA cap-binding protein. *J. Biol. Chem.* **260**, 7843-7849 (1985).

4. Sonenberg,N. & Dever,T.E. Eukaryotic translation initiation factors and regulators. *Curr. Opin. Struct. Biol. 2003. Feb. ;13(1):56. -63.* **13**, 56-63 (2003).

5. Jang,S.K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**, 2636-2643 (1988).

6. Pelletier,J. & Sonenberg,N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320-325 (1988).

7. Holcik,M., Sonenberg,N. & Korneluk,R.G. Internal ribosome initiation of translation and the control of cell death. *Trends Genet. 2000. Oct. ;16. (10):469. -73.* **16**, 469-473 (2000).

8. Pyronnet,S. & Sonenberg,N. Cell-cycle-dependent translational control. *Curr. Opin. Genet. Dev. 2001. Feb. ;11(1):13-8.* **11**, 13-18 (2001).

9. Vagner,S. *et al.* Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. *Mol. Cell Biol.* **15**, 35-44 (1995).

10. Rhoads,R.E., Joshi-Barve,S. & Rinker-Schaeffer,C. Mechanism of action and regulation of protein synthesis initiation factor 4E: effects on mRNA discrimination, cellular growth rate, and oncogenesis. *Prog. Nucleic Acid Res. Mol. Biol.* **46**, 183-219 (1993).

11. Rhoads,R.E. Cap recognition and the entry of mRNA into the protein synthesis initiation cycle. *Trends Biochem. Sci.* **13**, 52-56 (1988).

12. De Benedetti,A. & Graff,J.R. eIF-4E expression and its role in malignancies and metastases. *Oncogene 2004. Apr* **23**, 3189-3199 (2004).

13. Kozak,M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187-208 (1999).

14. Kapp,L.D. & Lorsch,J.R. The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem. 2004. ;73. :657. -704.* **73**, 657-704 (2004).

15. De Benedetti,A., Joshi-Barve,S., Rinker-Schaeffer,C. & Rhoads,R.E. Expression of antisense RNA against initiation factor eIF-4E mRNA in HeLa cells results in lengthened cell division times, diminished translation rates, and reduced levels of both eIF-4E and the p220 component of eIF-4F. *Mol. Cell Biol.* **11**, 5435-5445 (1991).

16. Borman,A.M., Michel,Y.M. & Kean,K.M. Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. *Nucleic Acids Res. 2000. Nov. 1;28. (21):4068. -75.* **28**, 4068-4075 (2000).

17. Wells,S.E., Hillner,P.E., Vale,R.D. & Sachs,A.B. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* **2**, 135-140 (1998).

18. Niedzwiecka,A. *et al.* Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol. 2002. Jun. 7. ;319. (3):615. -35.* **319**, 615-635 (2002).

19. Haghighat,A. & Sonenberg,N. eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. *J. Biol. Chem.* **272**, 21677-21680 (1997).

20. Pyronnet,S. *et al.* Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *EMBO J.* **18**, 270-279 (1999).

21. Scheper,G.C. & Proud,C.G. Does phosphorylation of the cap-binding protein eIF4E play a role in translation initiation? *Eur. J. Biochem. 2002. Nov. ;269. (22.):5350. -9.* **269**, 5350-5359 (2002).

22. Pyronnet,S. Phosphorylation of the cap-binding protein eIF4E by the MAPK-activated protein kinase Mnk1. *Biochem. Pharmacol. 2000. Oct. 15. ;60. (8.):1237. -43.* **60**, 1237-1243 (2000).

23. Haghighat,A., Mader,S., Pause,A. & Sonenberg,N. Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. *EMBO J.* **14**, 5701-5709 (1995).

24. Altmann,M., Schmitz,N., Berset,C. & Trachsel,H. A novel inhibitor of cap-dependent translation initiation in yeast: p20 competes with eIF4G for binding to eIF4E. *EMBO J.* **16**, 1114-1121 (1997).

25. Ptushkina,M. *et al.* Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5' cap in yeast involves a site partially shared by p20. *EMBO J.* **17**, 4798-4808 (1998).

26. Mader,S., Lee,H., Pause,A. & Sonenberg,N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol. Cell Biol.* **15**, 4990-4997 (1995).

27. Gingras,A.C. *et al.* Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes Dev. 2001. Nov. 1;15. (21):2852. -64.* **15**, 2852-2864 (2001).

28. Hanahan,D. & Weinberg,R.A. The hallmarks of cancer. *Cell 2000. Jan. 7. ;100. (1):57. -70.* **100**, 57-70 (2000).

29. Graff,J.R. & Zimmer,S.G. Translational control and metastatic progression: enhanced activity of the mRNA cap-binding protein eIF-4E selectively enhances translation of metastasis-related mRNAs. *Clin. Exp. Metastasis 2003. ;20. (3):265. -73.* **20**, 265-273 (2003).

30. Rinker-Schaeffer,C.W., Graff,J.R., De Benedetti,A., Zimmer,S.G. & Rhoads,R.E. Decreasing the level of translation initiation factor 4E with antisense RNA causes reversal of ras-mediated transformation and tumorigenesis of cloned rat embryo fibroblasts. *Int. J. Cancer* **55**, 841-847 (1993).

31. Koromilas,A.E., Lazaris-Karatzas,A. & Sonenberg,N. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *EMBO J.* **11**, 4153-4158 (1992).

32. Koch,G., Bilello,J.A., Kruppa,J., Koch,F. & Oppermann,H. Amplification of translational control by membrane-mediated events: a pleiotropic effect on cellular and viral gene expression. *Ann. N. Y. Acad. Sci.* **339**, 280-306 (1980).

33. Lodish,H.F. Model for the regulation of mRNA translation applied to haemoglobin synthesis. *Nature* **251**, 385-388 (1974).

34. Baserga,R. The cell cycle: myths and realities. *Cancer Res.* **50**, 6769-6771 (1990).

35. Pelletier,J. & Sonenberg,N. The involvement of mRNA secondary structure in protein synthesis. *Biochem. Cell Biol.* **65**, 576-581 (1987).

36. Thach,R.E. Cap recap: the involvement of eIF-4F in regulating gene expression. *Cell* **68**, 177-180 (1992).

37. Zimmer,S.G., Debenedetti,A. & Graff,J.R. Translational control of malignancy: the mRNA cap-binding protein, eIF-4E, as a central regulator of tumor formation, growth, invasion and metastasis. *Anticancer Res. 2000. May. -Jun. ;20. (3A.):1343. -51.* **20**, 1343-1351 (2000).

38. Huang,S. & Houghton,P.J. Targeting mTOR signaling for cancer therapy. *Curr. Opin. Pharmacol. 2003. Aug. ;3(4):371. -7.* **3**, 371-377 (2003).

39. De Benedetti,A. & Rhoads,R.E. Overexpression of eukaryotic protein synthesis initiation factor 4E in HeLa cells results in aberrant growth and morphology. *Proc. Natl. Acad. Sci. U. S. A* **87**, 8212-8216 (1990).

40. Lazaris-Karatzas,A., Montine,K.S. & Sonenberg,N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* **345**, 544-547 (1990).

41. Li,Y., DeFatta,R., Anthony,C., Sunavala,G. & De Benedetti,A. A translationally regulated Tousled kinase phosphorylates histone H3 and confers radioresistance when overexpressed. *Oncogene 2001. Feb. 8. ;20. (6.):726. -38.* **20**, 726-738 (2001).

42. Lazaris-Karatzas,A. & Sonenberg,N. The mRNA 5' cap-binding protein, eIF-4E, cooperates with v-myc or E1A in the transformation of primary rodent fibroblasts. *Mol. Cell Biol.* **12**, 1234-1238 (1992).

43. Miyagi,Y. *et al.* Elevated levels of eukaryotic translation initiation factor eIF-4E, mRNA in a broad spectrum of transformed cell lines. *Cancer Lett.* **91**, 247-252 (1995).

44. Anthony,B., Carter,P. & De Benedetti,A. Overexpression of the proto-oncogene/translation factor 4E in breast-carcinoma cell lines. *Int. J. Cancer* **65**, 858-863 (1996).

45. Rousseau,D., Gingras,A.C., Pause,A. & Sonenberg,N. The eIF4E-binding proteins 1 and 2 are negative regulators of cell growth. *Oncogene* **13**, 2415-2420 (1996).

46. Graff,J.R. *et al.* Reduction of translation initiation factor 4E decreases the malignancy of ras-transformed cloned rat embryo fibroblasts. *Int. J. Cancer* **60**, 255-263 (1995).

47. Nathan,C.A. *et al.* Elevated expression of eIF4E and FGF-2 isoforms during vascularization of breast carcinomas. *Oncogene* **15**, 1087-1094 (1997).

48. Nathan,C.A. *et al.* Detection of the proto-oncogene eIF4E in surgical margins may predict recurrence in head and neck cancer. *Oncogene* **15**, 579-584 (1997).

49. Defatta,R.J., Nathan,C.A. & De Benedetti,A. Antisense RNA to eIF4E suppresses oncogenic properties of a head and neck squamous cell carcinoma cell line. *Laryngoscope 2000. Jun. ;110. (6.):928. -33.* **110**, 928-933 (2000).

50. Kevil,C., Carter,P., Hu,B. & Debenedetti,A. Translational enhancement of FGF-2 by eIF-4 factors, and alternate utilization of CUG and AUG codons for translation initiation. *Oncogene* **11**, 2339-2348 (1995).

51. Zhu,N., Gu,L., Findley,H.W. & Zhou,M. Transcriptional repression of the eukaryotic initiation factor 4E gene by wild type p53. *Biochem. Biophys. Res. Commun. 2005. Oct. 7. ;335. (4):1272. -9.* **335**, 1272-1279 (2005).

52. Levine,A.J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323-331 (1997).

53. Jones,R.M. *et al.* An essential E box in the promoter of the gene encoding the mRNA cap-binding protein (eukaryotic initiation factor 4E) is a target for activation by c-myc. *Mol. Cell Biol.* **16**, 4754-4764 (1996).

54. Blackwell,T.K., Kretzner,L., Blackwood,E.M., Eisenman,R.N. & Weintraub,H. Sequence-specific DNA binding by the c-Myc protein. *Science* **250**, 1149-1151 (1990).

55. Kretzner,L., Blackwood,E.M. & Eisenman,R.N. Myc and Max proteins possess distinct transcriptional activities. *Nature* **359**, 426-429 (1992).

56. Polymenis,M. & Schmidt,E.V. Coordination of cell growth with cell division. *Curr. Opin. Genet. Dev.* **9**, 76-80 (1999).

57. Polymenis,M. & Schmidt,E.V. Coupling of cell division to cell growth by translational control of the G1 cyclin CLN3 in yeast. *Genes Dev.* **11**, 2522-2531 (1997).

58. Vousden,K.H. & Prives,C. P53 and prognosis: new insights and further complexity. *Cell 2005. Jan. 14;120. (1):7. -10.* **120**, 7-10 (2005).

59. Rosenwald,I.B., Rhoads,D.B., Callanan,L.D., Isselbacher,K.J. & Schmidt,E.V. Increased expression of eukaryotic translation initiation factors eIF-4E and eIF-2 alpha in response to growth induction by c-myc. *Proc. Natl. Acad. Sci. U. S. A* **90**, 6175-6178 (1993).

60. Saito,H., Hayday,A.C., Wiman,K., Hayward,W.S. & Tonegawa,S. Activation of the c-myc gene by translocation: a model for translational control. *Proc. Natl. Acad. Sci. U. S. A* **80**, 7476-7480 (1983).

61. Darveau,A., Pelletier,J. & Sonenberg,N. Differential efficiencies of in vitro translation of mouse c-myc transcripts differing in the 5' untranslated region. *Proc. Natl. Acad. Sci. U. S. A* **82**, 2315-2319 (1985).

62. Parkin,N., Darveau,A., Nicholson,R. & Sonenberg,N. cis-acting translational effects of the 5' noncoding region of c-myc mRNA. *Mol. Cell Biol.* **8**, 2875-2883 (1988).

63. Carter,P.S., Jarquin-Pardo,M. & De Benedetti,A. Differential expression of Myc1 and Myc2 isoforms in cells transformed by eIF4E: evidence for internal ribosome repositioning in the human c-myc 5'UTR. *Oncogene* **18**, 4326-4335 (1999).

64. Willis,A.E. Translational control of growth factor and proto-oncogene expression. *Int. J. Biochem. Cell Biol.* **31**, 73-86 (1999).

65. Pelengaris,S. & Khan,M. The many faces of c-MYC. *Arch. Biochem. Biophys. 2003. Aug. 15. ;416. (2):129. -36.* **416**, 129-136 (2003).

66. Levens,D.L. Reconstructing MYC. *Genes Dev. 2003. May. 1;17(9.):1071. -7.* **17**, 1071-1077 (2003).

67. Morrish,F. & Hockenbery,D. Myc's mastery of mitochondrial mischief. *Cell Cycle 2003. Jan. -Feb. ;2(1):11-3.* **2**, 11-13 (2003).

68. Morrish,F., Giedt,C. & Hockenbery,D. c-MYC apoptotic function is mediated by NRF-1 target genes. *Genes Dev. 2003. Jan. 15. ;17(2):240. -55.* **17**, 240-255 (2003).

69. Grens,A. & Scheffler,I.E. The 5'- and 3'-untranslated regions of ornithine decarboxylase mRNA affect the translational efficiency. *J. Biol. Chem.* **265**, 11810-11816 (1990).

70. Manzella,J.M. & Blackshear,P.J. Regulation of rat ornithine decarboxylase mRNA translation by its 5'-untranslated region. *J. Biol. Chem.* **265**, 11817-11822 (1990).

71. Shantz,L.M. & Pegg,A.E. Overproduction of ornithine decarboxylase caused by relief of translational repression is associated with neoplastic transformation. *Cancer Res.* **54**, 2313-2316 (1994).

72. Graff,J.R. *et al.* Translation of ODC mRNA and polyamine transport are suppressed in ras-transformed CREF cells by depleting translation initiation factor 4E. *Biochem. Biophys. Res. Commun.* **240**, 15-20 (1997).

73. Shantz,L.M. & Pegg,A.E. Translational regulation of ornithine decarboxylase and other enzymes of the polyamine pathway. *Int. J. Biochem. Cell Biol.* **31**, 107-122 (1999).

74. Hillary,R.A. & Pegg,A.E. Decarboxylases involved in polyamine biosynthesis and their inactivation by nitric oxide. *Biochim. Biophys. Acta 2003. Apr 11;1647. (1-2):161. -6.* **1647**, 161-166 (2003).

75. Moshier,J.A., Dosescu,J., Skunca,M. & Luk,G.D. Transformation of NIH/3T3 cells by ornithine decarboxylase overexpression. *Cancer Res.* **53**, 2618-2622 (1993).

76. Auvinen,M. *et al.* Human ornithine decarboxylase-overproducing NIH3T3 cells induce rapidly growing, highly vascularized tumors in nude mice. *Cancer Res.* **57**, 3016-3025 (1997).

77. Pegg,A.E. *et al.* Transgenic mouse models for studies of the role of polyamines in normal, hypertrophic and neoplastic growth. *Biochem. Soc. Trans. 2003. Apr;31. (2):356. -60.* **31**, 356-360 (2003).

78. Wallace,H.M. & Caslake,R. Polyamines and colon cancer. *Eur. J. Gastroenterol. Hepatol. 2001. Sep. ;13(9.):1033. -9.* **13**, 1033-1039 (2001).

79. Manni,A. Polyamine involvement in breast cancer phenotype. *In Vivo 2002. Nov. -Dec. ;16. (6.):493. -500.* **16**, 493-500 (2002).

80. Thomas,T. & Thomas,T.J. Polyamine metabolism and cancer. *J. Cell Mol. Med. 2003. Apr-Jun. ;7. (2):113. -26.* **7**, 113-126 (2003).

81. Seiler,N. Thirty years of polyamine-related approaches to cancer therapy. Retrospect and prospect. Part 1. Selective enzyme inhibitors. *Curr. Drug Targets. 2003. Oct. ;4(7.):537. -64.* **4**, 537-564 (2003).

82. Rosenwald,I.B. Growth factor-independent expression of the gene encoding eukaryotic translation initiation factor 4E in transformed cell lines. *Cancer Lett.* **98**, 77-82 (1995).

83. Rosenwald,I.B. *et al.* Eukaryotic translation initiation factor 4E regulates expression of cyclin D1 at transcriptional and post-transcriptional levels. *J. Biol. Chem.* **270**, 21176-21180 (1995).

84. Rosenwald,I.B. Upregulated expression of the genes encoding translation initiation factors eIF-4E and eIF-2alpha in transformed cells. *Cancer Lett.* **102**, 113-123 (1996).

85. Rousseau,D., Kaspar,R., Rosenwald,I., Gehrke,L. & Sonenberg,N. Translation initiation of ornithine decarboxylase and nucleocytoplasmic transport of cyclin D1 mRNA are increased in cells overexpressing eukaryotic initiation factor 4E. *Proc. Natl. Acad. Sci. U. S. A* **93**, 1065-1070 (1996).

86. Lai,H.K. & Borden,K.L. The promyelocytic leukemia (PML) protein suppresses cyclin D1 protein production by altering the nuclear cytoplasmic distribution of cyclin D1 mRNA. *Oncogene 2000. Mar. 23;19. (13):1623. -34.* **19**, 1623-1634 (2000).

87. Topisirovic,I. *et al.* Aberrant eukaryotic translation initiation factor 4E-dependent mRNA transport impedes hematopoietic differentiation and contributes to leukemogenesis. *Mol. Cell Biol. 2003. Dec. ;23(24.):8992. -9002.* **23**, 8992-9002 (2003).

88. Wang,T.C. *et al.* Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice. *Nature* **369**, 669-671 (1994).

89. Miyamoto,R., Uzawa,N., Nagaoka,S., Hirata,Y. & Amagasa,T. Prognostic significance of cyclin D1 amplification and overexpression in oral squamous cell carcinomas. *Oral Oncol. 2003. Sep. ;39. (6.):610. -8.* **39**, 610-618 (2003).

90. Hui,P. *et al.* Real-time quantitative RT-PCR of cyclin D1 mRNA in mantle cell lymphoma: comparison with FISH and immunohistochemistry. *Leuk. Lymphoma 2003. Aug. ;44. (8.):1385. -94.* **44**, 1385-1394 (2003).

91. Guo,S.S. *et al.* Frequent overexpression of cyclin D1 in sporadic pancreatic endocrine tumours. *J. Endocrinol. 2003. Oct. ;179. (1):73. -9.* **179**, 73-79 (2003).

92. Matsumura,I. *et al.* Transcriptional regulation of the cyclin D1 promoter by STAT5: its involvement in cytokine-dependent growth of hematopoietic cells. *EMBO J.* **18**, 1367-1377 (1999).

93. De Benedetti,A. & Harris,A.L. eIF4E expression in tumors: its possible role in progression of malignancies. *Int. J. Biochem. Cell Biol.* **31**, 59-72 (1999).

94. Kevil,C.G. *et al.* Translational regulation of vascular permeability factor by eukaryotic initiation factor 4E: implications for tumor angiogenesis. *Int. J. Cancer* **65**, 785-790 (1996).

95. Boudreau,N. & Myers,C. Breast cancer-induced angiogenesis: multiple mechanisms and the role of the microenvironment. *Breast Cancer Res. 2003. ;5(3):140. -6. Epub. 2003. Mar. 10.* **5**, 140-146 (2003).

96. Streit,M. & Detmar,M. Angiogenesis, lymphangiogenesis, and melanoma metastasis. *Oncogene 2003. May.* **22**, 3172-3179 (2003).

97. Bisacchi,D. *et al.* Anti-angiogenesis and angioprevention: mechanisms, problems and perspectives. *Cancer Detect. Prev. 2003. ;27. (3):229. -38.* **27**, 229-238 (2003).

98. Stein,I. *et al.* Translation of vascular endothelial growth factor mRNA by internal ribosome entry: implications for translation under hypoxia. *Mol. Cell Biol.* **18**, 3112-3119 (1998).

99. Stoneley,M., Paulin,F.E., Le Quesne,J.P., Chappell,S.A. & Willis,A.E. C-Myc 5' untranslated region contains an internal ribosome entry segment. *Oncogene* **16**, 423-428 (1998).

100. Pyronnet,S., Pradayrol,L. & Sonenberg,N. A cell cycle-dependent internal ribosome entry site. *Mol. Cell 2000. Apr;5(4):607. -16.* **5**, 607-616 (2000).

101. Kerekatte,V. *et al.* The proto-oncogene/translation factor eIF4E: a survey of its expression in breast carcinomas. *Int. J. Cancer* **64**, 27-31 (1995).

102. Scott,P.A. *et al.* Differential expression of vascular endothelial growth factor mRNA vs protein isoform expression in human breast cancer and relationship to eIF-4E. *Br. J. Cancer* **77**, 2120-2128 (1998).

103. Li,B.D., Liu,L., Dawson,M. & De Benedetti,A. Overexpression of eukaryotic initiation factor 4E (eIF4E) in breast carcinoma. *Cancer* **79**, 2385-2390 (1997).

104. Li,B.D., McDonald,J.C., Nassar,R. & De Benedetti,A. Clinical outcome in stage I to III breast carcinoma and eIF4E overexpression. *Ann. Surg.* **227**, 756-761 (1998).

105. Franklin,S. *et al.* Detection of the proto-oncogene eIF4E in larynx and hypopharynx cancers. *Arch. Otolaryngol. Head Neck Surg.* **125**, 177-182 (1999).

106. Rosenwald,I.B., Hutzler,M.J., Wang,S., Savas,L. & Fraire,A.E. Expression of eukaryotic translation initiation factors 4E and 2alpha is increased frequently in bronchioloalveolar but not in squamous cell carcinomas of the lung. *Cancer 2001. Oct. 15. ;92. (8.):2164. -71.* **92**, 2164-2171 (2001).

107. Seki,N. *et al.* Expression of eukaryotic initiation factor 4E in atypical adenomatous hyperplasia and adenocarcinoma of the human peripheral lung. *Clin. Cancer Res. 2002. Oct. ;8. (10):3046. -53.* **8**, 3046-3053 (2002).

108. Shevde,L.A. & Welch,D.R. Metastasis suppressor pathways--an evolving paradigm. *Cancer Lett. 2003. Jul. 30. ;198. (1):1-20.* **198**, 1-20 (2003).

109. Klein,P.S. & Melton,D.A. Induction of mesoderm in Xenopus laevis embryos by translation initiation factor 4E. *Science* **265**, 803-806 (1994).

110. Lazaris-Karatzas,A. *et al.* Ras mediates translation initiation factor 4E-induced malignant transformation. *Genes Dev.* **6**, 1631-1642 (1992).

111. Guba,M. *et al.* Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: involvement of vascular endothelial growth factor. *Nat. Med. 2002. Feb. ;8. (2):128. -35.* **8**, 128-135 (2002).

112. Luan,F.L. *et al.* Rapamycin is an effective inhibitor of human renal cancer metastasis. *Kidney Int. 2003. Mar. ;63. (3):917. -26.* **63**, 917-926 (2003).

113. Ramaswamy,S., Ross,K.N., Lander,E.S. & Golub,T.R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet. 2003. Jan. ;33. (1):49. -54. Epub. 2002. Dec. 9.* **33**, 49-54 (2003).

114. Tamm,I., Dorken,B. & Hartmann,G. Antisense therapy in oncology: new hope for an old idea? *Lancet 2001. Aug. 11;358. (9280.):489. -97.* **358**, 489-497 (2001).

115. Podsypanina,K. *et al.* An inhibitor of mTOR reduces neoplasia and normalizes p70/S6 kinase activity in Pten+/- mice. *Proc. Natl. Acad. Sci. U. S. A 2001. Aug. 28. ;98. (18.):10320. -5. Epub. 2001. Aug. 14.* **98**, 10320-10325 (2001).

116. Hengst,L. & Reed,S.I. Translational control of p27Kip1 accumulation during the cell cycle. *Science* **271**, 1861-1864 (1996).

117. Hidalgo,M. & Rowinsky,E.K. The rapamycin-sensitive signal transduction pathway as a target for cancer therapy. *Oncogene 2000. Dec. 27. ;19. (56.):6680. -6.* **19**, 6680-6686 (2000).

118. Peralba,J.M. *et al.* Pharmacodynamic Evaluation of CCI-779, an Inhibitor of mTOR, in Cancer Patients. *Clin. Cancer Res. 2003. Aug. 1;9. (8.):2887. -92.* **9**, 2887-2892 (2003).

119. Defatta,R.J., Li,Y. & De Benedetti,A. Selective killing of cancer cells based on translational control of a suicide gene. *Cancer Gene Ther. 2002. Jul. ;9. (7.):573. -8.* **9**, 573-578 (2002).

120. Marcotrigiano,J., Gingras,A.C., Sonenberg,N. & Burley,S.K. Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951-961 (1997).

121. Marcotrigiano,J., Gingras,A.C., Sonenberg,N. & Burley,S.K. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell* **3**, 707-716 (1999).

122. Tomoo,K. *et al.* Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m(7)GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region. *Biochem. J. 2002. Mar. 15. ;362. (Pt. 3):539. -44.* **362**, 539-544 (2002).

123. Matsuo,H. *et al.* Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nat. Struct. Biol.* **4**, 717-724 (1997).

124. Niedzwiecka,A., Stepinski,J., Balaspiri,L., Darzynkiewicz,E. & Stolarski,R. Thermodynamics of 7-methylguanosine cation stacking with tryptophan upon mRNA 5' cap binding to translation factor eIF4E. *Nucleosides Nucleotides Nucleic Acids 2003. May. -Aug. ;22. (5-8.):1557. -61.* **22**, 1557-1561 (2003).

Chapter 2: Expression, purification and crystallisation of eIF4E.

2.0 Introduction and aims.

A number of different purification protocols exist for eIF4E. Affinity chromatography using m7GTP agarose 4B affinity beads has been successfully used to extract eIF4E from lysed whole cell extract[1]. Other methods used to purify eIF4E have involved the use of MBP tagged eIF4E[2] or classical chromatography techniques such as ion exchange[3]. The main objectives for the purification protocol were:

- To provide large quantities of protein (on the scale of milligrams) for use in biophysical characterisation e.g. crystallography, mass spectrometry and tryptophan fluorescence experiments.

- To provide highly purified eIF4E that has not been contaminated by exposure to any ligands.

- The production of a wild-type protein free from recombinant tags that may interfere with its structure and function.

The techniques fulfilling the above objectives are classical chromatography techniques (such as ion exchange and hydrophobic interaction chromatography) and affinity chromatography exploiting the m7GTP affinity interaction, providing a KCl salt gradient is used for elution of bound protein. Neither of these techniques is dependent upon recombinant protein tags and both avoid using m7GTP in the elution phase. The use of m7GTP could potentially interfere in any type of ligand screening assay or co-crystallisation of a target ligand with eIF4E. Previous studies on eIF4E have mainly utilised protein production in *E.coli*. Advantages of using *E.coli* as an

expression host include high expression of protein and lack of post-translational protein modifications[4]. Expression of eIF4E in *E.coli* also has the advantage that it does not come into contact with any cap-analogues or capped mRNA as the m7Gppp 5'-5' modification is not found in bacteria [5].

2.1 Expression and refolding of full length human eIF4E.

2.1.1 Fundamental aspects of protein expression.

The full-length human eIF4E construct in the pET11d vector was kindly provided by Cyclacel Ltd., Dundee. The pET system has been developed by Novagen Ltd., for the cloning and expression of proteins in *E.coli*. Target genes are cloned in pET plasmids under control of strong bacteriophage T7 transcription and optional translation signals. Providing a source of T7 RNA polymerase in the host cell induces expression. T7 RNA polymerase is highly selective and in certain cases, in a few hours after induction, the desired product can comprise ~ 50% of total cell protein. It is possible to attenuate expression levels by decreasing the concentration of the inducer. Decreasing the expression level may enhance the soluble yield of some target proteins. The system also allows target genes to remain transcriptionally silent in the uninduced state. Target genes are initially cloned into hosts that do not contain the T7 RNA polymerase gene, thus eliminating plasmid instability due to the production of proteins potentially toxic to the host cell. Once established in a non-expression host target protein expression is usually initiated by transferring the plasmid into an expression host containing a chromosomal copy of the T7 RNA polymerase gene under *lacUV5* control. Expression is induced by the addition of IPTG to the bacterial culture [6].

Even in the absence of IPTG, there is some expression of T7 RNA polymerase from the *lac*UV5 promoter in DE3 lysogens and therefore basal expression of the target protein. Recombinant protein expressed in *E.coli* may or may not be toxic to the bacteria by interfering in the normal functions of the cell. The degree of toxicity will vary from protein to protein. If target genes are sufficiently toxic to *E.coli*, this basal level of expression can prevent vigorous growth and establishment of plasmids in λDE3 lysogens. Protein expression can be tightly controlled in the pET system by either using the T7/T7*lac* promoter, use of pLysS or pLysE hosts or by adding glucose to the medium depending on the characteristics of the target protein. The pET11d plasmid, which was used to express eIF4E, contains the T7/T7*lac* promoter[6].

T7*lac* allows control of basal expression of the target gene. These plasmids contain a *lac* operator sequence just downstream of the T7 promoter. They also carry the natural promoter and coding sequence for the *lac* repressor (lacI), orientated so that the T7*lac* and *lac*I promoter diverge. When this type of vector is used in DE3 lysogens, the *lac* repressor acts both at the *lac*UV5 promoter in the host chromosome to repress transcription of the T7 RNA polymerase gene by the host polymerase and at the T7lac promoter in the vector to block transcription of the target gene by any T7 RNA polymerase that is produced[6].

2.1.2 Expression of full length human eIF4E.

The full-length human pET11d eIF4E construct was transformed into competent Rossetta pLysS, Novagen, *E.coli* cells and plated out on ampicillin (100ug/ml)

plates. The Rosetta[TM] strains enhance expression of eukaryotic proteins that contain codons rarely used in E.coli. Most amino acids are encoded by more than one codon, and each organism carries its own bias in the usage of the 61 available amino acid codes. In each cell the tRNA population reflects the codon bias of the RNA population. When the mRNA of heterologous target genes is over expressed in E.coli, differences in codon usage can impede translation due to demand for one or more tRNAs that may be rare or lacking in that population. Insufficient tRNA pools can lead to translational stalling, premature translation termination, translation frame-shifting and amino acid misincorporation. Rosetta[TM] strains supply tRNAs for the codons AUA, AGG, AGA, CUA, CCC and GGA on a compatible chloramphenicol resistant plasmid. All of the B strains (BL21, Rossetta[TM], B834, Origami[TM], Tuner[TM], BLR) are deficient in the Ion protease and lack the ompT outer membrane protease that can degrade proteins during purification[6].

Rossetta pLysS variants were used as another method of tightly controlling basal expression of the target protein. pLysS host strains contain a compatible chloramphenicol-resistant plasmid that provides a small amount of T7 lysozyme, a natural inhibitor of T7 RNA polymerase. T7 lysozyme is a bifunctional protein: it cuts a specific bond in the peptidoglycan layer of E.coli cell wall, and it binds to T7 RNA polymerase, inhibiting transcription. Relatively high levels of T7 lysozyme can be tolerated by E.coli as the inner membrane prevents it reaching the peptidoglycan containing cell wall. pLysS has little effect on growth rate but does aid in lysing the bacterial cells. When the bacterial cells are initially broken up via

sonication, detergent treatment and other various methods T7 lysozyme is released and facilitates in lysing the cells[6].

Single colonies from successful transformation of pLysS Rosseta cells were used to inoculate 10ml LB (100µg/ml ampicillin) cultures and incubated over night at 37ºC. The overnight cultures were used to seed 800ml of LB (100µg/ml ampicillin) each in 2L shaker flasks. The bacterial cultures were grown to optical densities between 0.6 and 0.8 before induction. 1mM IPTG was used to induce protein expression, cultures were incubated at a temperature of 37ºC with shaking at 200 rpm. The cultures for full length human eIF4E were incubated for 3 hours before being centrifuged for 15 mins at 5,000rpm, the supernatant was discarded and the pellet re-suspended in 50mM pH 8.8 Tris and 10% w/v sucrose buffer and stored at -80 ºC. eIF4E is expressed predominately in inclusion bodies with approximately 10% in the soluble fraction (see figure 1). Several attempts were made to solubilise the protein by altering the expression conditions but these were ultimately unsuccessful. Expression conditions were changed by lowering the concentration of IPTG used for induction, lowering the induction temperature and inducing in the presence of 10% sucrose or 5% ethanol. Sucrose and ethanol cause the bacterial cells to grow in an osmotically stressful environment. This leads to expression of bacterial chaperones and heat shock proteins that in principle should aid solubilisation of the protein to be induced[7,8].
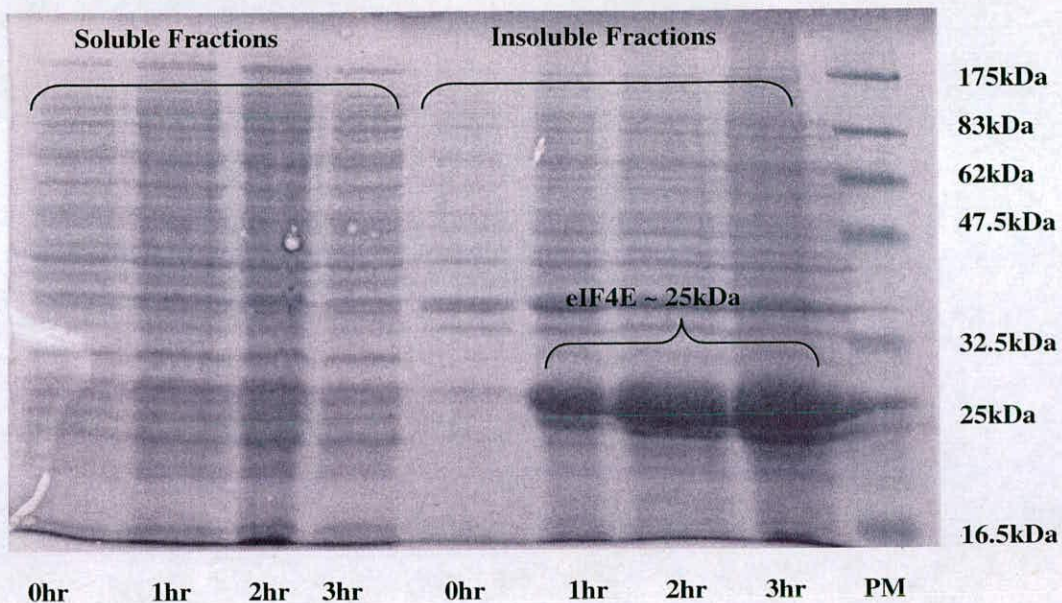
Figure 1: Denaturing SDS-PAGE GEL showing IPTG induced expression of full-length human eIF4E, in ROSETTA pLysS cells, over a period of 3 hours at 37°C. eIF4E is almost exclusively expressed in inclusion bodies. Lane PM contains pre-stained NEB marker proteins.

2.1.3 Purification and refolding of eIF4E from inclusion bodies.

Two possible approaches were available for purification of eIF4E after its expression; either it could be purified from the soluble fraction or the insoluble fraction. The main disadvantage of purifying from the soluble fraction was that very large volumes of *E.coli* culture would be required to produce the amount of protein needed for the planned experimental work. The option of purifying eIF4E from inclusion bodies was taken as it has several distinct advantages; it produces approximately ~30mg of eIF4E protein from 1l of culture, the protein is protected from the action of serine proteases due to the insoluble nature of inclusion bodies and also eIF4E is ~80% pure after being isolated from the inclusion bodies by a washing protocol[9].

39

After eIF4E had been isolated from its inclusion bodies and solubulised in 6M guanidine chloride, 50mM HEPES-KOH pH 7.6, 5mM DTT buffer it was refolded using a rapid dilution technique, whereby the protein was only exposed for a short period of time to intermediate denaturant conditions. DTT was added to reduce any misformed disulphide bonds[10]. Refolding was carried out by slow addition of the unfolded protein (~1 mg/ml) to a larger volume of refolding buffer (contains no denaturant) under vigorous stirring, leaving the protein at a final concentration of 0.1 mg/ml. By carrying protein refolding out under conditions of dilute protein concentration the chances of protein aggregation occurring should be decreased[10].

Apart from using the rapid dilution technique for protein folding there were other alternative methodologies that could have been used for refolding. These range from over-night dialysis into the refolding buffer or carrying out protein refolding on columns. There are three basic approaches: immobilisation of the denatured protein on a matrix and subsequent denaturant dilution to promote protein refolding; denaturant dilution using size exclusion chromatography (SEC); and immobilisation of folding catalysts onto chromatographic supports so that the column behaves like a catalytic folding reactor[10].

Immobilisation of denatured protein on a matrix can be achieved through non-specific interactions or through specific affinity interactions e.g. his-tag. The aim is to isolate individual proteins spatially, thus inhibiting aggregation. Refolding

conditions require careful optimisation because of non-specific protein-matrix interactions[10].

Refolding by SEC aims to inhibit aggregation by restricting the diffusion of various protein forms in the refolding mixture. High dilution factors are achieved during SEC refolding and consequently the final sample concentration is significantly reduced. SEC has a key advantage over simple dilution refolding – material leaving the column has been fractionated on the basis of size. It is therefore possible to obtain free refolded protein free from any contaminants present in the initial solubilsation solution, and also free of aggregates and any unfolded protein[10].

Catalytic column refolding exploits the action of naturally occurring enzymes[10]. Fersht and co-workers[11] developed an oxidative refolding column that has three components immobilised on agarose: GroEL mini-chaperone, which can prevent aggregation; DsbA, which catalyses the oxidation and shuffling of disulphide bonds and peptidyl isomerase[11].

2.1.4 Assessing refolding of eIF4E.

Analytical methods exist to analyse successful re-folding depending on the nature of the protein. Enzyme renaturation can be probed using a suitable activity assay. Therapeutic proteins can be assayed using immunoassays when a suitable antibody exists, or by using suitable bioassays. Proteins that possess internal aromatic amino acids can be probed using intrinsic tryptophan fluorescence. Protein aggregation can be determined by measuring turbidity at 450nm. Limited proteolysis can be used to

assess the compactness of refolded protein and can be used to compare refolding carried out in different environments. Circular dichroism (CD) can also be used to detect the presence of secondary structure. Correct folding of eIF4E was assessed by examining the protein solution for the presence of visible precipitate and by analysing whether it bound m7GTP 4B agarose beads or not using SDS-PAGE gel (see figure 2)[10].
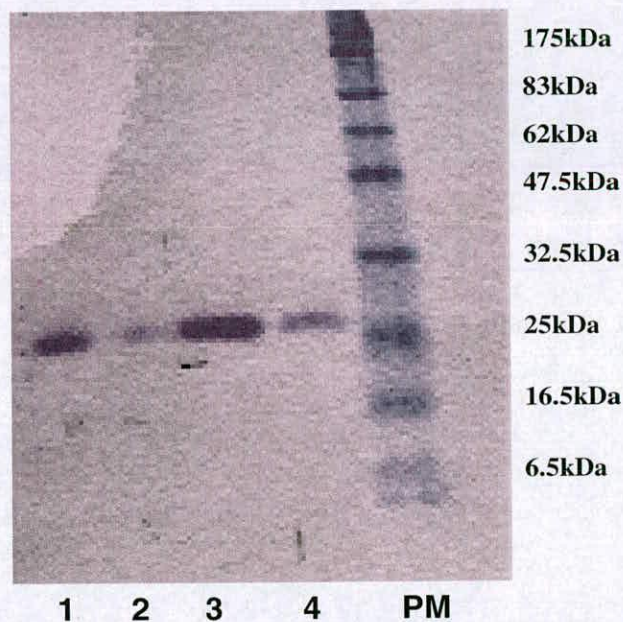
Figure 2: 12% SDS-GEL showing isolation of eIF4E with 0.8mls of m7GTP 4B agarose beads. Lane 1 is the sample of refolded eIF4E used, lane 2 is the elution fraction from the washing step, lane 3 is the elution fraction from the first wash with buffer A (see 7.2.3) containing 0.1mM m7GTP and lane 4 the elution fraction from the second wash with m7GTP containing buffer. Lane PM contains pre-stained NEB marker proteins.

In Figure 2 lanes 3 and 4 show a high proportion of eIF4E is capable of binding the m7GTP 4B agarose beads showing that the refolding protocol has produced a high yield of eIF4E. Lane 2 shows a small amount of eIF4E has not bound but this is probably due to the small amount of m7GTP beads being used.
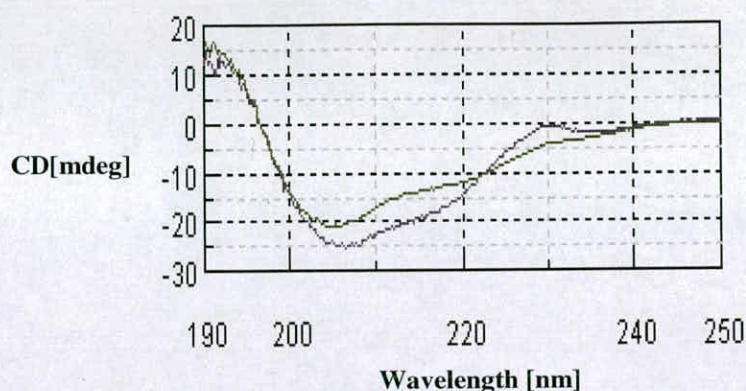
Figure 3: CD spectra of eIF4E with (green) and without the ligand m7GTP (blue) at 4°C. For experimental details see chapter 7.2.6.

The CD spectra, as shown above in figure 3, are highly similar to spectra seen in the literature[12,13]. The spectra were analysed using a suite of programs on Dichroweb[14]. The output from these programs forms a consensus that the spectrum of eIF4E is dominated by beta sheet (~42%) with some evidence of alpha helical secondary (~4%) structure, proportions that agree roughly with the crystal structures of eIF4E seen in the PDB[15-17], where the protein is shown to contain 8 beta sheets and 3 alpha helices. Also a conformation change is seen in the spectra upon binding of m7GTP, further indication that eIF4E is correctly folded and capable of binding the cap analogue.

2.2 Purification of Human Full Length eIF4E

2.2.1 Affinity purification using m7GTP 4B agarose beads.

Refolded protein was directly applied to a Tricorn HR 10/50 column packed with ~ 4mls of m7GTP 4B agarose beads. Both items were supplied by Amersham

biosciences. The column was pre-equilibrated in buffer A containing 20mM HEPES-KOH pH 7.6, 0.1M KCL and 1mM DTT. After application of the protein the column was washed with buffer A and the protein eluted with a block gradient of buffer B. Buffer B consisted of 20mM HEPES-KOH pH 7.6, 2M KCl and 1mM DTT. The block gradient was maintained until the UV trace returned to base line. Instead of using an elution buffer with a high ionic strength, protein can also be eluted with 0.1mM m7GTP or as reported elsewhere 0.6mM GTP[18]. After elution the column was equilibrated again with buffer A (see figure 4).

The purification produced eIF4E of >95% purity that was suitable for both crystallisation trials and biophysical experiments. However, even with the use of no ligand in the elution buffer there was contamination from m7GTP due to leaching from the column. This was noticed as the amount of protein that was eluted off in subsequent purifications was reduced until the column displayed no affinity for eIF4E. This residual contamination of the protein sample made eIF4E purified in this manner unsuitable for ligand binding/screening experiments.
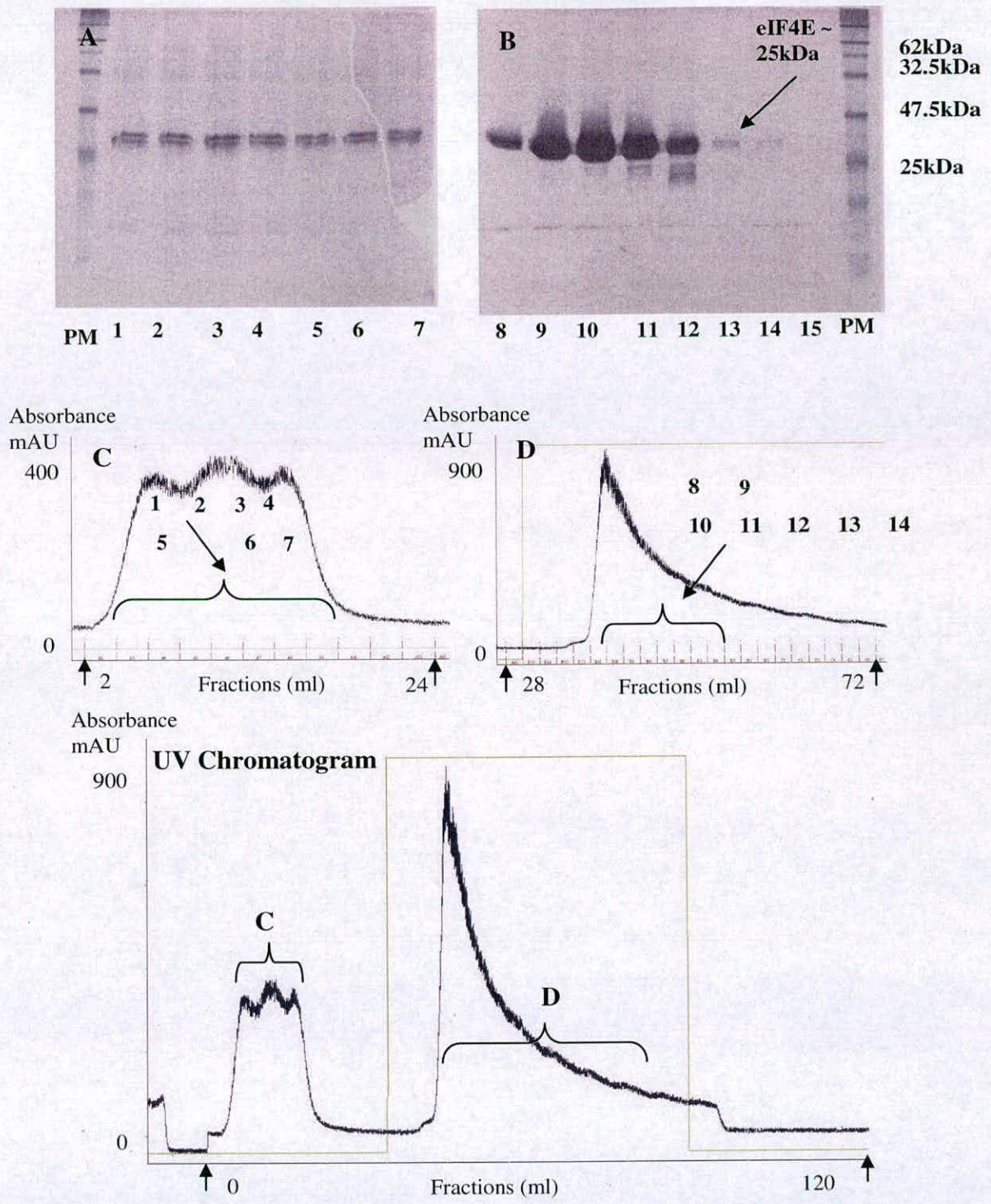
Figure 4: eIF4E was directly loaded onto a m7GTP agarose column, after refolding. The column was then washed using buffer A and the protein was eluted of as a sharp peak with a KCl block gradient. (A and B) shows a 15% SDS-PAGE analysis of the indicated fractions. (C) indicates which fractions from the flow-through were taken for analysis and (D) shows which elution fractions were taken for SDS-PAGE analysis. Flow-through (C) and elution (D) sections are shown on the UV chromatogram. Fractions 8, 9, 10 and 11 were taken for further experiments.

2.2.2 Anion exchange purification of eIF4E using monoQ15 beads.

Refolded eIF4E was concentrated and desalted using PD10 columns prior to loading on to a HR 10/10 monoQ column. Desalting was required to remove guanidine chloride that otherwise would prevent the protein binding to the column. The column was equilibrated in buffer A containing 20mM HEPES-KOH pH 7.6 and 1mM DTT. After application of the protein the column was washed with buffer A and then a KCl gradient was applied to the column. eIF4E eluted from the column at a salt concentration of 0.3M. After elution the column was equilibrated again with buffer A. eIF4E purified via this method was >95% pure despite containing some contaminants not seen in the affinity purification (see figure 5). The main advantage of this purification scheme is that the eIF4E has not come into any contact with m7GTP and as a result is not contaminated making it suitable for ligand binding/screening experiments.
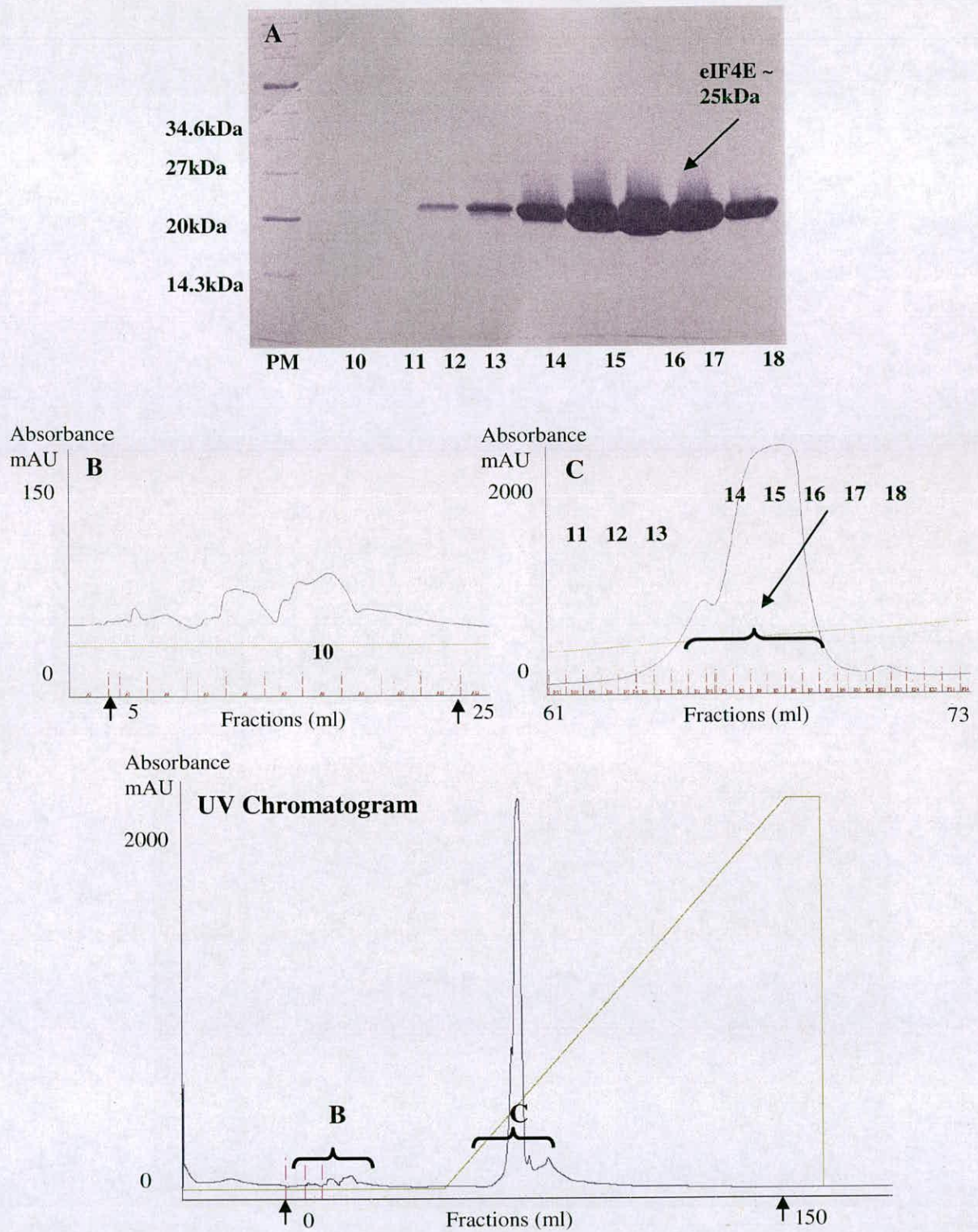
Figure 5: eIF4E protein was desalted using a PD10 column and loaded onto a monoQ column. The monoQ column was then washed using buffer A and the protein was eluted of as a sharp peak with a KCl gradient. (A) shows a 15% SDS-PAGE analysis of the indicated fractions. (B) indicates which fractions from the flow-through were taken for analysis and (C) shows which elution fractions were taken for SDS-PAGE analysis. Flow-through (B) and elution (C) sections are shown on the UV chromatogram. Fractions 14 to 18 pooled for further experiments.

2.3 Crystallisation trials of human full length eIF4E.

2.3.1 Principles of protein crystallisation.

Obtaining suitable single crystals is the least understood step in the X-ray structural analysis of a protein. Protein crystallisation is mainly a trial and error procedure in which the protein is slowly precipitated from solution. The presence of impurities, crystallisation nuclei and other unknown factors play a role in this process. As a rule of thumb, the purer the protein the better the chances of crystallisation[19].

The protein is usually dissolved in a suitable water based buffer system (e.g. HEPES, TRIS). Normally, the precipitant solution is usually added, but only to such a concentration that a precipitate does not develop. The solution is slowly brought to super-saturation. In this step small aggregates are formed, which are the nuclei for crystal growth. Spontaneous formation of nuclei is best achieved at a higher level of saturation. Once nuclei have formed, actual crystal growth can occur. The attachment of new molecules to the surface of a growing crystal occurs at steps on the surface. This is because the binding energy is larger at such positions than if the molecule attaches to a flat surface. These steps are either created by defects in the crystalline order or occur at nuclei formed randomly on the surface. To achieve crystal growth, super-saturation must be reduced to a lower level; maintaining a high level of super-saturation would result in the formation of too many small nuclei and therefore too many crystals (figure 6). Also crystals should grow slowly to reach a maximum degree of order in their structure[19].

Precipitation of the protein can be achieved in several ways. One method involves increasing the effective concentration of the protein, usually by adding a salt (salting out) or polyethyleneglycol (PEG). Either serves to immobilise water, thereby increasing the effective protein concentration. Some proteins are poorly soluble in water but do dissolve if a small amount of salt is added. By removing the salt the protein precipitates. One can regard salting in as the result of competition between charged groups on the surface of the protein molecule and the ions in solution. In the absence of solvent ions the protein precipitates by columbic attraction between opposite charges on different protein molecules. If ions are added they screen the charged groups on the protein and increase solubility. A second method of protein precipitation is to diminish or increase the attractive forces between protein molecules. These forces are: electrostatic, hydrophobic and hydrogen bonding. Additives and conditions that can be used and changed for this purpose are alcohols, salts, PEGs, pH and temperature[19].
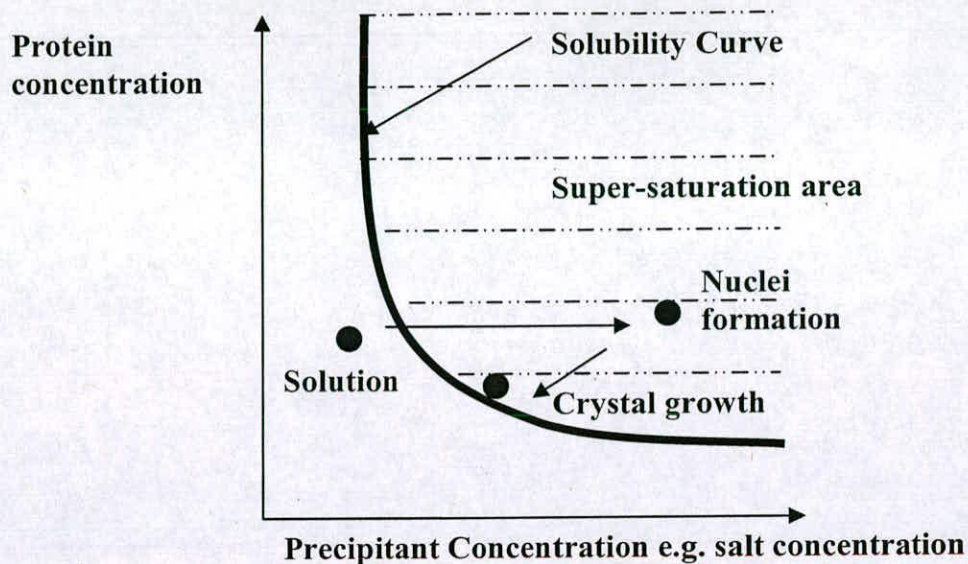
Figure 6: A typical solubility curve for a protein, as a function of the salt concentration or another parameter. Ideal crystal growth occurs at lower levels of super-saturation than is required for the formation of nuclei.

Several experimental methods exist for carefully controlling protein precipitation. These range from batch crystallisation, dialysis, sitting drop, liquid-liquid diffusion and hanging drop vapour diffusion.

In this work only the hanging drop method was used so only a brief description of this method will be included. In this method drops are prepared on a siliconised microscope glass cover by mixing X μl of purified protein, usually ~ 2.5 μl, with the same volume of precipitant solution. The cover slip is placed upside down over a well that contains the required precipitant solution (~ 1ml). The well is sealed by applying oil or grease to the circumference of the well before the cover slip is put into place[19].

$$[\text{ppt}]\text{drop} = \qquad [\text{ppt}]\text{drop} \sim \qquad [\text{ppt}]\text{drop} \sim$$
$$[\text{ppt}]_{\text{reservoir}}/2 \qquad [\text{ppt}]_{\text{reservoir}} \qquad [\text{ppt}]_{\text{reservoir}}$$

Reservoir solution Precipitant (ppt) | **Drop volume decreases, increasing concentration of both protein and precipitant** | **Thus resulting in crystal formation**
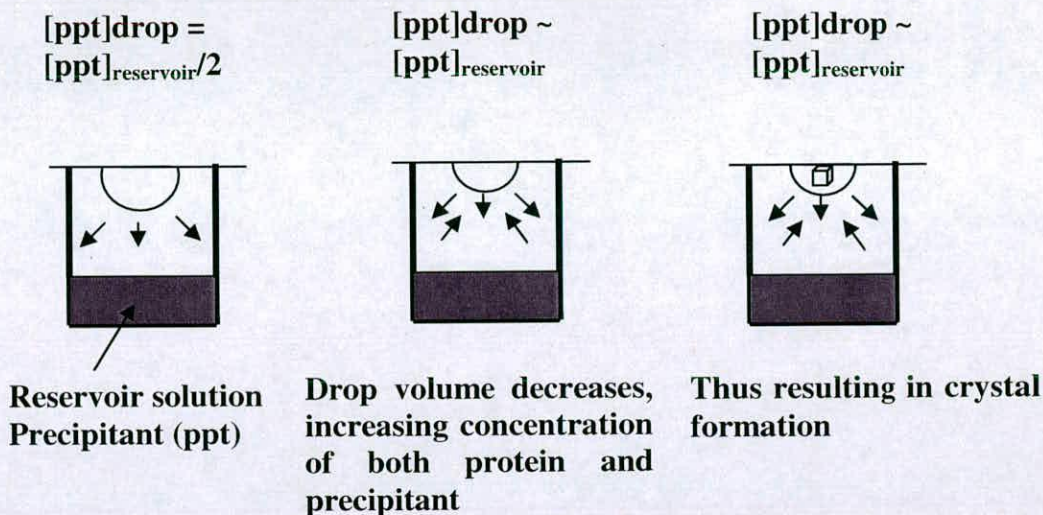
Figure 7: The process of vapour equilibration in a single well. Through loss of water vapour to the larger volume reservoir, the droplet comes into equilibration with the precipitant concentration in the reservoir, thereby inducing crystallisation.

This method works by adding precipitant to the aqueous protein solution just below the point required to precipitate the protein (see figure 7). The water is allowed to evaporate slowly, which gently raises the concentration of both the protein and the precipitant until precipitation occurs. At this point the protein hopefully forms crystals instead of brown amorphous precipitate. Crystal growth depends on many solution properties, including protein concentration, ionic strength etc. Finding the exact conditions to produce good crystals of a specific protein usually requires many careful trials[19].

2.3.2 Crystallisation trials of human full-length eIF4E with m7GTP.

Crystallisation trials were initially carried out with apo protein, however no promising hits were found. Screens using various PEGs (4000, 6000 and 8000) and ammonium sulphate at varying pH produced no promising hits. A brown amorphous precipitate was the most frequently seen result. This observation, supported by the conformational change seen in the CD spectra of eIF4E upon addition of m7GTP, led us to the conclusion that ligand is required for stabilisation and could be an important criteria for successful crystallisation. In the literature successful crystallisation of eIF4E has only been carried out with a cap-analogue present (see figure 8)[13,15,16].
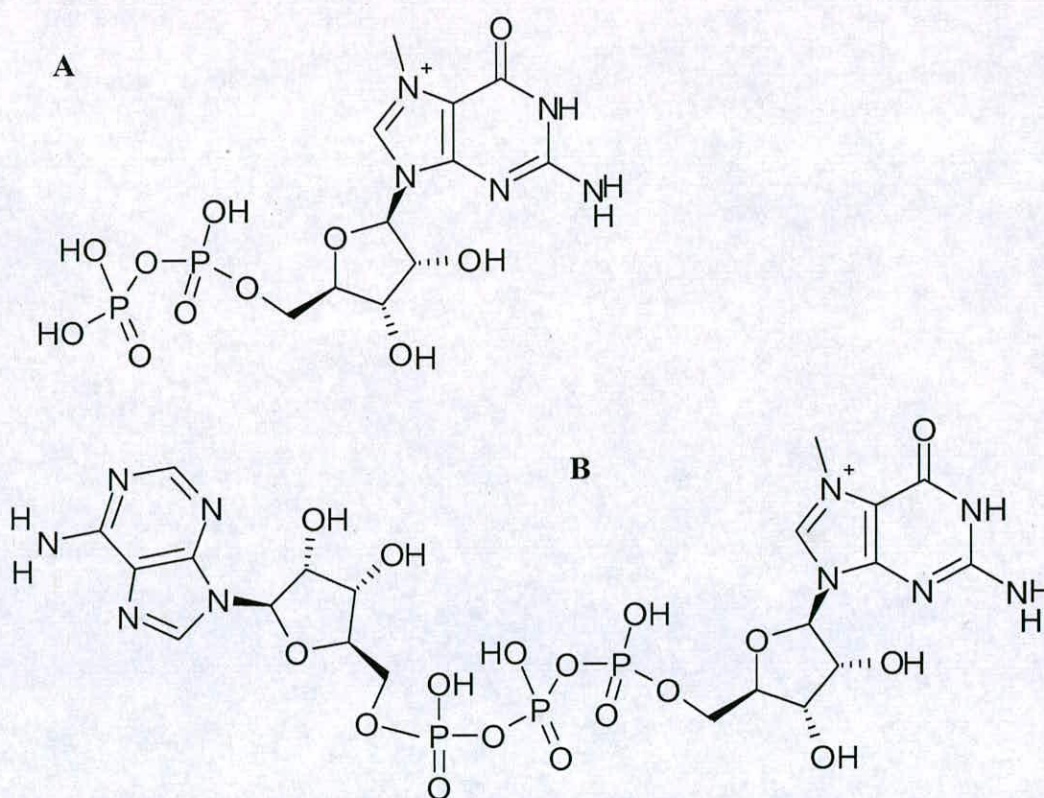
Figure 8: m7GDP (A) and m7GpppA (B) are cap-analogues that have been used in the crystallisation of murine eIF4E and human eIF4E respectively.

Preliminary crystallisation trials of eIF4E in complex with m7GTP were carried out with a PEG 6000/pH screen, an ammonium sulphate/pH screen and another screen based around crystallisation conditions described in Tomoo et al[15] (100mM HEPES-KOH, pH 7.5, 0.2M Ammonium Sulphate, 27% PEG 6000) using the hanging drop vapour diffusion method. Duplicate plates were setup at both 4°C and 18°C. Small needle like crystals were formed in all the screens except in the ammonium sulphate/pH screen where microcrystalline showers were observed. Results were similar at both temperatures. The screen based on previous crystallising conditions at 4°C produced a cluster of long thin needle crystals, similar to those shown in figure 9, that produced diffraction up to a resolution of 2.8-3Å at station 14.2, SRS, Daresbury.
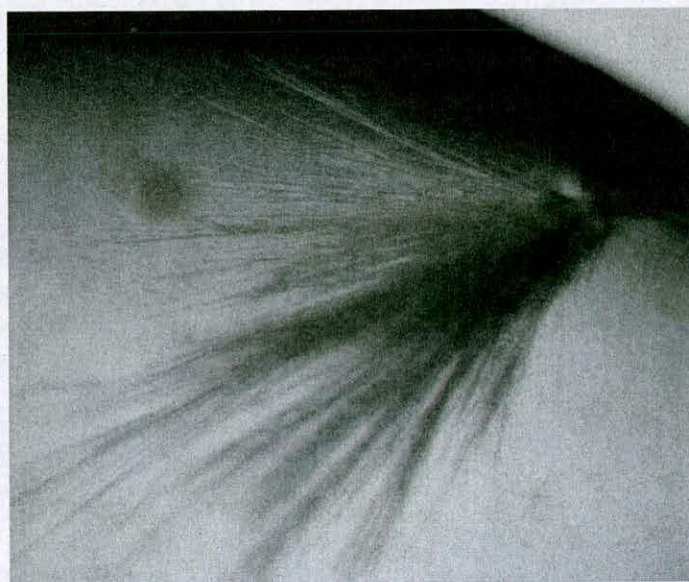
Figure 9: Cluster of needle-like crystals grown under conditions of 100mM HEPES-KOH pH 7.0, 20% PEG6000 and 0.1M ammonium sulphate.

The crystals that diffracted grew at 100mM HEPES, pH 7.0, 22% PEG 6000 and 0.2M ammonium sulphate at 4°C; unfortunately these conditions were not

reproducible. The PEG6000/pH screens were repeated in regions of pH (6.5-8.5) where crystals were seen at 0.1M, 0.2M and 0.3M ammonium sulphate concentrations. These screens produced numerous needle-like crystals. However all were too small to diffract. Crystal screens with a lower amount of protein in the drops were tried in order to decrease the amount of crystal nucleation sites. This worked to a certain extent producing slightly longer crystals, but the problem of too many crystals and crystal clusters forming remained. Also due to the amount of crystal growth being seen it was felt that the protein supply in the drop was being exhausted thus potentially preventing larger crystals from forming.

To control the large amount of nucleation being seen micro-seeding was carried out[20]. This technique involves eIF4E crystals being crushed and a set of serial dilutions being made from this stock. Cat whiskers were then used to seed pre-equilibrated drops with these serial dilutions. Several conditions where crystal growth had occurred were taken and the amount of protein halved in the drop. The aim of this technique is to bring the protein concentration in the drop into the meta-labile zone, where crystals are on the brink of forming. By seeding the drops nucleation sites are being inserted in order to 'kick start' crystal growth. The amount of crystals formed is dependent on how many nucleation sites are introduced into the drop. A single cluster of long needle-like crystals were observed at 10% PEG4000 and 100mM TRIS-HCL, pH 8.0 (as shown in figure 10).
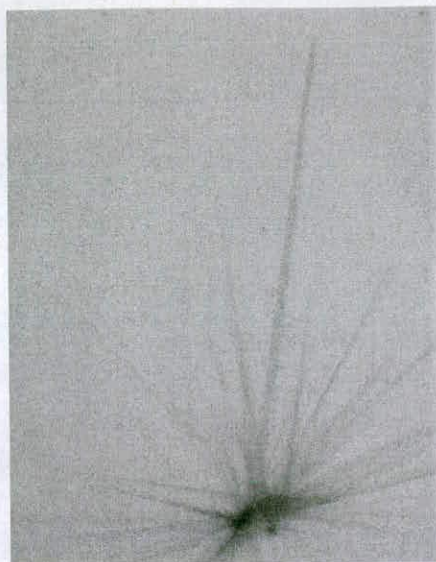
Figure 10: Crystals grown using the micro-seeding technique at 10% PEG4000 and 100mM TRIS-HCl, pH8.0.

Unfortunately a single needle broken off from the cluster did not diffract. In order to freeze these crystals the drops containing them were re-equilibrated with fresh mother solutions containing 50% PEG6000 or PEG4000 depending on which was the major precipitant. In the case of the needle cluster derived from micro-seeding other single needle-like crystals grew during this process. None of these crystals diffracted.

2.3.3 Crystallisation trials for human full-length eIF4E ternary complex with m7GTP and 4E-BP1 motif peptide.

In light of the poor success in crystallising full-length human eIF4E in complex with m7GTP alone, further trials were carried out with the 4E-BP1-motif peptide (RIIYDRKFLMECRN, provide by Cycalcel Ltd, Dundee). Preliminary trials were carried out with PEG/pH and ammonium sulphate/pH screens. These produced no crystals but did give grey crystalline precipitate across the pH range used (pH4-pH9).

Further trials were carried out using ammonium sulphate as an additive for the PEG screens and PEG 400 as an additive for the ammonium sulphate based screen at pH 7.0. Both screens produced crystals with the most promising in terms of size from the PEG screen (see figure 11B). These were thin plate-like crystals whilst those seen in the ammonium sulphate screen were small crystalline clusters appearing to nucleate from single points (see figure 11A).
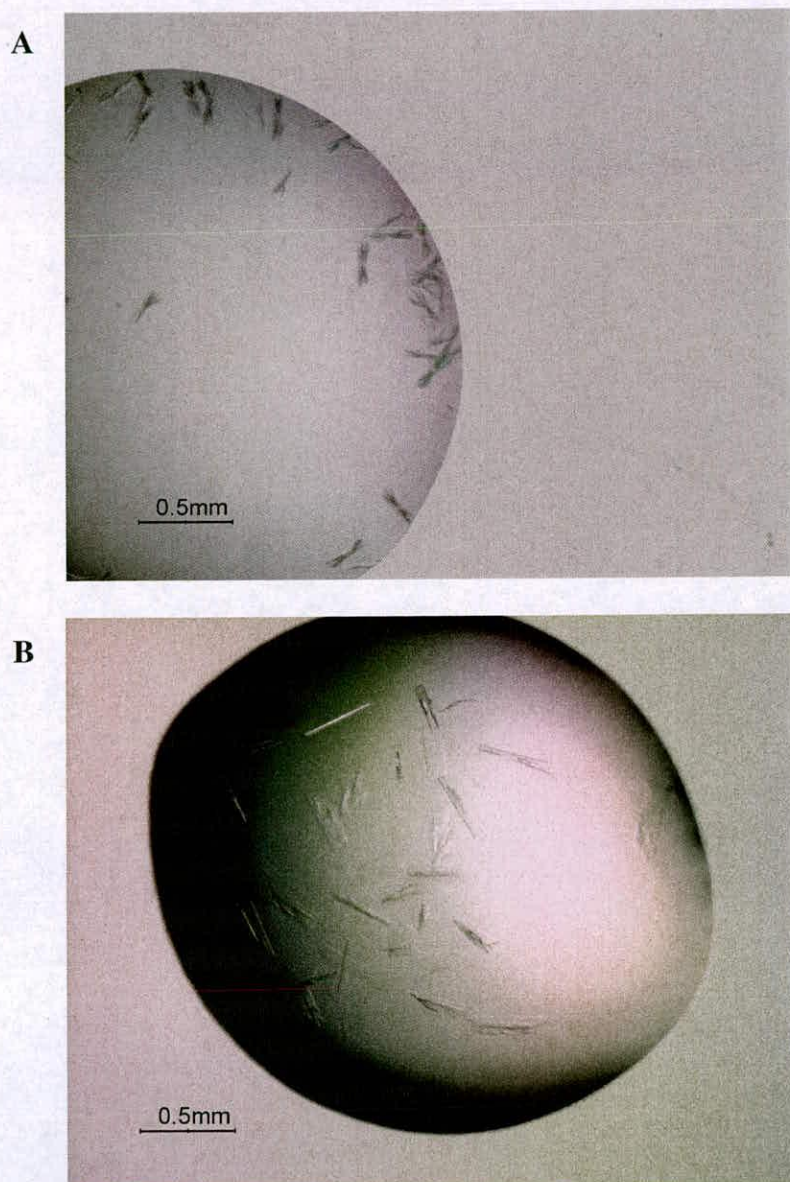
Figure 11: (A) Small crystalline clusters of eIF4E/m7GTP/4E-BP1 peptide complex grown at 60% ammonium sulphate, 4% PEG400, pH7.5 100mM HEPES-KOH. (B) Thin-plate like crystals of the eIF4E/m7GTP/4E-BP1 peptide complex crystallised at 25% PEG 6000, 5% ammonium sulphate, pH 7.0 100mM HEPES-KOH.

The crystals grown in the PEG screen, with ammonium sulphate as an additive, were analysed at the ID29 station at the ESRF, Grenoble where they diffracted to approximately 2Å (see figure 12). The crystal structure solution is discussed in chapter 3.

Figure 12: Diffraction pattern of the thin-plate like crystals of the eIF4E/m7GTP/4E-BP1 peptide complex (crystallised at 25% PEG 6000, 5% ammonium sulphate, pH7.0 100mM HEPES-KOH) as observed in figure 11B.

Crystallisation trials were also carried out with just full-length eIF4E and 4E-BP1-motif peptide. Crystals were observed at 60% ammonium sulphate with either 5% PEG400 or PEG600 at pH7.5 (100mM HEPES-KOH) (see figure 13). However it

was quickly noted that these crystals only grew after protein purification had been carried out using the m7GTP affinity column. It was concluded that m7GTP leeching from the column was responsible for the appearance of these crystals. The crystals grown were flash frozen in liquid nitrogen using immersion oil as a cryoprotectant. Diffraction was observed to 3Å, however after a rotation of 90° in the phi axis the spots become very diffuse and streaky.
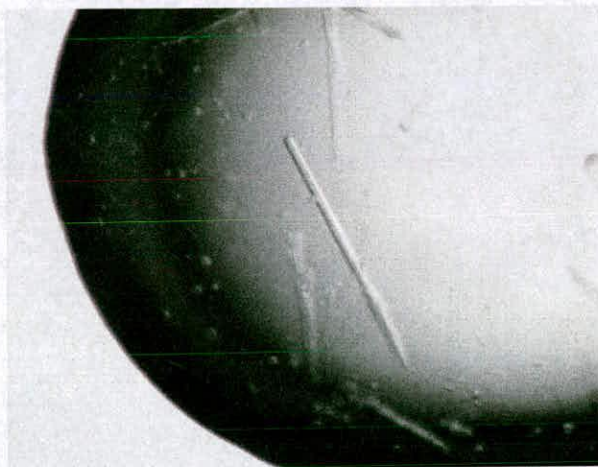
Figure 13: Crystals were observed at 60% ammonium sulphate with either 5% PEG400 or PEG600 at pH7.5 (100mM HEPES-KOH). Diffraction was observed to 3Å; however after a rotation of 90° in the phi axis the spots become very diffuse and streaky. Crystals were only seen after an affinity column preparation; highly suggestive of m7GTP leaching from the column and allowing formation of m7GTP/peptide complexed crystals.

2.4 Conclusions

eIF4E is principally expressed in inclusion bodies which allows it to be purified in a simple two step procedure: inclusion body wash followed by anion-exchange chromatography. This allows us to purify to a in large quantities (30 mg/l) of apo eIF4E to a high degree of homogeneity. eIF4E was readily crystallised in the ternary complex with m7GTP and the 4EBP1 peptide and produced crystals that diffracted to ~ 2Å. Trials of eIF4E with no m7GTP were unsuccessful. With the ability to purify and crystallise eIF4E rational drug design methodology is now open to us for the

search for anti-cancer drug leads. Virtual screens utilising *in silico* technologies such as DOCK[21] and LIDAEUS[22] can be used to investigate the cap-binding site and generate potential drug leads. Virtual hits can then be screened against eIF4E. Techniques for screening these hits will be discussed in chapter 5. Ligands that have then been shown to interact with eIF4E can be used in co-crystallisation trials or soaking trials with ternary crystals of eIF4E.

Reference List

1. Altmann,M., Edery,I., Sonenberg,N. & Trachsel,H. Purification and characterization of protein synthesis initiation factor eIF-4E from the yeast Saccharomyces cerevisiae. *Biochemistry* **24**, 6085-6089 (1985).

2. Morino,S. *et al.* Soluble expression of a synthetic gene for human translation initiation factor 4E in Escherichia coli. *Biol. Pharm. Bull.* **18**, 372-376 (1995).

3. Niedzwiecka,A. *et al.* Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol. 2002. Jun. 7. ;319. (3):615. -35.* **319**, 615-635 (2002).

4. Harris,T.J. & Emtage,J.S. Expression of heterologous genes in E. coli. *Microbiol. Sci.* **3**, 28-31 (1986).

5. Furuichi,Y. & Shatkin,A.J. Characterization of cap structures. *Methods Enzymol.* **180**, 164-176 (1989).

6. pET System Manual, 10th Edition.

7. Thomas,J.G. & Baneyx,F. Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant Escherichia coli. *Protein Expr. Purif.* **11**, 289-296 (1997).

8. Georgiou,G. & Valax,P. Expression of correctly folded proteins in Escherichia coli. *Curr. Opin. Biotechnol.* **7**, 190-197 (1996).

9. Georgiou,G. & Valax,P. Isolating inclusion bodies from bacteria. *Methods Enzymol.* **309**, 48-58 (1999).

10. Middelberg,A.P. Preparative protein refolding. *Trends Biotechnol. 2002. Oct. ;20. (10):437. -43.* **20**, 437-443 (2002).

11. Altamirano,M.M., Golbik,R., Zahn,R., Buckle,A.M. & Fersht,A.R. Refolding chromatography with immobilized mini-chaperones. *Proc. Natl. Acad. Sci. U. S. A* **94**, 3576-3578 (1997).

12. Shen,X., Tomoo,K., Uchiyama,S., Kobayashi,Y. & Ishida,T. Structural and thermodynamic behavior of eukaryotic initiation factor 4E in supramolecular formation with 4E-binding protein 1 and mRNA cap analogue, studied by spectroscopic methods. *Chem. Pharm. Bull. (Tokyo) 2001. Oct. ;49. (10):1299. -303.* **49**, 1299-1303 (2001).

13. Marcotrigiano,J., Gingras,A.C., Sonenberg,N. & Burley,S.K. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell* **3**, 707-716 (1999).

14. Lobley,A., Whitmore,L. & Wallace,B.A. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics. 2002. Jan. ;18. (1):211. -2.* **18**, 211-212 (2002).

15. Tomoo,K. *et al.* Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m(7)GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region. *Biochem. J. 2002. Mar. 15. ;362. (Pt. 3):539. -44.* **362**, 539-544 (2002).

16. Marcotrigiano,J., Gingras,A.C., Sonenberg,N. & Burley,S.K. Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951-961 (1997).

17. Matsuo,H. *et al.* Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nat. Struct. Biol.* **4**, 717-724 (1997).

18. Karim,M.M. *et al.* A quantitative molecular model for modulation of mammalian translation by the eIF4E-binding protein 1. *J. Biol. Chem. 2001. Jun. 8. ;276. (23):20750. -7. Epub. 2001. Mar. 2.* **276**, 20750-20757 (2001).

19. McPherson,A. Introduction to protein crystallization. *Methods 2004. Nov. ;34. (3):254. -65.* **34**, 254-265 (2004).

20. Bergfors,T. Seeds to crystals. *J. Struct. Biol. 2003. Apr;142. (1):66. -76.* **142**, 66-76 (2003).

21. Ewing,T.J., Makino,S., Skillman,A.G. & Kuntz,I.D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des 2001. May. ;15. (5):411. -28.* **15**, 411-428 (2001).

22. Wu,S.Y. *et al.* Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure. (Camb.) 2003. Apr;11(4):399. -410.* **11**, 399-410 (2003).

Chapter 3: Crystallographic and biophysical analysis of full length human eIF4E complexed with m7GTP and the 4E-BP1 motif peptide.

3.0 Introduction and aims.

The aims of this chapter are to provide crystallographic insight into the distinctive features of the cap-binding site and the peptide interactions of the 4E-BP1 peptide with eIF4E. The interactions of the 4E-BP1 peptide were further examined using surface plasmon resonance with the 4E-BP1 peptide immobilised on a streptavidin chip.

3.1 eIF4E's protein interactions.

For the efficient translation of mRNA, systematic interactions are required between the mRNA and the eukaryotic initiation factor proteins. These constitute eIF4E, eIF4A, eIF4B, eIF4H and eIF4G. eIF4A, eIF4E and eIF4G form a stable ternary complex in mammals termed eIF4F[1]. Cap-dependent translation begins with the eIF4F complex recognising the 7-methyl-G5'N cap of mRNA, via eIF4E, and recruiting the ribosome and the mRNA[1]. eIF4A is a DEAD family RNA helicase, which is hypothesised to unwind secondary structure in the mRNA 5' untranslated region to facilitate ribosome binding to the mRNA and its subsequent scanning to reach the initiation codon[2]. eIF4A and eIF4E interact with the large modular scaffolding protein eIF4G. eIF4G also interacts with the multi-subunit initiation factor eIF3 that binds directly to the ribosome, as well as the poly-A-binding protein (PABP)[1]. The PABP-eIF4G interaction brings about the circularisation of the mRNA[3].

Cap-dependent translation is inhibited by 4E-BP1, 2 and 3, without affecting cap recognition by eIF4E[4]. Biochemical studies have shown that eIF4Gs and 4E-BPs occupy overlapping sites on the surface of eIF4E[5]. Further work demonstrated that both sets of proteins converged on the same eIF4E binding mechanism, which employs a Tyr-X-X-X-X-Leu-Φ eIF4E recognition motif (X is variable and Φ is Leu, Met or Phe)[6]. The eIF4GII and 4E-BP1 recognition motifs have also been crystallised in a ternary complex with truncated mouse eIF4E and the cap-analogue m7GDP[7]. In both complexes the peptides were seen to adopt an L-shaped, extended chain/alpha-helical conformation that docks into an invariant portion of the convex surface of eIF4E (see figure 9). The 4E-BP1 peptide acts as a molecular mimic of eIF4GII and carries out inhibition by competing for the same binding site on the convex dorsum of eIF4E and blocking assembly of the translation machinery. Biophysical techniques have also demonstrated that these peptides are unordered in solution before becoming ordered upon binding[7,8]. Phosphorylation of 4E-BP1 leads to its dissociation from eIF4E, leaving eIF4E free to bind eIF4G and form the eIF4F complex.

Recently an NMR structure has been solved of the yeast eIF4E protein in complex with m7GTP and the eIF4GI binding domain[9]. Here they have described the N-terminal domain of eIF4E becoming ordered with the eIF4GI fragment forming a 'bracelet' around it and the remainder of the N-terminal forming a molecular 'fist' (see figure 1). From yeast to mammals, eIF4G mediates increased association of eIF4E with capped mRNA whereas the motif peptides alone are unable to cause such

an effect[9]. Gross et al[9] have proposed that the N-terminus acts as a 'remote sensor' for the binding slot. From SPR and ITC data for binding of eIF4GI to a series of N-terminally deleted eIF4E constructs they observed that a two state binding model described the results for Δ10 and Δ20 eIF4E and that a one state binding model described results for the Δ30 and Δ35 data[9]. From this they have proposed that the consensus peptide may participate in encounter complex formation as it binds to eIF4E which is followed by the formation of the interlocking interface. The interlocking interface promotes strong binding and slow dissociation kinetics[9].
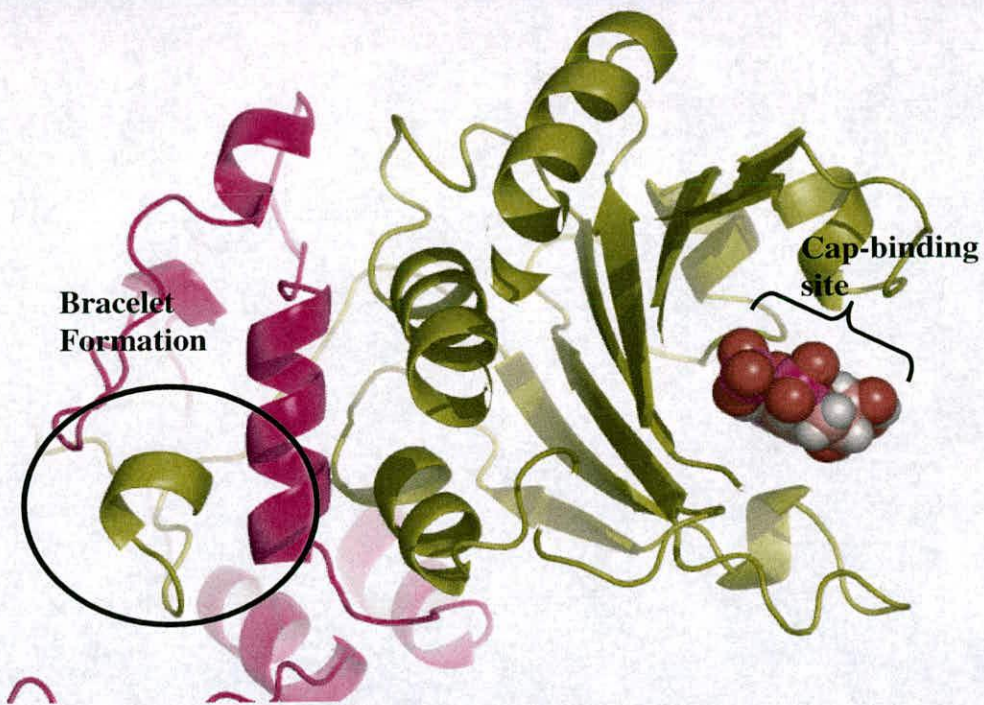
Figure 1: The N-terminal of eIF4E (yellow) becomes ordered when it interacts with the eIF4GI fragment (purple), forming a 'bracelet' around it with remainder of the N-terminal forming a molecular 'fist'[9].

3.2 Crystal structure and modelling of eIF4E ternary complex.

3.2.1 Data collection.
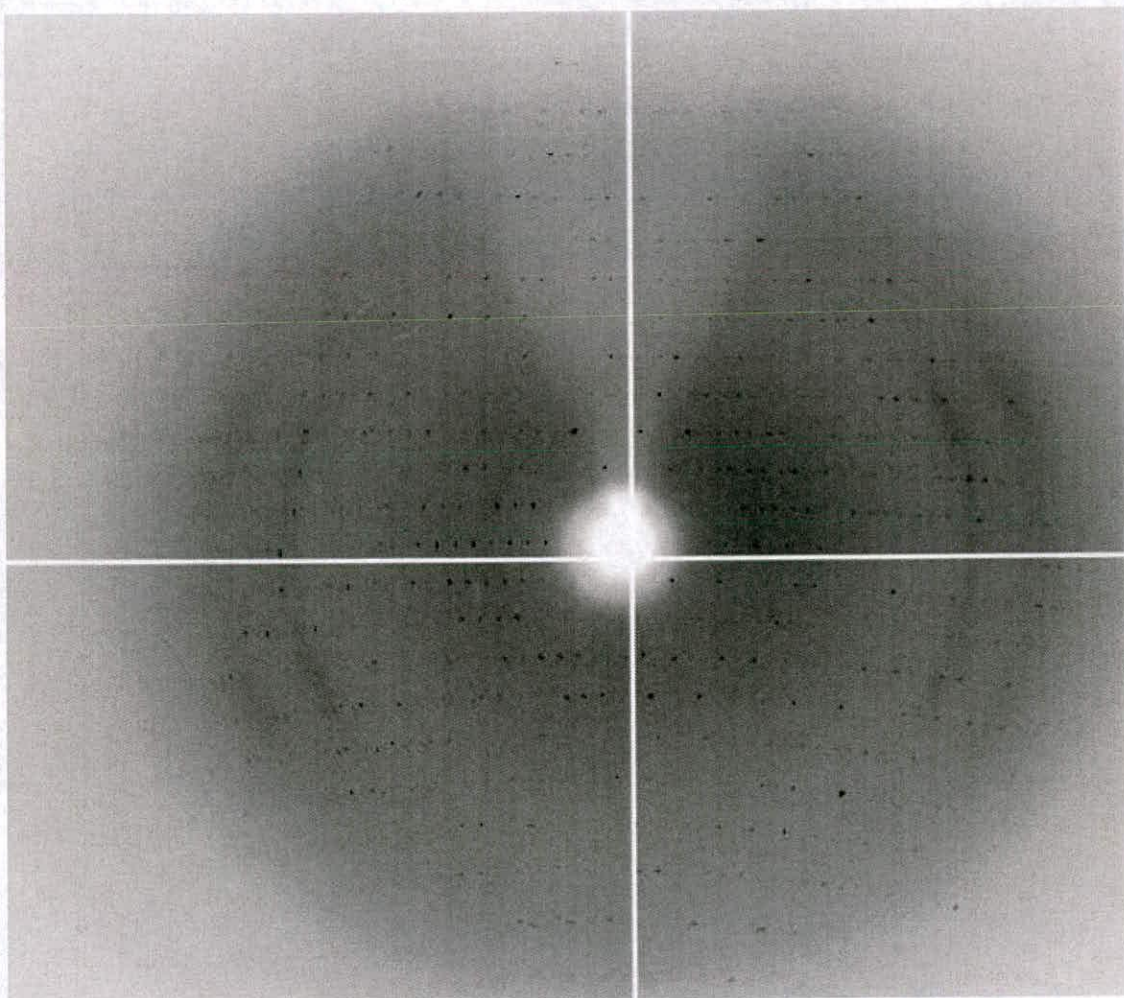
Thin-plate like crystals of the eIF4E/m7GTP/4E-BP1 peptide complex, crystallised at 25% PEG 6000, 5% ammonium sulphate, pH 7.0 100mM HEPES-KOH, were cryo-frozen using freezing solution containing 20% Glycerol, 5% ammonium sulphate, 25% PEG 6000 and pH 7.0 100mM HEPES-KOH. The eIF4E/m7GTP/4E-BP1 peptide crystal was mounted in a random orientation and 120 degrees of data were taken with an oscillation angle of $1°$ using the Phi rotation method at 100K at the ESRF ID29 beamline at Grenoble, France. Keeping the crystal at 100K increases the lifetime of the crystal in the X-ray beams. The X-ray beam generates heat and reactive free radicals. By freezing the crystal the diffusion of free radicals is limited and crystal deterioration is thus limited at sites were free radicals have been generated. Symptoms of x-ray damage are loss of resolution in the diffraction image and deterioration of scaling statistics when processing the data[10]. The goal of data collection is to determine the indices and record the intensities of as many reflections as possible. Diffraction data was indexed in MOSFLM[11] and scaled in SCALA[11] revealing the unit cell to be $P2_12_12_1$ (see table 1 for unit cell dimensions and scaling statistics).

3.2.2 Structure solution.

The Mathew's coefficient calculated using the CCP4 suite of programs[11] was 2.4 giving a solvent content 47.7% for the eIF4E/m7GTP/4E-BP1 peptide crystal. This result indicated the presence of two ternary complexes in the asymmetric unit of the

unit cell. The structure of the ternary complex containing full-length human eIF4E/4E-BP1/m7GTP was solved by molecular replacement using MOLREP[11]. The search model was the 2.2 angstrom structure of Δ27 murine eIF4E complexed with the 4E-BP1 peptide, with m7GDP and any structured waters removed. Data up to 3.5 angstrom was used for the MOLREP[11] solution. MOLREP[11] isolated the positions of two ternary complexes in the asymmetric unit. These initial co-ordinates were used to begin refinement.

3.2.3 Model building and refinement.

Prior to refinement, $2F_o.F_c$ and F_o-F_c maps were calculated using FFT in CCP4 using data in the resolution of 30-2.3 angstrom[11]. The model was examined and was found in electron density for the entire length, except for the N-terminal and a section of the C-terminal loop. The starting model was subjected to 10 cycles of rigid body refinement and followed by 10 cycles of restrained refinement in REFMAC 5.0[11,12]. This gave an R-factor of 28.3 and an R-free of 32.8. At this point waters were added using ARP/WARP[11,13]. After this the m7GTP ligand was used (REFMAC[11,12] library files generated using PRODRG[14]) in the refinement and placed into clearly visible electron density. This was then followed by continuous cycles of manual refinement of the model using COOT[15] and 10 cycles of restrained refinement using REFMAC 5.0[11,12].

Unit Cell Dimensions	a=37.9, b=99.6, c=135.3
	$\alpha = \beta = \gamma = 90°$
Space Group	$P2_12_12_1$
Temp(K)	100K
Number of collected reflections to 2.3Angstrom	112855 (12976)
Number of unique reflections to 2.3Angstrom	22656 (3030)
R sym	0.157 (0.574)
R factor	19.5
R free	24.9
RMS bonds	0.010
RMS angle	1.352
Average B value	26.3
% completeness	96% (89.5%)
Number of observed solvent molecules	300
Peptide	2
Ligand	2
Ramachandran data	92.2% residues in most favoured regions
	6.1% residues in additional regions
	1.2 residues in generously allowed regions
	0.6% residues in disallowed regions

Table 1: The table shows the unit cell dimensions and space group for the plate-like crystals of full length human eIF4E co-crystallised with m7GTP and the 4E-BP1 motif peptide. Also shown are the data collection statistics and the parameters of the final model.

3.3 Crystallographic analysis of the eIF4E ternary complex.

3.3.1 eIF4E general structure and fold.

The structures of truncated murine eIF4E bound to m7GDP/m7GpppG[16,17] and full length human eIF4E complexed with m7GTP/m7GpppA[18] have been solved by X-ray crystallography. The yeast homolog has also been elucidated by NMR studies[19]. These proteins have high sequence similarity (see figure 3) and resemble a cupped hand, consisting of a curved, eight-stranded anti-parallel beta sheet, with three long alpha helices packed behind them (Fig 2). The three long alpha helices lie along the beta sheet on the opposite side of the cap-binding site running almost parallel in direction to the strands (figure 2). The cleft that forms the cap-binding site is formed by the concave side of the beta sheet, the short alpha helix (A) above it and a loop (B) below (figure 2).
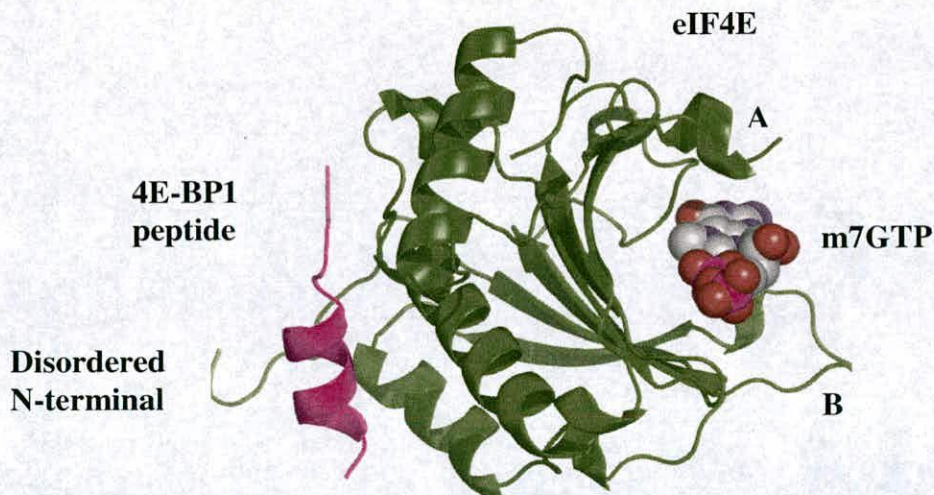
Figure 2: The crystal structure shown above is full length human eIF4E complexed with the 4E-BP1 motif peptide and m7GTP. eIF4E possesses an α/β fold typical of the RNP fold of RNA binding proteins. The 4E-BP1 motif peptide is located on the dorsal side of the protein with the cap binding side lying on the opposite side of the protein. A) and B) indicate the short helix and loop that together with the β sheet form the binding site for m7GTP.

```
            ....|....|  ....|....|  ....|....|  ....|....|
                5           15          25          35
Human       MATVEPETTP  TPNPPTTEEE  KTESNQEVAN  PEHYIKHPLQ
Mouse       MATVEPETTP  TTNPPPAEEE  KTESNQEVAN  PEHYIKHPLQ
Yeast       MSVEEVSKKF  EENVSVDDTT  ATPKTVLSDS  AHFDVKHPLN

            ....|....|  ....|....|  ....|....|  ....|....|
                45          55          65          75
Human       NRWALWFFK-  -NDKSKTWQA  NLRLISKFDT  VEDFWALYNH
Mouse       NRWALWFFK-  -NDKSKTWQA  NLRLISKFDT  VEDFWALYNH
Yeast       TKWTLWYTKP  AVDKSESWSD  LLRPVTSFQT  VEEFWAIIQN

            ....|....|  ....|....|  ....|....|  ....|....|
                85          95          105         115
Human       IQLSSNLMPG  CDYSLFKDGI  EPMWEDEKNK  RGGRWLITLN
Mouse       IQLSSNLMPG  CDYSLFKDGI  EPMWEDEKNK  RGGRWLITLN
Yeast       IPEPHELPLK  SDYHVFRNDV  RPEWEDEANA  KGGRWSFQLR

            ....|....|  ....|....|  ....|....|  ....|....|
                125         135         145         155
Human       KQQRRSDLDR  FWLETLLCLI  GESFDDYSDD  VCGAVVNVRA
Mouse       KQQRRSDLDR  FWLETLLCLI  GESFDDYSDD  VCGAVVNVRA
Yeast       --GKGADIDE  LWLRTLLAVI  GETIDEDDSQ  INGVVLSIRK

            ....|....|  ....|....|  ....|....|  ....|....|
                165         175         185         195
Human       KGDKIAIWTT  ECENREAVTH  IGRVYKERLG  LPPKIVIGYQ
Mouse       KGDKIAIWTT  ECENRDAVTH  IGRVYKERLG  LPPKIVIGYQ
Yeast       GGNKFALWT-  KSEDKEPLLR  IGGKFKQVLK  LTDDGHLEFF

            ....|....|  ....|....
                205         215
Human       SHADTATKSG  STTKNRFVV
Mouse       SHADTATKSG  STTKNRFVV
Yeast       PHSSANGRHP  QPSITL---
```

Figure 3: Sequence alignment of human, mouse and yeast eIF4E sequences. Tryptophans that stack with m7G are highlighted in cyan, the glutamate that forms hydrogen bonds with the m7G ring is in **bold**, the residues that form the phosphate binding pocket are highlighted in blue and the third tryptophan that the m7G ring packs against is shown in purple. The part of the mouse sequence not highlighted in red shows the construct used for the crystallisation of the mouse structure. The section in red also contains the majority of amino acid differences between the mouse and human sequences.

3.3.2. Cap-binding site.

Our structure is highly similar to the mouse, yeast and previously solved human structures[7,16,18,19]. The m7G moiety of the cap structure is sandwiched between two tryptophans (W56 and W102) forming π-cation interactions[20,21] (figure 4). The tryptophan residues are located in a narrow slot on the molecule's concave surface (see figure 2). The methyl group of the m7G moiety makes a van der Waals contact with a third conserved tryptophan (W166). The structure of the cap-binding pocket is conserved between human, mouse and yeast eIF4E. Two NH groups on the m7G moiety form H-bonds with the carboxyl group of E103 stabilising the ligand/protein interactions (figure 4). Our structure displays disorder in the C-terminal loop (Ala 203 to Thr 211) that in other structures containing the dinucleotide analogue m7GpppA is seen to be structured (figure 5)[18]. This stabilisation is brought about by hydrogen bonding between the adenosine ring with Thr 205 and Thr 211 via a water molecule and is stabilised by some Van der Waals contacts with Ser 207 and neighbouring residues (figure 5).

The α-phosphate has a direct electrostatic interaction with R157 and interacts with several conserved water positions (figure 6). The second phosphate group forms three direct contacts to K159, K162 and R157 (figure 6). The gamma phosphate forms a direct ionic contact but also interacts with other structured waters that mediate its interactions with the cap-binding site (figure 6). For further discussion of phosphate tail interactions see 6.3.4.

A

R112

K162

W166

W102

K159

E103

R157

W56
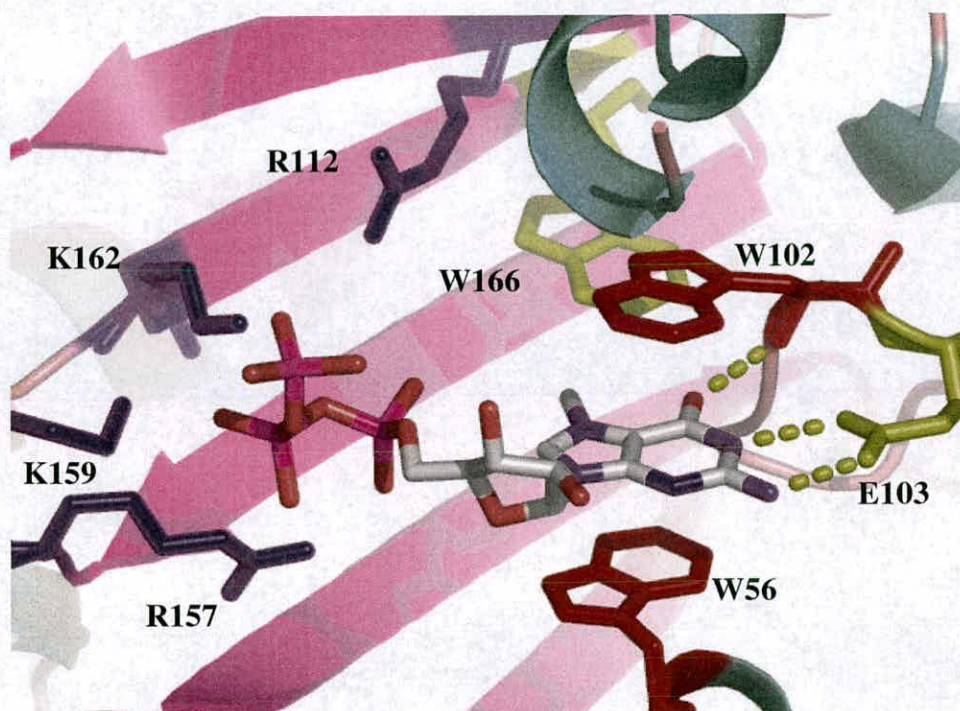
B

Negatively charged triphosphate tail

Tridentate HA and HD arrangement on m7G ring
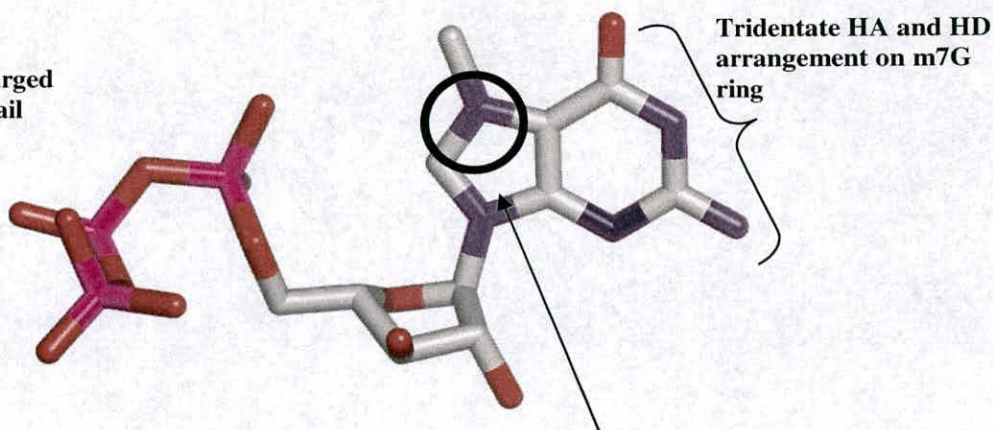
Positive delocalised charge on m7G ring

Figure 4: (A) The m7G moiety intercalates itself between the W102 and W56 side chains, the N7 methyl group forms a van der Waals contact with W166 and three hydrogen bonds are formed between the m7G ring and the E103 side chain and the backbone amide of W102. The triphosphate tail is recognised by R112, R157, K159 and K162 and via several water-mediated contacts (see figure 4). (B) Distinguishing feature of m7GTP specifically recognised by eIF4E.

A

Triphosphate tails packed against the positively charged phosphate binding site.

B

W102

m7G moiety intercalated between the tryptophans

W56

C

C-terminal loops becomes ordered in the (1IPB) eIF4E crystal structure complexed with m7GpppA

S207

K206

T205

T211

Adenine ring
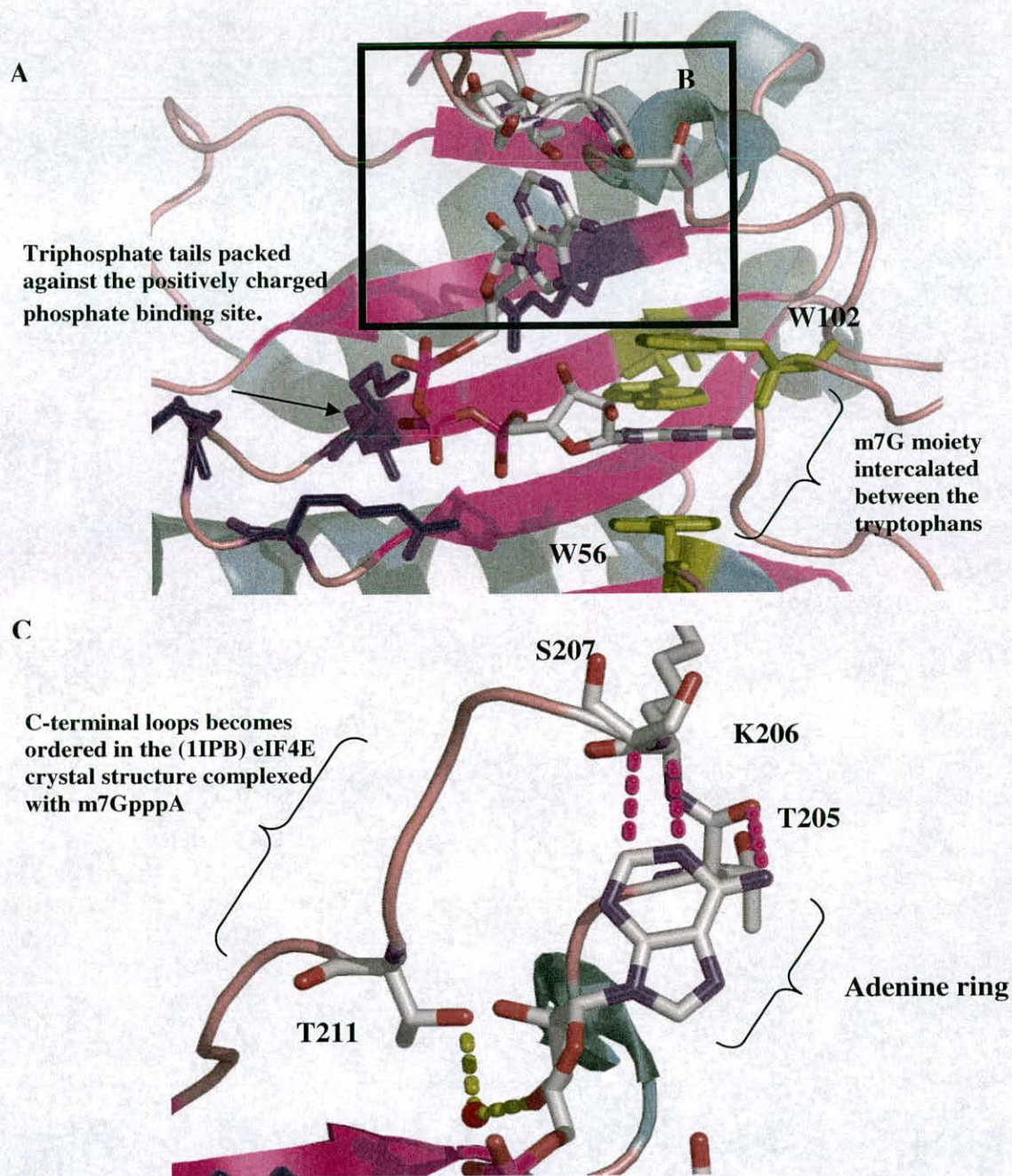
Figure 5: A) the m7G moiety intercalates between the tryptophans (yellow) as normal and the triphosphate tails resides on the positive binding pocket as described previously (figure 4). The area highlighted by the rectangle (B) in A is shown in C. Here it can be seen that the adenine moiety stabilises the C-terminal loop which is not seen in other eIF4E structures. The ribose rings forms a water mediated hydrogen bond to T211 (yellow dashes) whilst the adenine rings forms a variety of contacts with S207, K206 and T205 (magenta dashes).
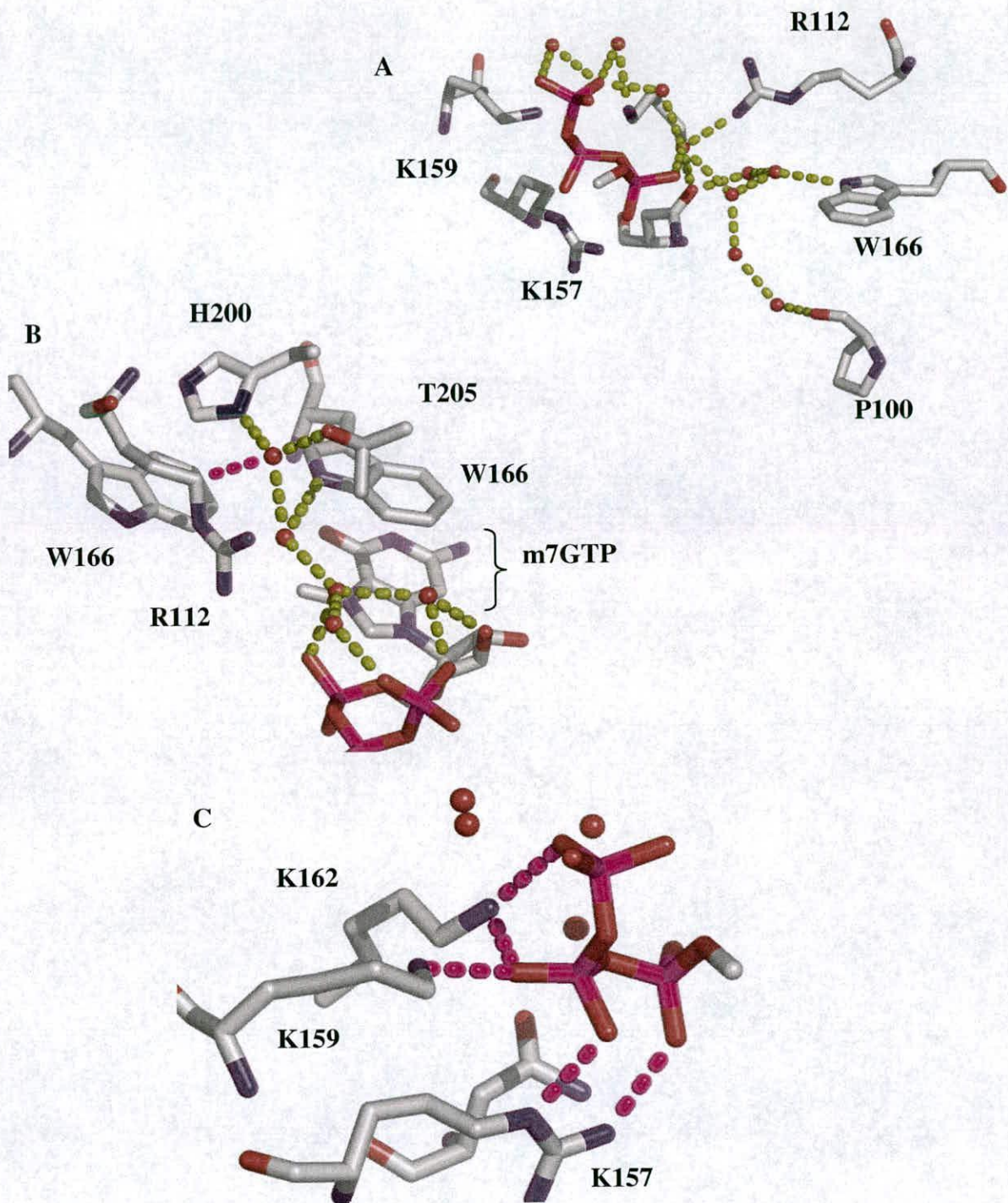
Figure 6: A) shows a water network between the triphosphate tail and eIF4E residues. B) is a water network that forms behind the cap binding site involving W102, the O' of the ribose ring, T205 and H200. Potential hydrogen bonds shown with yellow dashes and direct contacts shown with magenta dashes. Both these networks constitute part of a larger network at the protein/ligand interface. C) Direct contacts formed between the triphosphate tail and the residues R157, K159 and K162. K159 in other crystal structures is seen projecting into the solvent. Contacts shown with magenta dashes.

3.3.3 Other cap-binders in biology.

CBC is a cap-binding complex present in the nucleus and consists of two subunits CBP20 (the cap-binding protein) in association with an ancillary protein CBP80. The complex is thought to aid in the transport of pre-mRNAs through the different maturation processes[22]. Some viruses e.g. retroviruses, utilise the host cellular capping mechanism[23]. Most viruses have evolved their own virus specific capping system[23]. They can be divided into two groups.

The first group includes viruses such as poxviruses, coronaviruses, flaviviruses and reoviruses that replicate in the cytoplasm of the cell. Vaccinia virus (a double-stranded DNA pox virus) is a well studied example of such viruses. Thus VP39, the vaccinia viral protein 39, recognises its own capped mRNAs. After binding the cap structure, the enzyme acts as a 2'-O-methyltransferase transferring a methyl group to the 2' atom of the moiety of the first residue of the cap structure[23]. Baculoviruses, like flaviviruses, also appear to have evolved cap-dependent 2'-O-methyltransferases, but the mechanism of cap-binding by the flavivirus 2'-O-methyltransferase domain may differ from that of VP39[23].

The second group includes negative-strand RNA viruses replicating either in the nucleus, like orthomyxoviruses (e.g. influenza virus) or in the cytoplasm, like bunya viruses[23]. These viruses have not evolved capping enzymes. Instead they have evolved a mechanism to 'hijack' the cap structure from host cell mRNAs. The viral polymerase binds to cellular capped mRNA in infected cells. An endonuclease

activity associated with the polymerase then cleaves the capped oligonucloetides, which subsequently act as primers to initiate viral transcription[23].

Several isoforms of cap-binding proteins have also been found in *Leishmania* and *C.elegans*[24]. In *C.elegans* it has been discovered that cap-binders selectively lead to an increase in translation of a certain subset of mRNAs[25].

3.3.4 Comparison of m7GTP binding in VP39 and CBC.

Cap-binders that have successfully been structurally determined are VP39, CBC and eIF4E. The overall structure of the three cap-binding proteins eIF4E[16,18,19], CBC[26] and VP39[27] differ widely due to their evolutionary unrelated origin. However their cap-binding sites are essentially very similar. In addition to the two aromatic residues sandwiching the m7G moiety, they possess a basic area to accommodate the triphosphate moiety of the cap structure and acidic residues that recognises the HA and HD on the m7G ring.

All three proteins accommodate the m7G ring sandwiched between two aromatic residues (W102/W56 in eIF4E, Y20/Y43 in CBC and Y22/F180 in VP39) (figure 7). The presence of the aromatic sandwich was suggested from earlier crystallographic, spectroscopic and theoretical studies on small molecule models, like tryptophan containing peptides and/or indole derivatives[28-32].

A Guanine ring   m7G ring

Phosphate tail

Y138   Y43

Q135   D114

E133   D116

R127   Y20   R112

B m7G ring

Y204   Y22

E233

S205

Phosphate tail   D182

F180

E207

R177

C

Phosphate tail   R112

K162   W166   W102

K159

E103

R157   W56

m7G ring
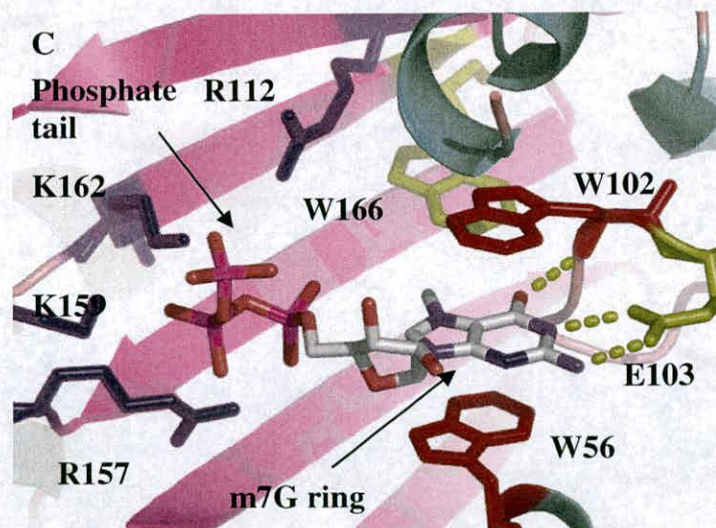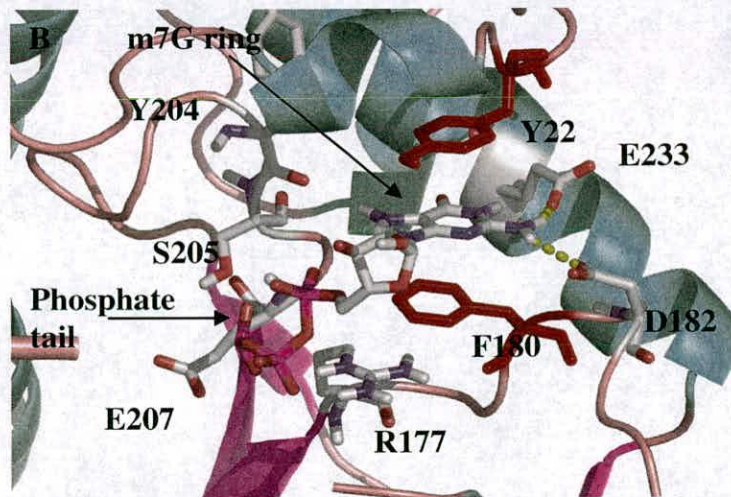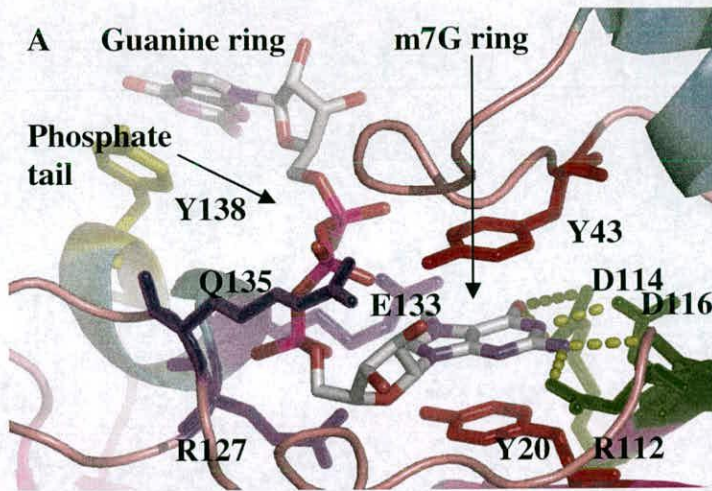
Figure 7: The figure displays the cap-binding site of VP39[27], CBC[26] and eIF4E and the common features used in recognition of the cap-analogue. The most distinctive feature is the intercalation of the cap-analogue between two aromatic residues (shown in red). (**A**) shows the cap-binding site in the CBP20 subunit of CBC. The dinucloetide (m7GpppG) intercalates itself between Y20 and Y43. D116, the carbonyl of W115 and D114 form hydrogen bonds with the m7G ring. The triphosphate section is recognised by R127, R135 and Q133. Y138 stacks against the guanine ring.

(**B**) shows the cap-binding site of VP39. Again the m7G ring stacks between two aromatic residues: Y22 and F180. The N2 on the m7G ring is again recognised by acidic side-chains: D182 and E233. The carbonyl on Y204 forms a contact with the methyl group. The triphosphate section is recognised by the amide of E207 and the O of the S205 side-chain. (**C**) The m7G moiety intercalates itself between the W102 and W56 side chains, the N7 methyl group forms a van der Waals contact with W166 and three hydrogen bonds are formed between the m7G ring and the E103 side chain and the backbone amide of W102. The triphosphate tail is recognised by R112, R157, K159 and K162 and via several water-mediated contacts (see figure 6).

The almost perfect alignment of the three aromatic rings, a nearly optimal inter-planar distance (between 3.2 and 3.6 angstrom), substantial areas of overlap in the stacked rings (see figure 8) and the delocalisation of positive charge arising from the N7-methylation of the guanine ring all contribute to strong interactions between the π-electrons of the stacked aromatic rings[33]. The strength of these specific interactions explains the low affinity of cap-binding proteins for non-methylated analogues (>100 fold difference in affinity compared with N7 methylated ones)[17].

Hydrogen bond formation occurs between N1 and/or N2 of m7G and the carboxylate groups of Glu or Asp in the cavity. This cavity contains at least one critical acidic amino acid – E103 in eIF4E, D116 in CBC[26] and E233 in VP39[27] (figures 7). Mutations of these residues to Ala decrease affinity as well as specificity[26,27,34]. There are differing results in the literature of the VP39 phenotype resulting from the mutation of E233 to Q[35]. Additional contacts are made between O6, N1 and N2 of m7G (figure 7) and other amino acids. W102 is involved in these additional contacts in the case of eIF4E (figure 7), although this residue is also involved in the stacking with m7G. D114, R112 and W115 are involved in these contacts in the case of CBC[26] (figure 7). D182 and water molecules make contacts with m7G in the case of VP39[27] (figure 7).

The triphosphate moiety is bound in the cavity by residues forming salt bridges with the oxygen atoms of the phosphate groups. R157, K162, K159, R112 and water molecules form a binding cavity in the case of eIF4E (figure 6 and 7). In CBC[26], R135, R127, Y20, Q133 and V134 are involved (figure 7), while in VP39[27], R177,

E207 and S205 form the cavity (figure 7). The N7 methyl group of the m7G moiety is stabilised by van der Waals or weak polar interactions with W166 in the case of eIF4E and Y204 in the case of VP39[27]. In CBC[26] there is no interaction with the N7-methyl group. Interactions with base 1 of the cap structure also contribute to binding, although non-specifically, to accommodate any base at that position. Base 1 (guanine, in CBC[26] and adenine in the eIF4E structure) is stacked with Y138 in CBC[26] (figure 7) or is stabilised by hydrogen bonds and electrostatic interactions with residues T205, S207 and K206 in eIF4E (figure 5)[18,26].
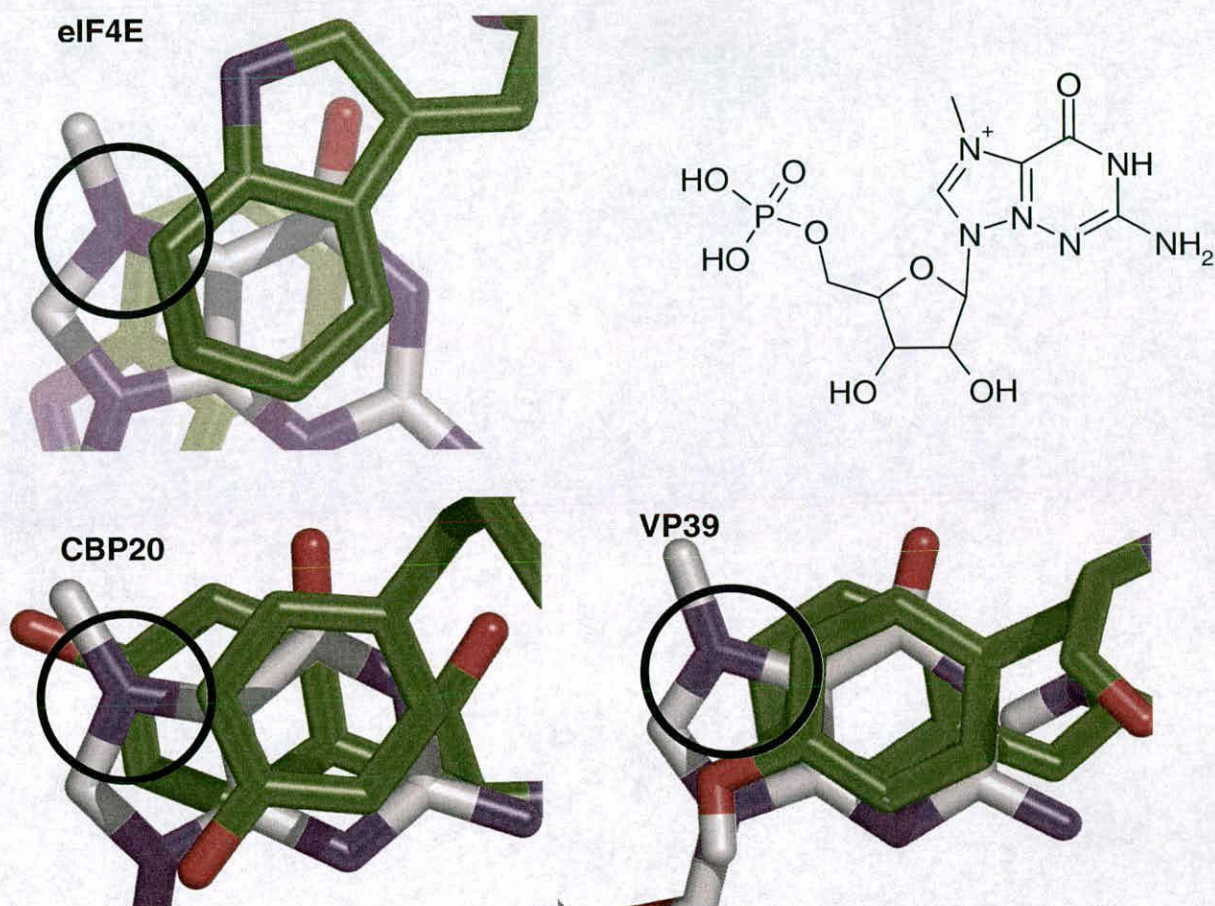
Figure 8: The figure shows intercalation of the m7G moiety between the aromatic residues of VP39, CBC and eIF4E. The structure of m7GMP, top right-hand corner, shows the position of the delocalised charge. If the charge is localised around the N7 atom[36] (m7G moiety shown in blue) it can be seen in all three cap binding proteins eIF4E, VP39 and CBP20 that it is off centre in respect to both aromatic residues involved in the cation-pi interaction. A situation similar to the optimal stacking of heterocyclic ring systems via π-π interactions. Also the aromatic side-chains principally overlap with the six-membered ring section of the m7G moiety.

3.3.5 4E-BP1 peptide: the molecular mimic

Source	PDB Details	Target	RMS backbone (37-200)
Chain A	Full length human eIF4E/m7GTP/4E-BP1 motif peptide	Chain B	0.72
1EJ1	N-terminal truncated murine eIF4E/m7GDP	Chain B	0.80
1IPB	Full human eIF4E/m7GTP	Chain B	0.74
1IPC	Full human eIF4E/m7GpppA	Chain B	0.72
1RF8	Solution structure of the yeast translation Initiation factor eIF4E In complex with m7GDP and eIF4GI residues 393 To 490	Chain B	0.73
1EJH: chain A	N-terminal truncated murine eIF4E/eIF4GII peptide/m7GDP	Chain B	0.91
1EJH: chain B	N-terminal truncated murine eIF4E/eIF4GII peptide/m7GDP	Chain B	0.63
1EJH: chain C	N-terminal truncated murine eIF4E/eIF4GII peptide/m7GDP	Chain B	0.67
1EJH: chain D	N-terminal truncated murine eIF4E/eIF4GII peptide/m7GDP	Chain B	0.54

Table 2: RMS values for backbone superimposition of available eIF4E structures in the PDB for residues 37-200 showing a high overall similarity for the structures. All structures were aligned on to chain B of our refined eIF4E structure. The N-terminal truncated murine 4E-BP1 peptide structure was not included.

The biological function of eIF4E is carried out when bound to eIF4G which then associates to the ribosomal apparatus and leads to mRNA translation. This process is inhibited by 4E-BP1 competing with eIF4G for binding to the dorsal side of eIF4E. The crystal structures of mouse eIF4E complexed with eIFGII (1EJH)[7] and 4E-BP1 motif peptides (1EJ4)[7] demonstrate that His 37, Val 69 and Trp 73 play an important role in binding eIFGII and 4E-BP1 [7]. The 4E-BP1 peptide containing the binding motif as shown in both the human full length/4E-BP1/m7GTP structure and

80

1EJ4 mimic the structure of the eIF4GII peptide by forming the same alpha helix with an extended strand conformation (figure. 9). The RMS differences (table 2) between the cap binding sites of peptide bound and peptide free structures are small and clearly show that peptide binding has no effect on the final configuration of the binding site. The shift seen in the position of K159 in our structures compared to others is not observed in the other peptide structures in the PDB[7] and is most likely an artefact of the crystallisation conditions.

4E-BP1 motif peptide   **eIF4GII motif peptide**
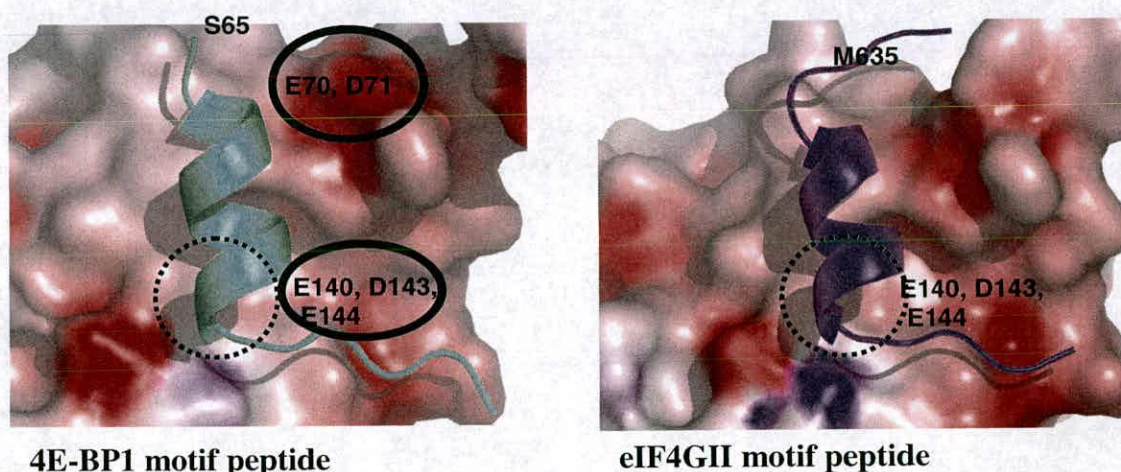
Figure 9: Picture demonstrating the adoption of the L-shaped conformer by the 4E-BP1 peptide that is used to imitate the eIF4GII peptide. Also indicated are areas of negative charge that presumably repel phosphorylated region of the 4E-BP1 protein (Bold circles). Dashed circle indicates position of hydrophobic residue in respect to the hinge area of the 'L' conformer. S65[37] is a residue (not present in the peptide used for crystallisation) that is phosphorylated on 4E-BP1 and is electrostatically repulsed by the eIF4E residues E70 and D71, and has a role in abrogating 4E-BP binding to eIF4E and relieving inhibition of cap-dependent translation.

```
                              54                60
Consensus SEQ                 Y-X-X-X-X-L-•

4E-BP1                        R-I-I-Y-D-R-K-F-L-M-E-C-R-N
M 4E-BP1                      R-I-I-Y-D-R-K-F-L-M-E-C-R-N-S-P-V
M eIFGII                      K-K-Q-Y-D-R-E-F-L-L-D-F-Q-F-M-P-A

                              624               630
```

Figure 10: Alignment of 4E-BP1 and eIF4GII peptides (used in crystallisation of eIF4E) and the binding consensus motif. Residues highlighted in red are strictly conserved while those in blue must be hydrophobic. M indicates eIF4GII and 4E-BP1 peptides used in other complex structures in the PDB[7].

3.3.6 Structural analysis of 4E-BP1 peptide binding to eIF4E.

In contrast to the previous published structure (1EJ4)[7] that contains only one ternary complex in the asymmetric unit, the X-ray structure of the human full length/4E-BP1/m7GTP contains two complexes within the asymmetric unit. Electron density is also clearly seen for all the peptide residues in both units.
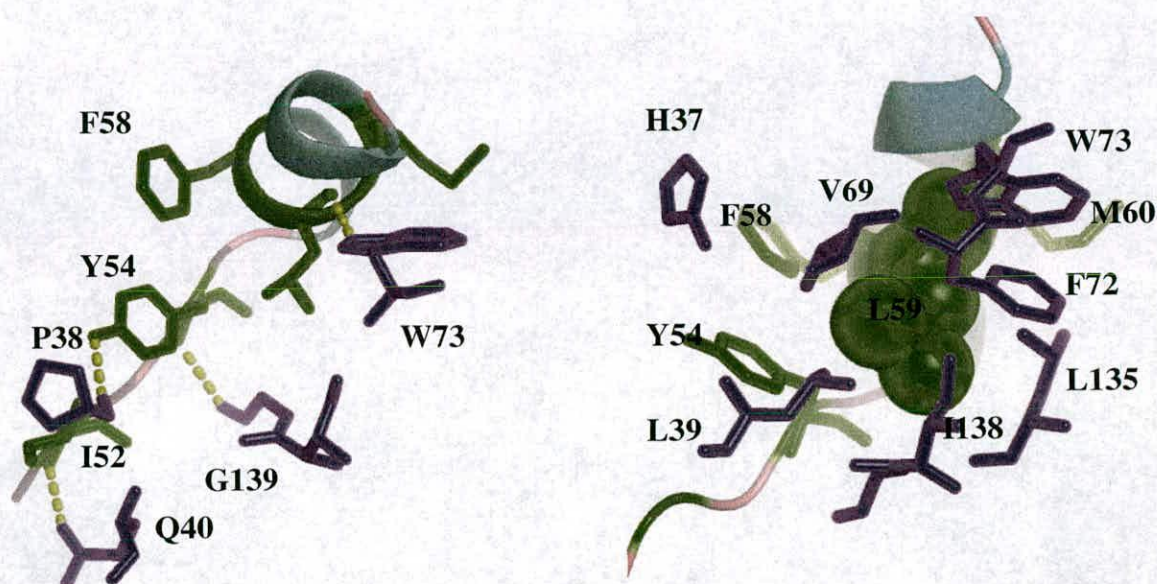
Figure 11: **(A)** Hydrogen bond interactions conserved in the crystal structures between eIF4E and the 4E-BP1 motif peptide are shown in yellow[7]. These are Y54 to G139 and P38 respectively, L59 to W73 and I52 to Q40. Peptide residues are coloured green whilst eIF4E residues are coloured blue. Note the 'T' shaped interaction between H37 and F58. The hydrogen bonds are predominately between the two strictly conserved residues of the peptide chain and eIF4E. **(B)** Diagram showing the hydrophobic cluster of residues at the protein/4E-BP1 peptide interface with eIF4E's P38, V69, W73, F72 I138 and L135 seemingly forming a cage around the L59 of the peptide. Peptide residues are coloured in green and eIF4E residues are coloured blue. L59 is shown with its Van der Waals radii as spheres.

Four hydrogen bonds are observed between eIF4E and the 4E-BP1 peptide that are found both in the full length human eIF4E complex and the N-terminal truncated murine complex (1EJ4)[7]. These hydrogen bonds are the OH group of Y54 to the carbonyl backbone of main chain P38 (2.45Å, figure 11A), the amide nitrogen of I52 to the main chain OE1 of Q40 (2.92 Å, figure 11A), L59 carbonyl with the NE1 of the main chain W73 (2.99Å, figure 11A) and finally the amide of Y54 to the main

82

chain carbonyl of G139 (3.12 Å, figure 11A). Y54 and L59 are strictly conserved in the binding motif of 4E-BP1 and eIFGII for eIF4E (figure 10).

When Tyr 624 is replaced with a Phe, in human eIF4GI, there is only a modest affect on binding [38]. If the tyrosine is then mutated to Ala in yeast eIF4G then it induces a lethal phenotype [39]. If Leu 629 and the hydrophobic residue at position 630 are also converted to Ala, in yeast eIF4G, then temperature sensitive phenotypes result [39]. This evidence shows the importance of these residues, and the interactions they are involved in, in the binding of the consensus motif peptide to eIF4E (figure 10). Apart from the direct hydrogen bond interactions there are two common hydrogen-bonded water interactions: the carbonyl oxygen of I52 bridged to the amide of eIF4E's Q40 and the carbonyl of eIF4E's I138. 1EJH contains 4 complexes in the asymmetric unit [7], and the same arrangement of hydrogen bond interactions are also seen most frequently (see figure 12A). The only minor differences observed are that the amide of K622 instead of I52 interacts with the OE1 of Q40 in eIF4E and that the water bridged hydrogen bonds are also forming with K622 instead of the I52.

The two conserved residues in the general peptide binding motif Y54 and L59 as well as being involved in hydrogen bond interactions are also excluded from the solvent substantially more than other peptide residues. L59 is surrounded by a cage of amino acid side chains essentially hydrophobic in character (figure 11B). The eIF4E residues involved are V69, G139, W73, I138 and L135 and are almost exclusively buried away from solvent. The peptide residues M60 and Y54 are also part of this hydrophobic cluster and complete the cage. This packing of hydrophobic resides together at the protein/peptide interface explains why the amino acid at

position 60 in the peptide must be hydrophobic (figure 11B). The hydrophobic cluster is also observed in the eIF4GII peptide/eIF4E interface and indicates its importance in the association of the peptide (figure 12B).
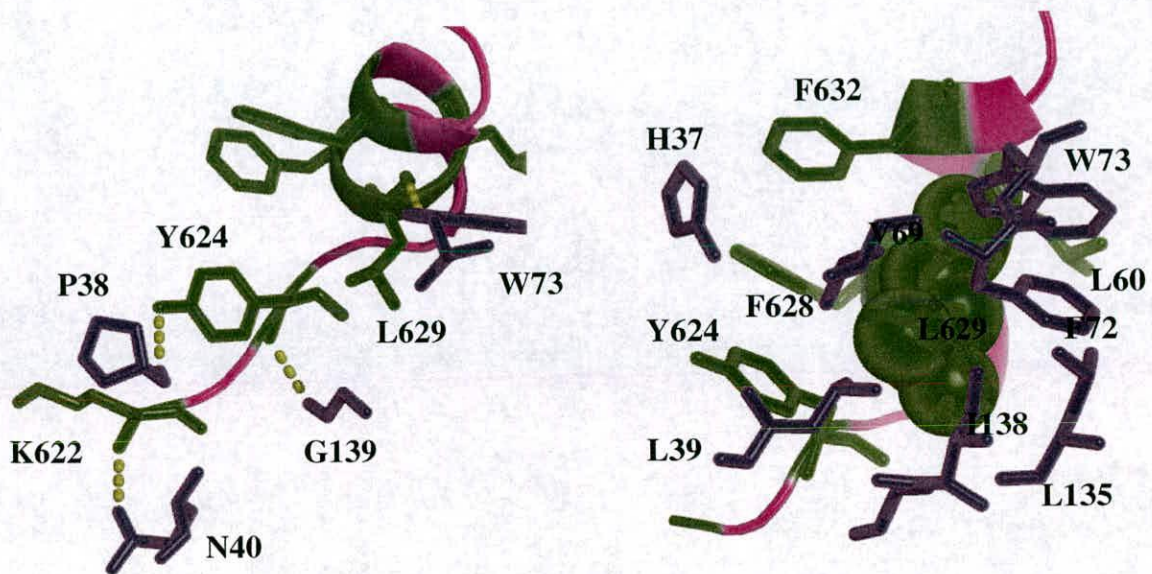
Figure 12: **(A)** Hydrogen bond interactions conserved in the crystal structure between eIF4E and the eIF4GII motif peptide[7] that are analogous to those in 4E-BP1 are shown in yellow. These are Y54 to G139 and P38 respectively, L629 to W73 and K622 to Q40. Peptide residues are coloured green whilst eIF4E residues are coloured blue. Note how the H37 and F628 and F632 residues form a trimeric ring arrangement. The hydrogen bonds are predominately between the two strictly conserved residues of the peptide chain and eIF4E. **(B)** Diagram showing the hydrophobic cluster of residues at the protein/eIF4GII peptide interface with eIF4E's P38, V69, W73, I138 and L135 seemingly forming a cage around the L629 of the peptide. Peptide residues are coloured in green and eIF4E residues are colured blue. L629 is shown with its Van der Waals radii as spheres.

If W73 in eIF4E is mutated to an Ala the interaction with 4E-BP1 or eIF4G1 is greatly reduced [40]. Other residues located on eIF4E do not have such a marked effect and this highlights the importance of W73 in the hydrophobic cluster. H37 appears to form a 'T' shaped interaction with eIF4E's F58 (figure 11A.) However in 1EJH[7] the residues F628 and F632 are seen together to be involved in forming hydrophobic interactions against H37 (figure 12A).

The three arginines found in the peptide are all involved in binding eIF4E. R51 interacts electro-statically with D143, D144 and D147, whilst R56 forms a salt bridge to E132. R63 seems to lay flat across the W73 and then projects away from eIF4E. However this data does not include the position of phospho-acceptor sites in 4E-BP1 relative to eIF4E. NMR data from Gross et al[9] have shown that eIF4GI (393-490) and the 4E-BP1 binding site on yeast eIF4E partially overlap in two regions containing conserved electrostatic residues: Glu 72, 73 and Glu 140/Asp 143/Glu 144 (see figure 9)[9]. These sites presumably repel phosphorylated 4E-BP1 due to electrostatic repulsion. In 1EJH[7] the peptide residue R626 interacts electro-statically with E132, as seen for the 4E-BP1 peptide. Q633 takes the place of R63 in the eIFG4II motif and carries out similar polar interactions against N77. Again another arginine residue is replaced with a conservative glutamine at position 633 in the eIF4GII peptide and interacts with residues E140 and D143 as in the eIF4E/4E-BP1/m7GTP structures[7].

3.3.7 4E-BP1 Peptide is independent of N-terminal stabilisation.

The N-terminus of eIF4E is absent from electron density in the human full length/4E-BP1/m7GTP crystal structure as well as all the other full length structures present in the PDB [18]. The peptide it would seem plays no role in stabilising or ordering the flexible N-terminus of eIF4E and that the N-terminus is highly mobile. Experiments have shown that $\Delta 35$ and $\Delta 30$ constructs of yeast eIF4E binds yeast eIF4G with a much lower affinity than the $\Delta 20$ construct and imply that the flexible N-terminus is important in binding eIF4G [9]. Our structure clearly shows that peptide formation is independent of N-terminal stabilisation and could initiate binding between 4E-BP1 and eIF4E. On the basis that yeast eIF4E has a high sequence similarity to human eIF4E and that the 4E-BP1 and the eIF4GI motif peptides are identical to their yeast counterparts the following conclusion can be drawn; that the structural evidence lends support to the hypothesis by Gross et al [9] that the binding of yeast eIF4GI to yeast eIF4E is initiated by a collision between eIF4E and the peptide motif where independent formation of the 'L' conformer occurs (see figure 9). This model then proposes that the N-terminus then interacts with eIF4G in a second step to complete the association process.

3.4 Surface plasmon resonance studies on the 4E-BP1 peptide interaction with eIF4E.

In this next section the kinetics of the 4E-BP1 peptide interaction with eIF4E using a BIAcore 2000 machine were studied. Surface plasmon resonance can be used to investigate the specificity of bimolecular interactions, the kinetics and affinity of an interaction and the concentration of specific molecules present in the sample. The section will begin with a brief description of the fundamentals of the surface plasmon resonance technique before an analysis of how 4E-BP1 interacts with eIF4E and what insights it may afford into inhibitor design.

3.4.1 Surface plasmon resonance basics.

Surface plasmon resonance can be used to monitor the interaction between molecules in real time. The technique involves attaching one interacting partner to the surface of a sensor chip, and then passing sample containing other interaction partners(s) over the surface. Binding of molecules to the surface produces a response which is proportional to the bound mass [41-43].

Surface plasmon resonance is a phenomenon that occurs in thin conducting films at an interface between media of different refractive index. The BIAcore 2000 system was used in the following experiments and the media used here are the glass of the sensor chip, the sample solution and the conducting film, which is a thin layer of gold on the sensor chip surface[41-43].

Under conditions of total internal reflection, light incident on the reflecting interface leaks an electric field intensity called an evanescent wave field across the interface into the medium of lower refractive index, without actually losing net energy. The amplitude of the evanescent field decreases exponentially with distance from the surface, and the effective penetration depth in terms of sensitivity to refractive index is about 20% of the wavelength of the incident light[41-43].

At a certain combination of angle of incidence and energy (wavelength), the incident light excites plasmons (electron charge density waves) in the gold film. As a result, a characteristic absorption of energy via the evanescent wave field occurs and SPR is seen as a drop in the intensity of the reflected light[41-43].

Due to penetration of the solution by the evanescent waves, conditions for this resonance effect are very sensitive to the refractive index of the solution within the effective penetration depth of the evanescent field. Changes in the solute concentration at the surface of the sensor chip causes changes in the refractive index of the solution which can be measured as changes in the SPR conditions. The penetration depth of the evanescent wave determines the thickness of the solution layer where refractives index changes are monitored: only refractive index changes close to the surface affect the surface plasmon resonance signal[41-43].

3.4.2 4E-BP1 peptide does not bind in a 1:1 Langmuir binding model.
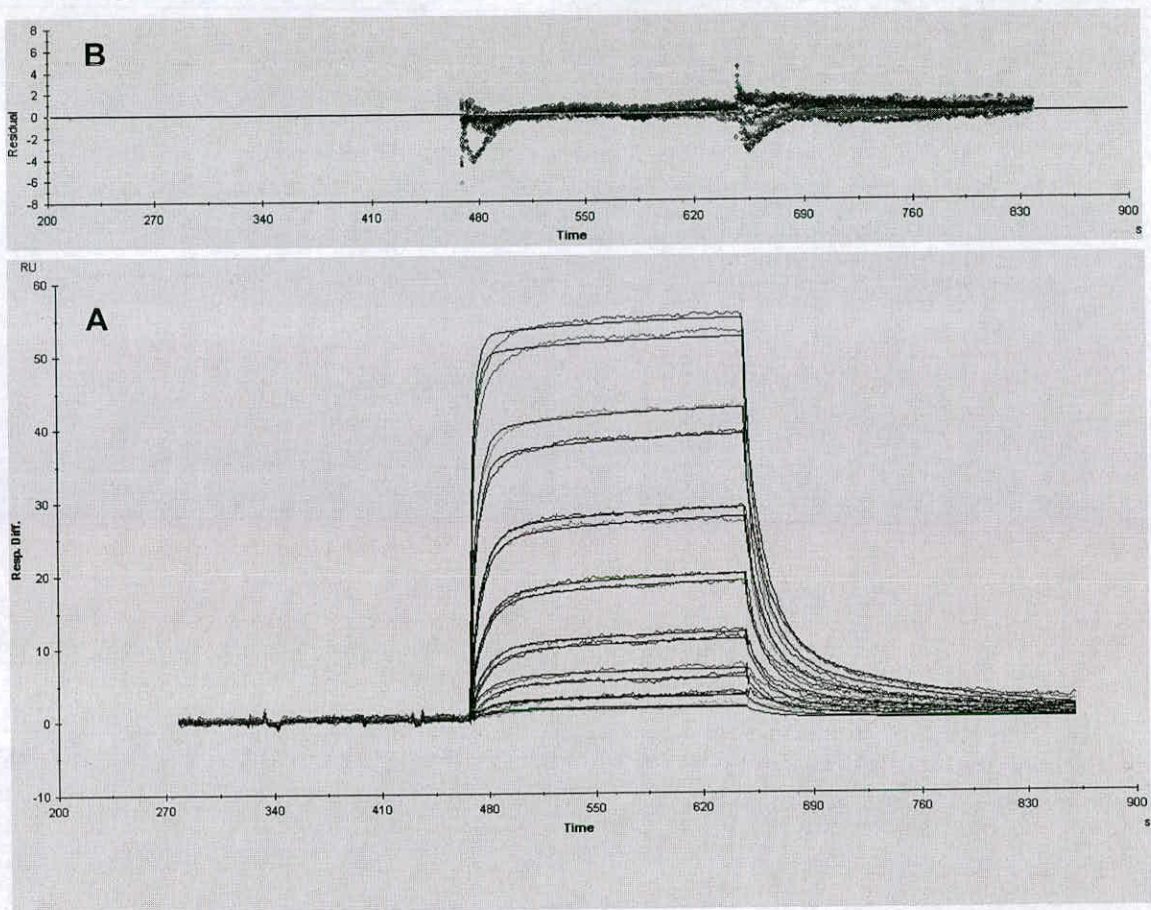
Figure 13 (A) The graph shows the fitting of the two step binding model via a conformational change to the BIAcore sensograms ([A] + [B] <> [A][B]* <> [A][B]). **(B)** The figure shows the residuals of the binding model's fit to the sensograms. The overall chi2 of the fit was 0.495.

BIAcore studies on the biotinylated version of the peptide used in the crystallisation of the ternary complex were carried out (biotin-β-Ala-β-Ala-RIIYDRKFLMECRN). eIF4E, at differing concentrations was injected over an SA chip (BIAcore, streptavidin chip) with the 4E-BP1 peptide immobilised on it. When the experiment was conducted with no m7GTP present there was a discrepancy in the results caused by degradation of eIF4E. This was overcome by having m7GTP present in a 5:1 ratio to protein. The surface plasmon resonance results from this experiment were fitted to a two step binding model using the BIAevaluation software (BIAcore) (figure 13A). The model

describes the situation where the analyte [A] binds the ligand [B] and forms a complex [A][B] that then undergoes a conformational change to [A][B]*. However the model makes the assumption that [A][B]* cannot dissociate directly to [A] + [B]. The two step binding model gives an overall of K_d of 700nM. This compares to values of 105nM for m7GTP saturated eIF4E and 94nM for apo eIF4E derived by steady state fluorescence [17]. The model fitted reasonably well with a chi2 value of 0.495 (fig 13B). The BIAcore results demonstrate a curve that does not fit 1:1 binding but follows a two step binding mode more closely. Such a model is supported by the fact that the binding peptide motifs (4E-BP1 and eIF4GII) are unstructured in solution[7][8] and undergo a conformational change when binding to eIF4E to form the 'L' conformer (figure 9). This may imply that the Y53 and L59 may bind first via specific hydrogen bond formation and surface area burial creating the turn between the extended strand and the alpha helix seen in the structure (figure 11A), and then recruiting M60 to complete the hydrophobic cage (figure 11B) which sets the stage for the other interactions to occur resulting in formation of the 'L' conformer (figure 9).

The experimental BIAcore setup used also provides a means to assay small molecule inhibitors of the protein interaction between eIF4E and its partner proteins in a competitive manner. The cluster of hydrophobic residues that occur at the eIF4E/peptide interface could be used as a template for the design of small molecular drug-like compounds.

This hypothesis could be further investigated and tested by using shorter peptides and looking for changes in binding affinity or mechanism of interaction. Crystallisation

studies could also be carried out side by side. Hopefully a minimal motif for binding could be structurally delineated and used as a refined hypothesis for structure-based drug design. An example of small molecules that target a protein-protein interaction are the nutlins which disrupt HDM2 binding to the tumour suppressor P53[44,45]

3.5 Conclusions.

The cap is recognised by a common fold that is easily seen in the many structures available of eIF4E and the crystal structures of VP39[27] and CBC[26]. The m7G moiety forms π-cation interactions with two aromatic residues, the m7G ring forms several hydrogen bonds usually with an acidic residue and the triphosphate tail is recognised by a set of positively charged side chains. Despite these common structural characteristics for the recognition of the cap-analogue none of these proteins share high sequence similarity or a common global fold. The 4E-BP1 peptide undergoes a disorder to order transition upon binding to eIF4E[46]. At the interface of the peptide and protein a hydrophobic cluster of residues is formed that appears critical to binding. Biacore studies suggest that this interaction occurs via a two step mechanism, which would be expected as the peptide orders itself upon binding eIF4E. Upon binding 4E-BP1 as in the case of the eIF4GII peptide forms an α-helix at the same site of interaction. 4E-BP1 therefore inhibits eIF4GII binding by molecular mimicry. Both these sites as discussed in chapter 1 present themselves as targets for structure-based drug design.

Reference List

1. Sonenberg,N. & Dever,T.E. Eukaryotic translation initiation factors and regulators. *Curr. Opin. Struct. Biol. 2003. Feb. ;13(1):56. -63.* **13**, 56-63 (2003).

2. Li,Q. *et al.* Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol. Cell Biol.* **19**, 7336-7346 (1999).

3. Kahvejian,A., Roy,G. & Sonenberg,N. The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb. Symp. Quant. Biol. 2001. ;66. :293. -300.* **66**, 293-300 (2001).

4. Haghighat,A., Mader,S., Pause,A. & Sonenberg,N. Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. *EMBO J.* **14**, 5701-5709 (1995).

5. Ptushkina,M. *et al.* Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5' cap in yeast involves a site partially shared by p20. *EMBO J.* **17**, 4798-4808 (1998).

6. Altmann,M., Schmitz,N., Berset,C. & Trachsel,H. A novel inhibitor of cap-dependent translation initiation in yeast: p20 competes with eIF4G for binding to eIF4E. *EMBO J.* **16**, 1114-1121 (1997).

7. Marcotrigiano,J., Gingras,A.C., Sonenberg,N. & Burley,S.K. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell* **3**, 707-716 (1999).

8. Shen,X., Tomoo,K., Uchiyama,S., Kobayashi,Y. & Ishida,T. Structural and thermodynamic behavior of eukaryotic initiation factor 4E in supramolecular formation with 4E-binding protein 1 and mRNA cap analogue, studied by spectroscopic methods. *Chem. Pharm. Bull. (Tokyo) 2001. Oct. ;49. (10):1299. -303.* **49**, 1299-1303 (2001).

9. Gross,J.D. *et al.* Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell 2003. Dec. 12. ;115. (6.):739. -50.* **115**, 739-750 (2003).

10. Nave,C. & Garman,E.F. Towards an understanding of radiation damage in cryocooled macromolecular crystals. *J. Synchrotron. Radiat. 2005. May. ;12. (Pt. 3):257. -60. Epub. 2005. Apr 14.* **12**, 257-260 (2005).

11. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760-763 (1994).

12. Murshudov,G.N., Vagin,A.A. & Dodson,E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240-255 (1997).

13. Perrakis,A., Sixma,T.K., Wilson,K.S. & Lamzin,V.S. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 448-455 (1997).

14. van Aalten,D.M. *et al.* PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des* **10**, 255-262 (1996).

15. Emsley,P. & Cowtan,K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr. 2004. Dec. ;60. (Pt. 12. Pt. 1):2126. -32. Epub. 2004. Nov. 26.* **60**, 2126-2132 (2004).

16. Marcotrigiano,J., Gingras,A.C., Sonenberg,N. & Burley,S.K. Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951-961 (1997).

17. Niedzwiecka,A. *et al.* Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol. 2002. Jun. 7. ;319. (3):615. -35.* **319**, 615-635 (2002).

18. Tomoo,K. *et al.* Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m(7)GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region. *Biochem. J. 2002. Mar. 15. ;362. (Pt. 3):539. -44.* **362**, 539-544 (2002).

19. Matsuo,H. *et al.* Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nat. Struct. Biol.* **4**, 717-724 (1997).

20. Gallivan,J.P. & Dougherty,D.A. Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. U. S. A* **96**, 9459-9464 (1999).

21. Zacharias,N. & Dougherty,D.A. Cation-pi interactions in ligand recognition and catalysis. *Trends Pharmacol. Sci. 2002. Jun. ;23(6.):281. -7.* **23**, 281-287 (2002).

22. Lewis,J.D. & Izaurralde,E. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247**, 461-469 (1997).

23. Fechter,P. & Brownlee,G.G. Recognition of mRNA cap structures by viral and cellular proteins. *J. Gen. Virol. 2005. May. ;86. (Pt. 5):1239. -49.* **86**, 1239-1249 (2005).

24. Jankowska-Anyszka,M. *et al.* Multiple isoforms of eukaryotic protein synthesis initiation factor 4E in Caenorhabditis elegans can distinguish between mono- and trimethylated mRNA cap structures. *J. Biol. Chem.* **273**, 10538-10542 (1998).

25. Dinkova,T.D., Keiper,B.D., Korneeva,N.L., Aamodt,E.J. & Rhoads,R.E. Translation of a small subset of Caenorhabditis elegans mRNAs is dependent on a specific eukaryotic translation initiation factor 4E isoform. *Mol. Cell Biol. 2005. Jan. ;25. (1):100. -13.* **25**, 100-113 (2005).

26. Mazza,C., Segref,A., Mattaj,I.W. & Cusack,S. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J. 2002. Oct. 15. ;21(20.):5548. -57.* **21**, 5548-5557 (2002).

27. Hodel,A.E., Gershon,P.D., Shi,X., Wang,S.M. & Quiocho,F.A. Specific protein recognition of an mRNA cap through its alkylated base. *Nat. Struct. Biol.* **4**, 350-354 (1997).

28. Ishida,T. & Inoue,M. An X-ray evidence for the charge-transfer interaction between adenine and indole rings: crystal structure of 1,9-dimethyladenine-indole-3-acetic acid trihydrate complex. *Biochem. Biophys. Res. Commun.* **99**, 149-154 (1981).

29. Ishida,T., Doi,M. & Inoue,M. A selective recognition mode of a nucleic acid base by an aromatic amino acid: L-phenylalanine-7-methylguanosine 5'-monophosphate stacking interaction. *Nucleic Acids Res.* **16**, 6175-6190 (1988).

30. Darzynkiewicz,E. & Lonnberg,H. Base stacking of simple mRNA cap analogues. Association of 7,9-dimethylguanine, 7-methylguanosine and 7-methylguanosine 5'-monophosphate with indole and purine derivatives in aqueous solution. *Biophys. Chem.* **33**, 289-293 (1989).

31. Stolarski,R. *et al.* 1H-NMR studies on association of mRNA cap-analogues with tryptophan-containing peptides. *Biochim. Biophys. Acta* **1293**, 97-105 (1996).

32. Ueda,H., Iyo,H., Doi,M., Inoue,M. & Ishida,T. Cooperative stacking and hydrogen bond pairing interactions of fragment peptide in cap binding protein with mRNA cap structure. *Biochim. Biophys. Acta* **1075**, 181-186 (1991).

33. Hu,G., Gershon,P.D., Hodel,A.E. & Quiocho,F.A. mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl. Acad. Sci. U. S. A* **96**, 7149-7154 (1999).

34. Morino,S. *et al.* Analysis of the mRNA cap-binding ability of human eukaryotic initiation factor-4E by use of recombinant wild-type and mutant forms. *Eur. J. Biochem.* **239**, 597-601 (1996).

35. Hsu,P.C. *et al.* Structural requirements for the specific recognition of an m7G mRNA cap. *Biochemistry 2000. Nov. 14;39. (45.):13730. -6.* **39**, 13730-13736 (2000).

36. Ruszczynska,K. *et al.* Charge distribution in 7-methylguanine regarding cation-pi interaction with protein factor eIF4E. *Biophys. J. 2003. Sep. ;85. (3):1450. -6.* **85**, 1450-1456 (2003).

37. Gingras,A.C. *et al.* Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes Dev. 2001. Nov. 1;15. (21):2852. -64.* **15**, 2852-2864 (2001).

38. Mader,S., Lee,H., Pause,A. & Sonenberg,N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol. Cell Biol.* **15**, 4990-4997 (1995).

39. Tarun,S.Z., Jr. & Sachs,A.B. Binding of eukaryotic translation initiation factor 4E (eIF4E) to eIF4G represses translation of uncapped mRNA. *Mol. Cell Biol.* **17**, 6876-6886 (1997).

40. Pyronnet,S. *et al.* Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *EMBO J.* **18**, 270-279 (1999).

41. Katsamba,P.S., Park,S. & Laird-Offringa,I.A. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods 2002. Feb. ;26. (2):95. -104.* **26**, 95-104 (2002).

42. von der,H.T., Hughes,J.M., Manjarul,K.M., Ptushkina,M. & McCarthy,J.E. Translation initiation and surface plasmon resonance: new technology applied to old questions. *Biochem. Soc. Trans. 2002. Apr;30. (2):155. -62.* **30**, 155-162 (2002).

43. Karlsson,R. SPR for molecular interaction analysis: a review of emerging application areas. *J. Mol. Recognit. 2004. May. -Jun. ;17(3):151. -61.* **17**, 151-161 (2004).

44. Kojima,K. *et al.* MDM2 antagonists induce p53-dependent apoptosis in AML: implications for leukemia therapy. *Blood 2005. Jul. 12. ;. .,* (2005).

45. Grasberger,B.L. *et al.* Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *J. Med. Chem. 2005. Feb. 24. ;48. (4):909. -12.* **48**, 909-912 (2005).

46. Fletcher,C.M. & Wagner,G. The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein. *Protein Sci.* **7**, 1639-1642 (1998).

Chapter 4: Probing the interactions of eIF4E cap analogue binding by mass spectrometry.

4.0 Introduction.

The aim of the following chapter is to characterise the cap-analogue interactions of the cap-binding site using mass spectrometry. The results of the experiments were compared and contrasted to solution based measurements in the literature to check the validity of gas-phase measurements and as a prelude to using mass spectrometry as a screening technique for eIF4E.

4.1 Electrospray ionisation mass spectrometry for the study of non-covalent complexes.

Weak non-covalent interactions play a fundamental role in the biological world. They regulate the interactions between substrate and enzymes, ligand and protein, protein and protein, and antibody and antigen. To successfully carry out mass spectrometry studies on non-covalent complexes the following criteria must be satisfied:

- The ionisation method must be capable of generating intact gas-phase protein ions from solutions that contain biological buffers.

- The internal energy transferred to the macromolecule during the ionisation process must be low enough to prevent dissociation of the complex.

- The MS instrumentation must be of sufficient mass range to observe the ionised complex.

Two ionisation techniques have been developed that fit the above criteria, electrospray ionisation (ESI) and matrix assisted laser desorption/ionisation (MALDI). ESI-MS only will be discussed as this was the technique used to carry out the mass spectrometry work in this chapter.

4.1.1 Electrospray ionisation (ESI).

ESI is an example of an atmospheric ionisation source. These types of sources ionise the sample at atmospheric pressure and then transfer the ions in to the mass spectrometer. An Electrospray is produced by applying a strong electric field, under atmospheric pressure, to a liquid passing through a capillary tube with a weak flux (normally 1-10ul min^{-1}) (see figure 1). The electric field is obtained by applying a potential difference of 3-6 kV between this capillary and the counter-electrode, separated by 0.3 to 0.2 cm, producing electric fields of the order of 10^6 V m^{-1}. This field induces a charge accumulation at the liquid surface located at the end of the capillary, which will break to form highly charged droplets. A gas injected coaxially at a low flow rate allows dispersion of the spray to be limited in space. These droplets then pass through either a curtain of heated gas, usually nitrogen, or through a heated capillary to remove the last solvent molecules[1].
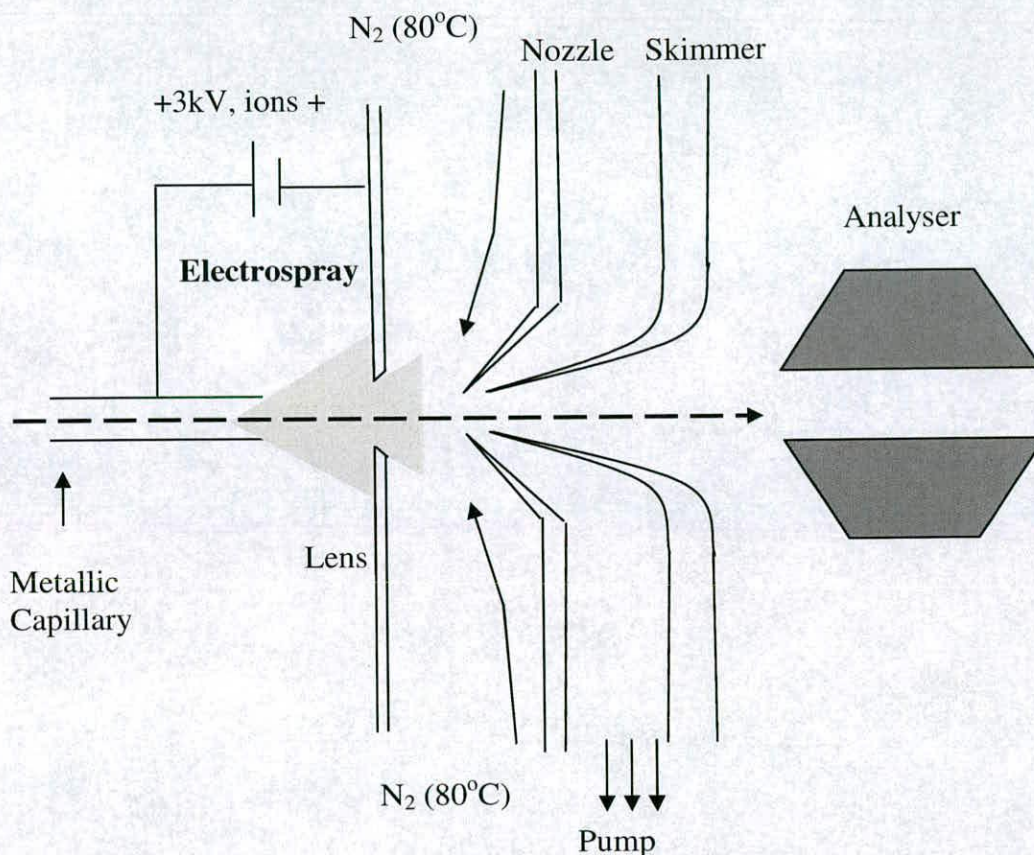
Figure 1: Diagram of ESI sorces, using skimmers for ion focalisation and a curtain of heated nitrogen gas for desolvation.

If one examines with a microscope the nascent drop forming at the tip of the capillary while increasing the voltage, at low voltages the drop appears spherical and then elongates under the pressure of the accumulated charges at the tip in the stronger electric field; when the surface tension is broken, the shape of the drop changes to a 'Taylor cone' and the spray appears (see figure 2)[1].

Figure 2: Effect of ESI potential on the drop at the tip of the capillary with increasing voltage. (left) At low voltage, the drop is almost spherical. (Centre) At about 1 or 2 kV but below the onset potential the drop elongates under the pressure of the charge accumulating at the tip. (Right) At onset voltage, the pressure is higher than the surface tension, the shape of the drop changes at once to a Taylor cone and small droplets are released. The droplets divide, producing spray.

The solvent contained in the droplets evaporates, which causes them to shrink to the point where the repelling coulombic forces come close to their cohesion forces, thereby causing them to break apart. Each rupture yields two charged droplets of similar size, but under the effect of the strong electrical field many smaller, highly charged droplets are produced too. When the electric field on the surface becomes large enough, desorption of ions from the surface occur. Charges in excess accumulate at the surface of the droplet. In the bulk, analytes as well as electrolytes (whose positive and negative charges are in equal number) are present. Desorption of charged molecules occurs from the surface, which means that sensitivity is higher for compounds whose concentration is higher at the surface. When mixtures of compounds are analysed, those present at the surface of the droplets can mask the presence of compounds that are more soluble in the bulk[1].

Electrospray is able to produce multiply charged ions from large molecules. The formation of ions is a result of the electrochemical process and accumulation of charge in the droplets. The Electrospray current is limited by the electrochemical process that occurs at the probe tip and is sensitive to concentration rather than to total amount of sample. This ionisation process is very soft and leaves the ion largely unfragmented; hence complexes involving non-covalent interactions can be studied. ESI experiments can be carried out in aqueous solution (near physiological pH) without the presence of organic solvent that tends to denature proteins[1,2].

Multiply charged ions for complexes such as protein-protein quaternary complexes exhibit relatively low charge at high m/z. The distribution of charge state is typically

low e.g., 3-5 charge states. For several tetrameric proteins, Light-Wahl et al.[3] noted that the average charge per subunit has the order: monomer > dimer > tetramer[3]. The amount of charging that a biomolecule exhibits in an ESI mass spectrum has been correlated to a global solution structure. It has been observed that disulphide containing proteins show less average charging than their disulphide reduced counterparts, presumably because either fewer charge sites are exposed or the Coulombic restraints restrict charging for a more compact structure[4-6]. The same rationale may apply to subunit protein complexes. The narrow charge distribution of a low charge state represents retention of the higher order structure of the native protein complex. If a protein is denatured a greater number of charge states with lower m/z are observed.

MS does not provide direct structural information but can provide important stoichiometric information, including systems where there may be heterogeneity in the stoichiometry of the complex. Specificity is a critical advantage of ESI-MS. Proteins affect each other and come together and recognise ligands via specific interactions, based on key structural and/or energetic features and non-specific aggregation. MS can monitor ligand binding by detecting the non-covalently bound ligand complexes. Smith and co-workers used this property to screen a 289-component library (based on benzene sulphonamides) against zinc loaded carbonic anhydrase[7]. The mass and relative abundance of a released ligand by a subsequent tandem MS experiment can provide information regarding its identity and relative binding constant[8].

4.1.2 Electrochemistry and electric fields as origins of multiply charged ions.

Charges of ions generated by ESI do not reflect the charge states of compounds in the sample but are the result of both charge accumulation in the droplets and charge modification by the electrochemical process at the probe tip (see figure 3). This is demonstrated clearly by Fenselau and coworkers[9]. They showed that negative ions of myoglobin can be observed at pH 3.0, whereas a calculation based on known pK values predicts that only one molecule per about 3500 would have one negative charge in the original solution. The results show the role of charge accumulation in droplets under the influence of the electric field on the formation of multiply charged ions. The "pumping out" of the negative charges can be performed only if, at the same time, the same number of positive charges is neutralised electrochemically at the probe tip. The negative ion spectrum of myoglobin at pH 3.0 has a better signal-to-noise ratio than the same spectrum at pH 10. This results from the fact that protons have a high electrochemical mobility, and is an indication of the importance of the reduction process, when negative ions are analysed at the probe tip[1].
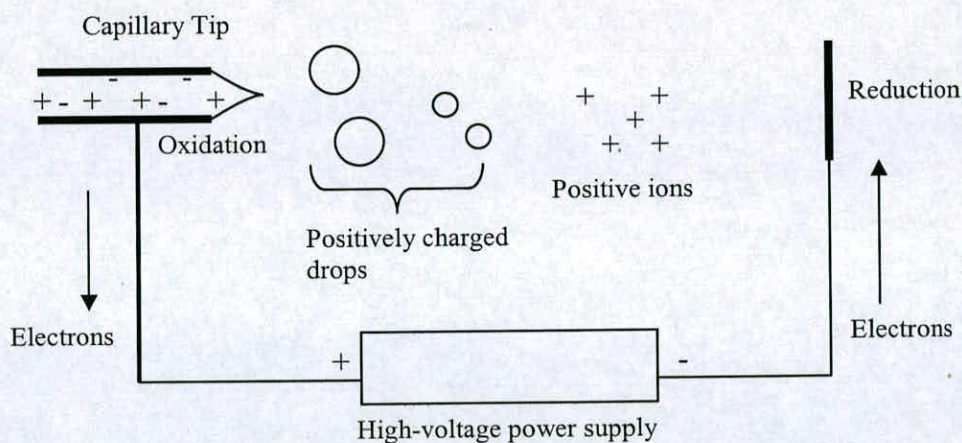
Figure 3: Schematic presentation of the electrochemical process in ESI (for positive ions).

As shown in figure 3, when positive ions are extracted for analysis electrons have to be provided in the circuit from the capillary. This will occur through oxidation of the species in the solution at the capillary tip, mainly ions having sufficient mobility. In other words, the same number of negative charges must be "pumped out" as positive charges are extracted to the analyser. ESI is an electrochemical process with its dependence on ion concentration and mobility as well as polarisation effects at the probe tip. For the detection of positive ions, electrons have to be provided by the solution and thus oxidation occurs. For negative ions, electrons have to be consumed and thus reduction occurs. A major consequence is that the total number of ions per unit time that can be extracted to the spectrometer is actually limited by the electric current produced by the oxidation or reduction process at the probe. This explains why ESI is concentration dependent, except at very low flow rates[1].

4.1.3 Can a gas-phase measurement reveal solution phase binding characteristics?

Gas phase studies need to be consistent with solution phase experiments in order for work done with ESI-MS to have some utility for the biochemist. Complex stoichiometry is easily provided and in most published reports the observed stoichiometry is consistent with the expected result[2]. Occasionally more ligand is observed to bind to the protein receptor perhaps due to non-specific gas phase aggregation[2]. Reducing the solution concentration of the analytes can decrease the prevalence of non-specific gas phase aggregation[10]. However numerous reports suggest that ESI-MS has some utility for studying non-covalent complexes[2].

The nature of the non-covalent interaction will have an effect on the mass spectrometry experiment. There are at least four types of non-covalent forces involved in protein folding and interactions: ionic interactions, hydrogen bonds, the hydrophobic effect and van der Waals forces. Going from solution to a solventless gas phase environment has many ramifications. Electrostatic interactions may be strengthened in vacuum. The binding of highly basic spermine to an acidic spermine-binding peptide containing 4 glutamic residues was studied in the gas-phase[2]. Despite a weak binding constant (10^4 M^{-1}) in solution, the spermine-SBP complex could be readily detected by ESI-MS. This complex is unusually stable in the gas-phase as demonstrated by collisionally activated dissociation (CAD) studies. The avidity of the gas-phase complex may or may not be similar to that found in solution. For the spermine-peptide complex, a solvent free environment may enhance Coulombic stabilisation of opposite charges.

Most gas-phase complexes are relatively fragile. ESI interface (atmosphere/vacuum) conditions need to be as gentle as possible to maintain the intact complex. Enough energy needs to be expended to reduce solvation of the complex for sensitive detection (see figure 4)[11,12].
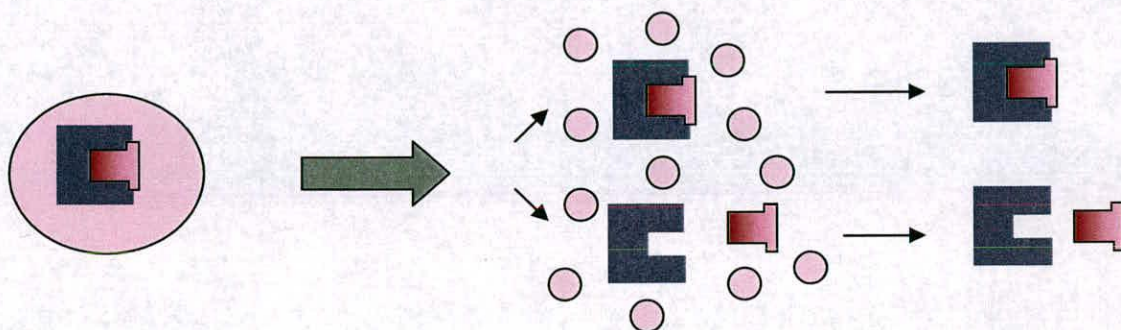
Figure 4: The process of solvent evaporation as the non-covalently bound complex is transferred from solution to the gas-phase. Gentle desolvation conditions are generally required to maintain the intact gas-phase complex. For several protein systems, the ESI-MS data may be consistent with the solution phase binding constants. Although some features of the solution structure may be preserved in the gas-phase ions, the stability of the gas-phase complex may not be reflected by the solution phase binding constant.

Depending on the interface, instrument and the complex there is a very fine line between sufficient desolvation of the gas-phase complex and dissociation of the complex. However there are examples at both ends of the scale. Protein complexes with oligonucleotides usually involve extensive electrostatic forces. Studies of protein-DNA complexes and protein-RNA complexes[2,11,13,14] are examples where dissociation of the gas-phase complex is very difficult: i.e. the multiply charged ions of the complex are stable at high interface energies. At the other extreme Robinson's work with acyl CoA analogues illustrates an example where the ESI-MS data do not agree with the solution phase data[15]. Acyl CoA analogues with different solution dissociation constants were not differentiated by the ESI-MS experiments. A

103

combination of hydrophobic, electrostatic and non-polar stacking interactions maintains the protein-ligand binding in this example. Changes in the length of the hydrocarbon acyl chain, although greatly affecting solution binding, did not appear to greatly affect the stability of the gas-phase complex. As noted by Robinson[15], "Where hydrophobic interactions have been shown to play a major role in the solution state of other systems, for example, the leucine zipper peptides and the receptor-ligand interactions, the fraction of the complexed species in the gas phase is low (typically ~ 10-20%) since such interactions are in large part attributed to the role of the solvent." Caution is needed in interpreting ESI-MS results but perhaps ESI-MS can be used to assess the type of bonding interaction that keeps complexes together.

It has yet to be conclusively proven that a gas-phase dissociation energy (or thermodynamics) can be used to predict or even to reflect the solution-phase binding strength e.g. studies on FKBP and rapamycin analogues in the gas-phase do not reflect solution based measurements[16]. ESI-MS studies on protein subunit complexes show gas-phase complexes that are not stable species in solution. However, recent binding studies have suggested that the stabilities of the gas-phase and solution-phase complexes are similar for selected systems e.g. studies on haem binding to variant forms of apomyoglobin and apocytochrome b_5[17].

4.1.4 Experimental consideration for ESI-MS experiments.

The solution conditions that are needed to maintain an intact complex are not optimal conditions for "normal" ESI operation. For maximum sensitivity, a solution pH of 2-

4 for positive ionisation and pH 8-10 for negative ion ESI are typical for polypeptide analysis. The addition of an organic modifier, such as methanol or acetonitrile, also enhances sensitivity and ion signal stability. These conditions are generally not tolerated when trying to maintain a non-covalent complex. Many protein complexes are denatured in solution at a pH value outside the pH range of 6-8 and the addition of a high concentration of an organic co-solvent can likewise disrupt the stability of the complex[2,11].

Quadrupole ESI mass analysers were the first systems used to detect non-covalent complexes; essentially all the major types of mass spectrometers, such as TOF ion traps etc have successfully been used. Analysers with high mass-to-charge range (greater than m/z 4000) have advantages for studying the full range of biochemical complexes because of the tendency of many non-covalent complexes to exhibit relatively low charge states. Every type of atmospheric pressure ionisation (API) interface has been utilised to observe such complexes (differentially pumped nozzle-skimmer interface, heated glass or metal capillary inlet with or without concurrent gas[2,11]).

Variables that affect droplet /ion desolvation and ion activation can affect the success of the experiment[2,11]. Solvent molecules need to be stripped from the ion prior to detection. Desolvation is achieved through a variety of methods, including the use of a counter-flow of nitrogen gas, heat (either a warm counter current gas, a heated capillary inlet or a warm interface chamber) and optionally a CAD downstream of the initial interface[2,11]. Some gas-phase complexes are extremely fragile and the

amount of activation (voltage difference between the capillary-skimmer or nozzle-skimmer lens elements) must be reduced to maintain the intact complex[2,11]. The gas-phase complex between acyl CoA binding protein and Acyl CoA derivatives was found to be sensitive to the ESI source temperature; increasing the temperature from 20-80°C reduced the proportion of complexed species to zero[15]. Best results were obtained by cooling the ESI nebuliser gas and analyte solution[15].

4.2 Results and discussion: Studying eIF4E cap analogue interactions using mass spectrometry.

4.2.1 Observation of the non-covalent complex.

The mass spectra of full-length human eIF4E were obtained under a variety of ESI conditions as shown in figure 5 and figure 6. The spectra from these different conditions are clearly distinct, indicating the protein retains a history of its solution environment in the gas phase. The spectrum obtained from the denaturing ESI-MS conditions, 50% methanol in water and 10mM ammonium acetate pH 3.0 and a source temperature of 60°C is shown in Figure 5A. These standard ESI-MS conditions usually denature the protein but allow accurate mass determination and confirm the molecular weight of the protein as ~ 24960 which differs from that anticipated from the amino acid sequence (25097.2). The difference between the calculated and observed mass is approximately 137. The mass difference probably corresponds to loss of the N-terminal methionine (131). The spectrum obtained from samples in 10mM ammonium acetate pH 5.0, figure 1B, exhibits a marked change in the charge state distribution from that observed in Figure 5A, with a shift of the charge state maximum from +25 to +11. This reduction in charge has been found to be a feature of spectra obtained from solutions in which proteins remain in a native-like conformation during introduction to the ESI source.

The spectrum obtained, under identical conditions to those of Figure 5B, but after addition of 3.7μM of m7GTP, is shown in Figure 5C. An additional series of peaks is observed, labelled with filled in circles, whose mass corresponds to the sum of m7GTP plus full-length eIF4E. The mass of the ligand derived from the spectra in

figure 5 is 539 and is within experimental error to the true value of 537.2. When this protein ligand complex was observed under denaturing ESI-MS conditions, ions corresponding to the mass of the complex were not seen.

eIF4E upon titration with m7GTP (figure 5B and 5C) exhibits no significant change in its charge species distribution, as might be expected with the addition of two negative charges. This lack of change in the eIF4E spectrum, upon binding negatively charge ligands, is seen for the monophosphate derivatives (see chapter 5, section 5.7.3), as well as m7GDP, GTP, GDP and GMP (not shown). Work by Samalikova et al[18] shows that by increasing the lysine content of RNase SA no major shift is seen in its spectrum. However they did show a shift of +1 in the spectrum of the peptide angiotensin upon substitution of an aspartate residue with an asparagine[18]. Other work by Wilson et al studying RNase A binding to 2'CMP and CTP also observed a similar lack of effect on the charge species distribution[19]. A plausible explanation for this is that the negative charge of the phosphate groups is being compensated for, in some manner, by the protein when bound, which may or may not result in altering the positive charge distribution on the protein. An alternative explanation is that the ligands are binding in their neutral form. However the pK_a of phosphate is very low and this would seem highly unlikely. Further evidence for this (results not shown) is that increasing the cone voltage does not generate a greater proportion of the apo peak in the complexed species indicating a strong electrostatic attraction is taking place.
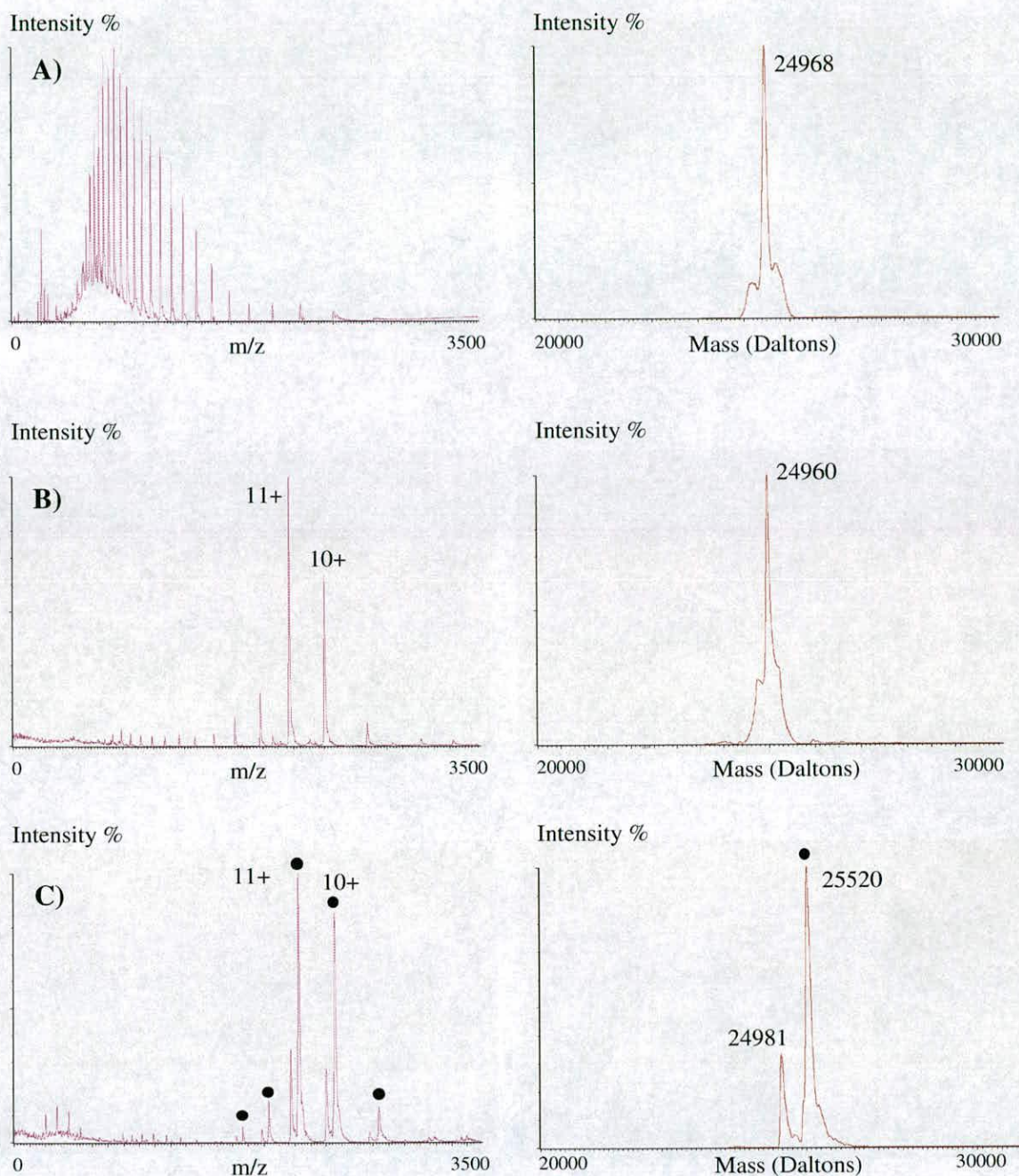
Figure 5: Positive ESI mass spectra of eIF4E under a variety of experimental conditions. The mass spectra on the left have been deconvoluted using a transform method in Mass Lynx to give the molecular masses on the right. The charge states and molecular masses marked with a filled in circle represent the eIF4E m7GTP gas-phase complex. The other peaks arise from free protein in solution. Experimental conditions: A) the running solvent and protein sample solution contained 50% methanol, 50% water and 10mM ammonium acetate pH 3.0 and the source temperature was set at 60°C, B) the running solvent and protein sample solution contained 10% methanol and 10mM ammonium acetate pH 5.0 and C) contained 10% methanol and 10mM ammonium acetate pH 5.0 and 3.76µM m7GTP. ESI source conditions were optimised for overall signal intensity. An additional series of peaks is observed in C, labelled with filled in circles, whose mass corresponds to the sum of m7GTP plus full-length eIF4E.

Figure 6 shows the spectra obtained from flying eIF4E at pHs 3.0, 4.0 and 5.0 with and without ligand. With out the addition of m7GTP gradual denaturation of the structure is observed by the accumulation of more charge. However in the presence of a saturating concentration of m7GTP the structure of eIF4E appears to be stabilised with a less noticeable shift in the charge distribution to lower m/z values. As the apo protein in figures 6A, 6C and 6E is gradually denatured the development of distinct charge distributions can be observed, as indicated in figure 6E. Partial unfolding or emergence of non-native conformations of proteins in solutions can be induced by a variety of agents, including variations in solvent composition (e.g., addition of chaotropes, such as detergents, alcohols and urea). pH adjustment within certain cellular compartments is a universal mechanism that often triggers complicated protein interaction cascades by inducing specific conformational changes[20]. The development of these different charge distributions is indicative of the presence of multiple conformational isomers of apo (see figure 6E) eIF4E coexisting in solution under equilibrium, and is consistent with observations by Wang[21] and Konermann[22] on myoglobin.

Spectra of apo-eIF4E and m7GTP bound to eIF4E exhibit tendencies that are consistent with Fenn's model of electrospray ionisation[23]. Tightly folded protein molecules are expected to have significantly smaller projected areas (as compared to less structured protein molecules), resulting in accommodation of a fewer number of charges on the protein surface at the ion evaporation stage. Most globular proteins are known to have multiple folding intermediates states (not just folded and unfolded states), some of which may have only minor structural differences. Subtle

differences are unlikely to lead to significant change in the overall shape of the protein. (e.g., projected areas, and, therefore, in the average number of charges accommodated upon protein ion desorption). Due to this the charge state distributions corresponding to various protein conformational isomers may be unresolved or poorly resolved (i.e. two or more different conformers may give rise to ions carrying the same charge). Close examination of the protein ion charge state distributions in figure 6 reveals in many cases the presence of multiple (>2) maximums on continuous distributions (e.g. 23+, 18+ and 10+ for apo eIF4E at pH 3.0; see figure 6E). This indicates the presence of at least three different protein conformations with significant Boltzman weight. This straight forward approach, however, can not determine the actual number of protein conformational isomers nor determine their characteristics.
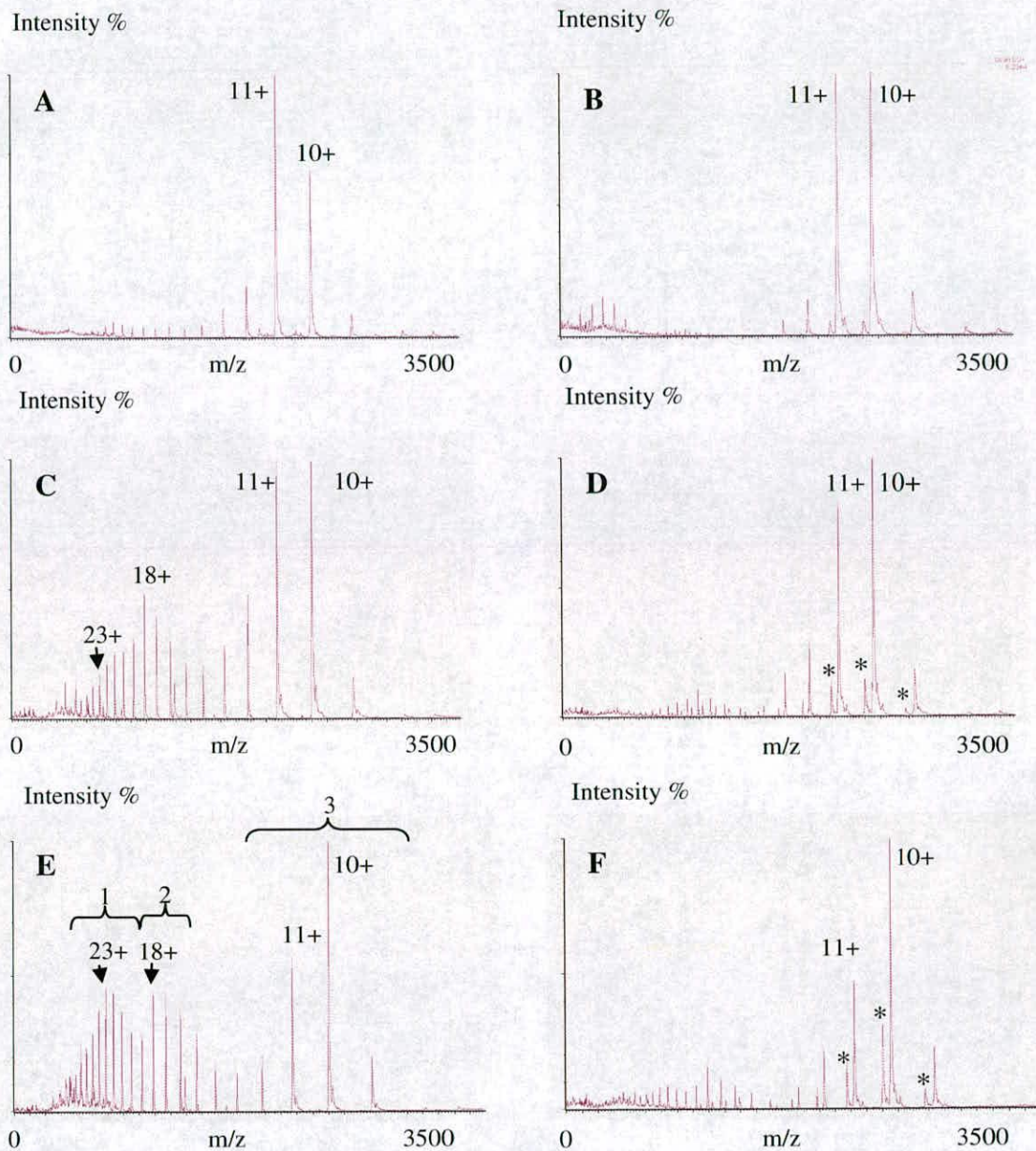
Figure 6: Panels A, C and E show the mass spectra for apo eIF4E at pHs 5.0, pH 4.0 and pH 3.0 (10mM ammonium acetate and 10% methanol). The adjacent panels B, D and F show mass spectra at identical conditions (pH 5.0, pH 4.0 and pH 3.0 10mM ammonium acetate and 10% methanol respectively) with eIF4E saturated with m7GTP at a concentration of 3.72μM. The 11+ and 10+ peaks are also labelled. For apo eIF4E there is a distinct change in the charge distribution as pH is decreased, with higher charged state species seen. At pH 4 and 3 new distributions of charged species can be seen which are analogous to different conformations of eIF4E. These are indicated in panel E. Bracket 3 corresponds to the correctly folded conformation seen alone at pH 5, whilst 2 and 3 indicate more unfolded conformations that are capable of becoming more highly charged. In panels B, D and E the presence of m7GTP stabilises the eIF4E in respect to pH and favours the more folded form. At lower pH a larger proportion of apo eIF4E (indicated by the asterisk) is seen suggesting that affinity for m7GTP does decrease. Also more peaks corresponding to higher charge states are seen as well.

112

Another pH experiment was carried out with eIF4E complexed with 3.53μM GTP at pHs 3.0, 4.0, 5.0, 6.0 and 7.0 (see figure 8) and an identical set of experiments with non-saturating levels of m7GTP (see figure 7). It was only with the presence of ligands (in this case m7GTP and GTP) that any signal in the mass spectra was seen at pHs 7.0 and 6.0. Apo eIF4E did not "fly" under these conditions. Under non-saturating conditions of m7GTP, the charge distribution associated with the native fold of eIF4 is not as strongly retained at lower pHs (figure 8) in comparison to saturating conditions of m7GTP (figure 6). Even under conditions of non-saturating levels of m7GTP the different charge distributions corresponding to different conformational isomers of eIF4E do not develop as strongly. Under denaturing conditions (figure 7F) the m7GTP eIF4E binding interaction is abolished. The spectra under denaturing conditions exhibit the two higher charge state distributions and thus indicate that the conformations of eIF4E responsible for these distributions have no affinity for m7GTP. Graph 7E shows the peak intensity for the apo and m7GTP complex peak (as indicated in figure 5) against pH. The 11+ peak and 10+ charge species were used for these calculations. It is immediately obvious that the affinity of eIF4E for m7GTP increases between the pH of 5.0 to 7.0. At these pHs (see figure 7) eIF4E exhibits the charge distribution centred around the 10+ and 11+ peaks and represents the eIF4E conformation that is in the correct fold to bind eIF4E.

The spectra for eIF4E against GTP at varying pHs show a similar type of behaviour to those with non-saturating levels of m7GTP (see figure 8). The most interesting difference is the presence of a third peak with the 11+ and 10+ charge species. This third peak corresponds to the association of two GTP molecules to eIF4E. However

under denaturing conditions (figure 8F) both peaks corresponding to the binding of one and two ligands are no longer present, suggesting that the interactions being observed are linked to the folded conformation of eIF4E. Again binding of GTP with eIF4E increases between pH 5.0 and 7.0 (see figure 8E). The binding of the 2L peak also shows an increase but with a slight decrease at pH 7.0.

The eIF4E/m7GTP complex is easily observed in the gas-phase (figures 5, 6 and 7) with distinct peaks forming for apo eIF4E and the complex, for each charge species. At a saturating concentration of m7GTP (see figure 7) the apo peak decreases reciprocally. GTP behaves quite differently in the gas-phase, when compared to m7GTP, where a peak corresponding to the binding of two ligands occurs for each charge species (see figure 8), as well as the apo and the complex peak seen with m7GTP. This third peak could either correspond to another binding site or be some form of non-specific aggregation taking place. Under denaturing conditions, for both the m7GTP/eIF4E and GTP/eIF4E complexes, all peaks corresponding to complex formation disappear. These observations imply that the formation of the eIF4E complexes are a function of the native fold. For both ligands the proportions of complex formation (see figures 7 and 8) increase as pH increases, indicating that at more physiological pHs eIF4E is either more correctly folded or a larger population of the protein is correctly folded and hence shows stronger affinities. The pH effect seen is more likely to be protein associated that any altered property of the ligands. The next section will further investigate the nature of the third peak seen with the eIF4E/GTP spectra (see figure 8).
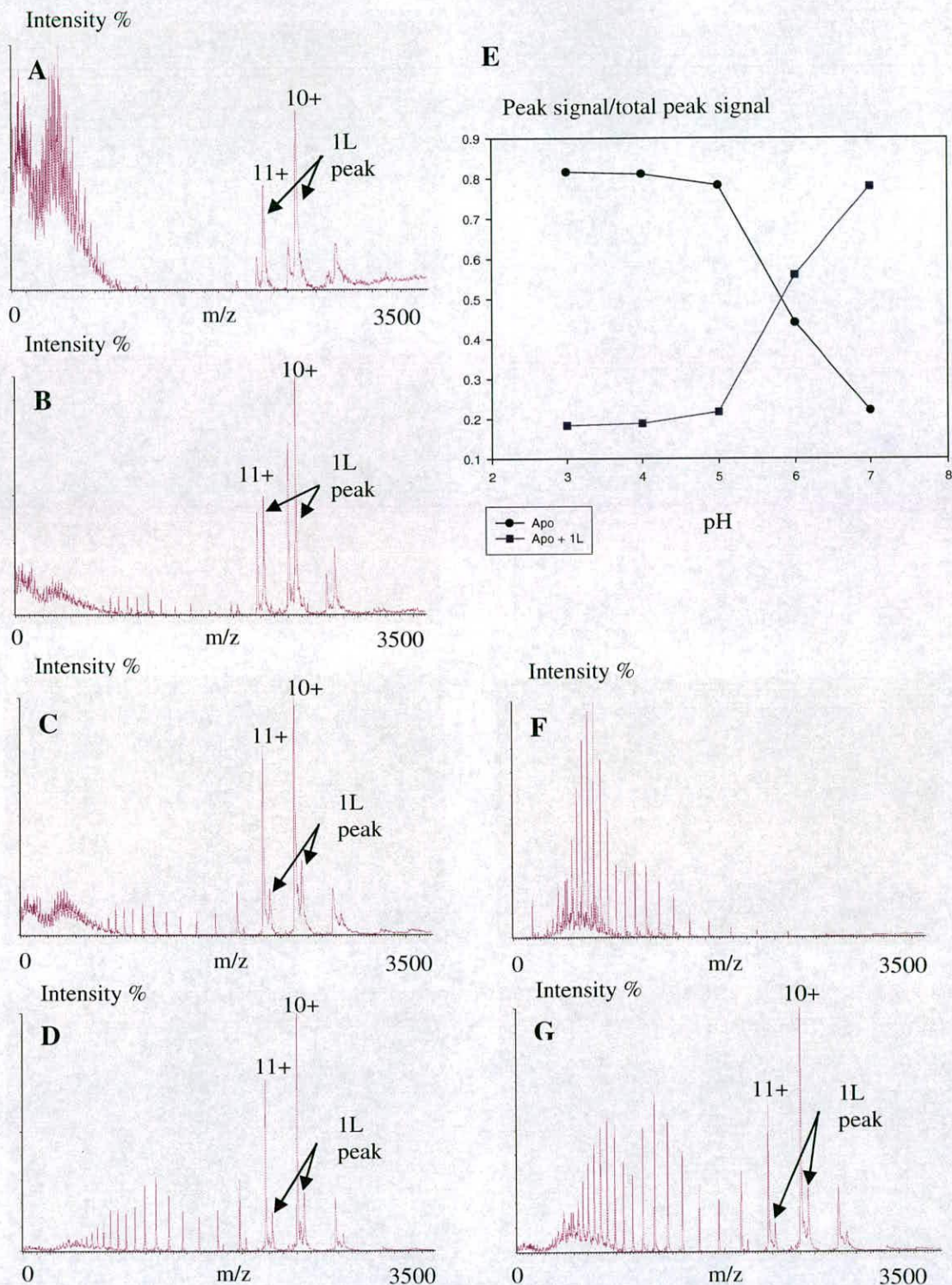
Figure 7: Mass spectra of eIF4E with 1.86μM m7GTP at several different pHs (A is pH 7.0, B is pH 6.0, C is pH 5.0, D is pH 4.0, G is pH 3.0, F is pH 3.0 with 50% methanol). Graph E demonstrates the behaviour of the apo and 1L peak (apo with one molecule of m7GTP associated) against pH. The numbers used for the peak signal over total signal intensity are the 11+ and 10+ charged species averaged out. In the spectra A, B, C, D and G the peak (1L) can be distinctly seen. However in the denatured spectra (F), the 1L peak is not seen and thus its association is linked with eIF4E retaining its fold at less denaturing conditions. The 1L peak increases significantly between pHs 5 and 7, which suggests that m7GTP's interaction with eIF4E is strongest at physiological pHs were eIF4E is known to function. The spectra show the characteristic reduction of charge species distribution as pH is increased. Spectra A, B and C (pHs 7.0, 6.0 and 5.0 respectively) show distributions associated with the folded protein conformation.

115
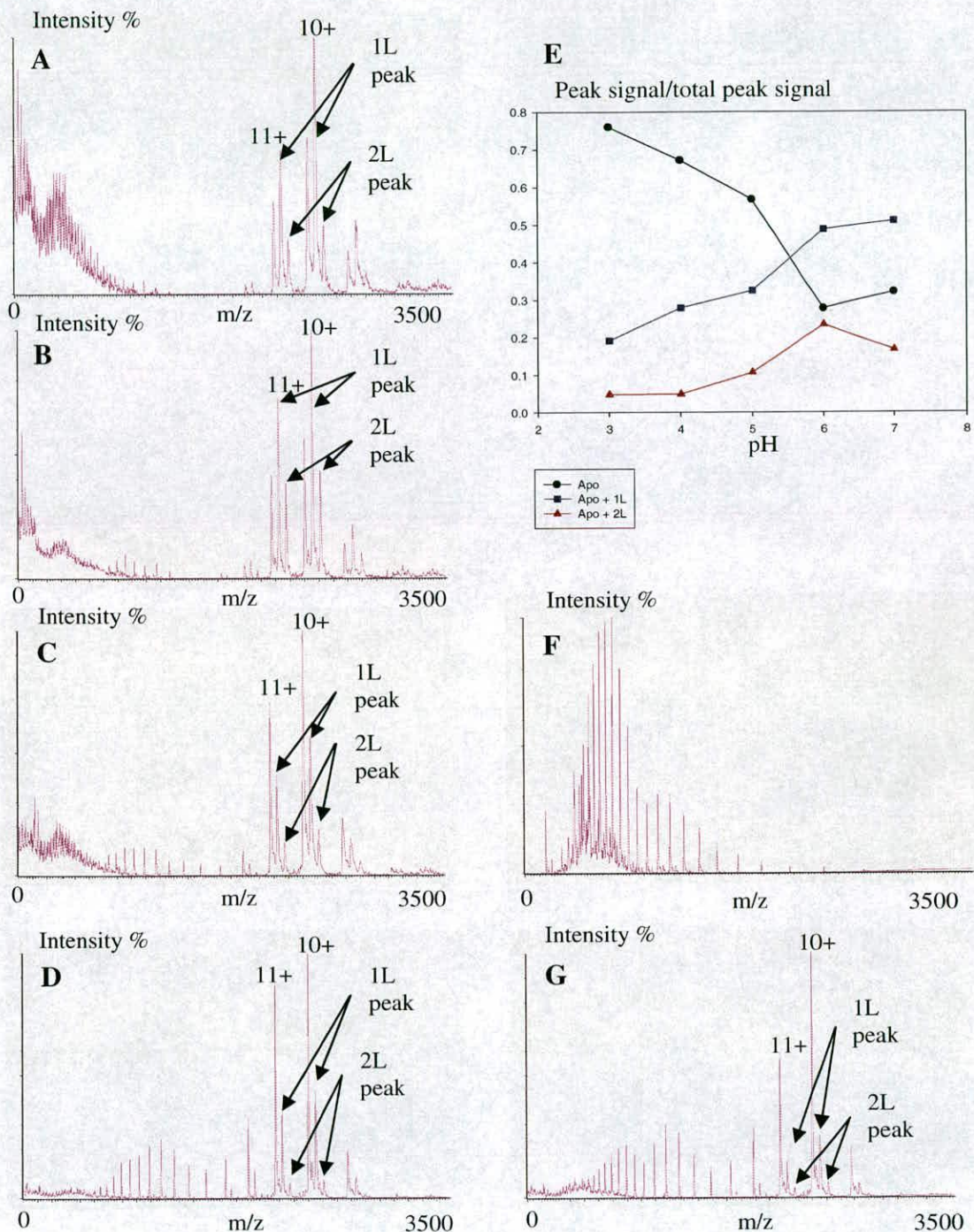
Figure 8: Mass spectra of eIF4E with 3.53μM GTP at several different pHs (A is pH 7.0, B is pH 6.0, C is pH 5.0, D is pH 4.0, G is pH 3.0, F is pH 3.0 with 50% methanol). Graph E demonstrates the behaviour of the apo peak, the 1L peak (apo with one molecule of GTP associated) and the 2L peak (apo with two molecules of GTP associated) against pH. The numbers used for the peak signal over total peak intensity are the 11+ and 10+ charged species averaged out. In the spectra A, B, C, D and G the two peaks (1L and 2L) associated with GTP binding of eIF4E can be distinctly seen. However in the denatured spectra (F) these are not seen and that their association is thus linked with eIF4E retaining its fold at less denaturing conditions. The 1L and 2L peaks increase as the pH increases which suggests that their interaction with eIF4E are strongest at physiological pHs were eIF4E is known to function. The spectra show the characteristic reduction of charge species distribution as pH is increased. Spectra A, B and C (pHs 7.0, 6.0 and 5.0 respectively) show distributions associated with folded protein conformation.

116

4.2.2 Mass spectrometry studies reveal that N7 methylated compounds interact specifically with eIF4E compared to a non-specific pattern of interaction seen in non-methylated compounds.

Titrations of molecules closely resembling the cap structure against eIF4E reveals the minimum criteria for ligands to interact with the cap binding site. The conditions chosen to conduct these experiments were 10% methanol and pH 5.0. The reason for the choice of these conditions were high signal intensity, the charge distribution of the native conformation was observed, interaction with m7GTP was preserved and that apo-eIF4E also "flies" at these conditions. When eIF4E is titrated against GTP the behaviour of the interaction in the gas-phase contrasts starkly with that of m7GTP. The apo peak of the 10+ and 11+ species is quickly saturated in the presence of m7GTP and there is a concurrent development of the ligand bound peak for both charge species (figure 10). In the case of GTP there is a quick development of peaks corresponding to the binding of one and two ligands whilst over the course of the titration the apo peak is never saturated (see figure 10). As the concentration of GTP is further increased there is a gradual development of peaks that are equivalent to the binding of three or more ligands. The m7GTP titration peaks corresponding to the binding of two ligands or more only occur after saturation of the apo peaks. m7GTP shows a much more specific interaction with the cap-binding site. Thus only when the cap binding site is saturated by m7GTP does the non-specific binding to other parts of the protein occur as shown by the development of the 2L peak (see figure 9). With GTP the cap binding site shows no strong preference for it and thus this non-specific interaction occurs before saturation of the apo peak (see figure 9 and 10).
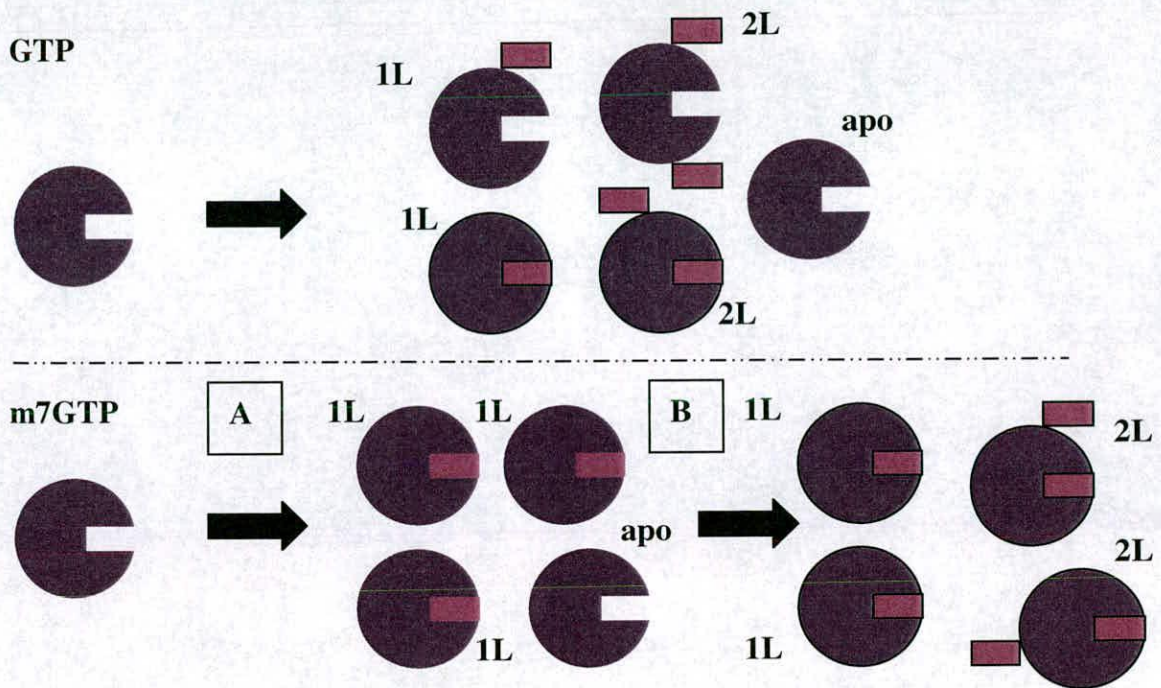
Figure 9: Diagram showing potential complexes in the m7GTP and GTP titrations in figures 10B and 10C respectively. In the GTP panel complexes are shown that could form in the gas-phase that correspond to 1L and 2L Note how the ligand when it does not bind in the cap-binding site may bind any where else on the surface. In the m7GTP section two potential samples of complexes (A and B) are shown from two points on the titration curve (figure 10). Section A corresponds to the part of the curve where only apo and the 1L peak are present, and section B corresponds to the part of the curve where the 2L peak (associated with non-specific binding) occurs. Note in section B saturation of the apo species and how the second ligand has the potential to interact anywhere on the protein surface.

The correlation between the proportion of the complex present in ESI spectra and the relative concentrations of protein and ligand in solution often suggests the possibility of non-specific protein–ligand interactions. However, non-specific interactions in the gas phase are also likely to alter the observed stoichiometry of the complex ion in a concentration-dependent manner as is seen for GTP. Interestingly it is occurring at a concentration where in the m7GTP titration complexes containing multiple ligands are not occurring. Also the GTP 1L and 2L peaks disappear under denaturing conditions (see figure 8F) which suggests that eIF4E has in contrast to its interaction with m7GTP a general non-specific interaction with GTP. Surface plasmon

resonance experiments with capped and non-capped mRNA have demonstrated a tight interaction between eIF4E with the capped variants and a non-specific interaction with the non-capped variants[24]. These observations support the hypothesis that eIF4E's non-specific behaviour with GTP may be an intrinsic mechanism that comes into play when binding the rest of the mRNA chain after cap-binding has occurred. From these titrations with m7GTP and GTP the presence of the delocalised charge on the m7G ring, induced by the methylation of the N7, is the defining feature that separates the specific and non-specific binding interactions.

.
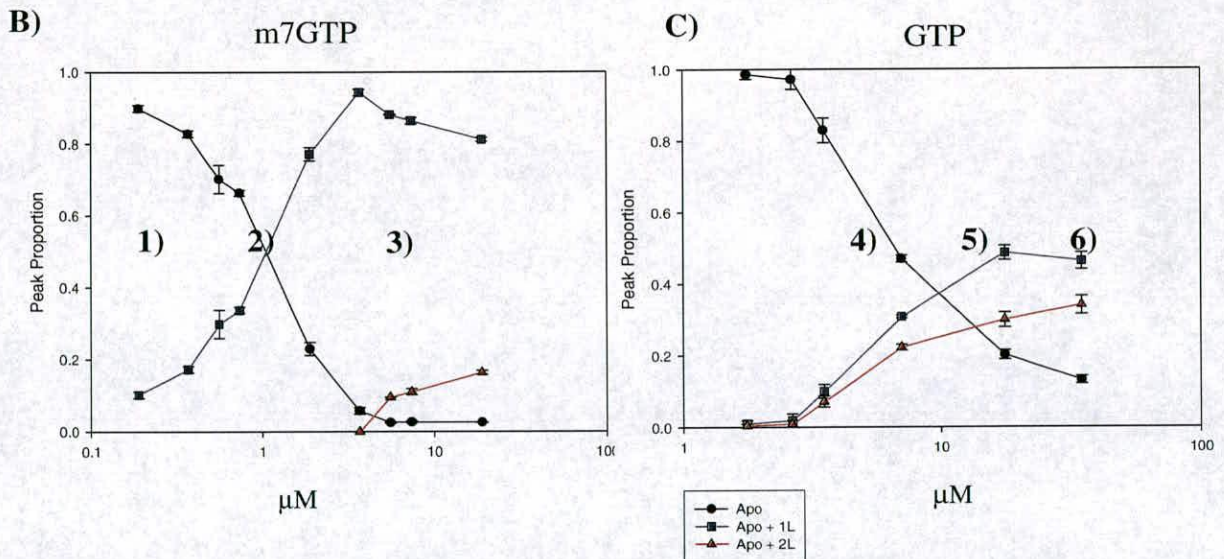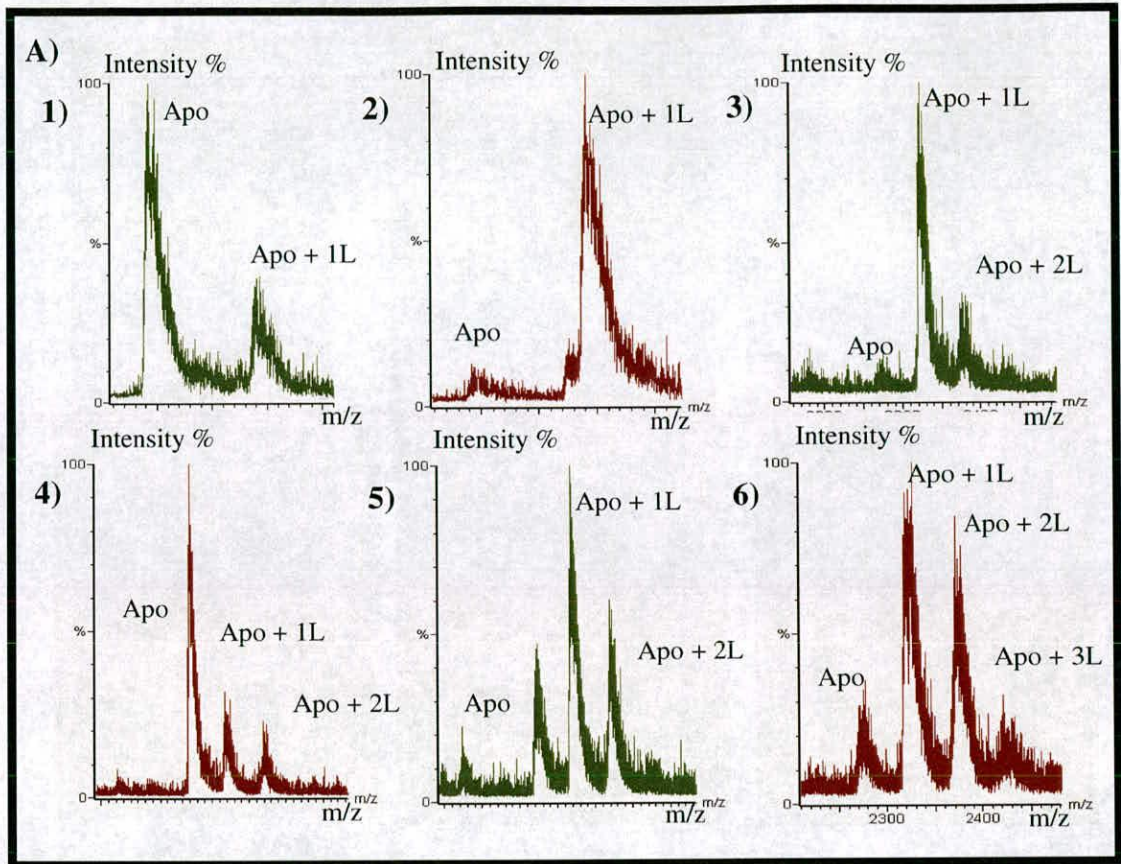
Figure 10: Titrations of m7GTP (panel B) and GTP (panel C) against eIF4E showing development of peaks corresponding to the binding of one (1L) and two (2L) ligand molecules. Panel A shows mass spectra (of the 11+ charge species) for the points indicated on the titrations in panel B and panel C. It is clearly seen in the m7GTP titration (panel B) that almost total saturation of the apo peak occurs before the occurrence of the peak corresponding to the binding of two ligands. In the GTP titration (panel C) the peaks corresponding to the binding of one and two ligands appear almost simultaneously and develop as a function of concentration. These titrations demonstrate that eIF4E specifically interacts with m7GTP and it is only at near saturation that non-specific aggregation as indicated by the third peak occurs. In the case of GTP, eIF4E only has a weak affinity and as a result non-specific aggregation occurs before near saturation of apo eIF4E.

A

Peak signal/total peak signal

GTP
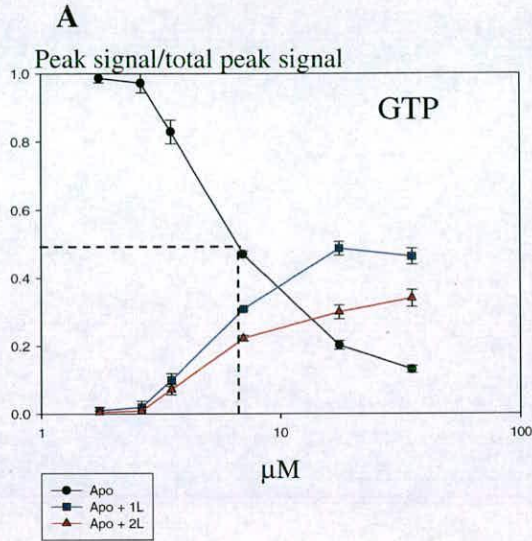
C

	GMP	GDP	GTP
50% Saturation of apo peak (50% sat)	~90μM (30%, 10%)	~20μM (35%, 10%)	~8μM (30%, 20%)

Values in brackets after the 50% sat. values are the proportion of the 1L and 2L peaks in % at these points in the titration. Dotted line in panel A demonstrates this for GTP.

B

Peak signal/total peak signal

GDP

D

Peak signal/total peak signal

GMP
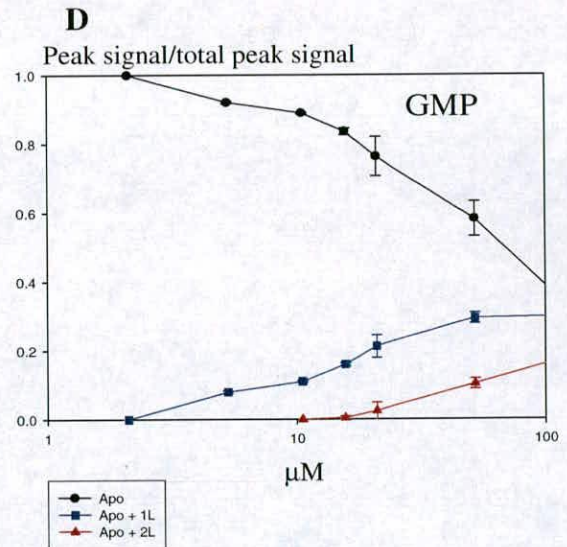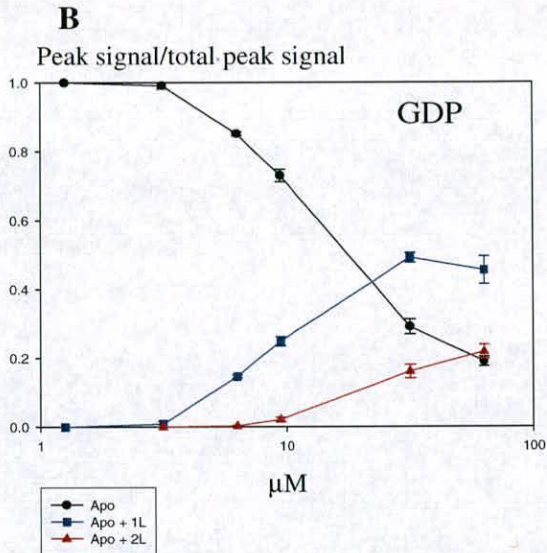
Figure 11: Titrations of GTP, GDP and GMP against eIF4E (A, B and D respectively) examining the effect of the non-specific binding described in figure 9. However as the length of the phosphate tail is decreased the strength of binding is decreased, but the non-specific interaction is still being observed. The concentrations at which 50% saturation of the apo peak occurs demonstrate this point (Table B). At these points of 50% saturation on the graphs, the proportion of the signal due to 1L and 2L complex species are similar (roughly 30% 1L and 10% 2L). The pattern of non-specific binding is related to the size of the negative charge in this set of ligands. Peak signal/ Total peak signal was calculated from both 11+ and 10+ charged species.

Titrations were then carried out with GDP and GMP to explore the effect of the length of the phosphate tail on the non specific interaction observed with GTP (figure 11). The type of interaction, as seen for GTP remains the same, with the gradual development of multiple ligand peaks with the apo peak never becoming saturated. The values at which the peak corresponding to the apo species is at 50% for these titrations (50% sat) clearly decrease as a function of the phosphate tail length (figure 10C). For GTP, GDP and GMP, when the apo peak is 50% saturated, the proportion of the signal corresponding to the binding of one and two ligands is approximately the same, ~30% and 10% (see figure 11C). This indicates that the concentration, at which development of the peaks with multiple ligands occurs, is correlated to the length of the phosphate chain. The non-specific interaction being observed for guanine with no methylation at the N(7) in the gas phase masks any weak binding that maybe occurring at the cap-binding site. Hence (figure 11) the development of peaks corresponding to multiple ligands before saturation of the apo peak, as opposed to the m7GTP/ eIF4E interaction (figure 10) where the binding of multiple ligands is only seen after the apo peak is almost saturated. Titration series carried out for m7G, guanosine and ribavirin, using the same concentration range for m7GTP and GTP, revealed no apparent binding in the gas phase. This lack of a visible interaction with m7G, ribavirin and guanosine confirms that for either the specific or non-specific pattern of interaction to occur the phosphate moiety is required.

The same sets of experiments were carried out with ATP, ADP and AMP (see figure 12). These compounds again displayed the same type non-specific interaction noted

for GTP, GDP and GMP. However ATP behaves quite differently with the development of the 2L peak occurring before the 1L peak (see figure 12A). ADP and AMP behave like their counterparts with the guanine ring moiety. Again the values for ADP and AMP, at which the peak corresponding to the apo species is at 50% for these titrations (50% sat), clearly decrease as a function of the phosphate tail length (figure 12D). At these points the signal contribution of the 1L and 2L peaks were ~ 30% and 10% respectively, highly similar to what occurs for the guanine based ligands. The interaction with ATP may be caused by eIF4E having two sites with a weak affinity for ATP, which might explain the early development of the 2L peak over the 1L peak. Support for this model comes from the fact that in the m7GpppA eIF4E crystal the adenine moiety is clearly seen (see chapter 3). However with increasing concentrations of ATP the development of multiple ligand peaks are seen, which show that eIF4E's interaction with ATP is mainly non-specific in character (see figure 10 for an example of this with GTP). The way in which the adenine and guanine ligands behave in a similar manner points towards a dominant role for the phosphate chain over the hydrogen bond/donor relationship on the ring moiety. AMP and ADP do have 50% saturation points that occur at higher concentration than their guanine counterparts which may indicate the fact that different hydrogen bond partners are found on the adenine ring. ATP has a very similar 50% sat point to GTP, however this is probably a reflection of the early 2L peak development.

A

Peak signal/Total peak signal

ATP

μM

- Apo
- Apo + 1L
- Apo + 2L

C

Peak signal/Total peak signal

ADP

μM

- Apo
- Apo + 1L
- Apo + 2L

B

Peak signal/Total peak signal

AMP

μM

- Apo
- Apo + 1L
- Apo + 2L

D

	AMP	ADP	ATP
50% Saturation of apo peak (50% sat)	~120μM (30%, 10%)	~90μM (35%, 10%)	~9μM (20%, 35%)

Values in brackets after the 50% sat. values are the proportion of the 1L and 2L peaks in % at these points in the titration. Dotted line in panel A demonstrates this for ATP.

Figure 12: The titrations of ATP, ADP and AMP (A, C and B respectively) against eIF4E overall show a similar trend to the GTP, GDP and GMP titrations., with the strength of binding related to the phosphate tail length. However there is one distinctive difference, in the case of ATP, where the 2L peaks starts to appear before the 1L peak. The reasons for such an interaction are not obvious but could be due to either ATP aggregating together and binding to the protein or location of another site away from the cap-binding site that may have an affinity for it. The first option appears to be unlikely as this does not occur in titrations of ADP or AMP.

4.2.3 m7GTP and m7GDP interact identically with human full length eIF4E in the gas phase.

An identical titration was carried out with m7GDP to the m7GTP one to investigate the effect of phosphate chain length, when a positive delocalised charge is present on the m7G ring (see figure 13). As with the m7GTP, in the m7GDP titration eIF4E is near saturation before development of the 2L peak. However where a clear effect was seen for decreasing the phosphate tail length in the guanine and adenine based ligands, there is no such difference between m7GDP and m7GTP.

An explanation for this behaviour in the gas-phase can be seen in the crystal structure of full-length human eIF4E with m7GTP (see figure 13C). The β phosphate of the triphosphate tail has four close contacts with the residues in the phosphate-binding pocket (indicated by the yellow dashes) whilst the γ phosphate has only one contact. The γ phosphate mainly interacts with eIF4E through water-mediated interactions (discussed in chapter 3 and 6), which are expected not to be present in the gas-phase. It is the absence of these structured waters seen in the crystal structure that may offer a reason for the similar behaviour of m7GTP and m7GDP with eIF4E as revealed by mass spectrometry studies. To further investigate this model a titration of m7GMP against eIF4E could verify it i.e. the β phosphate would no longer be forming any contacts and a much weaker interaction would be expected in the gas-phase for m7GMP. However m7GMP is not available commercially. The seemingly identical titration curves of m7GTP and m7GDP were investigated by seeing how their interactions behaved in respect to increasing concentrations of ammonium acetate (figure 13D), whilst keeping the concentration of m7GTP to m7GDP identical. By

increasing the concentration of ammonium acetate, a volatile buffer, the ionic strength of the sample being flown is also increased. This should have the effect of attenuating the ionic part of the m7GTP/m7GDP interaction and differentiate them. The graph in figure 10D confirms that the ionic interaction between eIF4E and two ligands are identical.

A

Peak signal/Total peak signal

m7GTP

µM

B

Peak signal/Total peak signal

m7GDP

µM

- Apo
- Apo + 1L
- Apo + 2L

C

Lys162

γ

β

Lys159

α

Arg157

D

Peak signal/Total peak signal

- m7GDP
- m7GTP

mM (ammonium acetate)
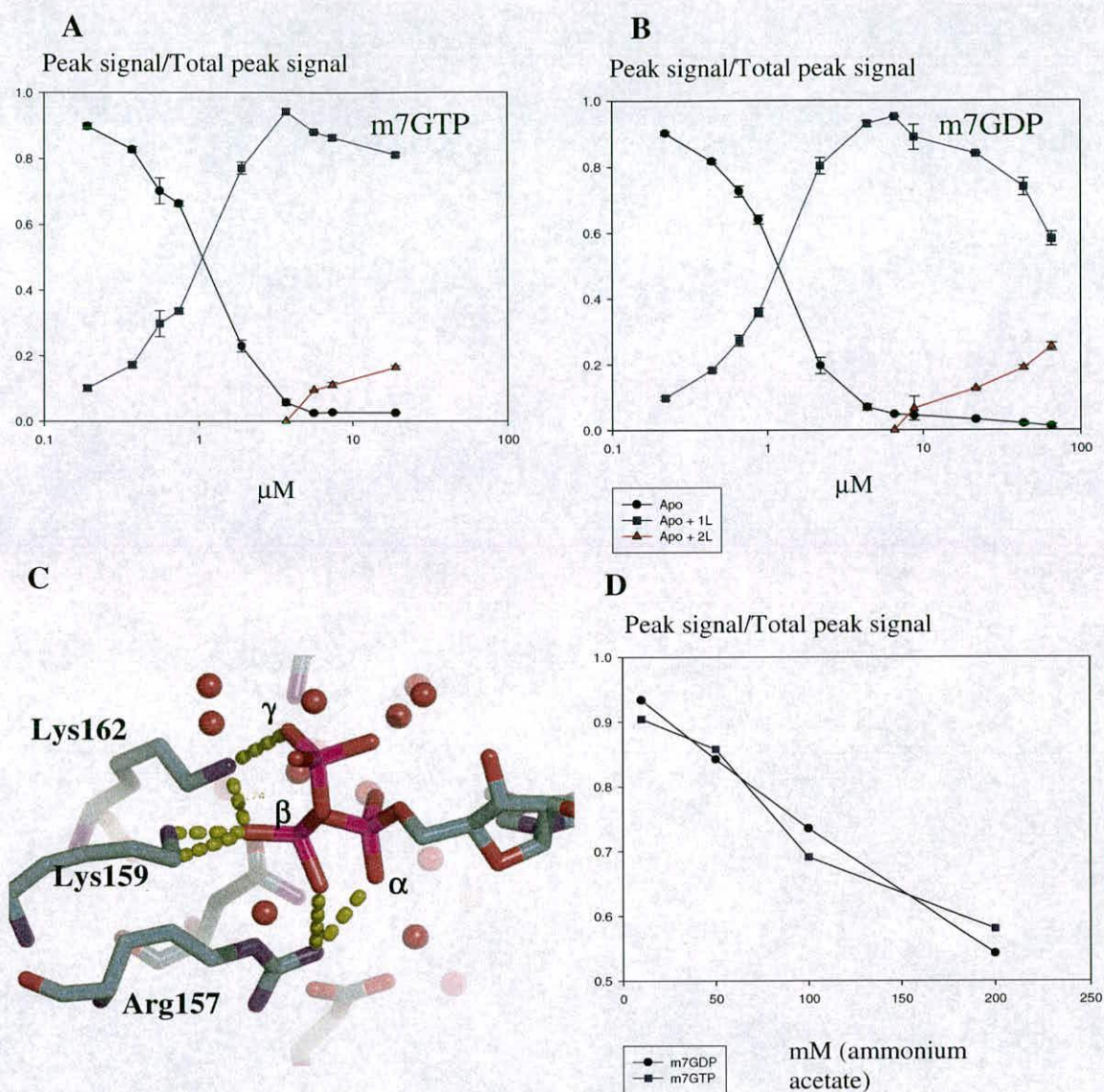
Figure 13: Titrations of m7GTP (A) and m7GDP (B) against eIF4E. Both ligands demonstrate a highly specific interaction with eIF4E. However, both curves are highly similar with 50% saturation of the apo peak occurring at ~ 1µM and the onset of the non-specific aggregation peak (2L) at ~7µM and 9 µM respectively. This interaction was further investigated by monitoring the peak proportion (average of 11+ and 10+ peaks) at increasing concentrations of ammonium acetate (D). By increasing the ammonium acetate concentration the ionic strength of the sample to be run on the mass spectrometer is being effectively increased. This should attenuate the ionic strength of the ligands (m7GTP and m7GDP) interaction with eIF4E and allow it to be assessed. The ammonium acetate curves in graph (D) reveals no intrinsic difference in the ionic interaction between m7GTP and m7GDP. An explanation for this behaviour in the gas-phase can be seen in the crystal structure of full-length human eIF4E with m7GTP (C). The β phosphate of the triphosphate tail has four close contacts with the residues in the phosphate-binding pocket (indicated by the yellow dashes) whilst the γ phosphate has only one contact (which is not seen in the other m7GTP structure in the PDB). The γ phosphate mainly interacts with eIF4E through water-mediated interactions (discussed in chapter 3), which are expected not to be present in the gas-phase. It is the absence of these structured waters seen in the crystal structure in the gas phase that may offer a reason for the similar behaviour of m7GTP and m7GDP with eIF4E as revealed by mass spectrometry studies. All samples were flown at 10% methanol, 3.2µM eIF4E and 10mM ammonium acetate pH 5.0.

4.2.4 Insights into the eIF4E cap-analogue binding mechanism from mass spectrometry.

From the mass spectrometry studies eIF4E only shows a specific interaction when a phosphate group and a delocalised charge on the m7G moiety is present in the gas-phase. When no delocalised charge is present on the m7G ring a more non specific pattern of interaction seen, with the presence of multiple ligand peaks at high concentrations of ligand. This is further supported by Kds, determined in the fluorescence studies by Niedzwiecka et al[25], which vary from 0.01 uM for specific 7-substituted cap analogues to 10000uM for 7-unsubsituted cap analogues. The results of Niedzwiecka[25] et al also reveal that although unmethylated guanosine phosphates are potentially capable of forming three Watson-Crick like hydrogen bonds that this does not occur. As shown by $\Delta G^O(GMP) = -3.09(+/-0.90)$ kcal/mol[25] when compared to the contribution of the single α-phosphate ($\Delta\Delta G^O = -3.01 +/-0.16$)[25]. It is only when the guanine is N(7) substituted that the hydrogen bonds are formed. Hence the cation-π stacking enhancement is a precondition for hydrogen bond formation, and a double role can be prescribed to the presence of the conserved residues 56 and 102 in the eIF4E binding site: stabilization of the 7-methyl guanine ring by stacking itself and enabling 7-methylguanine to form the hydrogen bonds to the site-chain of Glu103 and the amide backbone of Trp102[25].

The non-specific interaction observed in the gas-phase with GTP may indicate that eIF4E has a general promiscuity to negatively charged compounds, which may be a mechanism involved in interacting with the rest of the mRNA chain during translation via the phosphate backbone. VP39 has successfully been crystallised

with an RNA hexamer that interacts exclusively with the protein via the phosphate backbone[26] (see figure 14).
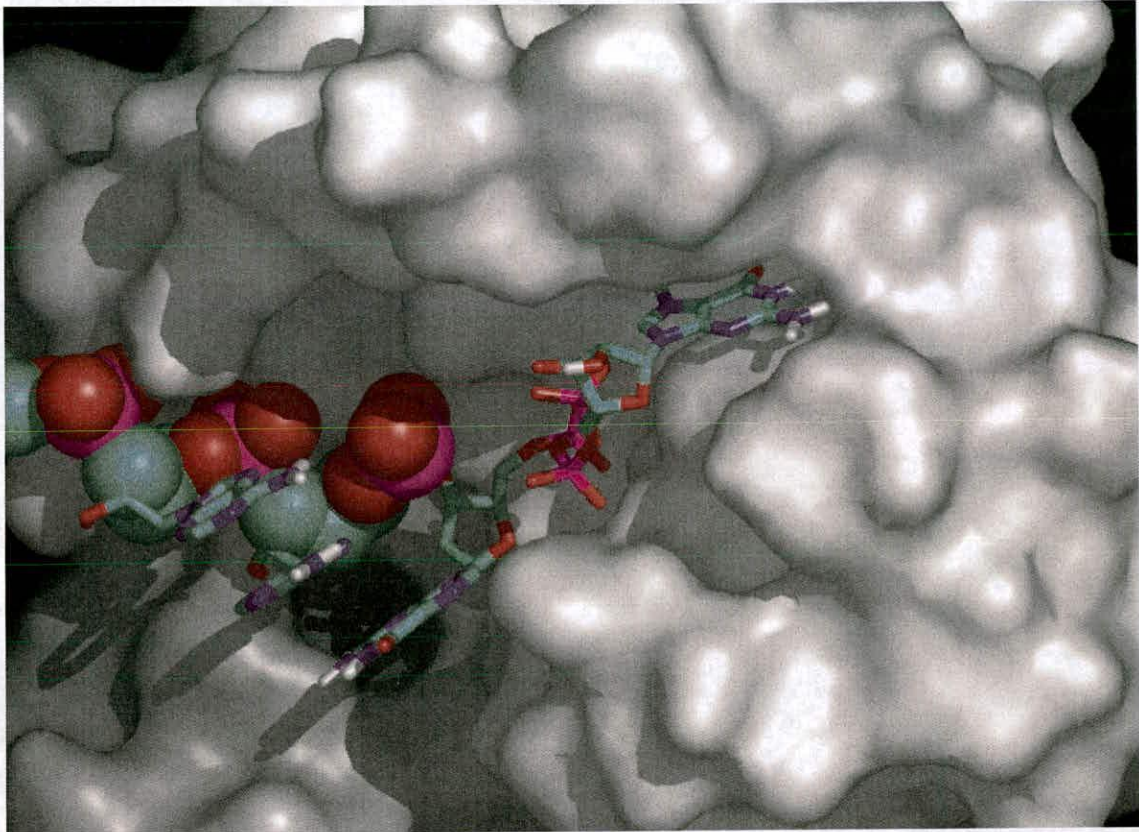
Figure 14: The m7G ring binds in a small pocket between the two tryptophans in VP39. The rest of the RNA is recognised in sequence independent manner via interactions with the phosphate backbone (shown in spheres). This mechanism could apply to eIF4E and may offer an explanation for the non-specific interaction observed for GTP in the gas-phase against eIF4E.

Another element important for specific eIF4E-cap binding is the negative electrostatic charge of the phosphate chain, which depends upon the phosphate number. The increase of the K_d with single step-wise reduction of the phosphate chain in the series of 7-methylated mononucleotides, as demonstrated by Niedzwiecka et al[25], is followed by the same changes of Kd for the unmethylated compounds (~five and ~25 fold for the removal of the γ–phosphate and the β–

phosphate, respectively). Removal of the α-phosphate leads to a major increase of K_d which is 175-fold higher for m7G than that of m7GMP[25]. These results show much agreement with what was observed in the gas-phase. That is if the phosphate tail length is decreased in the case of the non-methylated ligands an attenuation in the non-specific interaction is seen (figure 11). In the gas-phase no binding is seen with m7G which corresponds to the observation by Niedzwiecka et al[25] that upon the removal of the α phosphate there is a significant drop in binding. However they observe a stronger interaction for m7GTP over m7GDP with eIF4E than appears to be the case in the gas-phase. The model proposed in the last section of structured waters having an important role in the interaction of the γ-phosphate (figure 13) explains this discrepancy. Steady-state fluorescence measurements are solution based, where water molecules will be present to participate in any interactions, whilst gas-phase measurements are not.

eIF4E requires a negative phosphate to be present on the cap-analogue for guanosine to have any type of interaction and for m7G and Ribavirin to intercalate between W102 and W56. These experimental results indirectly show that in the gas phase it is more than likely that the phosphate tail anchors the m7G(p)^n molecule to eIF4E and that the m7G moiety then intercalates itself between the tryptophans in the cap-binding slot. Experiments by Blachut-Okrasinska et al[27] observed a two-step binding mechanism for the cap-analogue eIF4E association by means of stopped flow spectroscopy. Thus it seems plausible that the phosphate tail serves as a molecular anchor enabling the m7G moiety to form further contacts within the binding site.

4.2.5 How does the mechanism of cap-binding in eIF4E compare to other cap-binders?

VP39 is another well-studied cap-binding protein that has been characterised via crystallographic techniques and fluorescence methods[28,29]. VP39 can bind a variety of methylated nucleobases in the pocket of the binding site for the m7G moiety of the cap dinucleotide m7G(5')pppG and its nucleoside or nucleotide derivatives. Methylated compounds that were successfully soaked in to the VP39 crystals in contrast to their non-methylated counterparts were m3Ade, m3Cyt, m1Ade and m1Cyt. These results again demonstrate how intercalation of compounds between the two aromatic resides requires a delocalised positive charge.

In contrast to eIF4E where the cap-analogue seems to bind by a two-step mechanism the interaction with VP39 seems to be a single step (stop-flow studies carried out on mutant VP39 with Phe180 replaced by a tryptophan). This is consistent with crystallographic data of VP39 that shows no structural rearrangement following binding of m7G and several other cap analogues to VP39 and other mutants studied[28]. The uppermost association rate constant (k_1 of $\sim 3 \times 10^7 \, m^{-1} \, s^{-1}$) observed at pH 5.5 and 6.0 is about two orders magnitude lower than the diffusion limit, which appears fast for a binding process dependent on the precise and tight insertion of the m7Gua nucleobase between two aromatic side-chains (Trp180 and Tyr22)[29]. The electron rich π clouds of both aromatic side chains, which possess a permanent electrostatic quadrupole, must therefore impart a strong attractive force on the electron deficient m7Gua moiety. It is this force that has a significant effect on cap-binding by cap-specific proteins that rely on stacking interactions.

In the case of VP39 the presence of a positive delocalised charge on the ring, supported by mass spectrometry studies and other fluorescence studies in the literature on eIF4E, is critical to success in crystallographic soaking experiments[28]. However intercalation between the tryptophans in VP39, as shown by the crystallographic experiments, can occur with methylated bases without a phosphate group present[28]. This situation is in stark contrast to eIF4E, where any interaction in the gas-phase requires a phosphate to be present (fluorescence studies in the literature also report much weaker k_ds when all the phosphates are removed[25]). It would seem that in VP39 that the aromatic cap binding residues are fairly rigid. In eIF4E, considering that no apo structure exists and that the m7GTP must be orientated by the phosphate tail to allow it to interact, that the tryptophans probably become rigid upon interacting with the m7G ring.

4.3 Conclusion.

Mass spectrometry studies on eIF4E and various cap- analogues reveal interesting and complementary data to solution based studies. Also mass spectrometry with knowledge from the crystal structure has allowed us to analyse the contributions of the γ- and β phosphates in the gas-phase, and explain why m7GTP and 7GDP behave the same in the gas-phase. Mass spectrometry can also differentiate between GTP and m7GTP, allowing a specific interaction to be differentiated from a non-specific interaction. This offers the potential to use mass spectrometry as a potential screening tool to verify virtual screening hits against eIF4E.

Reference List

1. Hoffmann,E. & Strooban,V. Mass Spectrometry: Principles and Applications. (2002).

2. Loo,J.A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1-23 (1997).

3. Light-Wahl,K.J., Schwartz,B.L. & Smith,R.D. Observation of the noncovalent quaternary associations of proteins by electrospray mass spectrometry. *J. Am. Chem. Soc.* **116**, 5271-5278 (1994).

4. Loo,J.A., Edmonds,C.G. & Smith,R.D. Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* **248**, 201-204 (1990).

5. Loo,J.A., Edmonds,C.G., Udseth,H.R. & Smith,R.D. Effect of reducing disulfide-containing proteins on electrospray ionization mass spectra. *Anal. Chem.* **62**, 693-698 (1990).

6. Katta,V. & Chait,B.T. Hydrogen/deuterium exchange electrospray ionization mass spectrometry -- A method for probing conformational changes in solution. *J. Am. Chem. Soc.* **115**, 6317-6321 (1993).

7. Cheng,X. Using electrosprsay ionization FTICR mass spectrometry to study competitive binding of inhibitors to carbonic anhydrase. *J. Am. Chem. Soc.* **117**, 8859-860 (1995).

8. Cheng,X. Using electrosprsay ionization FTICR mass spectrometry to study competitive binding of inhibitors to carbonic anhydrase. *J. Am. Chem. Soc.* **117**, 8859-860 (1995).

9. Kelly,A.M., Vestling,M.M., Fenselau,C.C. & Smith,P.B. Electrospray analysis of proteins: A comparison of positive-ion and negative-ion mass spectra at high and low pH. *Mass Spectrom. Rev.* **27**, 1143-1147 (1992).

10. Smith,R.D. & Light-Wahl,K.J. Perspectives -- the observation of non-covalent interactions in solution by electrospray ionization mass spectrometry -- promise, pitfalls, and prognosis. *J. Am. Chem. Soc.* **22**, 493-501 (1993).

11. Heck,A.J. & Van Den Heuvel,R.H. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev. 2004. Sep. -Oct. ;23(5):368. -89.* **23**, 368-389 (2004).

12. Loo,J.A. *et al.* Biophysical characterization of zinc ejection from HIV nucleocapsid protein by anti-HIV 2,2'-dithiobis[benzamides] and benzisothiazolones. *J. Med. Chem.* **39**, 4313-4320 (1996).

13. Hanson,C.L. & Robinson,C.V. Protein-nucleic acid interactions and the expanding role of mass spectrometry. *J. Biol. Chem. 2004. Jun. 11;279. (24.):24907. -10. Epub. 2004. Mar. 31.* **279**, 24907-24910 (2004).

14. Hofstadler,S.A. & Griffey,R.H. Analysis of noncovalent complexes of DNA and RNA by mass spectrometry. *Chem. Rev. 2001. Feb. ;101. (2):377. -90.* **101**, 377-390 (2001).

15. Robinson,C.V. *et al.* Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-CoA ligand binding and assembly. *J. Am. Chem. Soc.* 8646-8653 (1996).

16. Li,Y.T. *et al.* Analysis of the energetics of gas-phase immunophilin ligand complexes by ion spray mass spectrometry. *J. Am. Chem. Soc.* 7487-7493 (1994).

17. Hunter,C.L., Mauk,A.G. & Douglas,D.J. Dissociation of heme from myoglobin and cytochrome b5: comparison of behavior in solution and the gas phase. *Biochemistry* **36**, 1018-1025 (1997).

18. Samalikova,M. & Grandori,R. Role of opposite charges in protein electrospray ionization mass spectrometry. *J. Mass Spectrom. 2003. Sep. ;38. (9.):941. -7.* **38**, 941-947 (2003).

19. Zhang,S., Van Pelt,C.K. & Wilson,D.B. Quantitative determination of noncovalent binding interactions using automated nanoelectrospray mass spectrometry. *Anal. Chem. 2003. Jul. 1;75. (13):3010. -8.* **75**, 3010-3018 (2003).

20. Dobo,A. & Kaltashov,I.A. Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS. *Anal. Chem. 2001. Oct. 15. ;73. (20.):4763. -73.* **73**, 4763-4773 (2001).

21. Wang,F. & Tang,X. Conformational heterogeneity of stability of apomyoglobin studied by hydrogen/deuterium exchange and electrospray ionization mass spectrometry. *Biochemistry* **35**, 4069-4078 (1996).

22. Konermann,L. & Douglas,D.J. Unfolding of proteins monitored by electrospray ionization mass spectrometry: a comparison of positive and negative ion modes. *J. Am. Soc. Mass Spectrom.* **9**, 1248-1254 (1998).

23. Fenn,J.B. Ion formation from charged droplets: roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry* **4**, 524-535 (1993).

24. von der,H.T. & McCarthy,J.E. Studying the assembly of multicomponent protein and ribonucleoprotein complexes using surface plasmon resonance. *Methods 2003. Feb. ;29. (2):167. -74.* **29**, 167-174 (2003).

25. Niedzwiecka,A. *et al.* Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol. 2002. Jun. 7. ;319. (3):615. -35.* **319**, 615-635 (2002).

26. Hodel,A.E., Gershon,P.D., Shi,X., Wang,S.M. & Quiocho,F.A. Specific protein recognition of an mRNA cap through its alkylated base. *Nat. Struct. Biol.* **4**, 350-354 (1997).

27. Blachut-Okrasinska,E. *et al.* Stopped-flow and Brownian dynamics studies of electrostatic effects in the kinetics of binding of 7-methyl-GpppG to the protein eIF4E. *Eur. Biophys. J. 2000. ;29. (7.):487. -98.* **29**, 487-498 (2000).

28. Hu,G., Gershon,P.D., Hodel,A.E. & Quiocho,F.A. mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl. Acad. Sci. U. S. A* **96**, 7149-7154 (1999).

29. Hu,G., Tsai,A.L. & Quiocho,F.A. Insertion of an N7-methylguanine mRNA cap between two coplanar aromatic residues of a cap-binding protein is fast and selective for a positively charged cap. *J. Biol. Chem. 2003. Dec.* **278**, 51515-51520 (2003).

Chapter 5: Identifying initial drug leads for eIF4E and development of an overall structure-based drug design strategy.

5.0 Introduction.

In this chapter the aims are to identify lead ligands for eIF4E by using virtual docking methods[1], to synthesise compounds similar to the cap-analogue and to use other methodologies to search for compounds with an analogous N7 delocalised positive charge. The classic drug design cycle involves the ability to produce large amounts of protein that can be screened by various methods to find target ligands. In the case of eIF4E mass spectrometry is a technique that lends itself easily to ascertaining the activity of a potential ligand. There is also the potential of exploiting the intrinsic tryptophan fluorescence of eIF4E to screen for hits [1]. Successful hits can then be used in co-crystallisation trials. From binding data and successful structure elucidation a binding model can be developed that can be used in further rounds of drug design e.g. step-wise modifications of hits and refinement of virtual screening models.

5.1 Screening for potential inhibitors using chemical libraries.

The antithesis of rational methods for drug design is the reliance on chance discoveries of active ligands. High throughput screening is a chance-based method. The limitations of high throughput screening are the volume of the screen that can be achieved within commercial constraints and the theoretical coverage within the diversity of the compound set being screened[2]. If the number of compounds in existence were 10^8 and the number of drug like compounds that could be made is 10^{30}, then the probability of a

compound set of a size currently in existence showing useful statistical coverage in chemical space is minuscule[3].

Historically the number of compounds that could be synthesised by a chemist has been at most a few hundred. However, new technologies such as combinatorial chemistry and high throughput screening (HTS) offer a much broader range of possibilities and the capacity to synthesise many more compounds, even to contemplate numbers in the millions[4]. This new technology requires chemists to confront a large and diverse "chemical landscape"[3].

There are many chemical libraries that a trained chemist could reasonably hope to synthesise. Each library can, in principle, contain a large number of compounds. Combinatorial chemists have demonstrated in several systems that libraries containing 1,000-100,000 compounds can be assembled. Figure 1 shows a 1,4 benzodiazepine scaffold along with components that this scaffold might also be modified with[5]. The components or "building blocks" in many cases are very simple and can be readily bought and/or synthesised. From such a set of compounds it is easy to imagine that a library could contain 10^9 or more possible compounds. It is reasonable to presume that a "a virtual chemistry space" exists that perhaps contains 10^{100} possible molecules[2,3].

Combinatorial chemistry and HTS have advanced rapidly[3,4]. Many chemistries are still not amenable to rapid synthesis. Also screening a million compounds and following up hits is still a considerable effort. Both HTS and combinatorial chemistry labs are

confronted with the need to miniaturise and automate as ways to control costs, save time, reduce the volume of waste materials, etc[3,6,7].
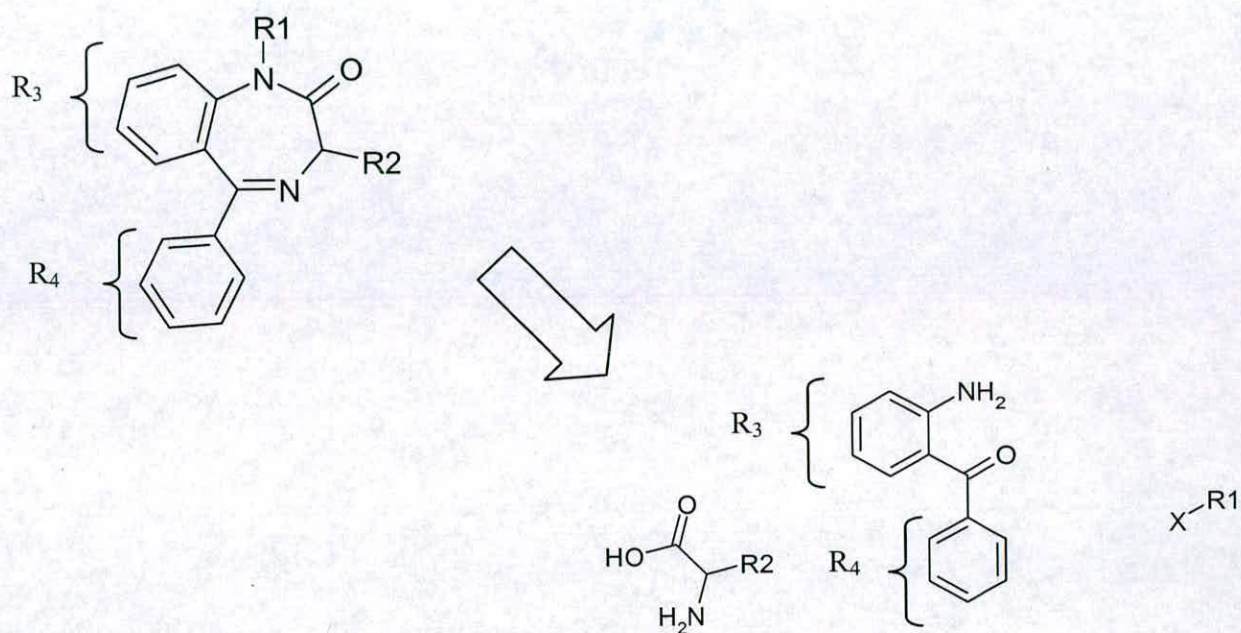
Figure 1: Generic 1,4-benzodiazepine showing locations of four variable R groups and the components from which the combinatorial library might be constructed[3].

5.2 Virtual screening.

To circumvent these problems a plethora of docking algorithms have been written that perform docking of large virtual libraries to structurally known sites and thus save on unnecessary compound screening[2,3,8]. Docking algorithms allow the user to search large virtual compound sets that have the right geometric and electronic features to fit the designated sites. This allows the user to commit fewer resources until after the in silico search has produced a refined set of compounds that can then be screened experimentally[6,8]. When the initial set of virtual hits is produced it is important to check the compounds for ease of chemical synthesis and modification and their stability. If they do not meet these criteria and produce positive results in the experimental screen they represent a dead end in the drug design cycle and use up valuable time and resources.

With the ability to assay ligand binding in eIF4E using either steady state fluorescence or mass spectrometry based assays, virtual screening could now be used to produce viable hits from commercially available combinatorial libraries. In this study, LIDAEUS[9] was used to carry out virtual screening. Docking protocols can be described as a combination of two components: a search strategy and a scoring function. The search algorithm should generate an optimum number of configurations that include the experimentally determined binding mode[8,10].

5.2.1 Search algorithms.

A rigorous search algorithm in an ideal world would exhaustively elucidate all possible binding modes between the ligand and the receptor. All six degrees of translational and rotational freedom of the ligand would be explored along with the internal conformational degrees of freedom of both the ligand and the protein[2,10]. However this is impractical due to the size of the search space. As a result only a small amount of conformational space can be searched through, and so a balance must be reached between the computational expense and the amount of the search space sampled[10].

Carrying out such an exhaustive search involves the sampling of many high energy unfavorable states which can restrict the success of an optimisation algorithm. To sample such a large conformational space the computational expense is limited by applying constraints, restraints and approximations to reduce the dimensionability of the problem in an attempt to locate the global minimum as efficiently as possible[10]. An early common approximation used in docking algorithms was to treat the ligand and the protein as rigid bodies and only explore the six degrees of translational and rotational freedom. The limitation of this rigid body docking method was that the ligand conformation must be close to the experimentally observed conformation when bound to the target. A more realistic approach in modelling molecular flexibility is to consider only the conformational space of the ligand, assuming a rigid receptor throughout the docking protocol[2,8,10].

5.2.2 Scoring functions.

Generating a diverse range of binding modes is ineffective without a model to rank each conformation that is both accurate and efficient. The scoring function should ideally be able to differentiate the experimental binding modes from all other modes sampled through the searching algorithm[10]. A rigorous scoring function will usually be expensive and so often the complexity of the function is reduced, with a resulting loss of accuracy. Scoring methods range from molecular mechanics force fields such as AMBER[11], CHARMM[12], etc, through to empirical free energy scoring functions or knowledge based functions[6].

Most current docking methods utilise the scoring function in one of two methods. One approach is to use the scoring function to rank a protein ligand conformation[6,10]. The system is then modified by the search algorithm, and the same scoring function is again applied to rank the new structure. The other method is to use a two stage scoring function[10]. Here a reduced function is used to direct the search strategy and a more rigorous scoring function is then used to rank the resulting structures. These directed methods make assumptions about the energy hypersurface, omitting computationally expensive terms such as electrostatics and considering only a few types of interactions such as hydrogen bonds[10]. Algorithms such as these are therefore directed to areas of importance as determined by the reduced scoring function. Examples of directed methods are Gold[13] and Dock[13].

A limitation in many existing scoring functions is to either neglect solvation effects or

use solvent models in a snap-shot fashion[10]. A snap-shot method involves the generation of structures in vacuo, that are then ranked with a scoring function that includes a solvent model. The search function is therefore directed to the conformational space which favours the in vacuo conformations. The role of bound solvent molecules and ions is usually not considered[10].

5.3 Chemical synthesis using the cap analogue as a template.

An alternative route to the design of suitable drug leads has been the stepwise modification of the natural ligand. In the literature there is a large body of information on nucleotide chemistry and on the modification of the guanine ring[14]. Of particular interest is the large body of work on N7 derivatives and their inhibition effects on eIF4E. These molecules offer potential insights into cap-binding interactions and may offer alternative directions into successful drug design[15].

5.3.1 N7 Guanine monophosphate derivatives .

It has been shown in the literature that eIF4E binds more tightly to GMP derivatives that contain benzene moieties instead of smaller alkyl substitutions such as –CH3 at the N7 position[16]. The first study that showed the preference of eIF4E for 7bzGMP over other N7 substituents such as ethyl, butyl, propyl, cyclopentyl etc, used an in vitro translation assay were inhibition of globin synthesis was monitored[16]. In the same report these results were also supported by UV cross-linking studies of radiolabeled mRNA to cap-binding proteins[16]. Further papers in the literature have reported, using the same in vitro translation assay, that modifying the benzyl ring at the para position with chloride
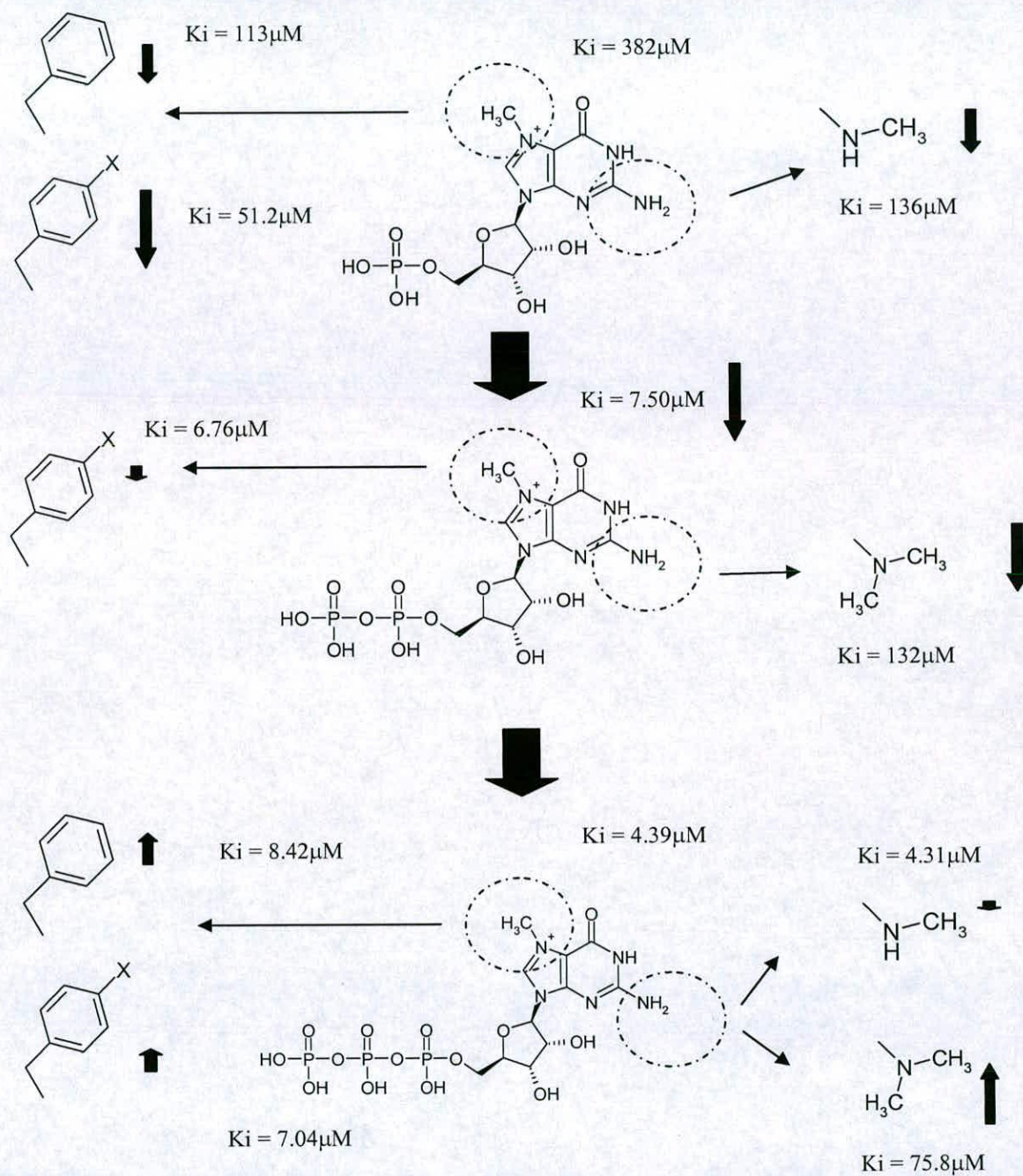
Figure 2: The above diagram based on results from the literature[15] shows the affects of N7 substitution and methylation of the imino group on the activity of the cap-analogue to eIF4E. For m7GMP if the benzylation and methylation are carried out together the derivative shows the most potency towards eIF4E (K_i = 13.9µM). X on the benzyl derivative represents the halogen fluorine or chlorine (K_is given for chlorine derivatives). Other N7 substitutions apart from the benzyl derivatives that increased potency are phenyl ethyl, ethyl, butyl, propyl and cyclopentyl[16]. The potency of m7GMP increases with phosphate addition[15].

increases potency of inhibition two-fold (see figure 2 and 3)[15]. The same paper also showed that substituting at the para position with a methyl or nitro group made the cap analogue a poorer inhibitor[15]. The para fluoride modification was not included in the monophosphate series analysis. Substituting 2-phenylethyl at the N7 position also improved efficacy over m7GMP.

The exocyclic amino group N2 of the 7-methylguanine moiety is known to be important for binding, since cap-analogues containing 7-methyllinosine or 7-methylxanthosine fail to inhibit mRNA binding to ribosomes[17]. Other studies have shown that a single methyl subsistent is tolerated[17,18]. Cai et al[15] shows that the addition of a single methyl group at N2 increases the effectiveness of cap-analogues as translational inhibitors. This "rule" holds for 7-methyl substituted nucleoside monophosphates, 7-(2-phenylethyl) - substituted nucleoside monophosphate, 7-benzyl-substituted nucleoside triphosphate, 7-(p-fluorobenzyl) nucleoside triphosphate, 7-(p-clorobenzyl) nucleoside triphosphate and others. The report by Cai et al[15] also confirms that modification of the exocyclic amino group by two methyl groups is detrimental to inhibition of the translational assay. Combining the positive effect of 7-(p-chlorobenzyl) substituents with the positive effect of N2-methyl substituents produced a compound with improved potency[15].

Compound	Reported K_i (μM)	Reference
7-methyl GMP	382	15
7-benzyl GMP	113	15
7-(p-chlorobenzyl) GMP	51.2	15
2-methyl-7-(p-chlorobenzyl) GMP	13.9	15
7-methyl GTP	4.39	15
7-benzyl GTP	8.42	15
7-benzyl -(p-chlorobenzyl) GTP	7.04	15
7-benzyl -(p-fluorobenzyl) GTP	7.35	15

Figure 3: Selected K_i values (Cai et al[15]) for N7 substituted m7G derivatives. Values were derived from a rabbit reticulate based lysate assay. The benzyl derivatives of GMP give a significant decrease in Ki against eIF4E over m7GMP. However in the case of m7GTP such a decrease is not observed.

Substitution of the benzyl group for the methyl at the N7 position was not beneficial in the triphosphate analogue (see figure 2 and 3)[15]. A marginal improvement was seen with p-halogen substitutions (chloride and fluoride) but these were also less effective than the 7-methyl analogue. In the diphosphate series, the 7-(p-clorobenzyl) substitution causes only marginal improvement. For the N2 –methyl nucleoside triphosphate series, the effect of substituting aryl for methyl at the N7 position produced a small improvement. However this is negligible compared to the monophosphate series[15].

Essentially, 7-aryl substitutions have a strong positive influence on the weaker nucleoside monophosphate inhibitors (as long as the ring possess no methyl or nitro groups), but the effect of 7-aryl substituents on the stronger nucleoside diphosphates and triphosphates is marginal and varies from compound to compound[15] (see figure 2). However, the 2-methyl-7-(p-chlorobenzyl) analogue was shown by Cai et al to be the best inhibitor in the nucleoside mono-, di- and triphosphate series[15] (see figure 2).

Later work by Ghosh et al[19] using a competition assay based on m7GTP resin beads report that increasing the bulk of the N7 substitution with either athyl, allyl or benzyl groups reduced inhibitor potency by 2- to 4- fold, in contrast to the results from Cai at al[15] for the monophosphate series.. The reason they suggest for this disagreement is that it may stem from differences in the conformation of eIF4E bound to 7-methyl GTP resin than when it is actively engaged in translation[19].

5.4 Virtual screening with LIDAEUS.

5.4.1 Fundamentals of LIDAEUS.

LIDAEUS[9] is a software tool written by P Taylor at the ISMB, University of Edinburgh. It works by using a solved ligand protein crystal structure to define the site on the protein to be used for virtual screening. When LIDAEUS[9] has been used to define the site to be searched it generates site points where the ligand is located. These can either be designated as hydrogen bond acceptor, hydrogen bond donor, hydrophobic and especially hydrophobic sites. These are used primarily for docking compounds. Compounds from the virtual library to be screened are docked into the sites in succession. LIDAEUS[9] attempts to fit the ligand on to the site points and scores each pose. After a set number of attempts if the compound does not meet a threshold score it is rejected. In the next stage the ligand is more rigorously fitted into the active site and scored. LIDAEUS[9] retains a specified number of the best scoring compounds.

5.4.2 LIDAEUS virtual screening results.

The full-length human eIF4E in complex with m7GTP, crystallised by Tomoo et al[20] (1IPC), was used for the LIDAEUS[9] screening run. The m7GTP ligand was used to target the cap-binding slot in order to generate the site points to be used in the LIDAEUS[9] docking protocol. LIDAEUS[9] was then used to screen the Maybridge combinatorial chemistry library. The initial virtual screen returned 250 ranked compounds. The top 30 compounds were then analysed to reveal what features seemed to be important in selection of compounds by LIDAEUS[9]. The hits were examined on the basis of what features they contained that mimicked those found in m7GTP. These

146

ranged from satisfying the hydrogen bond donors and acceptors of the m7G ring, whether or not a group of some kind overlaid the triphosphate tail, whether or not there was any chemical group occupying space at the back of the binding site and what type of ring system was intercalating between the two aromatic residues. Figure 4 gives graphical examples of the types of ligands and features selected by LIDAEUS[9], respectively.

In the figure 4 the compounds fall into several families that satisfy some or all of the hydrogen bonding partners that recognised the m7G ring. These families are ring systems that intercalate between the aromatic residues, ring systems that intercalate and have a chemical group of some type occupying the position in space of the m7GTP triphosphate tail, compounds that have no ring system sandwiched between the two tryptophans but do have a group in the triphosphate tail position and compounds that have a chemical group that project into the back of the cap-binding pocket (see figure 4E). The fourth family does have some overlap with the first family: compounds that intercalate and project into the back of the pocket (see figure 4D). The definition for ring systems used here is reasonably loose but encloses phenyl, napthyl and morpholene. Figures 4 and 5 show several of the chemical group structures that were orientated in the phosphate-binding site (i.e. overlapping the phosphates in m7GTP) and the ring systems intercalated between the two ring systems, respectively. The potential phosphate mimics identified by LIDAEUS (figure 5) could be used in the design of future ligands, if a drug lead is identified. In figure 6 the ring systems shown all have the potential of forming some of the hydrogen bonds formed by m7G and are capable of π-π stacking

interactions with W56 and W102. However none possess a positive delocalised charge on the ring, like that seen on m7GTP, but some of the ring systems identified by LIDAEUS[9] can be alkylated to induce such a charge on the ring system. The ligands that do possess a nitrogen containing ring systems may be protonated at physiological pHs that are found in the cell, which would also induce a positive delocalised charge on the ring system. However it is highly difficult and unreliable to predict whether or not these compounds could be protonated at physiological pHs.

The classical approach for calculating dissociation constants (pK_a) from molecular structure is given by the Hammet equation, which is based on a separation of the compound of interest into a suitable parent structure and substituents with associated increment values[21]. An alternative approach is given by knowledge based systems that make use of increment values for more generally defined features[21]. However these methods are limited by requiring a suitable catalogue of experimental data on related compounds[21]. In other words, pK_as can only be reliably predicted for compounds very similar to those in the training set. This limitation also applies to semi-empirical treatments using quantum chemical continuum-solvation methods[22]. pK_as can also be calculated using *ab initio* methods, however these would be very computationally expensive and time consuming[22]. Especially if they were to be used on large databases of molecules with diverse structure and size.

A

B

C

D

E

A) ((NH$_2$)$_2$CH$_2$)NHNHCH$_2$CH$_2$NH(CH$_2$(NH$_2$)$_2$))

B) C$_8$O$_2$NH$_5$C$_4$OH$_2$NO$_2$

C) NO$_2$C$_6$H$_4$CH$_2$NHNHCOCH$_2$SO$_2$CH$_2$C$_6$H$_4$CFL$_3$

D) FlC$_6$H$_4$CH$_2$C$_3$N$_2$H$_2$CH$_3$NO$_2$

E) NH$_2$NHCOCH$_2$NHCOC$_6$H$_4$NO$_2$
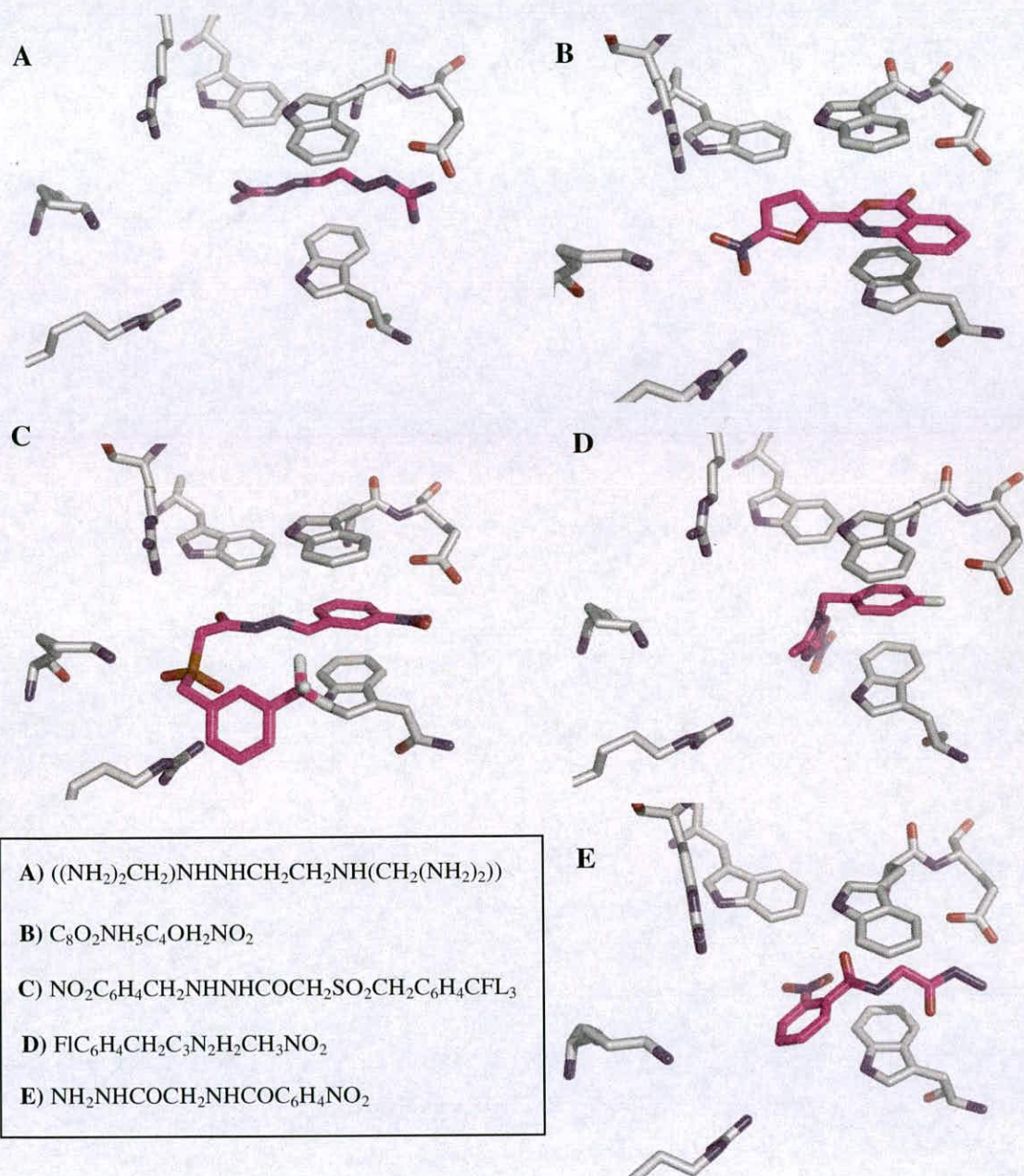
Figure 4: Virtual hits representative of some of the apparent families of ligands being selected by LIDAEUS. Panels A and E show ligands that possess no intercalating ring system, but still sandwich themselves between W102 and W56, and project into the back of the binding. Panels B and D show ligands that have a ring system intercalated between the two tryptophans and that also have a chemical group projecting into the cavity of the cap-binding site. Panel C depicts a ligand that has a chemical group in the same position as the triphosphate tail of m7GTP and again that intercalates a ring system between the two tryptophan. Molecular formulae for virtual hits shown in the inset box.

149
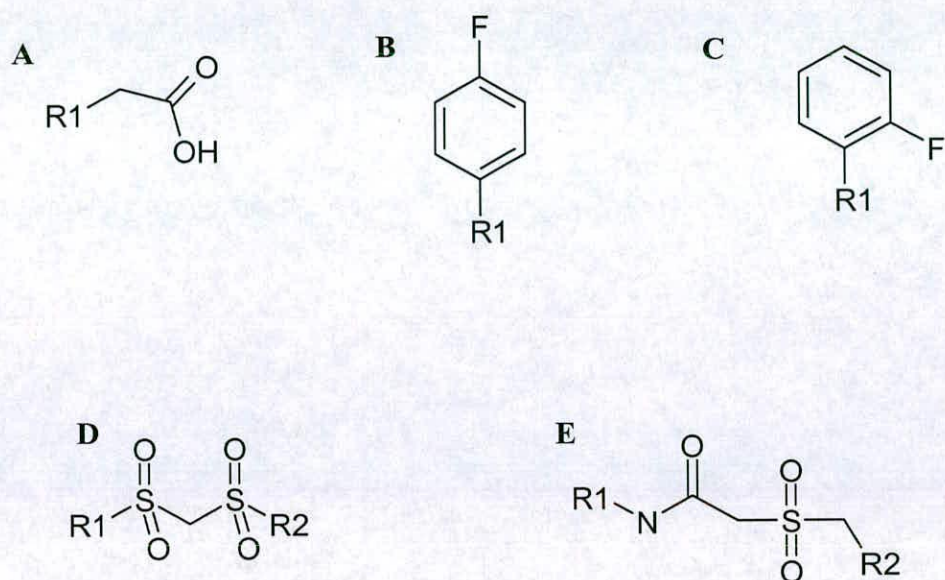
Figure 5: Above are shown the chemical groups that LIDAEUS overlaid with the triphosphate tail of m7GTP. In future ligand design efforts these groups could be useful potential phosphate mimics to investigate. Ligands that contain phosphate groups have poor ADME properties and are not membrane permeable, which makes them poor candidates for drug development.
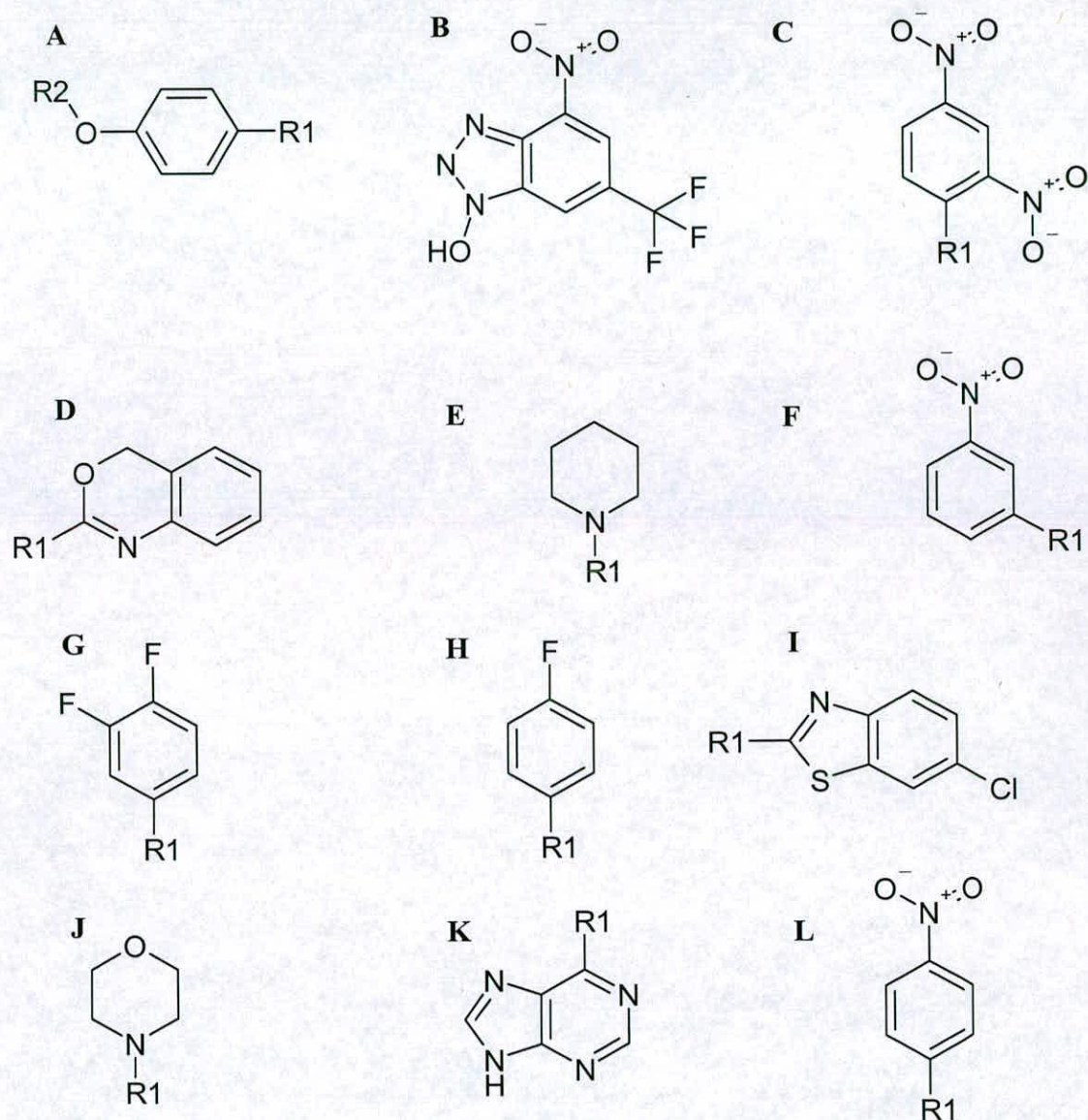
Figure 6: Above are shown ring systems that LIDAEUS intercalated between the W56 and W102 trptophans. The majority are planar ring systems except for E and J. All possess features that allow them to form potential hydrogen bonds to E103 and the backbone amide of W102. All the planar ring systems are capable of forming π stacking interactions with the two tryptophan residues. However none possess a delocalised positive charge on the ring system. Some of the nitrogen containing heterocyclic ring systems, such as B, I and K could be alkylated to induce a positive delocalised charge on the ring systems.

5.5 Ligand synthesis of cap-analogues.

Maintenance of the N7 positive delocalised charge on the guanine ring and the presence of a phosphate group seem to be the two deciding factors on whether a ligand binds to eIF4E or not based on the mass spectrometry studies (see chapter 4). With this information a set of ligands were synthesised using different scaffolds with N7 equivalent modifications in order to give them a positive delocalised charge. The scaffolds used for the N7 substitutions were guanine monophosphates, quinolene, isoquinolene and benzothiazole. The modifications themselves ranged from ethyl to benzyl groups. Figure 8 to 11 show the N7 analogue compounds that were successfully synthesised.

The scaffolds were chosen for several reasons. Guanine monophosphate was chosen so that the original features of the cap structure were retained: the HA/ HD pattern on the guanine ring and the presence of a phosphate group. The quinolene and quinene ring systems were chosen to explore whether inhibition could be achieved with no phosphate group and no HD/HA present. Also these ring systems gave us the opportunity to investigate whether the positioning of the N on the double ring systems and the subsequent reorientation of any N7 modification was important. Benzothiazole was chosen for the same reasons as well as the fact that it contained a sulphur atom that could be used as a marker in any resulting electron density map if the compound proved active. Also by carrying out these reactions the applicability of this reaction for modifying basic ring structures was being investigated.

The groups chosen to be substituted on to the heterocyclic ring systems and GMP were selected for one of two reasons: either there was literature precedent or modeling studies carried out using LUDI[10] indicating whether or not there was space to accommodate the modification in eIF4E's cap binding site. The benzyl modification was of particular interest as it has been shown to improve the inhibitory properties of m7GMP in relation to eIF4E.

The modifications were made to the respective ring systems by using an alkylation reaction[17]. The halide alkyl was incubated with the scaffold to be modified for a period of 48 hours under continuous mixing for a period of two days in DMSO or DCM. The halide alkyl modifies the ring system of the scaffold by an S_N2 substitution reaction (see figure 7).
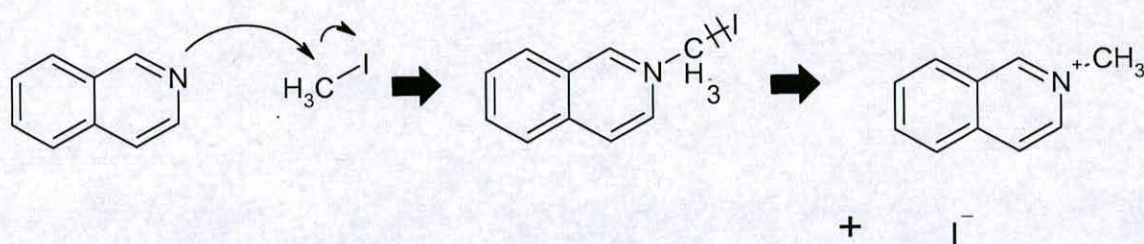
Figure 7: Example of an S_n2 substitution using quinolene and iodomethyl. The nucleophile, in this case the nitrogen containing ring system, attacks the alkyl halide (ICH$_3$) 180° away from the halogen. The nucleophilic attack generates a transition state with simultaneous bond breaking and bond formation occurring. The transition state then breaks down to form the substituted N7 analogue compound and a free halide ion.

After incubation the reaction was quenched with ice-cold acetone or distilled water and centrifuged. The top layer was decanted of and the bottom layer was washed via several

153

rounds of centrifugation with ethyl ether. The resulting crude extract was dissolved in 90% H_2O and 10% methanol and analysed using an electrospray mass spectrometer to ensure the compound being synthesised had been made. The compounds described, due to the presence of a positive charge, should fly at their exact molecular weight. In the case of the GMP derivatives they will fly in the form where the phosphate tail is fully protonated (see figure 11).

Methyl quinolene

Calculated mass = 144

Benzyl quinolene

Calculated mass = 220
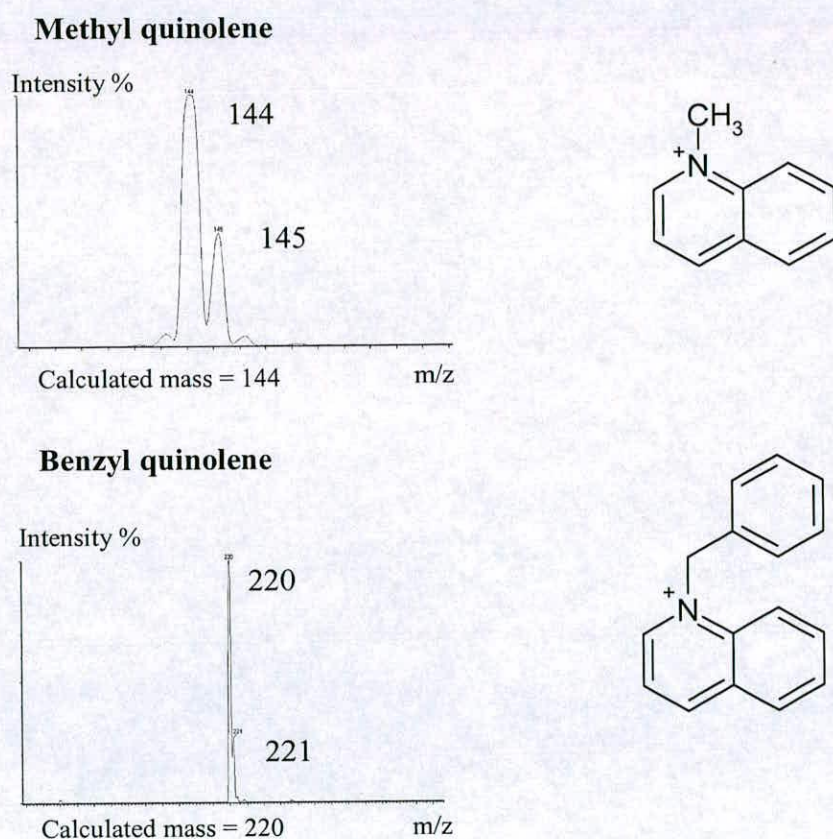
Figure 8: The figure shows the compounds based on the quinolene scaffold structure that were successfully synthesised, with the corresponding experimental mass spectra confirming their molecular mass and presence in the reaction mix. Due to the presence of a positive charge on the nitrogen the quinolene based compounds flew at their calculated molecular weight.

154

Ethyl isoquinolene

Intensity %

158

159

m/z
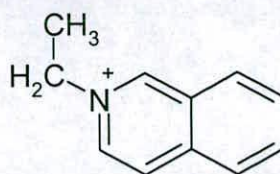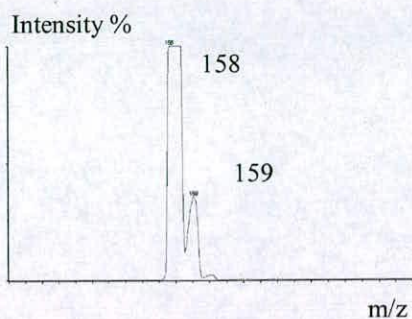
Calculated mass = 158

Methyl isoquinolene

Intensity %

144

145

m/z
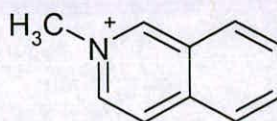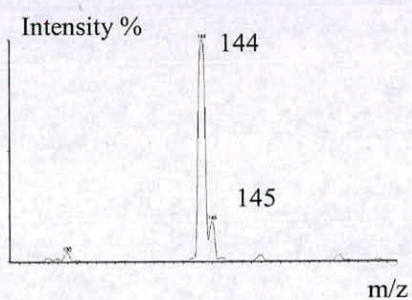
Calculated mass = 144

Benzyl isoquinolene

Intensity %

220

221

m/z
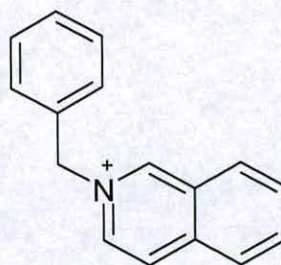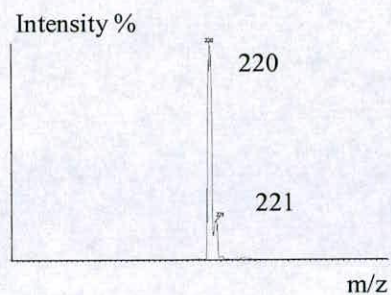
Calculated mass = 220

Figure 9: The figure shows the compounds based on the isoquinolene scaffold structure that were successfully synthesised, with the corresponding experimental mass spectra confirming their molecular mass and presence in the reaction mix. Due to the presence of a positive charge on the nitrogen the isoquinolene based compounds flew at their calculated molecular weight.

155

Ethyl benzothiazole
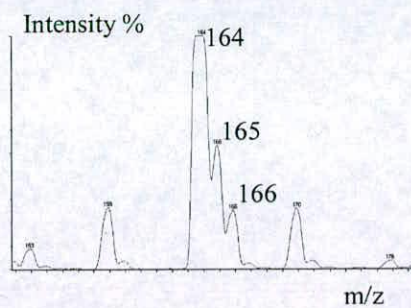
Calculated mass = 164
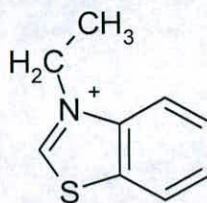
Benzyl benzothiazole
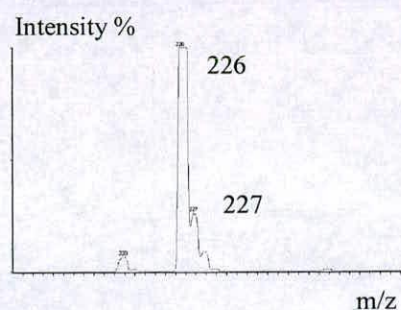
Calculated mass = 226
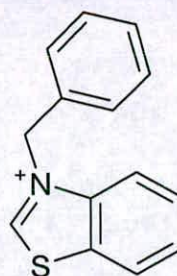
Figure 10: The figure shows the compounds based on the benzothiazole scaffold structure that were successfully synthesised, with the corresponding experimental mass spectra confirming their molecular mass and presence in the reaction mix. Due to the presence of a positive charge on the nitrogen the benzothiazole based compounds flew at their calculated molecular weight.

7-benzyl guanine monophosphate

Calculated mass = 454

7-methyl guanine monophosphate

Calculated mass = 378   m/z

7-*p-fluoro*benzyl guanine monophosphate

Calculated mass = 472   m/z

7-ethyl guanine monophosphate

Calculated mass = 392   m/z
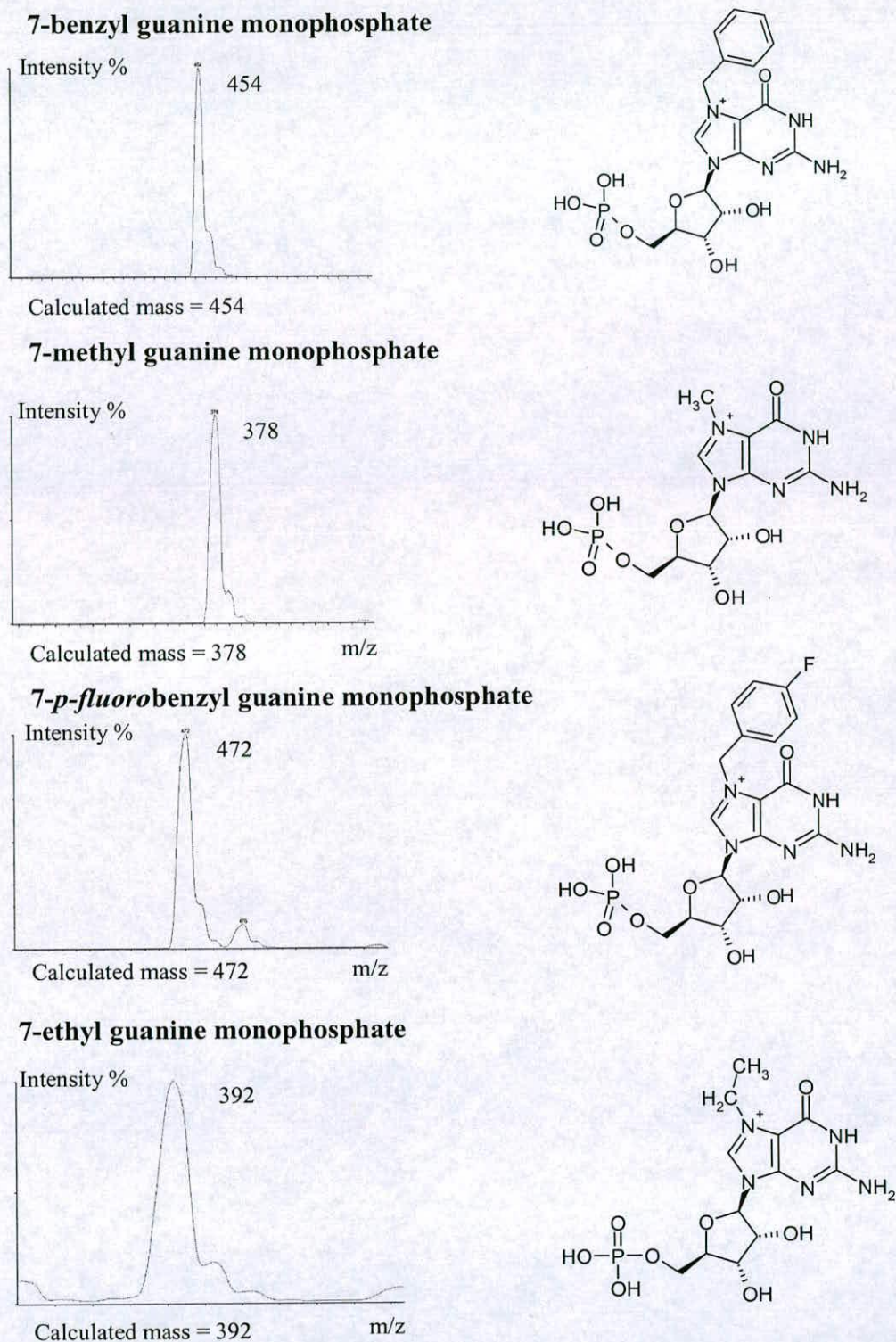
Figure 11: The figure shows the compounds based on the GMP scaffold structure that were successfully synthesised, with the corresponding experimental mass spectra confirming their molecular mass and presence in the reaction mix. Due to the presence of a positive charge on the nitrogen the GMP based compounds flew at their calculated molecular weight for the form where the phosphate group is fully protonated as shown above.

157

5.6 Data-mining for potential ligands with an analogous charge to the N7 nitrogen on the m7G moiety using ISISBASE.

As well as synthesizing this small family of potential inhibitors of eIF4E a variety of compounds were purchased that contained an N7 equivalent delocalised positive charge. Compounds similar to m7GTP can also be searched for using 2D-similarity methods e.g. ISISBase (MDL information systems, Inc.). 2D methods are rapid and allow a small set of compounds to be selected quickly and efficiently based on certain features. Apart from the requirement for some type of quaternary cation, not necessarily located in a ring system, there were no other limits set on the search. Surprisingly few compounds were found in the search especially cationic nitrogen containing ring systems. In figure 12 the compounds identified, apart from containing a delocalised positive charge, are on the whole bulky, contain no obvious phosphate mimics or numerous hydrogen bond donors or acceptors.
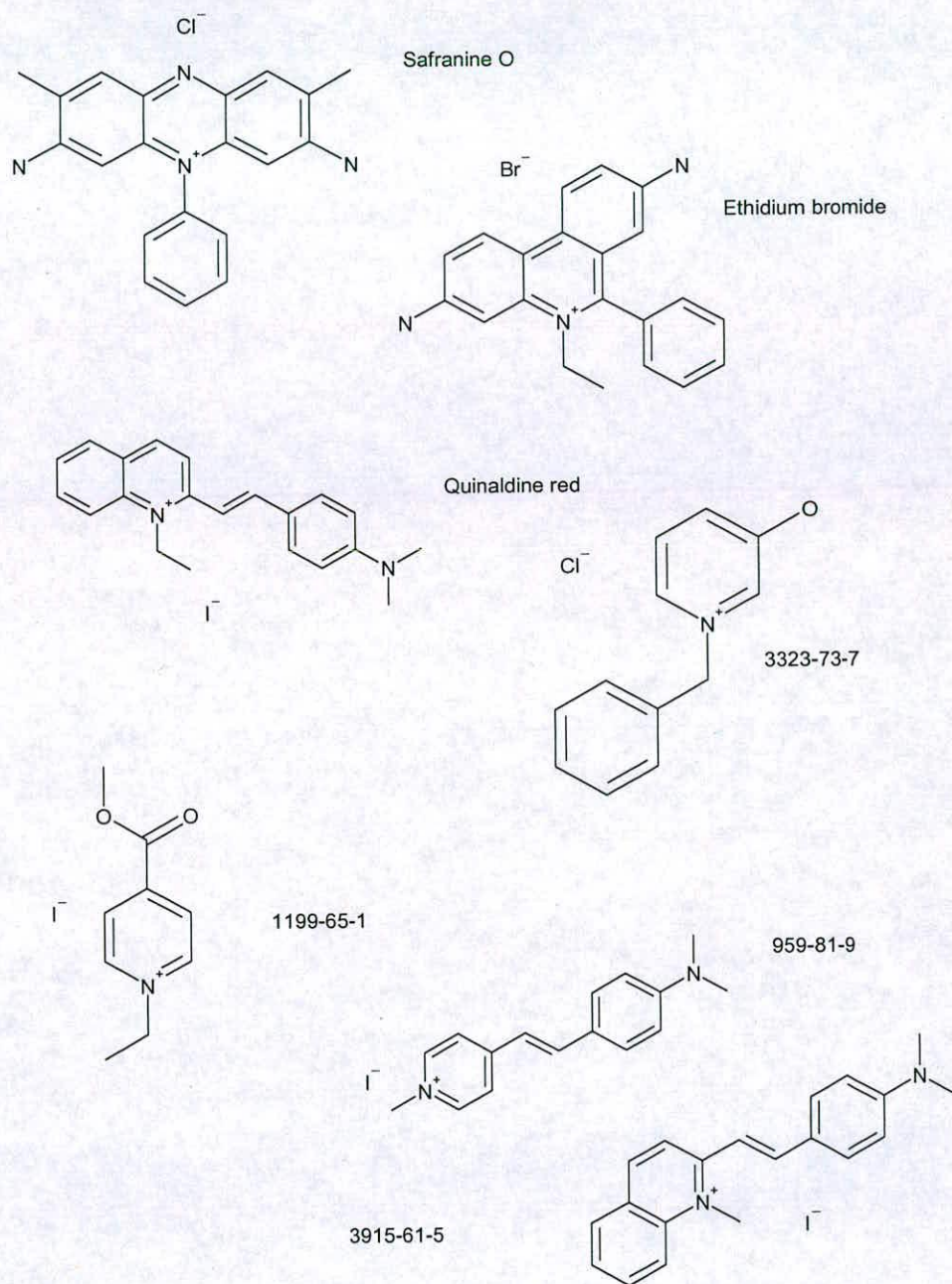
Figure 12: The figure shows a selection of compounds identified through ISISBase. Search criteria consisted of searching for compounds from the ACD that had at least had a positively charged nitrogen on the ring system. These compounds on the whole are bulky, do not possess any groups that are known phosphate mimics and have few hydrogen bond donors or acceptors.

159

5.7 Experimental screening using mass spectrometry.

Mass spectrometry was used as the primary screening technique on the basis that it has a greater potential of screening compounds more quickly and robustly than intrinsic tryptophan fluorescence. The disadvantages of intrinsic tryptophan fluorescence for screening of eIF4E are; many of the potential ligands absorb at the wavelength required for excitation, they can also fluoresce in the same wavelength regions as tryptophan itself, solubility issues can give rise to light scattering effects adding to the absorbance problem and also solvents used to dissolve these ligands can denature the protein and alter its fluorescence properties.

5.7.1 Mass spectrometry screening.

Mass spectrometry as shown in the previous chapter can be used to detect m7GTP and binding of other related nucleobases, and differentiate between what we have termed "specific" and "non-specific" binding interactions. Mass spectrometry lends itself easily to ligand screening for eIF4E in contrast to quenching of intrinsic tryptophan fluorescence. Unlike protein fluorescence mass spectrometry does not require as many control experiments e.g. ligand fluorescence, ligand absorbance. The only control experiments needed are to fly eIF4E with whatever solvent the ligand has been dissolved in e.g. methanol, DMSO. In the case of methanol if it is maintained at a constant 10% in the sample to be analyzed then all the control experiments have been carried out in chapter 4. These control experiments are still required for DMSO. For these experiments methanol was kept at 10% and the addition of DMSO with ligand to an absolute minimum ($\leq$1%). The main effect from addition of DMSO was a shift in the

charge distribution seen at pH 5.0 from the 10+ charge species to the 8+ charge species (see figure 13). With addition of DMSO to the standard sample conditions (10% methanol and 10mM ammonium acetate pH 5.0) the GTP and m7GTP eIF4E complexes are still detectable (figure 13). However there is a noticeable drop in signal intensity presumably due to the viscous DMSO interfering with the process of ionisation in the electrospray source. Also the spectra with DMSO have more noise (see figure 5B and 5C) that can make identification of bound ligand difficult. The cause of the noise is presumably due to aggregation with DMSO that has not desolvated completely during ionization. However by keeping DMSO levels to an absolute minimum, ligands that require it can be screened against eIF4E using mass spectrometry.
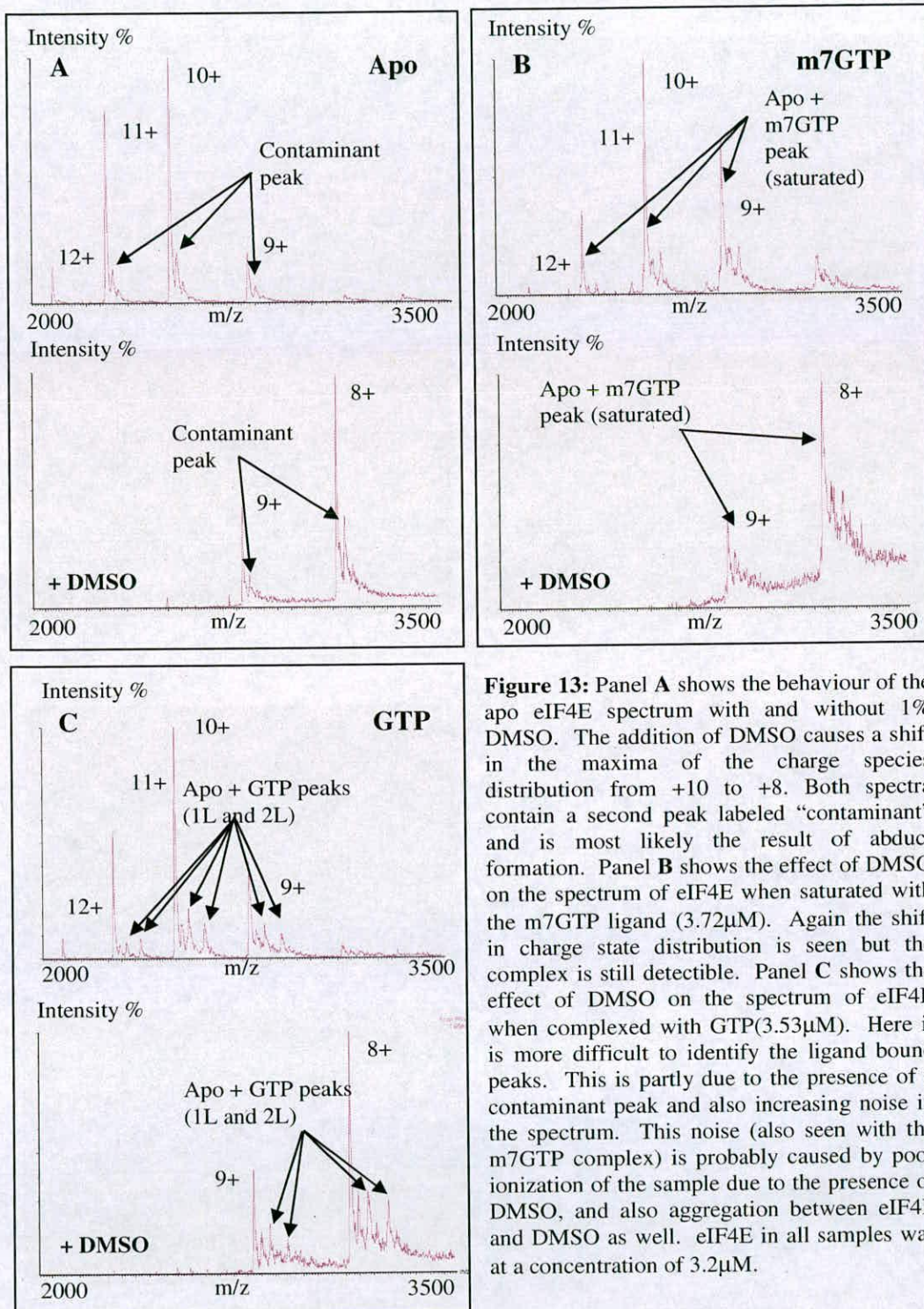
Figure 13: Panel **A** shows the behaviour of the apo eIF4E spectrum with and without 1% DMSO. The addition of DMSO causes a shift in the maxima of the charge species distribution from +10 to +8. Both spectra contain a second peak labeled "contaminant" and is most likely the result of abduct formation. Panel **B** shows the effect of DMSO on the spectrum of eIF4E when saturated with the m7GTP ligand (3.72μM). Again the shift in charge state distribution is seen but the complex is still detectible. Panel **C** shows the effect of DMSO on the spectrum of eIF4E when complexed with GTP(3.53μM). Here it is more difficult to identify the ligand bound peaks. This is partly due to the presence of a contaminant peak and also increasing noise in the spectrum. This noise (also seen with the m7GTP complex) is probably caused by poor ionization of the sample due to the presence of DMSO, and also aggregation between eIF4E and DMSO as well. eIF4E in all samples was at a concentration of 3.2μM.

162

A criterion was needed to differentiate between specific and non-specific binders in the mass spectrometry screening experiments. It was decided using the GTP, GDP and GMP titration experiments as a guide that 50% sat should not be equal to 30% 1L and 10% 2L peaks i.e. the 1L% should be higher and the 2L% lower for the interaction with eIF4E to be considered specific. If a ligand meets this requirement then the ligand can be recognized as a valid hit against eIF4E and used in co-crystallisation experiments.

5.7.2 Mass spectrometry screening results.

All ligands identified from the virtual screens were dissolved in DMSO as a 10mM stock. Compounds from the mini-library were dissolved in a 100% methanol at 20mg/ml. The concentrations of these compounds was unknown as they were never purified to a 100% purity, so any concentrations mentioned in respect to this library are only apparent. Thus mass spectrometry in these cases can only be used to see if binding occurs and not to assess the strength of the interaction. All compounds identified using ISISBase were also dissolved in 100% methanol as a 10mM stock.

The compounds from the virtual screening and ISISBase searches were found not to bind eIF4E under the screening conditions used, which were 10mM ammonium acetate pH5.0 10% methanol (for mass spectrometer parameters see methods and materials). Compounds from the mini-library based on the quinolene, isoquinolene and benzothiazole ring systems also showed no interaction with eIF4E of any kind in the gas-phase. These results are not surprising considering the results seen in chapter 4 that indicate a positive delocalised charge and a negative phosphate tail are required for

binding to occur to eIF4E. The hits generated from virtual screening, as stated early, did not contain a positive delocalised charge on a ring system, the hits from the ISISBase searches were probably either too bulky and also contained no phosphate mimic group as did the compounds from the mini-library mentioned. However N7 substituted GMP analogues did exhibit binding in the gas phase with the following group: methyl, ethyl, benzyl and *para*fluoro-benzyl (see figure 14). These compounds unlike the others screened possessed a positive delocalised charge on a ring system (guanine in this case) and a single negatively charged phosphate group.
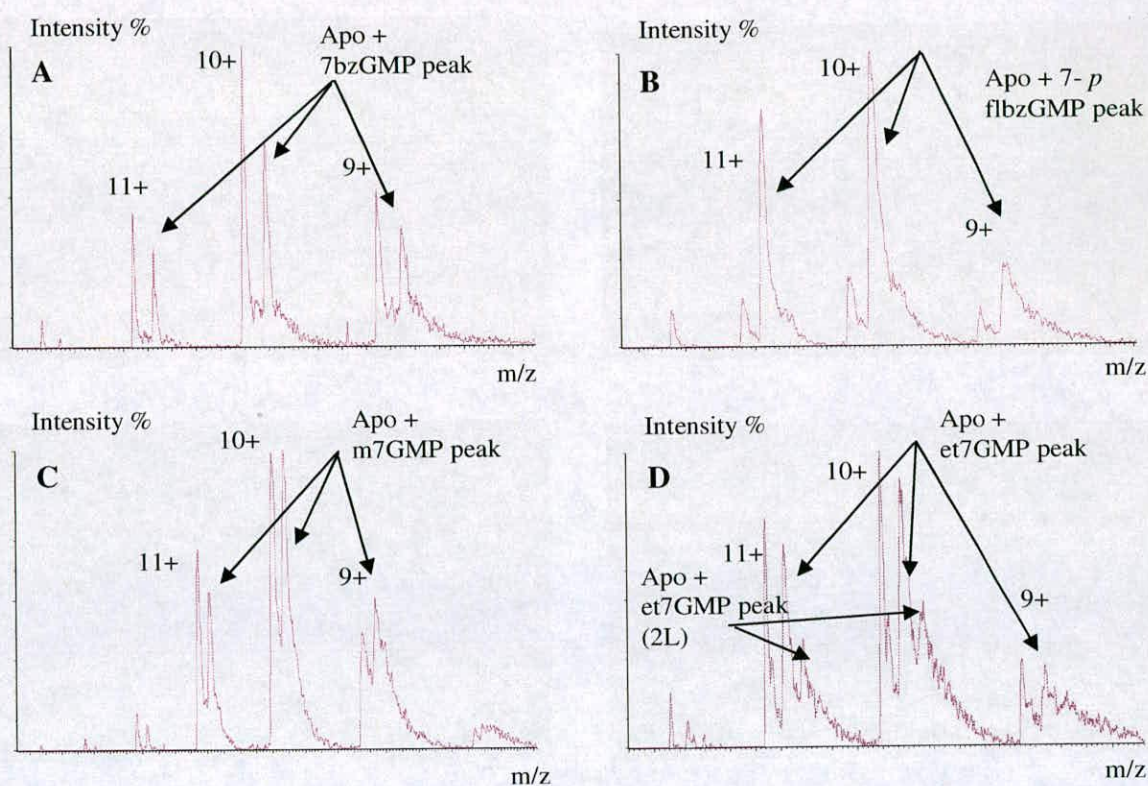
Figure 14: The figure shows the positive hits from the mass spectrometry screening. Panel **A** shows 7-benzyl GMP (~4.4μM) saturating eIF4E. Panel **B** shows 7-(*p*-fluorobenzyl) GMP (~4.23μM) binding to eIF4E and Panel **C** shows m7GMP (~5.29μM) interacting with eIF4E. Panel **D** shows 7-ethyl GMP (~50μM) binding to eIF4E. These hits all saturate the apo peak to ~50% with out giving rise to the 2L peak, associated with the non-specific effect discussed in chapter 4. The exception to this is et7GMP that exhibits development of a 2L peak. All samples were run with 3.2μM of eIF4E.

The hits did not exhibit the presence of the apo + 2L peak associated with non-specific binding (see chapter 4), except et7GMP as indicated in figure 14D. The occurrence of this 2L peak could possibly indicate that out of the compounds which have shown an interaction with eIF4E that et7GMP one has the weakest interaction with the protein.

5.8 Conclusion.

Virtual screening methods selected a variety of compounds that initially appeared likely candidates to bind to eIF4E, possessing features ranging from ring intercalation between the W102 and W56 and phosphate mimicry. The majority possessed features that could mimic the hydrogen bonds that recognised the m7G ring. However none of these compounds showed binding in the mass spectrometry based screens. Also no activity was seen for compounds extracted from the ACD using ISISBase or from the compounds based on the quinolene, isoquinolene and benzothiazole rings. These results support the observations made in chapter 4 that a phosphate group and a delocalised charge are required for successful interaction with eIF4E. Features that the N7 substituted GMP analogues possessed and which were screened successfully against eIF4E.

It would appear that virtual screening results have not succeeded in producing viable hits against eIF4E; however there are several possible avenues to pursue to overcome this apparent dead end. The initial library used for screening could be pruned to contain compounds that possess a delocalized positive charge on a ring system or that possess

ring systems that could be chemically modified to induce such a charge. Nitrogen containing ring systems may be protonated at physiological pHs which would also give the ring a positive delocalised charge. The methods required to predict this are either not reliable enough if there is no prior experimental data or computationally expensive. It would be far easier just to screen and see if an interaction with eIF4E occurs. There are positives from the LIDAEUS screen; several possible groups appear in the screen that could be possible mimics for the phosphate groups in m7GTP and several ring systems that could be alkylated to give a positive charge have also been recognised. Another valid approach to attempt in order for LIDAEUS to produce hits with a bigger chance of interacting with eIF4E is incorporation into the scoring model of the cation-π interaction.

Mass spectrometry has been shown to be a tool that can be used to screen eIF4E for potential cap-analogue mimics. It managed to identify all the N7 subistituted GMP analogues as potential leads which have reported K_is of 382μM to 13.9 μM. Also the alklylation reaction has been shown as a tool, that as well as modifying the GMP scaffold (as previously described), it can also be used to modify other ring systems such as quinolene, isoquinolene and benzothiazole. The hits for 7bzGMP, 7- *pf*lbzGMP and m7GMP can now be forwarded to cocrystallisation trials with eIF4E.

The main aim for any further drug design efforts is to modify potential scaffolds in order to possess a positive delocalised charge and move away from the need to have a

phosphate group.   The most apparent being to try and find chemical groups that can

mimic the phosphate.

Reference List

1. Carberry,S.E., Rhoads,R.E. & Goss,D.J. A spectroscopic study of the binding of m7GTP and m7GpppG to human protein synthesis initiation factor 4E. *Biochemistry* **28**, 8078-8083 (1989).

2. Shoichet,B.K. Virtual screening of chemical libraries. *Nature 2004. Dec. 16. ;432. (7019.):862. - 5.* **432**, 862-865 (2004).

3. Walters,W.P., Stahl,M.T. & Murcko,M.A. Virtual screening - an overview. *Drug Discovery Today* **3**, 160-178 (1998).

4. Goodnow,R.A., Jr., Guba,W. & Haap,W. Library design practices for success in lead generation with small molecule libraries. *Comb. Chem. High Throughput. Screen. 2003. Nov. ;6. (7.):649. - 60.* **6**, 649-660 (2003).

5. Bunin,B.A., Plunkett,M.J. & Ellman,J.A. The combinatorial synthesis and chemical and biological evaluation of a 1,4-benzodiazepine library. *Proc. Natl. Acad. Sci. U. S. A* **91**, 4708-4712 (1994).

6. Schneider,G. & Bohm,H.J. Virtual screening and fast automated docking methods. *Drug Discov. Today 2002. Jan. 1;7. (1):64. -70.* **7**, 64-70 (2002).

7. Battersby,B.J. & Trau,M. Novel miniaturized systems in high-throughput screening. *Trends Biotechnol. 2002. Apr;20. (4):167. -73.* **20**, 167-173 (2002).

8. Lyne,P.D. Structure-based virtual screening: an overview. *Drug Discov. Today 2002. Oct. 15. ;7. (20.):1047. -55.* **7**, 1047-1055 (2002).

9. Wu,S.Y. *et al.* Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure. (Camb.) 2003. Apr;11(4):399. -410.* **11**, 399-410 (2003).

10. Taylor,R.D., Jewsbury,P.J. & Essex,J.W. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des 2002. Mar. ;16. (3):151. -66.* **16**, 151-166 (2002).

11. Cornell,W.D. *et al.* A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **117**, 5179-5197 (1995).

12. Brooks,B.R. *et al.* Charmm - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **4**, 187-217 (1983).

13. Jones,G., Willett,P., Glen,R.C., Leach,A.R. & Taylor,R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267**, 727-748 (1997).

14. Darzynkiewicz,E., Ekiel,I., Lassota,P. & Tahara,S.M. Inhibition of eukaryotic translation by analogues of messenger RNA 5'-cap: chemical and biological consequences of 5'-phosphate modifications of 7-methylguanosine 5'-monophosphate. *Biochemistry* **26**, 4372-4380 (1987).

15. Cai,A. *et al.* Quantitative assessment of mRNA cap analogues as inhibitors of in vitro translation. *Biochemistry* **38**, 8538-8547 (1999).

16. Darzynkiewicz,E. *et al.* Inhibition of eukaryotic translation by nucleoside 5'-monophosphate analogues of mRNA 5'-cap: changes in N7 substituent affect analogue activity. *Biochemistry* **28**, 4771-4778 (1989).

17. Adams,B.L., Morgan,M., Muthukrishnan,S., Hecht,S.M. & Shatkin,A.J. The effect of "cap" analogs on reovirus mRNA binding to wheat germ ribosomes. Evidence for enhancement of ribosomal binding via a preferred cap conformation. *J. Biol. Chem.* **253**, 2589-2595 (1978).

18. Darzynkiewicz,E. *et al.* Beta-globin mRNAs capped with m7G, m2.7(2)G or m2.2.7(3)G differ in intrinsic translation efficiency. *Nucleic Acids Res.* **16**, 8953-8962 (1988).

19. Ghosh,P. *et al.* Synthesis and evaluation of potential inhibitors of eIF4E cap binding to 7-methyl GTP. *Bioorg. Med. Chem. Lett. 2005. Apr 15. ;15. (8.):2177. -80.* **15**, 2177-2180 (2005).

20. Tomoo,K. *et al.* Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m(7)GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region. *Biochem. J. 2002. Mar. 15. ;362. (Pt. 3):539. -44.* **362**, 539-544 (2002).

21. Schuurmann,G., Cossi,M., Barone,V. & Tomasi,J. Prediction of the pK(a) of carboxylic acids using the ab initio continuum-solvation model PCM-UAHF. *Journal of Physical Chemistry A* **102**, 6706-6712 (1998).

22. Chaudry,U.A. & Popelier,P.L. Estimation of pKa using quantum topological molecular similarity descriptors: application to carboxylic acids, anilines and phenols. *J. Org. Chem. 2004. Jan. 23;69. (2):233. -41.* **69**, 233-241 (2004).

Chapter 6: Crystallisation of eIF4E with N7 substituted cap analogues and insights into structure based drug design for eIF4E.

6.0 Introduction.

The N7 substituted cap-analogues, synthesised and screened in the last chapter, were used for co-crystallisation trials. The aims were to discover how the large bulky benzyl group was accommodated in the cap-binding site and how the monophosphate group also interacted with eIF4E. Successful crystallisation would also allow the reported tighter binding of the *para* halogen substituted benzyl derivatives over the non-substituted variants to be structurally rationalised.

6.1 Prior proposed model for the binding of 7-bzGMP and 7-*p*clbzGMP to eIF4E.

As discussed in chapter 5 it has been reported that the benzyl substitution of the N7 position in GMP increases the strength of its interaction 3-fold (K_i = 113µM) with eIF4E as compared to m7GMP (K_i = 382µM), and that the *p*-chloro-benzyl substitution (K_i = 51.2) results in a two-fold increase in affinity compared to the benzyl only substitution[1]. It has also been reported that the higher affinity the benzyl substitutions have for eIF4E arises from a favourable entropy change[2]. Several explanations for the strong changes in binding entropy of the benzylated cap-analogues been proposed. The most plausible being that the positive entropy change is caused by expulsion of a few water molecules from the depth of the cap-binding slot into the bulk solvent, proportional to the size of N(7)-substituent. In the case of the *p*-cloro-benzyl substitution it has been proposed that the large and electronegative

chloride atom at the *N*7-benzyl ring can attenuate the van der Waals interaction with Trp166 and destroy the specific water network inside the eIF4E cap-binding centre due to steric and electrostatic effects[2].

6.1.2 Modelling the benzyl derivatives into the cap-binding site of eIF4E.

The full-length human m7GTP eIF4E complexed to the 4E-BP1 motif peptide crystal structure was used to model in N7 benzylated eIF4E. Modelling was carried out manually using PYMOL[3] to orientate the benzyl group into the cap-binding site.
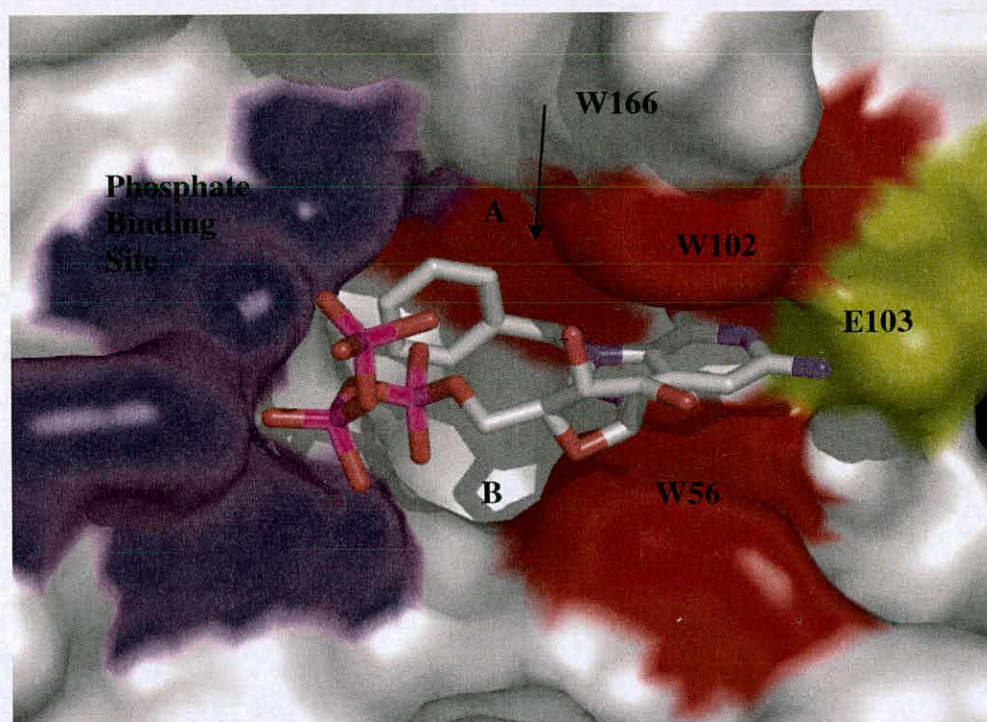
Figure 1: Manual modelling of N7 benzylated GTP into the cap-binding slot of eIF4e. The benzyl group could not be orientated pointing into the cavity (B) or up behind the W102 (red) residue (A) without steric clashes against the surface of eIF4E. The position shown above appeared to be the most plausible orientation for the benzyl group. However this assumes there is no conformational change in the residues in the active site and that the individual phosphate positions also remain the same for the monophosphate, diphosphates and triphosphate derivatives.

Modelling was carried out by adding a phenyl group to the N7 methyl on the GTP molecule, under the assumption that the phosphate positions remain the same in the

m7GMP, m7GDP and m7GTP and other N7 derivatives when interacting with eIF4E. Another inherent assumption is that the cap-binding site undergoes no conformational change. Modelling with PYMOL[3] showed that the pose (shown in figure 1) was the most convincing; all other poses had steric clashes with the protein surface of eIF4E. However the benzyl group is in close proximity to the negative phosphate tail and would be energetically unfavourable. The pose shown in figure 1, agrees with the main points of the proposed model by Niedzwiecka et al[2], showing contacts between the benzene ring and W166 and possible disruption of any structured waters in the cavity. To orientate the benzyl group into other positions in the cavity would entail disrupting the planar stacking arrangement of the tryptophan side-chains and m7G and weaken the hydrogen bonds to E103 and the backbone amide of W102.

6.2 Co-crystallisation of 7bzGMP, *p*flbzGMP and m7GMP with full-length human eIF4E and the 4E-BP1 motif peptide.

Preliminary screens were carried out around the conditions required for m7GTP/eIF4E/4E-BP1 crystal growth. Protein solution for crystallisation was prepared by concentrating eIF4E to 8mg/ml and then adding an equivalent volume 4E-BP1 peptide solution (0.0032M) to produce a sample with eIF4E at 4mg/ml and the peptide at 0.0016M. The reaction product, from the alkylation reactions that produced m7GMP, 7bzGMP and *p*flbz-GMP (see figure 2 for chemical structures), were dissolved in 0.1 M KCL 100mM HEPES-KOH pH 7.6 at a concentration of 50mg/ml. Ligand solution was added in a 1/6 dilution to the prepared eIF4E/4E-BP1 solution. Vapour diffusion was carried out as described in the material and methods. Crystals were produced at ~7% ammonium sulphate, 25% PEG6000 and 100mM HEPES-KOH at pHs 6.8 and 7.0 or 100mM MOPS pH 6.4 (see figure 2). The m7GMP/eIF4E/4E-BP1 (figure 2C) screens produced crystals that looked more disordered and were more irregularly shaped than the 7bzGMP/eIF4E/4E-BP1 (figure 2A) and 7-*p*flbzGMP/eIF4E/4E-BP1 (figure 2B) co-crystals. The 7bzGMP/eIF4E/4E-BP1 and 7-flbzGMP/eIF4E/4E-BP1 co-crystals were very similar in morphology to the plate-like crystals of m7GTP/eIF4E/4E-BP1 originally observed in chapter 2. Crystallisation trials with ribavirin, GTP and m7G proved unsuccessful highlighting the need for a phosphate group and a delocalised charge on the guanine ring system.
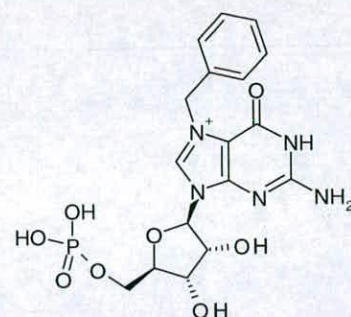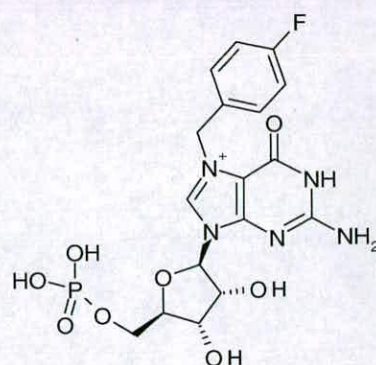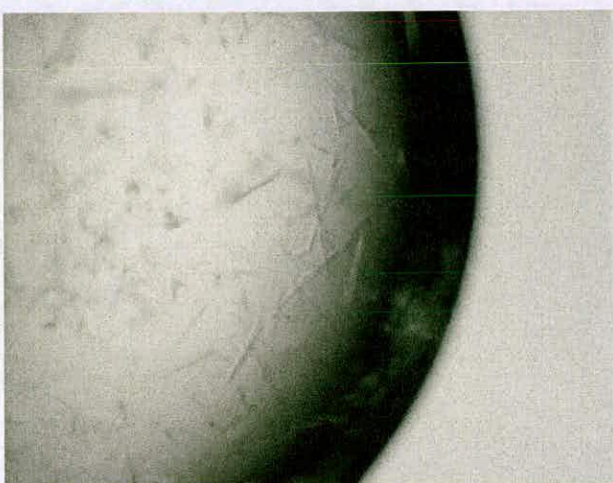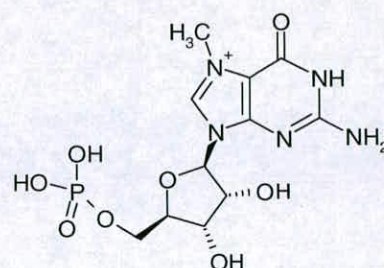
7bzGMP

7-pflbzGMP

m7GMP

Figure 2: Panel A shows co-crystals of the 7-bzGMP/eIF4E/4E-BP1 motif peptide complex and Panel B shows co-crystals of the 7-*p*fLbzGMP/eIF4E/4E-BP1 motif peptide complex. These co-crystals exhibit the same crystal morphology as the m7GTP variants discussed in chapter 3. Panel C shows crystals from the crystal screening trials with m7GMP. These trials produced lower quality crystals that appeared to be more disordered and showed poor diffraction.

6.2.1 Structure solution of 7BzGMP/eIF4E/4E-BP1 and 7FlBzGMP/eIF4E/4E-BP1 co-crystals.

The thin-plate like crystals of the eIF4E and N7 substituted derivatives peptide complexes (figure 2) were cryo-frozen using freezing solution containing 20% glycerol, 5% ammonium sulphate, 25% PEG 6000 and 100mM of the buffer used in the well they were harvested from. The co-crystals were then mounted in a random orientation and 120 degrees of data were taken with an oscillation angle of 1 degree using the Phi rotation method at 100K at the 14.1 PX station at Daresbury.

The structures of the ternary complex containing N7 substituted derivatives were solved by molecular replacement using MOLREP[4], the search model was the 2.3 angstrom structure of full length human eIF4E complexed with the 4E-BP1 peptide solved in chapter 3. m7GTP and any structured waters were removed from the search model. Data up to 3.5 angstrom was used for the MOLREP[4] solution. MOLREP[4] isolated the positions of the two ternary complexes in the asymmetric unit. These initial co-ordinates were used to begin refinement (see table 1 for data collection statistics).

6.2.2 Refinement and validation of the 7bzGMP/eIF4E/4E-BP1 and 7-pflbzGMP/eIF4E/4E-BP1 complex structures.
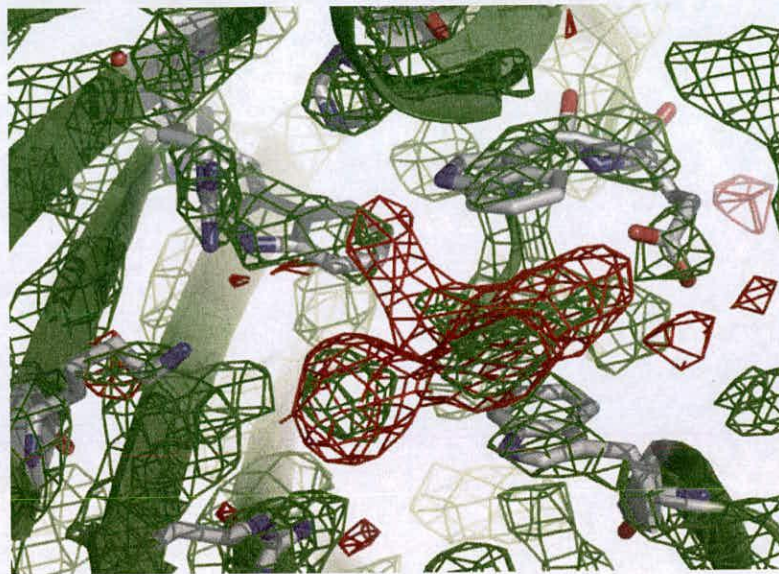
Figure 3: Initial F_o-F_c density map showing clear electron density for the 7bzGMP ligand (sigma 2.8). The density map was calculated from a crystal collected on the home source whose data was processed up to a resolution of 3.2 angstrom.

Prior to refinement, $2F_o$-F_c and F_o-F_c maps were calculated using FFT in CCP4i[4]. The models were both examined and were continuous in density for the entire length, except for the N-termini and the C-terminal loop. The starting models was subjected to 10 cycles of rigid body refinement and followed by 10 cycles of restrained refinement in Refmac 5.0[4,5]. At this point the F_o-F_c were recalculated for both structures revealing large peaks of density in the cap-binding site for both non-crystallographic symmetry related complexes (see figure 3). These areas of density correspond in shape to guanine mono-phosphate showing clear density for the α-phosphate. The R-factors at this point were 24.6 for the 7bzGMP and 23.6 for the 7-pflbzGMP ligand structure. The corresponding R-free values were 29.1 and 28.9. After this the N7 substituted guanine monophosphates derivatives were used

(refmac[4,5] library files generated using PRODRG[6]) in the refinement and placed into clearly visible electron density. At this point waters were added using ARP_Warp[7]. This was then followed by continuous cycles of manual refinement of the models using Coot[8] and 10 cycles of restrained refinement using refmac 5.0[4,5]. Structure was validated using PROCHECK[4] (see table 1 for final model statistics).

	7-bzGMP co-crystal structure	7-*p*-flbzGMP co-crystal
Unit Cell Dimensions	A=38.2, b=100.2, c=135.2 $\alpha = \beta = \gamma = 90$	A=38.4, b=100.7, c=135.3 $\alpha = \beta = \gamma = 90$
Resolution	2.3Å	2.1Å
Space Group	$P2_12_12_1$	$P2_12_12_1$
Temp(K)	100	100
Number of collected	77817	111399
Number of unique reflections	21598	28510
R sym	11.9(46.6)	12.5(52.7)
R factor	19.9	19.9
R free	25.6	24.9
RMS bonds	0.008	0.016
RMS angle	1.187	1.725
Average B value	31.234	26.088
% completeness	91%(87%)	91.1%(87.6%)
Number of observed solvent molecules	353	313
Peptide	2	2
Ligand	2	2
Ramachandran data: Residues in most favoured regions	93.4%	92.5%
Residues in most favoured regions	5.8%	6.4%
Residues in generously allowed regions	0.9%	0.6%
Residues in disallowed regions	0%	0.6%
I/sigma	5.8(1.6)	5.0(1.4)

Table 1: Data collection and final model statistics for the 7-benzyl GMP and 7-*p*fl-benzyl GMP co-crystal structure.

6.3 Structural analysis of N7 substituted cap-analogue crystal structures.

6.3.1 Flipping of the tryptophan 102 allows the N7 benzyl group to be accommodated in the cap-binding site.

The m7G moiety of the cap structure is sandwiched between two tryptophans (W56 and W102) forming π-cation interactions (see figure 4). The cap analogue makes three hydrogen bonds, involving a backbone amino group of W102 and the side chain of the conserved E103 (see figure 4). The structural interactions of the cap-binding pocket with m7GDP/m7GTP are conserved between human, mouse and yeast eIF4E.
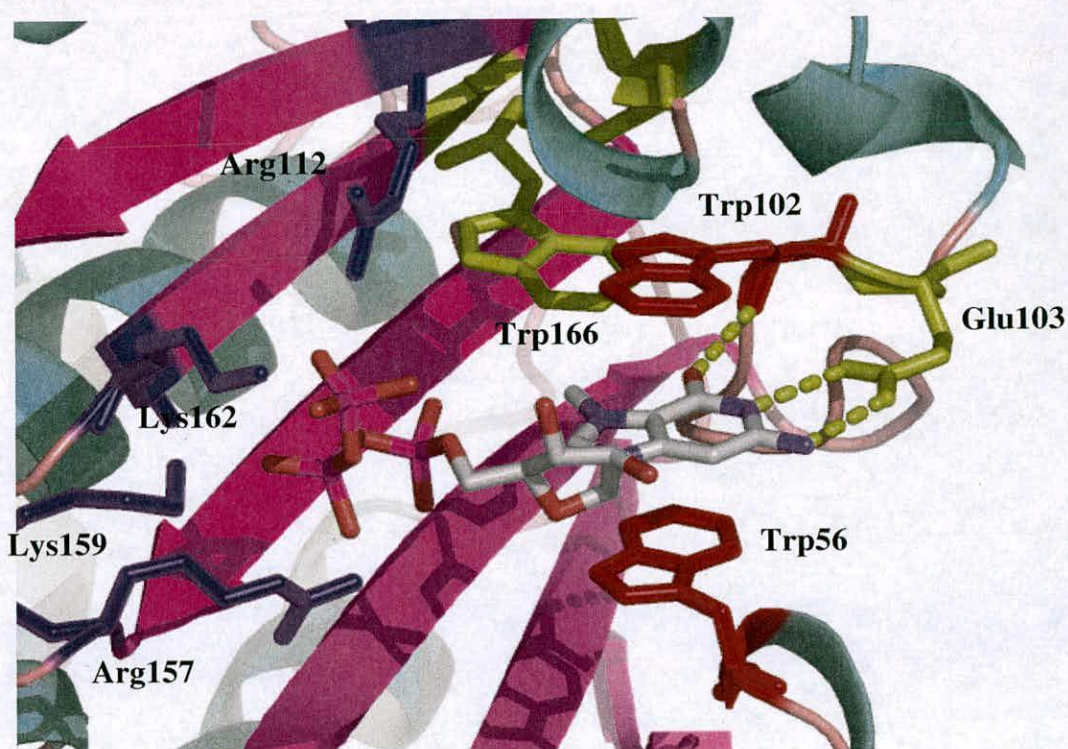
Figure 4: Graphic summarising the interactions made by m7GTP with eIF4E. The cap-analogue is intercalated between tryptophans W102 and W56, forms three hydrogen bonds with the side-chain of E103 and the amide backbone of W102 and the triphosphate tail is recognised by R157, R112, K162 and K159. The m7G ring interacts with the two tryptophans via a cation-π interaction.

In the co-crystal structures containing 7-bzGMP and 7-$p$flbzGMP a conformational change is induced in the cap binding site that allows the benzyl ring to pack efficiently into the cap-binding site. The W102 residue flips through 180° and allows the benzyl group to orientate itself into the cavity behind the two tryptophan residues (see figure 5).
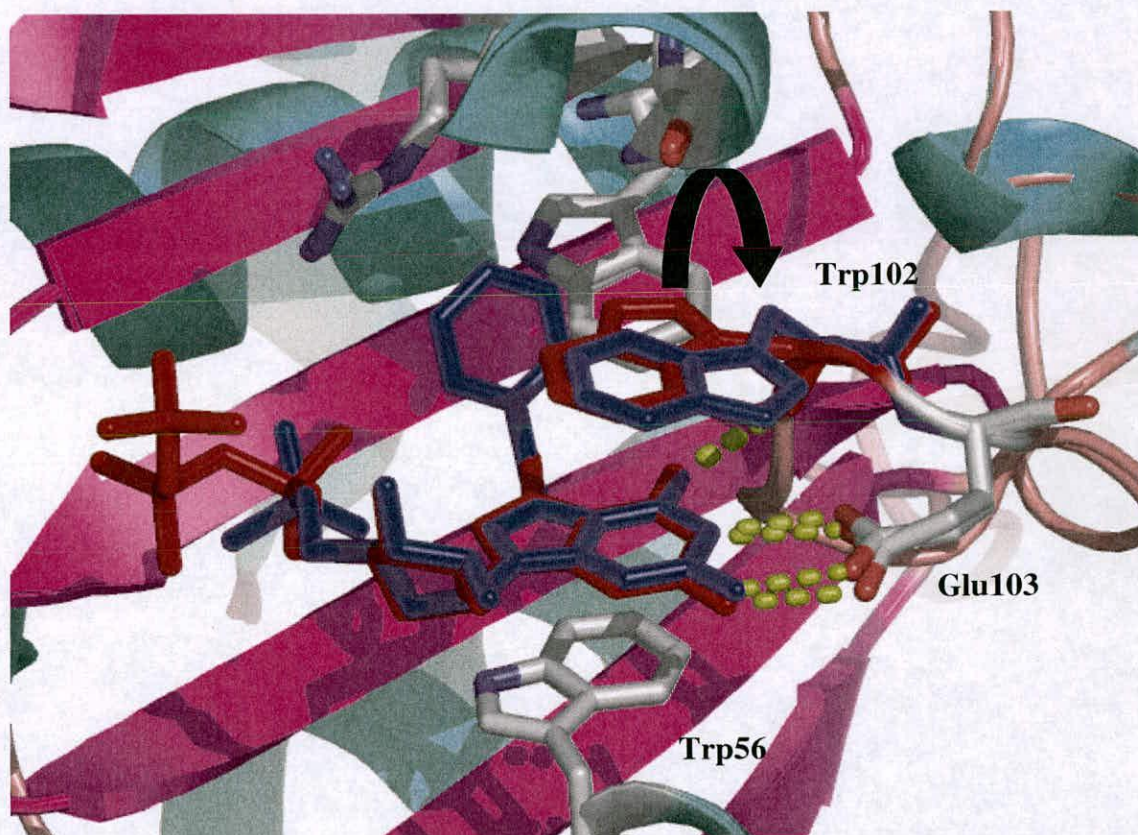
Figure 5: The co-crystal structure of 7-bzGMP overlaid over the m7GTP co-crystal structure revealing the flipping of the side–chain of W102 through 180°. It is this flipping that allows the N7 benzyl group to orientate itself behind W102. The 7bzGMP ligand is shown in blue as well as the corresponding W102 side chain and the position it adopts in this crystal structure. The m7GTP ligand and the former position of the tryptophan are shown in red. Modelling studies failed to reveal this mode of binding due to steric clashes between the W102 side-chain, in the position as found in the m7GTP structure, and the benzyl group in that particular orientation.

Modelling studies in section 6.1.2 could not orientate the benzyl group into the cavity, despite there being sufficient space, with out forming bad contacts in the cap-binding site or losing the planar arrangement between the side-chains of W102 and

W56 and the m7G moiety. The flipping of W102 does not disturb the tridentate hydrogen bonding arrangement to the m7G moiety, the planar packing arrangement of the ring systems in the cap binding site or the overlap between the tryptophans and the guanine ring (see figure 6).
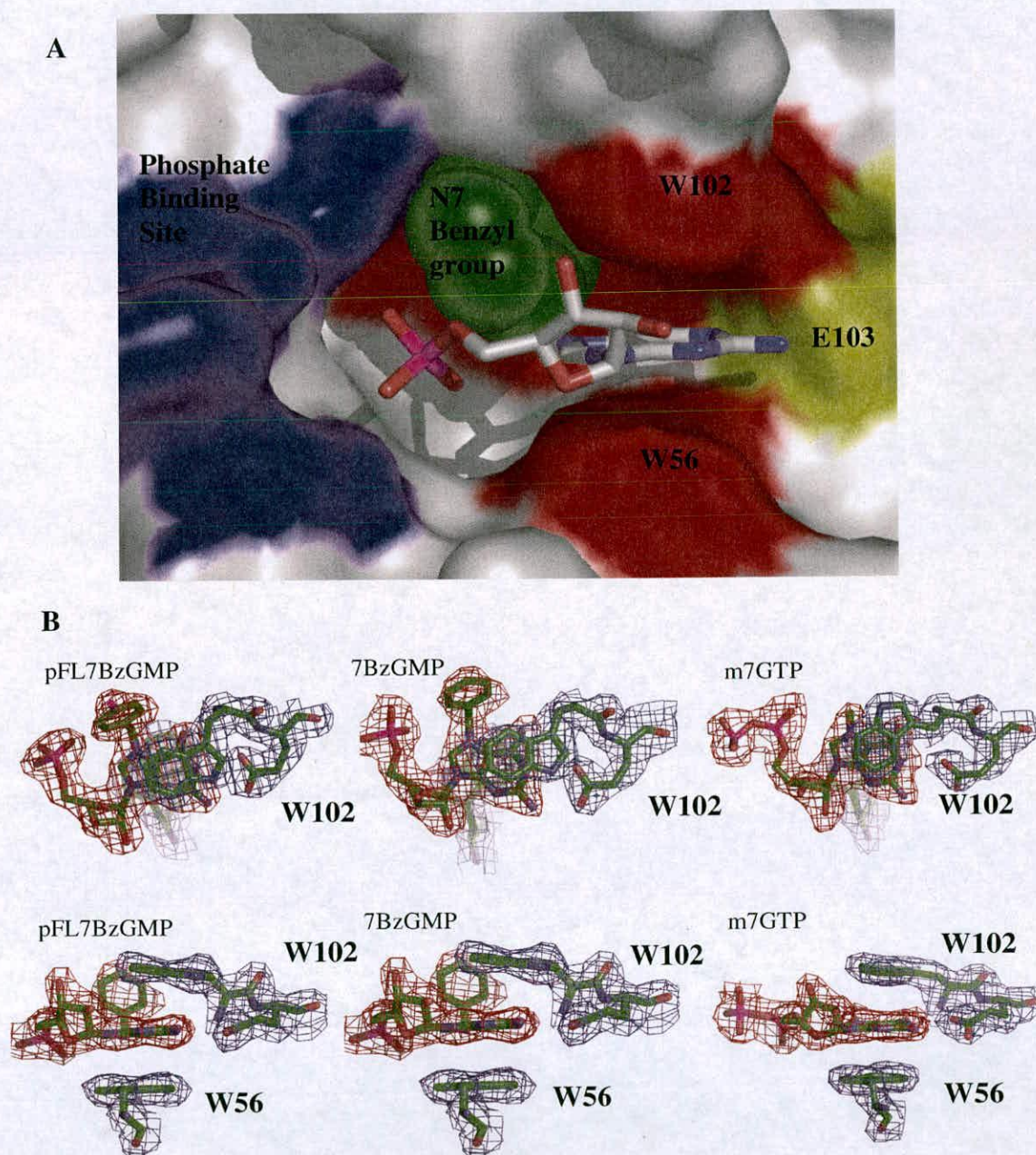
Figure 6: Panel A shows the surface of eIF4E and how the benzyl group fits into the back of the cap-binding slot with a space filling model. If the tryptophan was flipped, in the position seen in the m7GTP co-crystal structure, a steric clash would occur. The flipping of the tryptophan allows the benzyl group to form a maximum number of contacts with eIF4E, the m7G ring to remain planar in respect to W102 and W56, ensure that the overlap between the ring systems is approximately the same and that the three hydrogen bonds recogonising the ring system are still formed. Panel B shows the maintenance of ring planarity and how the areas of ring overlap approximately remain the same between the ligands m7GTP, 7-BzGMP and 7-Bz-pFL-GMP.

The flipping of the W102 also ensures that as well as maintaining the amount of overlap with the m7G moiety it also overlaps with the same part of the molecule (see figure 6). The part of the m7G moiety that W102 overlaps in both configurations is the pyrimidine section rather than the imidazole section. Studies of model compound systems also indicate a stronger stacking interaction with the pyrimidine portion of the guaninium and adenine rings than with the imidazole portion[9,10]. For VP39 it has also been observed that F180, as well as having a larger area of overlap with m7G than Y22, also overlaps more substantially with the pyrimidine portion of the m7G ring system. This is consistent with their finding that the F180A mutation led to a significantly greater reduction in VP39 methyltransferase activity than the Y22A mutation.

Figure 7: m7GTP (red) co-crystal structure overlaid with the NCS related ternary complex of the 7-pflbzGMP(green) which showed partial density for the W102 side chain. This non-crystallographic symmetry unit is showing that the W102 side chain is sampling multiple conformations (as shown by the lack of density) and that the p-fl-benzyl group is orientating itself in such a position as to sterically hinder the freedom of the side-chain of W102. This potential steric clash is shown by the close proximity of the red tryptophan to the N7 derivative. The red tryptophan corresponds to the m7GTP structure and the green tryptophan to the 7-Bz-pFLGMP structure.

181

In the other non-crystallographic related ternary complex located in the asymmetric unit, for both the 7-bzGMP and the 7-*p*flbzGMP co-crystal structures, there is partial electron density for the W102 and poor density (occupancy) for the N7 substitution on the m7G ring. However in the 7-*p*flbzGMP co-crystal structure the conformations of the N7 substituted cap-analogue differ (see figure 8). The phosphate tails are in identical positions but the orientation of the benzyl group has changed and the positions of the m7G ring and the ribose ring do not overlay precisely (see figure 8). The configuration of the N7 GMP derivative supports the two step binding mechanism with the phosphate binding first (their positions in figure 8 are identical) and then the rest of the ligand orientating itself between the two tryptophans (these do not overlay identically in figure 8). If this ternary complex from the 7-*p*flbzGMP structure is overlaid with either of the ternary complexes in the m7GTP structure (see figure 7) the benzyl ring and the W102 from the m7GTP structure clearly clash. This non-crystallographic symmetry unit is showing that the W102 side chain is sampling multiple conformations (as shown by the lack of density) and that the *p*-fl-benzyl group is orientating itself in such a position as to sterically hinder the freedom of the side-chain of W102. It is this steric clash, shown by the close proximity of the *p*-fl-benzyl to the overlaid W102 in the m7GTP co-crystal structure, which is responsible for the side-chain of W102 stacking on the m7G ring in an alternate conformation. Evidence that the arrangement described in the other eIF4E molecule (chain A) is the final packing arrangement is supported by other published structures of eIF4E and other cap-binders, where for ligands ranging from m7GTP to dinucleotide derivatives

the planar parallel packing arrangement of the two tryptophans with the ligand is maintained.
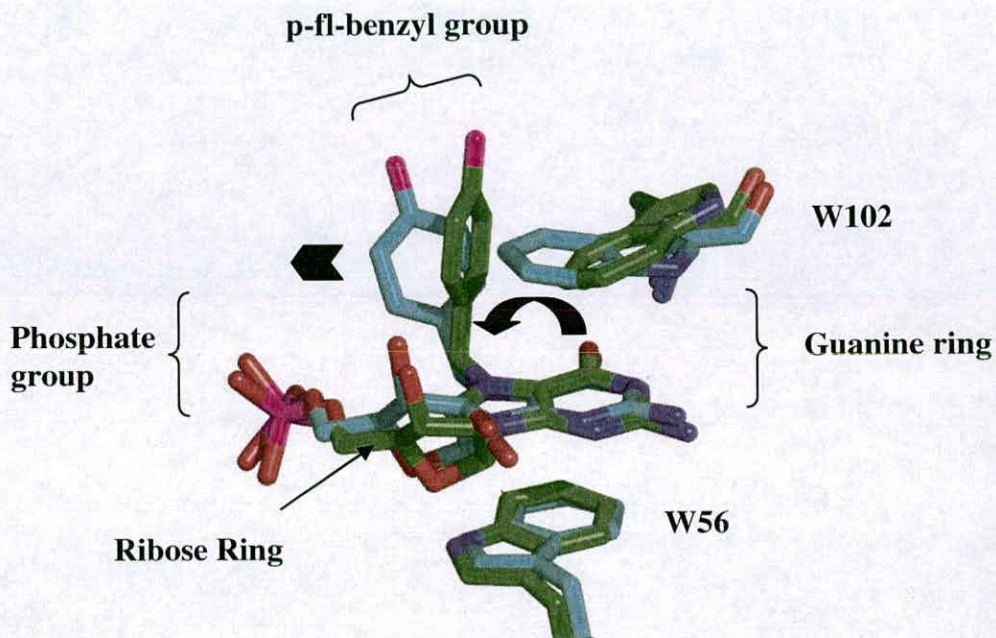
Figure 8: The figure shows the conformational differences between the 7-*p*flbzGMP (green) bound ligand, in the NCS ternary complex showing partial density for the W102 side chain, compared to opposing ternary complex that shows good density. Other bound ligands overlay almost identically over the 7-*p*flbzGMP ligand shown in turquoise. The 7-*p*flbzGMP ligand (green) has not yet adopted the final planar arrangement noted for the other ligands and the benzene ring has not adopted its final orientation. The configuration of the 7-*p*flbzGMP derivative supports the two step binding mechanism with the phosphate binding first (the positions of the above are identical) and then the rest of the ligand orientating itself between the two tryptophans (these do not overlay identically). To move the m7G moiety into the final bound position is a simple procedure. The block arrow shows the direction the m7G ring (green) needs to roll to adopt the final ligand (turquoise) position, and the chevron shows the direction the benzene rings needs to move to adopt the final bound position.

6.3.2 The large bulky benzyl group expels two structured waters from the active

Site and alters the water network in the cap-binding cavity.
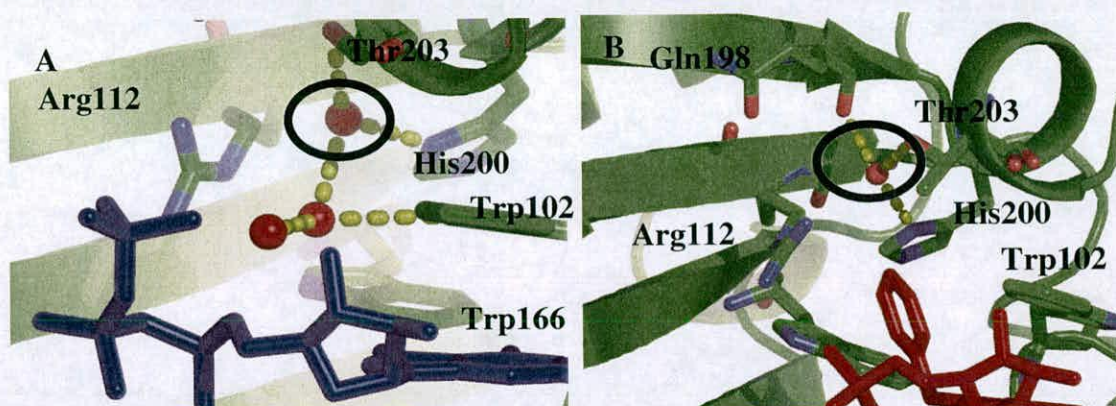
Figure 9: Panel A shows the cavity which the benzyl group packs into in the m7GTP co-crystal structure. Panel B shows the change in the water network, in the cavity behind the two tryptophans, where the benzyl causes two structured waters to be expelled and creates a new structured water position. This new position is in close proximity to the side-chains of T203 and H200 and the carbonyl backbone of Q198.

In the m7GTP structure the nitrogen atom of W102 is pointing in towards the cavity of the cap binding site and forms a direct hydrogen bond to the structured water (see figures 9 and 10). This structured water interacts with a further two: one of which is involved in interactions with the triphosphate tail of the ligand whilst the other interacts with T203 and H200 (see figures 9 and 10). Out of this set of three structured water molecules two are expelled by the benzylated GMP derivative and the other, at the back of the pocket, shifts its position (see figures 9 and 10). The structured water that shifts maintains the h-bonds to T203 and H200 (see figure 9 and 10). These structured water positions are observed in both non-crystallographic symmetry related units. The two structured waters that are expelled by the bulky benzyl group confirm the hypothesis (Niedzwiecka et al) for why the energetics of the N7 benzyl substituted derivatives in relation to eIF4E are entropy driven. H200 has been modelled into the position shown in figures 9 and 10 to maximise the

number of hydrogen bonds formed with the ε^2 nitrogen interacting with a structured water in the cap-binding slot and the δ^1 nitrogen interacting with another structured water outside the cap-binding site.
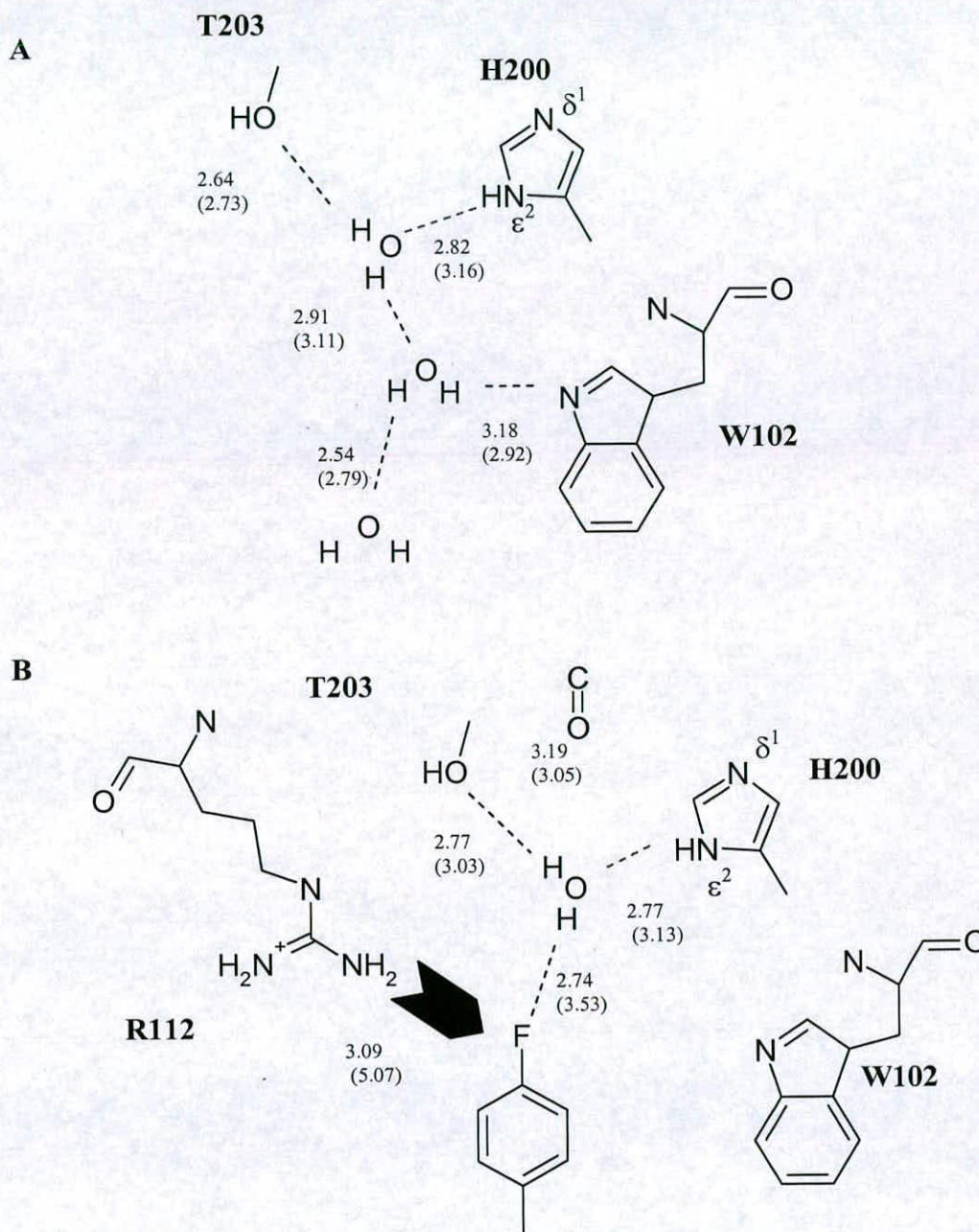
Figure 10: In the m7GTP structure the nitrogen atom of W102 is pointing in towards the cavity of the cap binding site and forms a direct hydrogen bond to the structured water (Panel A). This structured water interacts with a further two: one of which is involved in interactions with the triphosphate tail of the ligand whilst the other interacts with T203 and H200) (Panel A). Out of this set of three structured water molecules two are expelled by the benzylated GMP derivative and the other, at the back of the pocket, shifts its position (see Panel B). The structured water that shifts maintains the h-bonds to T203 and H200. The fluorine of the para-halide substituted benzene ring interacts with the structured water that shifts its position (Panel B). There is also an arginine residue in close proximity (R112), which interacts electrostatically with the fluoride. This set of interaction explains the increased potency of halide substituted benzyl ring systems reported by Cai et al[1]

186

6.3.3 N7 –*p*fl-benzyl modification interacts with the altered water network.
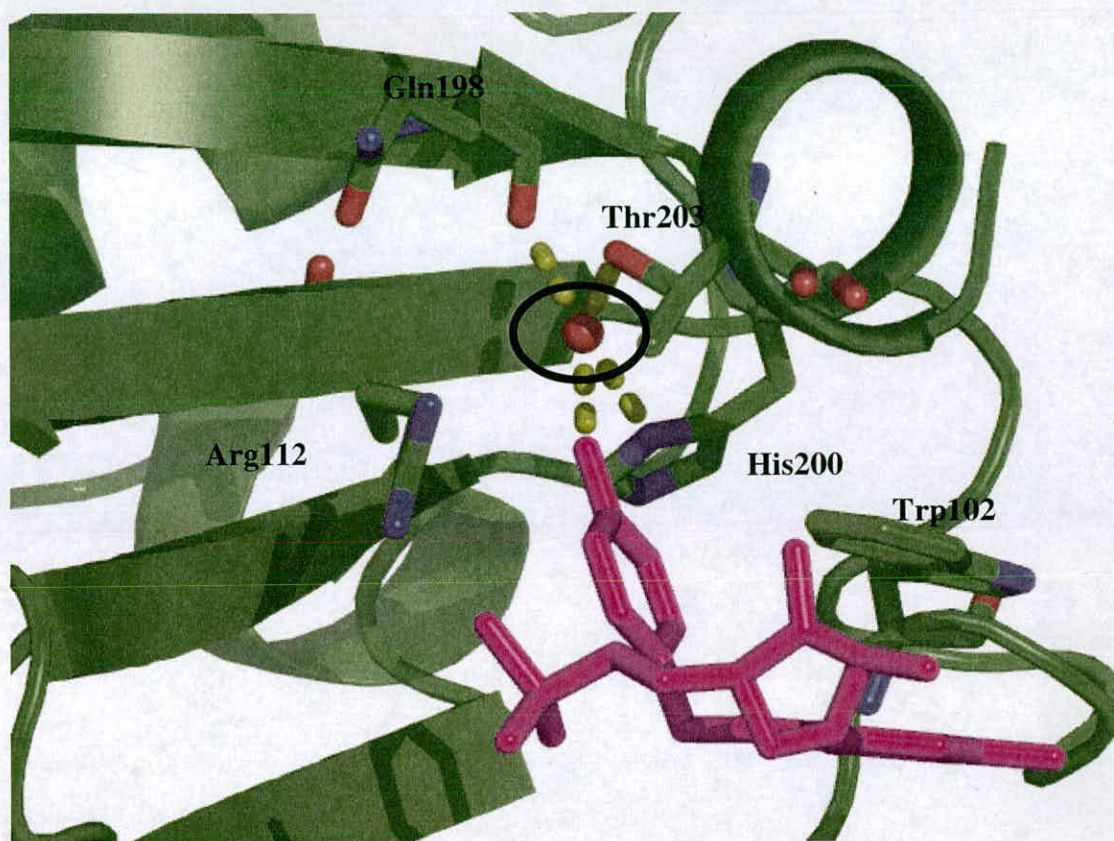
Figure 11: The para fluoride atom on the benzyl ring forms a hydrogen bond interaction with the structured water at the back of the binding pocket The structured water is also in close proximity to the side-chains of T203 and H200 and the carbonyl backbone of Q198. It is the bulk of the benzyl substitution that causes the structured water to occur in this shifted position as discussed in the previous section. The fluoride atom is also situated near a positively charged arginine where an electrostatic interaction occurs.

The same rearrangement of the cap-binding site occurs with the para-fluoro substituted benzyl with the W102 carrying out the same movement. However the fluorine interacts with the structured water that shifts its position, in the small hydrogen bond network, to the back of the binding cavity. The structured water still interacts with T203 and H200. Apart from these interactions there is an arginine residue in close proximity (R112), which interacts electrostatically with the fluoride. This set of interactions explains the increased potency of halide substituted benzyl ring systems reported by Cai et al[1]

In the other eIF4E molecule, in the asymmetric unit, density for the fluoride atom is not in the same orientation to form such tight interactions. However the structured water still forms the interactions described at the back of the cap-binding slot (see figure 10 and 11). The disruption of the two structured waters observed in the m7GTP co-crystal structure by the presence of the large N7 substitution is the driving factor for the positioning of the shifted structured water. The hydrogen bonds formed by the para-fluoro halide are longer and weaker ~ 3.5Å in length. The fluoride is now also not in such close proximity to R112 ~ 5Å instead of 3.09Å.

6.3.4 Monophosphate tails interactions.

The N7 GMP derivatives are the first monophosphate molecules to be co-crystallised with eIF4E. The α-phosphate has a direct electrostatic interaction with Arg 157 and interacts with two conserved water positions (see figure 12A). One water position interacts with K162 and the other forms hydrogen bonds with R112 and another structured water (see figure 12A). This structured water eventually feeds into a water molecule that interacts with the indole ring of W166 and another that interacts with the carbonyl backbone group of V153. Upon the addition of a second phosphate group (as seen in the m7GTP co-crystal structure) the water bridging to K162 is displaced and three direct contacts are made to K159, K162 and R157 (see figure 12B). An interaction with a structured water occurs outside the cap-binding site on the lip of the pocket. The gamma phosphate, in the m7GTP structure, forms only one direct ionic contact with R157 and interacts with a structured water (see figure 12B). This structured water forms an h-bond with another water that mediates a hydrogen

bridge between the ribose ring and the carbonyl backbone group of T203. These results show that the phosphate binding pocket is optimised to bind the beta phosphate which has the most direct contacts to eIF4E compared to the alpha and gamma phosphates that mainly interact through water mediated bridges.
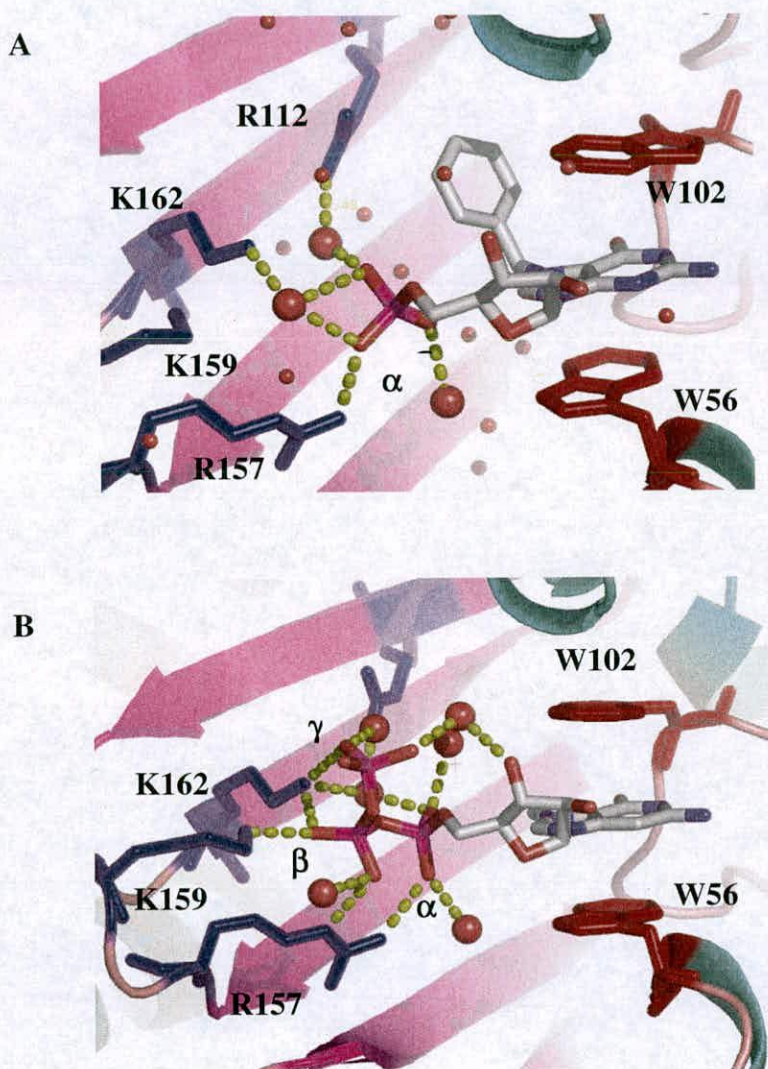
Figure 12: Panel A shows interactions between the α-phosphate of the GMP derivatives and the phosphate binding pocket. A direct interaction is formed between the α-phosphate and R157. Water mediated hydrogen bridges occur with K162 and R112. Another interaction is formed to structured water that interacts with other waters in the binding pocket. Panel B shows the interactions between the triphosphate tail and the phosphate binding pocket. The β-phosphate replaces the structured water and the interactions it made between the α-phosphate and K162 and K159 respectively. The γ-phosphate forms a direct contact to K162 and two hydrogen bonds to two structured waters: one of which forms an interaction to the 3' hydroxyl group on the ribose ring.

6.5 Future drug design development for eIF4E.

6.5.1 Design principles.

The N^7-methyl group of the m^7G moiety points towards a large hydrophobic cavity that forms the floor of the eIF4E cap-binding slot and only forms a van der Waals contact with W166. The N7 benzyl substituted ligand structures reveal that the benzyl substituent is accommodated into the cap-binding site by flipping of the W102 tryptophan side-chain. Possible future pharmacophore subunits could be identified that optimally occupy the position of the benzyl in order to provide the bulk of the binding energy. The presence of the strongly negatively charged phosphate groups in potent cap analogues is the main reason for the lack of membrane permeability and bioavailability of these compounds (see figure 14).
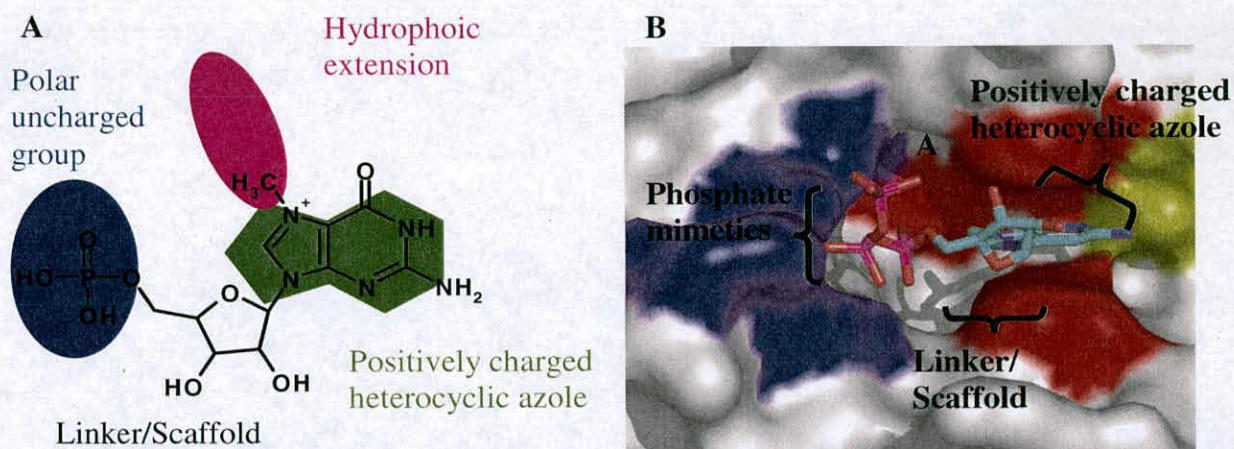
Figure 14: The figure shows the functional moieties that m7G phosphate family can be broken down into and that need to be chemically explored in the design of an ideal drug candidate. Panel A shows the three functional groups of the generic cap-analogue and the role ribose plays as a linker and scaffold functionality. Panel B shows how these relate to the active site with the tryptophans (W56, W102 and W166) shaded red, E103 shaded yellow and the phosphate binding site shown in blue. The A label shows the area where the benzyl group packs into which would be a site for exploration of other chemical groups.

190

Phosphate mimetic groups that are highly polar, but uncharged, should be investigated. These include polyhydroxylated systems, trifluoromethylsulfonyl groups[11], and squaryldiamides[12]. For ribose replacements emphasis should be on optimal spacing between the heterocyclic and the phosphate mimetic pharmacophore groups. Since the ribose group is not involved directly in interactions with eIF4E except via a single water mediated hydrogen bond, linker diversity can be explored to achieve optimal pharmaceutical properties and conformational constraint between the heterocycle- and phosphate-mimetic groups (see figure 14). The β-phosphate would seem an optimal position for placement of a phosphate mimetic group because it forms more direct contacts to eIF4E that the other phosphate positions (see section 6.3.4). The other phosphates groups of m7GTP interact mainly though water mediated contacts. Isosteric and isoelectronic replacements for the modified nucleobase should be sought, with heterocyclic azoles capable of delocalising a positive charge as in m^{7}G being investigated. Potentially useful systems are purines, as well as thiazoles, imidazoles, pyrazoles, indazoles, triazoles, and their benzo- and pyrido-derivatives. At present the only known system other than m7G is based on benzimidazole[13]. The amidotriazole system present in ribavirin also possesses a delocalized charge that has been reported to interact with eIF4E[14]. However recent reports contradict this[15,16].

Certain analogies with protein kinases, where mononucleotide (*i.e.* ATP)-antagonist inhibitor design has been highly successful[17], can be discerned. In both cases the mononucleotide-binding site is not occupied fully by the natural ligand, the

191

triphosphate moiety contributes significantly to binding, and the nucleobase interacts with the protein through a tridentate H-bonding system.

A

7-Benzyl GMP

B

Idealised target compound arising from step-wise redesign of 7-Benzyl GMP

C

Potential starting point for chemistry arising from virtual screens
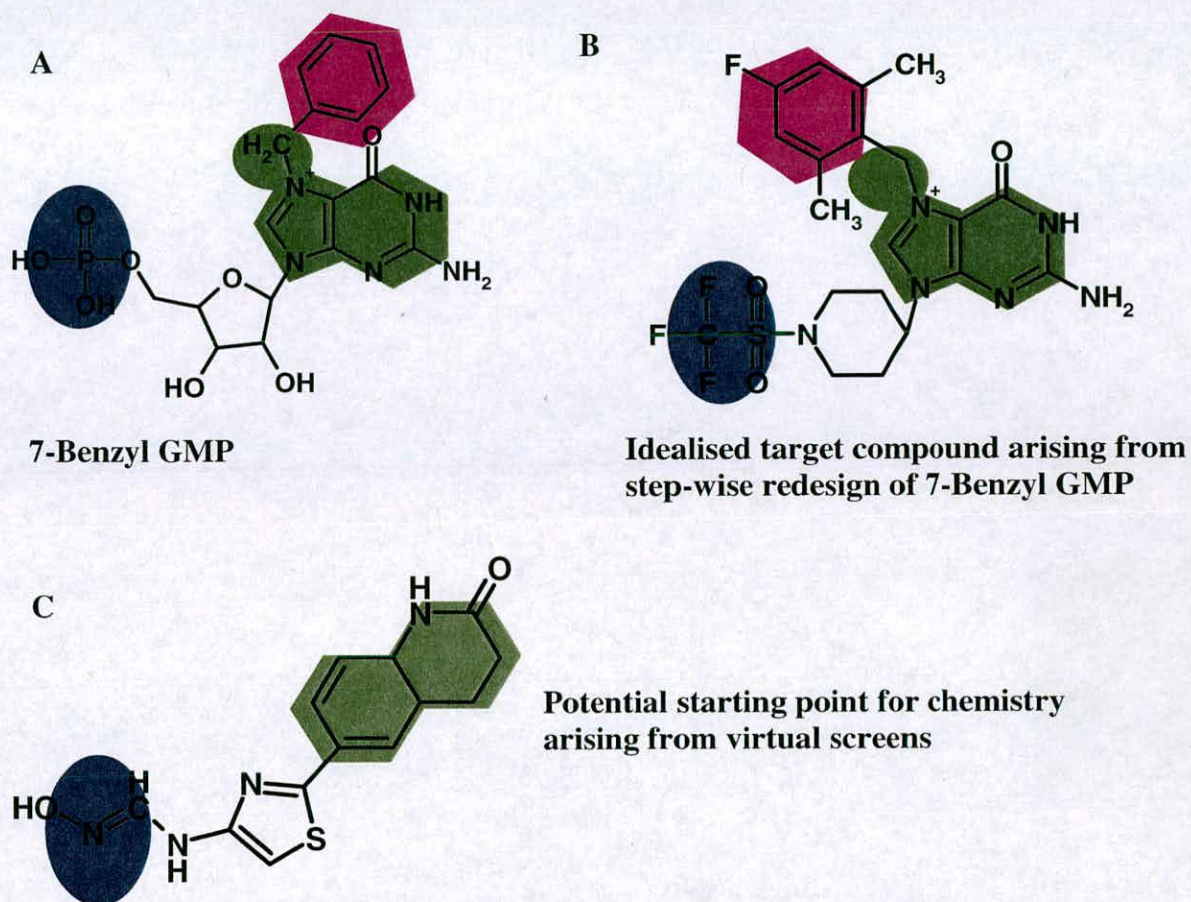
Figure 15: Panel A shows the 7-benzyl GMP ligand structure, which has been shown as a potent binder of eIF4E in the literature. 7-benzyl GMP can be redesigned using a step-wise strategy to synthesis a compound with more favourable ADMET properties. A compound that might arise from such an approach is illustrated in panel B. As well as using the step-wise strategy convincing hits from virtual screening can also be used as potential chemical starting points. Panel C shows an example of such a starting point. In these cases an obvious strategy would be to try an introduce some type of delocalised positive charge on the ring system to mimic the charge found on the m7G ring system. The example in panel C has tail which could be a potential phosphate mimic (highlighted in blue).

In the case of kinases numerous ADMET-favourable pharmacophores unrelated to ATP have been identified and this has enabled the development of kinase-inhibitor drugs. Selectivity should be less of an issue with cap antagonists than with kinase inhibitors. Thus there is the potential of using these kinase antagonists by inducing some type of positive delocalised charge on them, and then assessing whether they are capable of then binding eIF4E.

6.5.2 Medicinal chemistry based on m7GTP.

In the first instance the above design principles can be applied in the context of 7BzGMP (see figure 15), one of the most potent nucleoside monophosphate cap antagonists known to date which we have solved the ligand structure for[1]. Modifications of single pharmacophore groups (see figure 14) can be introduced initially in order to assess SARs at each position. This is convenient since established cap analogue chemistry can be applied (refer retrosyntheses in figure 16). Optimal individual solutions for the four pharmacophoric groups could then be combined using synthetic methods; a possible target structure is shown in figure 15 (retrosynthetic analysis to 5,6-dinitrobenzimidazole not shown).

De novo design. Additionally, inhibitor candidates with chemical complementarity to the cap-binding site could also be designed without explicit reference to the natural ligand. Initial feasibility work suggests that application of our LIDAUES high-throughput docking program[18] to the eIF4E–m^{7}GTP interaction can suggest plausible pharmacophore starting points (see figure 15 and chapter 5). This approach can be refined, especially in order to account for the m^{7}Gua stacking interaction. Also the new eIF4E ligand structures with 7bzGMP and 7-pfLbzGMP have provided a

different conformation for virtual screening studies, which allows for the accommodation of large bulky hydrophobic groups. It is reasonable to expect that this approach could yield alternative pharmacophore hypotheses for medicinal chemistry.

SARs of hydrophobic extension
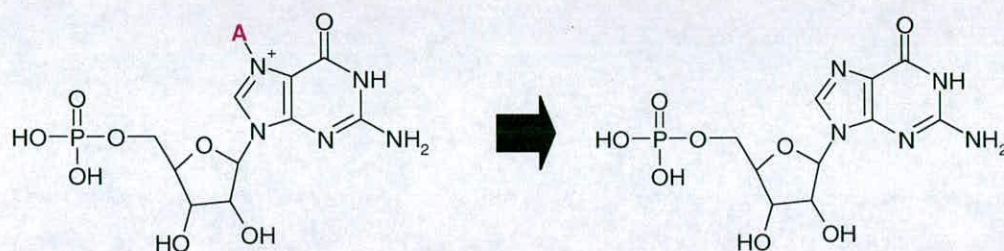
Positive delocalised charge on ring and presence of phosphate group, will specifically bind to eIF4E. Optimal N7 substitutions found so far are para halide substituted benzyl groups.

No positive delocalized charge on ring, due to no N7 alkylation, will cause weak binding to eIF4E.

SARs of phosphate replacements
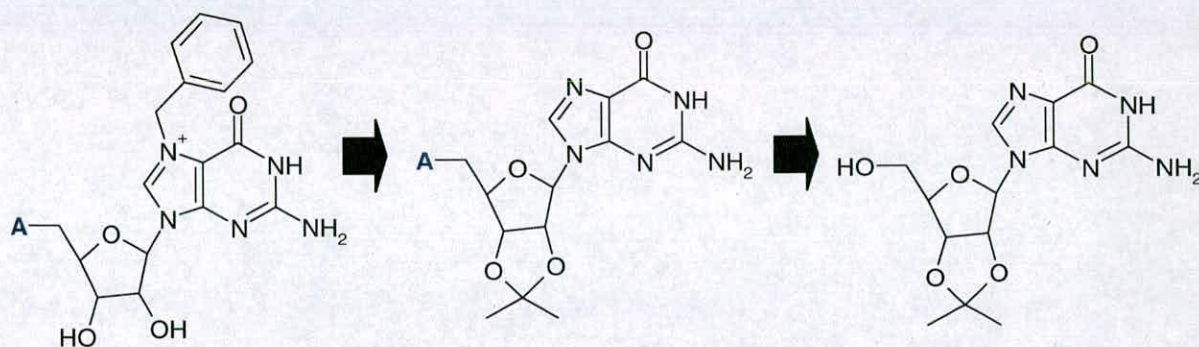
Positive charge on ring system with successful phosphate mimic will interact with eIF4E

No positive charge on ring system with successful phosphate mimic will interact weakly with eIF4E

No positive charge on ring system with no phosphate mimic will have little or no interaction with eIF4E

SARs of ribose replacements
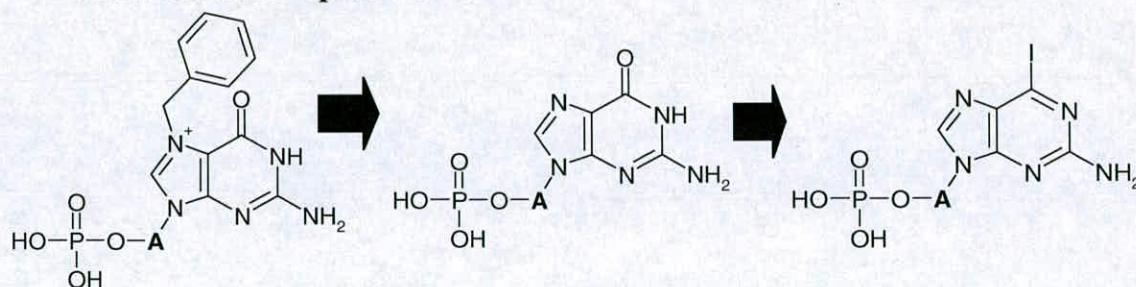
Positive charge on ring system with optimal linker to place the ribose will bind eIF4E.

No positive charge on ring system with optimal linker to place the ribose will bind weakly with eIF4E.

No positive charge on ring system with optimal linker to place the ribose will bind weakly or not at all with eIF4E.

Figure 16: Modifications of single pharmacophore groups can be introduced initially in order to assess SARs at each positions. This is convenient since established cap analogue chemistry can be applied. Above are shown potential series of compounds that could be designed around different functionalities to investigate there effects on binding, with commentary based on the model of interaction proposed in chapter 4 on whether these ligand types will bind or not.

6.5.3 X-ray crystallography.

High-resolution X-ray crystal structures of recombinant human eIF4E constructs complexed with 4E-BP peptides and a reasonably potent cap-analogue, as has been shown, can be co-crystallised. Co-crystals have been grown for m7GTP, 7bzGMP and 7-*p*flbzGMP. Both medicinal chemistry based on m^{7}GTP and *de novo* design can be directed through the study of the structural consequences of modifications on eIF4E binding. Fragment-based crystallography screening[19], which has been used successfully in kinase inhibitor discovery, is probably not applicable to eIF4E. With eIF4E the apo form has never successfully been crystallised and the crystals that have been grown have required the presence of a cap-analogue. It is this presence of the cap-analogue that would interfere with a fragment based crystallography screen. There is the potential to soak out the cap-analogue and then attempt to use the fragment based screen approach. However dialysing out cap-analogues such as m7GTP and m7GDP from eIF4E has not been successful. If eIF4E is in solution with 0.6mM m7GTP and dialysed overnight and analysed via mass spectrometry afterwards the protein is still saturated. The tight attraction that eIF4E has for cap-analogues and its need for them to be present when being crystallised are disadvantages against using the fragment based screening approach

The fragment based screening approach could be used with VP39[20] to identify systems that could intercalate between the two tryptophans. The advantage of VP39[20] over eIF4E is that it has been crystallised with the two aromatic residues required for cap-binding in the apo form. This makes it ideal for a fragment based screening approach. In the literature it has already been shown that methylated bases

can be soaked into VP39[20]. A further option could be to mutate both aromatic residues to tryptophan. The strategy would be to soak VP39[20] crystals with pools of low-MW fragments potentially capable of intercalating between the aromatic residues of the cap-binding site..

6.5.4 Biological assays.

Potential ligands can be screened via mass spectrometry (see chapter 4 and 5) and positive hits can be further investigated utilising intrinsic tryptophan fluorescence to characterise its behaviour in solution. Direct competitive binding assay based on streptavidin-immobilised biotin-derivatised m^7GMP and a fluorescence polarisation assay using γ-phosphate-labelled m^7GMP[21] could also be developed. Functional activity of the test compounds can be assessed using *in vitro* and cellular transcription-translation assays based on a bicistronic dual luciferase assay system as described[22] in order to be able to distinguish between effects on cap- and IRES-dependent translation initiation. The antiproliferative effects of test compounds could be examined using standard cell viability assays in a panel of transformed and normal human cell lines. Another technique that could be used to investigate the apparent flexibility of eIF4E is hydrogen-exchange monitored via mass spectrometry. This would provide direct evidence of whether eIF4E is indeed less constrained when ligand free and offer an explanation to why it does not crystallise in the apo form.

6.6 Conclusion.

This body of work demonstrates that eIF4E can be easily purified in the large amounts required for successful structure based drug-design. High resolution diffracting co-crystals can be produced to investigate the interactions of any lead hits produced. Lead hits can be identified using mass-spectrometry as a primary screen and their behaviour in the gas-phase can then be further investigated. This work also describes the behaviour of eIF4E and the cap-analogues in the gas-phase for the first time. N7 derivatives which from previous work were known to be good binders of eIF4E were shown to interact in the gas-phase, and their interactions rationalised structurally using crystallography. This work also shows that for successful drug design to be applied to eIF4E, the interactions of the phosphates and the delocalised positive charge on the m7G ring need to be accounted for; as shown by mass spectrometry and the failure to crystallise compounds that do not contain both of these features. Another interesting characteristic to be revealed by crystallographic analysis is that the cap-binding site is conformationally flexible (shown by flipping of the W102 side-chain), a feature that has been implied by a number of studies. Using the knowledge gained from crystallisation of N7 substituted cap-analogues and applying it to virtual screening and medicinal screening efforts a more drug-like target compound should be producible.

Reference List

1. Cai,A. *et al.* Quantitative assessment of mRNA cap analogues as inhibitors of in vitro translation. *Biochemistry* 38, 8538-8547 (1999).

2. Niedzwiecka,A., Darzynkiewicz,E. & Stolarski,R. Thermodynamics of mRNA 5' cap binding by eukaryotic translation initiation factor eIF4E. *Biochemistry 2004. Oct. 26. ;43. (42.):13305. -17.* 43, 13305-13317 (2004).

3. DeLano,W.L. The PyMOL User's Manual. 2002.

4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* 50, 760-763 (1994).

5. Murshudov,G.N., Vagin,A.A. & Dodson,E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* 53, 240-255 (1997).

6. van Aalten,D.M. *et al.* PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des* 10, 255-262 (1996).

7. Perrakis,A., Sixma,T.K., Wilson,K.S. & Lamzin,V.S. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D. Biol. Crystallogr.* 53, 448-455 (1997).

8. Emsley,P. & Cowtan,K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr. 2004. Dec. ;60. (Pt. 12. Pt. 1):2126. -32. Epub. 2004. Nov. 26.* 60, 2126-2132 (2004).

9. Pullman,B. & Pullman,A. Electron-Donor and Electron-Acceptor Properties of Biologically Important Purines, Pyrimidines, Pteridines, Flavins, and Aromatic Amino Acids. *Proceedings of the National Academy of Sciences of the United States of America* 44, 1197-1202 (1958).

10. Ishida,T., Shibata,M., Fujii,K. & Inoue,M. Structural Studies of the Interaction Between Indole-Derivatives and Biologically Important Aromatic-Compounds .10. Intermolecular and Intramolecular Stacking Interaction Between Indole and Adeninium Rings. *Biochemistry* 22, 3571-3581 (1983).

11. Huang,P. *et al.* Structure-based design and discovery of novel inhibitors of protein tyrosine phosphatases. *Bioorg. Med. Chem. 2003. Apr 17;11(8.):1835. -49.* 11, 1835-1849 (2003).

12. Sato,K., Seio,K. & Sekine,M. Squaryl group as a new mimic of phosphate group in modified oligodeoxynucleotides: synthesis and properties of new oligodeoxynucleotide analogues containing an internucleotidic squaryldiamide linkage. *J. Am. Chem. Soc. 2002. Oct. 30. ;124. (43.):12715. -24.* 124, 12715-12724 (2002).

13. Chlebicka,L. *et al.* Synthesis and Properties of Messenger-Rna 5'-Cap Analogs with 7-Methylguanine Replaced by Benzimidazole Or 3-Methylbenzimidazole. *Nucleosides & Nucleotides* 14, 771-775 (1995).

14. Kentsis,A., Topisirovic,I., Culjkovic,B., Shao,L. & Borden,K.L. Ribavirin suppresses eIF4E-mediated oncogenic transformation by physical mimicry of the 7-methyl guanosine mRNA cap. *Proc. Natl. Acad. Sci. U. S. A 2004. Dec. 28. ;101. (52.):18105. -10. Epub. 2004. Dec. 15.* 101, 18105-18110 (2004).

15. Yan,Y., Svitkin,Y., Lee,J.M., Bisaillon,M. & Pelletier,J. Ribavirin is not a functional mimic of the 7-methyl guanosine mRNA cap. *RNA. 2005. Aug. ;11(8.):1238. -44*. 11, 1238-1244 (2005).

16. Westman,B. *et al.* The antiviral drug ribavirin does not mimic the 7-methylguanosine moiety of the mRNA cap structure in vitro. *RNA. 2005. Aug. 30. ;. .*, (2005).

17. Fischer,P.M. The design of drug candidate molecules as selective inhibitors of therapeutically relevant protein kinases. *Curr. Med. Chem. 2004. Jun. ;11(12.):1563. -83*. 11, 1563-1583 (2004).

18. Wu,S.Y. *et al.* Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure. (Camb.) 2003. Apr;11(4):399. -410*. 11, 399-410 (2003).

19. Gill,A., Cleasby,A. & Jhoti,H. The discovery of novel protein kinase inhibitors by using fragment-based high-throughput x-ray crystallography. *Chembiochem* 6, 506-512 (2005).

20. Hu,G., Gershon,P.D., Hodel,A.E. & Quiocho,F.A. mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl. Acad. Sci. U. S. A* 96, 7149-7154 (1999).

21. Natarajan,A. *et al.* Synthesis of fluorescein labeled 7-methylguanosinemonophosphate. *Bioorg. Med. Chem. Lett. 2004. May. 17;14(10):2657. -60*. 14, 2657-2660 (2004).

22. Poulin,F., Gingras,A.C., Olsen,H., Chevalier,S. & Sonenberg,N. 4E-BP3, a new member of the eukaryotic initiation factor 4E-binding protein family. *J. Biol. Chem.* 273, 14002-14007 (1998).

Chapter 7: Materials and methods.

7.1 Expression and protein induction methods.

7.1.1 Expression of the full length eIF4E pET11d clone.

Rossetta pLysS competent bacteria were transformed with the pET11d expression plasmid containing the full length human eIF4E clone, both materials were provided by Cyclacel Ltd., Dundee. Bacterial transformation was carried out via a typical heat shock procedure (Molecular Cloning: A laboratory Manual (3rd edition) Sambrook and Russel, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, 2001) and plated out on 100μg/ml ampicillin LB Petri dishes and incubated at 37°C overnight. A single colony was picked from the plate and used to inoculate 10mls of LB media, containing 100μg/ml ampicillin, and grown overnight in a shaker (200 rpm) at 37°C. The inoculated 10mls was then used to seed 800ml LB media, also containing 100ug/ml ampicillin, in a 2L conical flask and grown to an optical density between 0.4 and 0.6 whilst on a shaker (200 rpm) at 37°C. Induction of Protein expression was started by addition of IPTG up to a concentration of 1mM. The culture was immediately incubated on a shaker for 3 hrs at 37°C (200 rpm). Afterwards the culture was placed in 250ml centrifuge bottles and spun at 5,000 rpm for 15mins at 4°C. The cell pellets were then re-suspended in 50mM Tris pH8.0, 10% sucrose (re-suspension buffer), 1ml for every 250mls of culture. Re-suspended cell pellets were placed in 1.5ml eppendorfs and frozen at 70°C.

7.1.2 Analysis of soluble and insoluble protein fractions from expression studies.

Centrifuge 1ml samples from set time points during expression of full length human eIF4E. Spin samples for 15 minutes at 2.5 krpm on a table top centrifugue. Discard the supernatant and resuspend the pellet in 50μL H_2O. Sonicate the sample for 10 to 20 seconds, and then centrifuge at 13krpm for 15 minutes. Separate the supernatant (soluble fraction) from the pellet and re-suspend the pellet in 40μL H_2O (insoluble fraction). Analyse samples on SDS-PAGE gel (Molecular Cloning: A laboratory Manual (3rd edition), Sambrook and Russel, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, 2001).

7.2 Protein purification methods.

7.2.1 Purification of full-length eIF4E from inclusion bodies.

Two 1.5ml eppendorfs containing re-suspended cell pellets were thawed under running cold water. 40μl of Lysozyme at 10mg/ml and 60μl of 4M NaCl were added, and then incubated on ice for 45 minutes. Following incubation 5ml of re-suspension buffer was added and the sample sonicated. This was carried out on ice in 6x20sec periods, with rest intervals of 20sec in between each sonication. The sonicated sample was then centrifuged for 10 min at 10,000 rpm at 4°C. Resulting pellet was re-suspended in 5ml Tris/Triton buffer (50mM Tris pH 8.0, 2mM EDTA, 100mM NaCL, 0.5% Triton X-100) and homogenized with a glass homogenizer. The sample was then centrifuged at 12,000 rpm for 15min at 4°C. The pellet again was re-suspended in Tris/Triton buffer and the steps described repeated. The remaining pellet was then re-suspended in 7ml Guanidine/Buffer A (6M Guanidine Chloride, 50mM Hepes-KOH pH 7.6, 5mM DTT).

The protein concentration of the sample was measured via the Bradford assay and the concentration of the sample then adjusted to 1mg/ml by addition of Guanidine/Buffer A.

7.2.2 Re-folding of eIF4E by rapid dilution.

The sample from the "purification from inclusion bodies" protocol was then stirred for 2hrs at 4°C. Rapid re-folding via a 1/10 dilution was then carried out. The supernatant from the centrifuged sample was filtered via a 0.2μm filter and then slowly injected into pre-chilled refolding buffer (10 times the volume of the sample) under magnetic stirring. Refolding buffer consisted of 20mM Hepes-KOH pH 7.6, 100mM KCL and 1mM DTT. After addition of the protein sample to be refolded, magnetic stirring was carried on for a further ten minutes. The diluted refolded protein was then left at 4°C for 2hrs. The dilute refolded protein, at a concentration of ~ 0.1mg/ml, was then concentrated to a volume of 2.5ml, using Vivaspins with a 10,000KDa cut off, and run over a Amersham PD10 desalting column in order to remove residual Guanidine Chloride. The sample was desalted into refolding buffer.

7.2.3 m7GTP 4B sepharose beads analysis.

0.8mls of m7GTP beads were placed in a 12ml column and equilibrated with 20ml of Buffer A (20mM Hepes-KOH pH 7.6, 100mM KCL, 1mM DTT). 1mg total of refolded eIF4E was added to the column. The column was rotated for 2hrs before the liquid was allowed to drip off. This was then followed with a 10ml wash with Buffer A. After this step two elution washes were carried out with 3mls of Buffer A containing 0.1mM

m7GTP. The column was then washed with 10mls of Buffer A and then stored with 0.01% sodium azide.

7.2.4 Purification of eIF4E via anion exchange on a high resolution MonoQ column.

An Amersham MonoQ column was used on a FLPC AKTA (Amersham, Ltd.) machine in a refrigerated cabinet at 4°C. The column was equilibrated with 3 volumes of Buffer A (20mM Hepes-KOH pH 7.6, 1mM DTT) and the prepared eIF4E sample was loaded on. A washing step was then carried out that involved a 3 volumes wash of the column using Buffer A. The protein was then competed off with an increasing KCL gradient of up to 1M over ten volumes. This was held for a further one volume wash and then dropped down to a salt concentration of 0M to ensure all protein loaded was washed off. 0.5ml Fractions were collected while the gradient was run over the column. Peaks in the UV chromatogram were analysed by SDS-PAGE gel (Molecular Cloning: A laboratory Manual (3rd edition), Sambrook and Russel, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, 2001).

7.2.5 Affinity purification of eIF4E via m7GTP 4B sepharose column.

5 mls of m7GTP 4B sepharose beads were packed into a Tricorn HR10/50 column, both supplied from Amersham Ltd. The column was run on a FLPC AKTA machine in a refrigerated cabinet at 4°C. The column was equilibrated with 5 volumes of Buffer A (20mM Hepes-KOH pH 7.6, 1mM DTT) and the prepared eIF4E sample was loaded on. A washing step was then carried out that involved a 5 volumes wash of the column using Buffer A. The protein was then competed off with a block gradient of 0.1mM m7GTP

over ten volumes (0.1mM m7GTP, 20mM Hepes-KOH pH 7.6, 1mM DTT). 0.5ml Fractions were collected while the gradient was run over the column. The column was then re-equilibrated with a five volume wash with buffer A. Peaks in the UV chromatogram were analysed by SDS-PAGE gel (Molecular Cloning: A laboratory Manual (3rd edition), Sambrook and Russel, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, 2001)..

7.2.6 CD spectroscopy.

Sample solutions were adjusted to 0.3mg/ml by phosphate buffer (10mM sodium phosphate, pH 7.0, 1mM EDTA) through buffer exchange using an Amersham PD10 desalting column. CD spectra were measured using a JASCO-810 spectropolarimeter, reading in the region of 190nm-260nm, with a scan speed of 10nm/min, in a quartz cell of path length 0.02mm. Measurements were carried out at a constant temperature of 4°C. m7GTP did not exhibit any spectra in the region of 190nm to 260nm. Samples that were complexed with the ligand contained 0.1mM m7GTP.

7.3 Crystallisation methods.

7.3.1 Protein preparation and hanging drop vapour diffusion method for crystallisation studies.

For preliminary crystallisation screens involving:

Apo eIF4E: eIF4E from monoQ preparations was concentrated, using a 6ml vivaspin tube with a 10 000kDA cut-off membrane, to a final protein concentration of 10mg/ml. The sample before crystallisation trials was spun at 13krpm for 30 minutes on a tabletop centrifuge.

eIF4E complexed to m7GTP: eIF4E from monoQ preparations was pooled and m7GTP added to a concentration of 0.6mM (stock solution 10mM in buffer A) It was then concentrated, using a 6ml vivaspin tube with a 10 000kDA cut-off membrane, to a final protein concentration of 10mg/ml. The sample before crystallisation trials was spun at 13krpm for 30 minutes on a tabletop centrifuge.

eIF4E complexed to m7GTP and 4E-BP1 Motif Peptide: eIF4E from monoQ preparations was pooled and m7GTP added to a concentration of 0.6mM (stock solution 10mM in buffer A) It was then concentrated, using a 6ml vivaspin tube with a 10 000kDA cut-off membrane, to a final protein concentration of 8mg/ml. To the eIF4E sample solution an equivalent volume of 4E-BP1 motif peptide solution was added. Lyophilised peptide was provided by Cyclacel, Ltd., Dundee and was dissolved in buffer A to make a stock

solution at a concentration of 0.0032M. The sample before crystallisation trials was then spun at 13krpm for 30 minutes on a tabletop centrifuge.

apo eIF4E complexed to 4E-BP1 Motif Peptide: eIF4E from monoQ preparations was pooled. It was then concentrated, using a 6ml vivaspin tube with a 10 000kDA cut-off membrane, to a final protein concentration of 8mg/ml. Alternatively eIF4E eluted from the m7GTP 4B agarose beads using a KCL gradient was concentrated instead. To the eIF4E sample solution an equivalent volume of 4E-BP1 motif peptide solution was added. Lyophilised peptide was provided by Cyclacel Ltd., Dundee, and was dissolved in buffer A to make a stock solution at a concentration of 0.0032M. The sample before crystallisation trials was then spun at 13krpm for 30 minutes on a tabletop centrifuge.

eIF4E complexed to N7 derivatives and 4E-BP1 Motif Peptide: eIF4E from monoQ preparations was pooled. It was then concentrated, using a 6ml vivaspin tube with a 10 000kDA cut-off membrane, to a final protein concentration of 8mg/ml. To the eIF4E sample solution an equivalent volume of 4E-BP1 motif peptide solution was added. Lyophilised peptide was provided by Cyclacel Ltd., Dundee, and was dissolved in buffer A to make a stock solution at a concentration of 0.0032M. N7 derivatives were added in a 1/6 dilution from a stock of 50mg/ml dissolved in buffer A. The sample before crystallisation trials was then spun at 13krpm for 30 minutes on a tabletop centrifuge.

The hanging drop vapour diffusion method was used for all crystallisation studies. The drop size consisted of 1.5ul of eIF4E mixed with 1.5ul of liquid from the mother solution in the 24 well Linbro plates. Hanging drops were equilibrated with 1000ul of mother solution. Cover-slips were sealed to the wells using paraffin grease.

7.3.2 Protein crystallisation screens.

Preliminary screens carried out were structure base I and structure base II screens from Molecular Dimensions, Ltd. PEG screens against pH were also carried out corresponding to table 1 shown on the next page. These screens were carried out for PEGs 4000, 6000 and 8000.

	pH 4.0 100mM Sodium Citrate	pH 5.0 100mM Sodium Citrate	pH 6.0 100mM Bis-Tris	pH 7.0 100mM Hepes-KOH	pH 8.0 100mM Tris-HCL	pH 9.0 100mM Tris-HCL
5% PEG						
10% PEG						
20% PEG						
40% PEG						

Table 1: Grid used for initial PEG based crystallisation screens.

Ammonium Sulphate	pH 4.0 100mM Sodium Citrate	pH 5.0 100mM Sodium Citrate	pH 6.0 100mM Bis-Tris	pH 7.0 100mM Hepes-KOH	pH 8.0 100mM Tris-HCL	pH 9.0 100mM Tris-HCL
10%						
20%						
30%						
40%						

Ammonium Sulphate	pH 4.0 100mM Sodium Citrate	pH 5.0 100mM Sodium Citrate	pH 6.0 100mM Bis-Tris	pH 7.0 100mM Hepes-KOH	pH 8.0 100mM Tris-HCL	pH 9.0 100mM Tris-HCL
50%						
60%						
70%						
80%						

Tables 2 and 3: Grids used for initial ammonium sulphate based crystallisation screens.

A similar screen to the PEG based screen against pH was also carried out using ammonium sulphate as shown above (tables 2 and 3). Any crystalline type precipitates observed from the initial screens were further optimised by generating a finer grid around the conditions used. In the case of the ammonium sulphate and PEG- based screens a range of additives were also explored. These exploratory screens were carried out by

keeping the pH constant and varying the major precipitant and the additive being used.

Additives used are shown in the table 4 below:

Major Precipitant	Additive	[Major Precipitant]	[Additive]
Ammonium Sulphate	PEG 400-2000, MPD, ethanol, methanol	2.0-4.0M	6%-0.5%
PEG 1000-20000	Ammonium Sulphate, Sodium Chloride, Sodium Formate	40-50%	0.2-0.6M

Table 4: Additives used with major protein precipitating agents used in the preliminary protein crystallisation screens

These screens were performed for each protein preparation described in section 7.3.1. All stock solutions of PEG 4000-8000, 1M buffers and saturated ammonium sulphate were filtered using a 0.2μm filter.

7.3.3 Micro-seeding technique.

Drops containing crystals to be used were pippetted into 10ul of harvesting solution. Harvesting solution has a 10% higher PEG concentration than the mother solution in the well. Crystals in the harvesting solution were then crushed with a pipette with its tip melted closed. Three serial dilutions were then made using harvesting solution of 1/10, 1/100 and 1/1000. Pre-equilibrated hanging drops were then seeded using cat whiskers. The conditions used for micro-seeding were decided by choosing conditions were crystals had been observed in earlier screens and then halving the protein concentration in the drop to be seeded.

7.3.4 Cryo-freezing of protein crystals.

The freezing solutions that were used, were based on the mother liquor of the well in which the crystal grew in, with the addition of 20% glycerol to prevent ice formation. Crystals were manipulated using a nylon loop to transfer the crystal from the hanging drop to a 5ul drop of freezing solution (on a siliconised cover slip). Crystals were soaked in the drop of freezing solution for approximately 5s and then manually plunged into liquid nitrogen using the nylon loop. Crystals grown in high concentrations of ammonium sulphate were frozen using immersion oil.

7.4 Data collection and crystal structure solution methods.

7.4.1 Data collection.

Thin-plate like crystals of the eIF4E/m7GTP/4E-BP1 peptide complex, crystallised at 25% PEG 6000, 5% ammonium sulphate, pH 7.0 100mM Hepes-KoH, were cryo-frozen using freezing solution containing 20% Glycerol, 5% ammonium Sulphate, 25% PEG 6000 and 100mM Hepes-KOH, pH 7.0. The eIF4E/m7GTP/peptide crystal was mounted randomly and 120 degrees of data were taken with an oscillation angle of 1 degree using the Phi rotation method at 100K at the ESRF ID29 beamline at Grenoble.

Co-crystals of the N7 derivatives complexed with the eIF4E/4E-BP1 peptide complex were crystallised at 25% PEG 6000, ~7% ammonium sulphate either at pHs 7.0 or pH 6.8 100mM Hepes-KOH or pH 6.4 100mM MOPS. The thin-plate like crystals of the eIF4E and N7 substituted derivative peptide complexes were cryo-frozen using freezing solution containing 20% Glycerol, 5% ammonium sulphate, 25% PEG 6000 and 100mM of the

buffer used in the well they were harvested from. The co-crystals were then mounted randomly and 120 degrees of data were taken with an oscillation angle of 1 degree using the Phi rotation method at 100K at the 14.1 PX station at Daresbury.

7.4.2 Molecular replacement and structure solution.

m7GTP/eIF4E/4EBP1 co-crystal structure: Molecular replacement solutions were carried out using MOLREP in CCP4 v5.0[1]. The Mathews coefficient was estimated using the CCP4 suite of programs, using a molecular weight of 27393kDa for the m7GTP/eIF4E/4EBP1 peptide complex. This gave a unit cell with a solvent content of 47.7% and a V_m of 2.4 corresponded to two ternary complexes in the asymmetric unit. The 2.2 angstrom structure of $\Delta27$ murine eIF4E complexed with the 4E-BP1 peptide was used as the search model in MOLREP[1], all waters and the m7GDP ligand were removed from the model. Data up to 3.5 angstrom was used for the MOLREP[1] solution. MOLREP[1] isolated the positions of two ternary complexes in the asymmetric unit. These initial co-ordinates were used to begin refinement.

N7 derivatives complexed with the eIF4E/4E-BP1 peptide complex: The 2.3 angstrom structure of full length eIF4E with the 4E-BP1 peptide (solved in chapter3) was used as the search model in MOLREP[1], all waters and the m7GTP ligand were removed from the model. Data up to 3.5 angstrom was used for the MOLREP[1] solution. MOLREP[1] isolated the positions of two ternary complexes in the asymmetric unit. These initial co-ordinates were used to begin refinement.

7.4.2 Model building and refinement.

Prior to refinement, $2F_0$-F_c and F_0-F_c maps were calculated using FFT in CCP4i[1]. The models for the co-crystal structures were examined to see whether density was continuous for the entire length of the polypeptide chains. The starting models were subjected to 10 cycles of rigid body refinement and followed by 10 cycles of restrained refinement in Refmac 5.0[1,2]. At this point waters were added using ARP_Warp[3]. After this the different ligands for the co-crystals were used (refmac library files generated using PRODRG[4]) in the refinement and placed into clearly visible electron density. This was then followed by continuous cycles of manual refinement of the model COOT[5] and 10 cycles of restrained refinement using Refmac 5.0[1,2]. Positions of structured water were also verified and added using COOT[5]. Models were validated using PROCHECK[1].

7.4.3 Structural analysis.

Final models were analysed using COOT[5], PYMOL[6] and INSIGHTII (Acceryls, Inc.). All structural superimpositions carried out were in INSIGHTII (Acceryls, Inc.) overlaying the α-carbon backbone. Superimposition between the co-crystals structures of m7GTP and the N7 derivatives complexed with full-length eIF4E and the 4E-BP1 peptide were overlaid in COOT[5],.

7.5 Surface plasmon resonance (SPR) studies.

The SPR assay was performed on a BIAcore 2000 (BIAcore, UK). The Sensorchip SA (BIAcore) was used to immobilise the biotin labelled 4E-BP1 peptide and the full-length human eIF4E was injected over the chip in the presence of m7GTP (in a ratio of 5:1 protein concentration). 600 RUs of biotinylated peptide were immobilised on the chip. Each cycle consisted of a 150μl injection of buffer C (10mM HEPES pH 7.5, 300mM KCL, 50μM EDTA and 0.005% surfactant p20 (BIAcore) followed by a 90μl injection of eIF4E in buffer C. eIF4E was injected over the chip in a series of injections that covered a concentration range from 0.005 to 5 μM. This was then followed by an injection of 450μl of buffer C to allow dissociation of eIF4E. The chip was regenerated after each cycle using 50mM glycine pH 3.0 buffer. All measurements were performed at a flow rate of 30μl/min at 25°C. The resulting sensograms were evaluated using the BIAevaluation package. The response from the SA chips (control) without immobilised 4E-BP1 motif peptide was subtracted from the chips with the peptide immobilised. The control chips were coated with free biotin to prevent non-specific binding of eIF4E. The resulting data was analysed using a two step model.

7.6 Mass spectrometry Materials and methods.

7.6.1 Protein preparation for mass spectrometry studies.

eIF4E purified via a monoQ column was adjusted to a concentration of 2mg/ml and dialysed over night, using slide-a-lyzers from Pierce, Ltd, into 10mM ammonium acetate pH 6.5. The protein concentration was then measured at A_{280} using an extinction coefficient of 53440. All the samples were run on the mass spectrometer at a protein concentration of $3.2\mu M$.

7.6.2 Electrospray ionisation mass spectrometry.

ESI mass spectra were acquired using a single quadropole mass spectrometer (Micromass, Wyntheshawe, UK) equipped with a Z-spray electrospray ionisation source. The mass spectrometer has a m/z range of up to 3500. Samples were injected directly into the source using a Harvard Model 22 syringe at a flow rate of $5\mu L\ min^{-1}$. The conditions used to obtain mass spectra of eIF4E and the various ligand complexes were capillary voltage, 3.5kV; cone voltage, 45V;source block temperature, 60°C; desolvation temperature, 90°C; extraction voltage, 4V; RF lens, 0.3V. Spectra were acquired in positive ion mode over a m/z range of 500 to 3500 in continuous mode, with a scan time of 12s. The ESI spectra shown in this work and those used for measurements of the intensities of ion peaks were subjected to background subtraction and were smoothed. This data manipulation was carried out in Mass Lynx version v3.5 (Micromass, Wyntheshawe, UK).

7.6.3 Preparation of cap-anlogue/eIF4E complexes.

Cap-analogues (m7GTP, m7GDP, GTP, GDP, GMP, ATP, ADP, AMP, ribavirin, m7G and guanosine) were ordered and supplied from Sigma-Aldrich, Ltd. The ligands were dissolved in 10mM ammonium acetate pH 5.0 at stock concentration of 30mM. Serial dilutions in 10mM ammonium acetate pH 5.0 were made to prepare cap-analogues for use in ligand titrations against eIF4E. Samples to be run on the mass spectrometer were prepared as follows:

Reagent	Amount	[Concentration]
eIF4E	XµL	3.2µM
Ligand	XµL	XµM
Methanol	10µL	10%
Buffer (ammonium acetate)	XµL	10mM
Distilled water	XµL (make up to a 100µL)	

In experiments were ammonium acetate concentrations were varied a 1M stock was used at pH 5.0 and the required amounts added to generate concentrations between 10mM and 500mM. In experiments were pH was varied different ammonium buffers at different pHs were used. Glacial acetic acid was used to pH the ammonium acetate buffers. eIF4E was run under denaturing conditions on the mass spectrometry by increasing the concentration of methanol to 50% and using an ammonium acetate buffer at pH 3.0.

7.6.4 Mass spectrometry based screening.

Compounds were screened at two different concentrations against eIF4E at 1:3 and 1:5 ratios unless described otherwise. The concentration of eIF4E being analysed in these samples was kept at 3.2μM. Running parameters of the mass spectrometer were maintained as in section 7.6.2. Compounds that were soluble in methanol were added to the sample as described in section 7.6.3 but the methanol concentration was maintained at 10%. For compounds only soluble in DMSO, methanol was maintained at a constant 10% in the samples to be analysed, whilst DMSO concentration were kept as low as possible without precipitation occurring. DMSO control curves were run on apo eIF4E, GTP complexed with eIF4E (3.53μM) and m7GTP (3.72μM) complexed eIF4E at 1% DMSO. Spectra were acquired as mentioned in section 7.6.2.

7.6.5 Mass spectrometry data analysis.

The experimental data, from the cap-analogue titrations against eIF4E, were shown by plotting ΣPeak intensity/(Σapo peak intensity+ Σcomplex peak intensity) against the concentration of the ligand being used. **Peak** corresponds to the peak of the charge species being analysed, i.e. apo or apo + one ligand or apo + two ligand. The term **complex peak** includes all peaks corresponding to the binding of ligand molecules in the charge species being analysed. The height of the peak intensities of the 11+ and 10+ charge species were used for the plot described and averaged out. These charge species were used as they comprised > 85% of the total signal seen for eIF4E at pH 5.0 and peak shifts corresponding to ligand binding were easily resolved. Intensity values were only

taken after acquired mass spectra had been treated with a polynomial background subtraction and smoothed using default values in Mass Lynx v3.5. Plots were created using SigmaPlot 9.0 (Systat Software Inc.). The same procedure was applied to studying the effect of pH on ligand binding. Σ**Peak** intensity/(Σ**apo peak** intensity+ Σ**complex peak** intensity) was plotted against pH instead of concentration. All data points for titrations concerning m7GTP, m7GDP, GTP, GDP and GMP were done in triplicate and error bars shown. All other analyses were not duplicated and no error bars shown.

7.7 Data-mining for potential drug leads.

7.7.1 Virtual screening of the eIF4E cap-binding site with LIDAEUS.

The full-length human eIF4E structure complexed to m7GTP (1IPB) solved by Tomoo et al[7] was used as the structural template for virtual screening runs. All waters and ligand molecules must be removed. A program called MAPSITE is used to generate map sites that LIDAEUS[8] uses in the initial stages of docking. MAPSITE requires a mol2 file of the ligand specifying where to make the site points and a PDB of the protein under study. Default parameters were used to run MAPSITE. The mol2 file produced by MAPSITE was then converted to a .clu file using the mol2clu utility program. The parameters of the pose.dat file, which controls the tolerance of the fit to the site points generated, was also maintained at the default settings. LIDAEUS[8] was then run and the top 250 best scoring compounds retained.

7.7.2 ISISBase searching for analogous quaternary ammonium ions to N7 in m7GTP.

ISISBase was used to search for compounds within the ACD (Available Chemical Database) by designing template ring systems that contained a charged nitrogen and then using the similarity search option to search for potential compounds. Levels of similarity specified in the search ranged from 30% to 50%.

7.8 Synthesis of ligand cap-analogues and verification with mass spectrometry.

7.8.1 Materials.

Guanosine 5'-monophosphate, quinolene, isoquinolene and benzothiazole were ordered from Sigma-Aldrich, Ltd. All alkyl and aryl halides were also bought from Sigma-Aldrich, Ltd. All chemical were of reagent grade or higher quality.

7.8.2 Synthesis of cap-analogues.

The 7-substituted GMP cap analogues were synthesized from GMP (Na^+, 0.28mM) and the corresponding alkyl halide (1mM) by the general method described earlier[9,10]. Reactants were incubated at room temperature for 48 h in 5ml of DMSO, or in the case of quinolene, isoquinolene and benzothiazole in dichloromethane (DCM). Upon completion of the reaction, acetone was added, and free halide was removed by several washes with ethyl ether. After each wash the sample was centrifuged and the supernatant decanted off.

7.8.3 Verification of reaction products using mass spectrometry.

Reaction product was taken and dissolved in 90% H_2O and 10% methanol and analysed on a single quadropole mass spectrometer (Micromass, Wyntheshawe, UK) equipped with a Z-spray electrospray ionisation source. The mass spectrometer has a m/z range of up to 3500. Samples were injected directly into the source using a Harvard Model 22 syringe at a flow rate of $20\mu L$ min^{-1}. The conditions used to obtain mass spectra of the reaction products were capillary voltage, 3.5kV; cone voltage, 45V; source block temperature, 60°C; desolvation temperature, 90°C; extraction voltage, 4V; RF lens, 0.3V. Spectra were acquired in positive ion mode over a m/z range of 100 to 1000 with scan time of 5s. The ESI spectra shown in this work and those used for measurements of the intensities of ion peaks were subjected to background subtraction and were smoothed. These data manipulations were carried out using Mass Lynx, version 5 (Micromass, Ltd.).

Reference List

1. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760-763 (1994).

2. Murshudov,G.N., Vagin,A.A. & Dodson,E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240-255 (1997).

3. Perrakis,A., Sixma,T.K., Wilson,K.S. & Lamzin,V.S. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 448-455 (1997).

4. van Aalten,D.M. *et al.* PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des* **10**, 255-262 (1996).

5. Emsley,P. & Cowtan,K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr. 2004. Dec. ;60. (Pt. 12. Pt. 1):2126. -32. Epub. 2004. Nov. 26.* **60**, 2126-2132 (2004).

6. DeLano,W.L. The PyMOL User's Manual. 2002.

7. Tomoo,K. *et al.* Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m(7)GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region. *Biochem. J. 2002. Mar. 15. ;362. (Pt. 3):539. -44.* **362**, 539-544 (2002).

8. Wu,S.Y. *et al.* Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure. (Camb.) 2003. Apr;11(4):399. -410.* **11**, 399-410 (2003).

9. Darzynkiewicz,E., Ekiel,I., Lassota,P. & Tahara,S.M. Inhibition of eukaryotic translation by analogues of messenger RNA 5'-cap: chemical and biological consequences of 5'-phosphate modifications of 7-methylguanosine 5'-monophosphate. *Biochemistry* **26**, 4372-4380 (1987).

10. Darzynkiewicz,E. *et al.* Inhibition of eukaryotic translation by nucleoside 5'-monophosphate analogues of mRNA 5'-cap: changes in N7 substituent affect analogue activity. *Biochemistry* **28**, 4771-4778 (1989).