Evaluating the Conformation and Binding Interface of Cap Binding Proteins and Complexes via Ultraviolet Photodissociation Mass Spectrometry

*John P. O'Brien[1], Laura K. Mayberry[1,2], Patricia A. Murphy[1,2], Karen S. Browning[1,2] and Jennifer S. Brodbelt[1]**

[1]Department of Chemistry and Biochemistry and
[2]Institute for Cell and Molecular Biology
The University of Texas at Austin
1 University Station A5300
Austin, TX, USA 78712

SUPPLEMENTAL INFORMATION

Mass Spectrometry

All experiments using eIF4E and mini-eIF4F were performed on a Bruker HCTUltra 3-D quadrupole ion trap mass spectrometer (Billerica, MA), and experiments with eIFiso4E and eIF4F were performed on a Thermo Fisher Velos Pro linear ion trap mass spectrometer (San Jose, CA). Both instruments were modified to allow photodissociation as described previously.[1,2] UVPD-MS experiments were performed using a Coherent ExciStar-XS excimer laser (Santa Clara, CA) operated at 351 nm at 500 Hz. The UVPD mass spectra were acquired using twenty 3 mJ pulses.

Liquid Chromatography

All eIF4E and mini-eIF4F protein digests were analyzed by liquid chromatography (LC) performed using a Dionex 3000 capillary LC system (Sunnyvale, CA). An Agilent ZORBAX 300 Extend-C18 column (Santa Clara, CA) (150 × 0.3 mm, 3.5 μm particle size) was used for all separations. Mobile phase A (MPA) consisted of 0.1% formic acid in water and mobile phase B (MPB) 0.1% formic acid in acetonitrile. The flow rate was 4 μL/min, and a linear gradient from

2% MPB to 40% MPB over 60 minutes was applied. The column was flushed with 80% MPB for 10 minutes and re-equilibrated using 2% MPB for an additional 10 minutes. Injections of approximately fifty picomoles were used for each protein digest. For all LC-CID-MS/MS runs, the first event was the full mass scan (*m/z* range of 400 – 2800). The five most abundant ions from the full mass scan were selected for CID (0.5 V). For all LC/UVPD runs, the first event was the full mass scan (*m/z* range of 400 – 2800) followed by consecutive UVPD events on the five most abundant ions from the ESI-MS survey scan.

eIFiso4E and eIF4F digests were analyzed using a Dionex nanoRSLC nanoLC (Sunnyvale, CA) with mobile phases consisting of A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. The flow rate was set to 300 nL/min and used a similar gradient as described above. Injections of approximately 1 picomole were used for each protein digest. For all data dependent LC-CID-MS/MS runs, the first event was the full mass scan (*m/z* range of 350 – 1800). The ten most abundant ions from the full mass scan were selected for CID (NCE 35%). For all LC/UVPD runs, the first event was the full mass scan (*m/z* range of 350-1800) followed by consecutive UVPD events on the ten most abundant ions from the full mass scan.

Database Searching

Processed CID mass spectra were searched using MassMatrix, a free online database search algorithm for peptide MS/MS data (www.massmatrix.net).[3–6] MassMatrix was modified to include the sequences of eIF4E, isoeIF4E and eIF4G in a custom protein database. Peptides with up to four possible missed cleavages were specified as a search parameter. A peptide mass accuracy of 2.0 Da was used for full mass spectra and a peptide fragment mass accuracy of 0.8 Da was used for the MS/MS spectra. Dynamic chemical modifications corresponding to M-oxidation, S,T,Y-phosphorylation and NN addition were included as variable search parameters.

Peptide matches were identified using three scoring parameters: pp and pp2 cut-off values of 5.0, and a ppTag cut-off value of 1.3 was used to positively confirm peptide matches within the database. In addition all matched MS/MS mass spectra were manually inspected to confirm identification.

Derivatization and Sample Preparation

For all NN reactions, 10 nmol protein was reacted with NN at a 20:1 protein:NN mole ratio (with NN at 20 mM in DMSO) in 200 μL PBS buffer (pH 7.3 - 7.5) with 10 μM m^7GTP ligand and 0.1 mM DTT. The reaction was allowed to proceed for 30 minutes at room temperature prior to removal of unreacted NN by using a 10 kDa MWCO filter. These reaction conditions were chosen based on the inspection of ESI mass spectra of the intact NN-modified proteins (**Supplemental Figure 9**) and circular dichroism measurements of the same solutions (**Supplemental Figure 10**) to ensure that the protein did not exhibit signs of denaturation after the NN reactions. The modified proteins were diluted in 75 μL of 100 mM NH_4HCO_3 and digested overnight at 37 ^oC with trypsin (1 mg/mL in 1 mM HCl) using a 1:20 w/w ratio of protein to protease. All protein reactions and digests were analyzed immediately and performed in triplicate. To examine eIF4E and eIFiso4E in a completely reduced state, the NN reactions were undertaken in an identical manner as the reactions described above but 1 mM DTT was added to the PBS reaction buffer. The full 186 kDa eIF4F protein complex was reacted at a slightly larger molar ratio of 1:30 protein:NN (to compensate for the greater number of lysines in the larger 165 kDa eIF4G protein) using 0.37 μL of 20 mM NN in a total reaction volume of 30 μL in PBS buffer. Proteolysis of the eIF4F complex was performed using trypsin, glu-C or chymotrypsin and reacting 15 ug of protein using a protein ratio of 1:20 (w/w). Trypsin (1 mg/mL in 50 mM acetic acid) digests were performed in 75 μL reaction mixtures containing 100

mM NH_4HCO_3 (pH 8.5) buffer, GluC (1 mg/mL in water) digests were performed in 75 µL containing 50 mM Tris-HCl (pH 8), 0.5 mM Glu-Glu (a dipeptide that facilitates the proteolytic activity of GluC), and chymotrypsin (0.5 mg/mL in 1 mM HCl) digestions were performed in 75 µL containing 100 mM Tris-HCl (pH 8), 2 mM $CaCl_2$. Trypsin and GluC digests were incubated at 37^{o}C, and chymotrypsin digestions were incubated at room temperature. Due to limited quantity of the eIF4F complex, reactions and digestions were performed once.

Circular Dichroism

Circular dichroism (CD) experiments were performed using a Jasco J-815 CD spectrometer (Easton, MD), and the instrument was operated at ambient temperature under a stream of nitrogen gas. Spectra were acquired between 190 nm and 260 nm using a spectral scanning speed of 50 nm/min and each spectrum acquired was an average of three scans. Protein solutions were prepared by diluting the samples to 1.0 µM using 10 mM KH_2PO_4, 100 mM NaF buffer (pH 7.4) and analyzed using a 1 mm quartz cell.

Determination of Primary Amine Reactivities

The reactivity of each lysine residue (α-amine) and the N-terminus was calculated based on assessment of the peak areas of all modified and unmodified peptides identified in the digests from the extracted ion chromatograms (EIC). The integrated areas for the EIC of each peptide were tabulated using the Data Analysis program in the Bruker software or QualBrowser in the Thermo software package. The reactivities were then calculated by dividing the sum of the areas of all peptides containing each modified residue by the sum of the areas of all the peptides (both modified and unmodified) containing the same residue as summarized below.

$$Residue\ Percent\ Reactivity = \frac{\sum Area\ of\ all\ peptides\ containing\ modified\ residue\ N}{\sum Area\ of\ all\ peptides\ containing\ residue\ \ N}\ x\ 100$$

Lysine and N-termini solvent accessibilities were calculated using protein structures and the protein structure viewing program PyMOL.[7] Briefly, the areas of each lysine-containing residue were calculated using the "get_area" function in PyMOL using a dot_density of 4. Solvent accessibilities (SA) were calculated by dividing the tabulated lysine solvent exposed areas by the area of a completely exposed lysine residue (as found in peptide Gly-Lys-Gly)[8] and multiplying by 100. Examples of the SA calculations are found in **Supplemental Table 9**. A higher SA value signifies greater exposure of the residue. The general correlation between the NN percent reactivity and solvent accessibility is anticipated for most residues but is not expected to be entirely quantitative. In brief, the SA calculated for each amine residue is based on a water molecule interacting with each residue's surface area. Several factors are expected to affect NN reactivity, including the chemical properties of the probe, such as size and polarity, as well as the chemical properties of the residue sites like pKa, nucleophilicity, and polar contacts with other interacting residues. It should be noted that all of the SA calculations were performed for structures with the m^7GTP cap removed, a factor which could presumably affect the chemical probe reactivity within the binding region.

Primary Amine Reactivities and Solvent Accessibilities of eIF4E in complex with eIF4G peptide (mini-eIF4F complex)

The reactivity of K56 was expected to decrease in the protein complex as K56 is near the cap binding region, but in fact a decrease in reactivity was not observed for K56 in the mini-complex. This suggests that K56 exhibits a stronger salt bridge interactions with T65

(**Supplemental Figure 5A**) while in complex with the full length eIF4G protein compared to the mini-eIF4G protein. K182 is in close proximity to the eIF4G binding interface, and it would not be unexpected that the larger eIF4G protein interacts with K182, creating a less solvent exposed structural environment not induced by the smaller eIF4G peptide. K210 and K191 were anticipated to react with the NN probe due to their high predicted solvent accessibilities, but these residues did not react with NN in the truncated versions of the eIF4E protein and eIF4F mini-complex. However, the K191 and K210 residues within the full eIF4F complex did react with NN. The lack of reactivity of K191 and K210 in the monomeric and mini-complex states suggest that K191 and K210 help stabilize the tertiary structure of eIF4E in the absence of the large eIF4G protein. This result implies that K210 and K191 are reactive in the full eIF4F complex due to the ability of the larger eIF4G protein to stabilize eIF4E and increase the exposure of K210 and K191.

Homology Model of eIFisoE

Wheat eIF4E proteins are unique because they contain two isoforms with the same biological activity (mRNA cap binding) but with different amino acid sequences. The second isoform of eIF4E, eIFiso4E, has 209 amino acids including 18 lysines and the N terminus. There are no known crystal structures of eIFiso4E. The sequence identity of eIF4E and eIFiso4E is 52% (91 of 177 residues of eIF4E are identical to the eIFiso4E sequence calculated via BLAST, www.blast.ncbi.nlm.nih.gov) (see **Figure 3**), suggesting that the tertiary structure of eIF4E can be used to model the structure of eIFiso4E with a root mean square (RMS) error of ~1 Å.[9] SWISS-MOL was thus utilized to create a structure of eIFiso4E by imposing the primary amino acid sequence of eIFiso4E onto the solved X-ray crystal structures of m^7GTP bound. The homology models of eIFiso4E produced by SWISS-MOL are illustrated in **Supplemental**

Figure 3 and provide a template for calculation of the solvent accessibilities of the lysine amines. Similar to the experimental workflow and data analysis described above for eIF4E, the primary amine reactivities of eIFiso4E in reducing and non-reducing environments were determined using the NN-chemical probe and MS/MS methods. The results are reported in **Table 1** (and as histograms in **Supplemental Figure 2)** along with the calculated solvent accessibility values. The solvent accessibilities of A1 and K29 were not calculated because these residues are not in the homology model (which is based on the structure of truncated eIF4E).

NN Reactivities and Structural Comparison of eIF4E and eIFiso4E

Several interesting differences in NN reactivities between eIF4E and eIFiso4E for conserved residues (highlighted gray in **Table 1**) are evident. The structure of eIF4E and the homology crystal model of eIFiso4E (**Figure 1B** and **Supplemental Figure 3B,** respectively) provide insight into the structural differences between the two isoforms and help rationalize the differences in NN reactivities between the proteins, thus showing that this method can accurately report structural differences. K56 of eIF4E is less reactive than the eIFiso4E K47/K49 residues. This outcome can be explained because K56 in eIF4E is hydrogen bonded to T65 (**Supplemental Figure 5A**), an interaction that is not observed for either K47 or K49 of eIFiso4E (**Supplemental Figure 5B**). A similar difference is noted for K107 in eIF4E and K99 in eIFiso4E. In the eIF4E structure, K107 forms an ion pair interaction with E105 at a distance of 3.2 Å (**Supplemental Figure 6A**). In the eIFiso4E structure, K99 can similarly ion pair with E97 (**Supplemental Figure 6B**), but there is more open space around K99 as the residue I66 (in eIF4E) is replaced with L59 which leads to rotation of the side chain away from K99, creating easier access to the chemical probe. The K110 residue of eIFiso4E is more exposed than K118 in eIF4E. In the eIF4E structure, K118 forms hydrogen bonds with water molecules near the cap

binding site (**Supplemental Figure 7A)**. As expected K118 (in eIF4E) did not react with the probe, but some reactivity was observed for the analogous K110 residue in eIFiso4E. This difference in reactivity can be explained by the replacement of P209 in eIF4E by R203 in eIFiso4E (**Supplemental Figure 7B)**. It is plausible that the R203 residue can orient itself near K110, thus facilitating a stronger reaction with NN or alternatively R203 could push K110 into a more exposed position. In the eIF4E structure, K182 forms a hydrogen bond with S179 (**Supplemental Figure 8A**), but this residue changes to D172 in the eIFiso4E model which gives the potential for K175 to create a salt bridge with D172 (**Supplemental Figure 8B**) which can suppress the reactivity of K175.

References

(1) Shaw, J.; Madsen, J.; Brodbelt, J. S. Systematic Comparison of Ultraviolet Photodissociation and Electron Transfer Dissociation for Peptide Anion Characterization Journal of The American Society for Mass Spectrometry. *J Am Soc Mass Spectrom* **2012**.

(2) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11*, 1329–1334.

(3) Xu, H.; Freitas, M. A. A Mass Accuracy Sensitive Probability Based Scoring Algorithm for Database Searching of Tandem Mass Spectrometry Data. *BMC Bioinformatics* **2007**, *8*, 133.

(4) Xu, H.; Zhang, L.; Freitas, M. A. Identification and Characterization of Disulfide Bonds in Proteins and Peptides from Tandem MS Data by Use of the MassMatrix MS/MS Search Engine. *J. Proteome Res.* **2008**, *7*, 138–144.

(5) Xu, H.; Freitas, M. A. Automated Diagnosis of LC-MS/MS Performance. *Bioinforma. Oxf. Engl.* **2009**, *25*, 1341–1343.

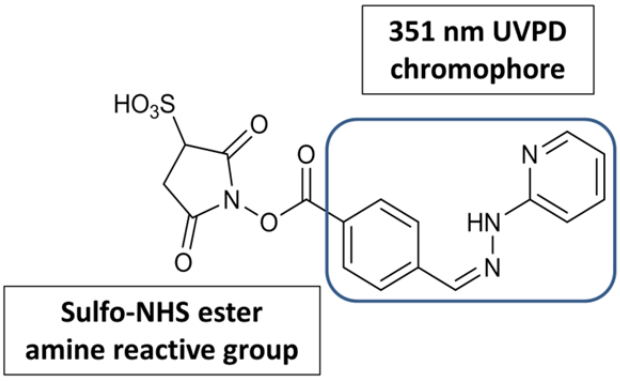
(6) Xu, H.; Wang, L.; Sallans, L.; Freitas, M. A. A Hierarchical MS2/MS3 Database Search Algorithm for Automated Analysis of Phosphopeptide Tandem Mass Spectra. *Proteomics* **2009**, *9*, 1763–1770.

(7) Schrödinger, L. The PyMOL Molecular Graphics System, Version 1.3r1 **2010**.

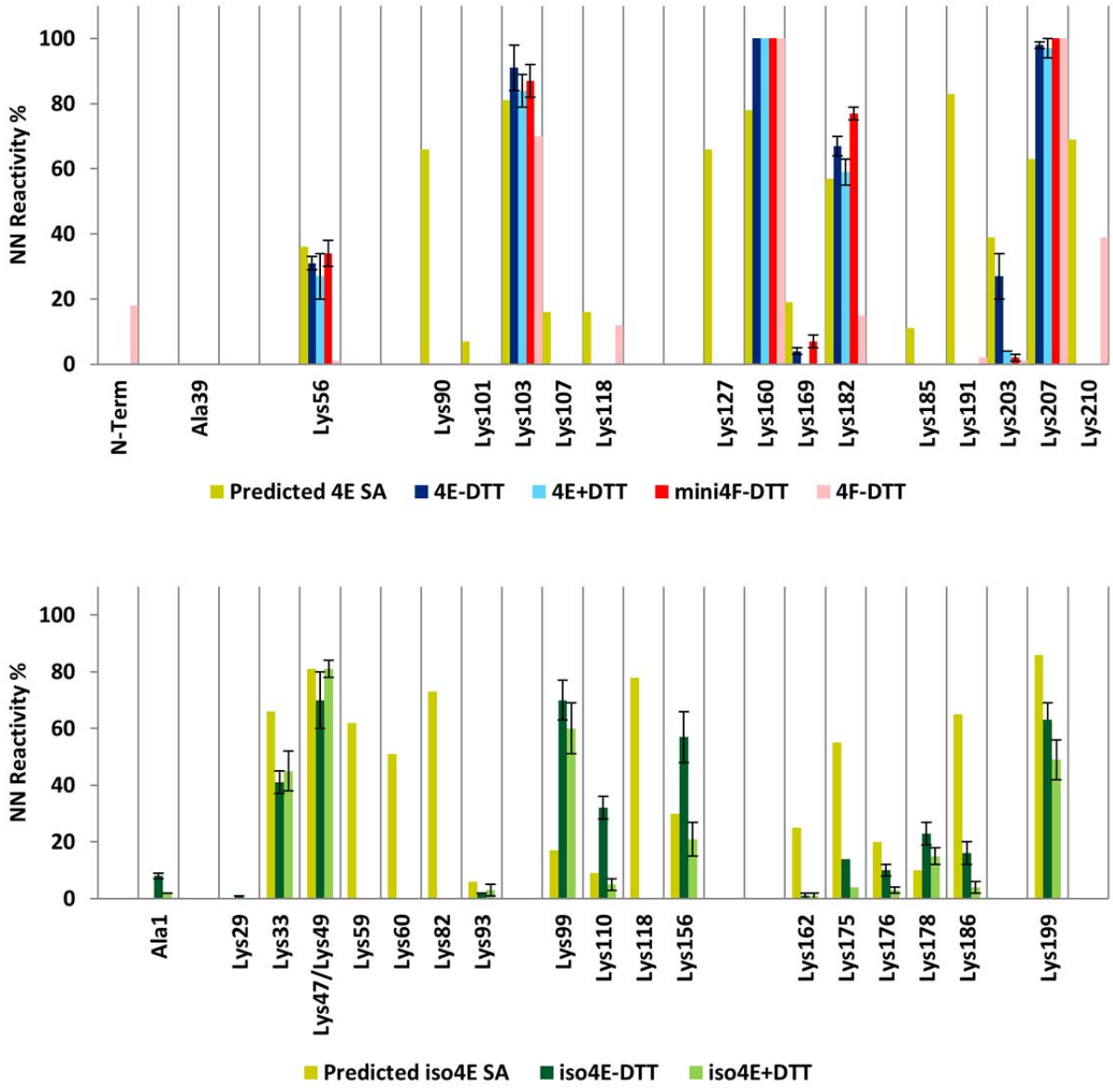
(8) Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. Interior and Surface of Monomeric Proteins. *J. Mol. Biol.* **1987**, *196*, 641–656.

(9) Baker, D.; Sali, A. Protein Structure Prediction and Structural Genomics. *Science* **2001**, *294*, 93–96.

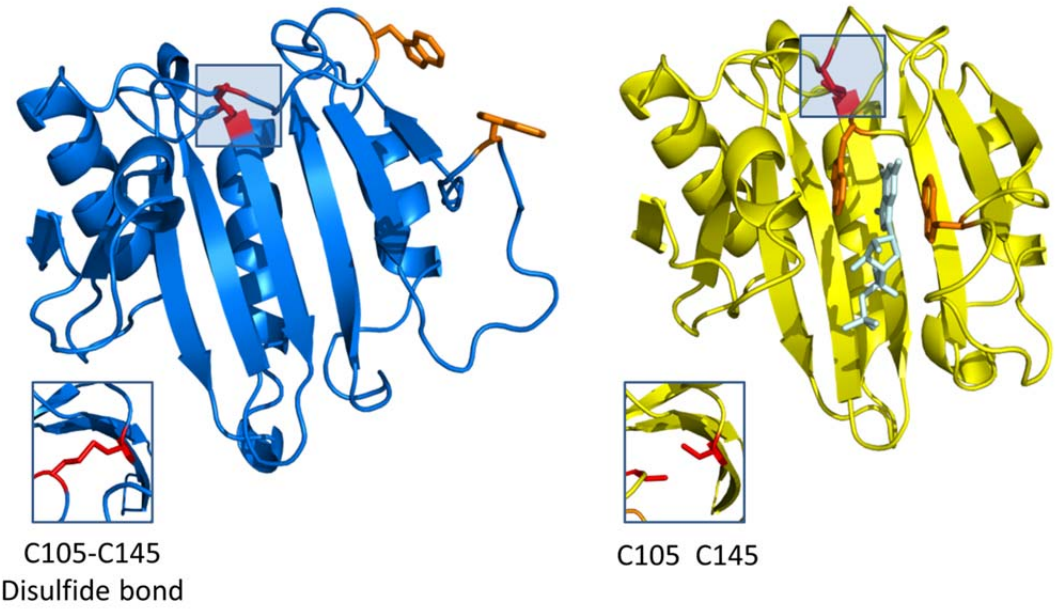
Supplemental Figures:



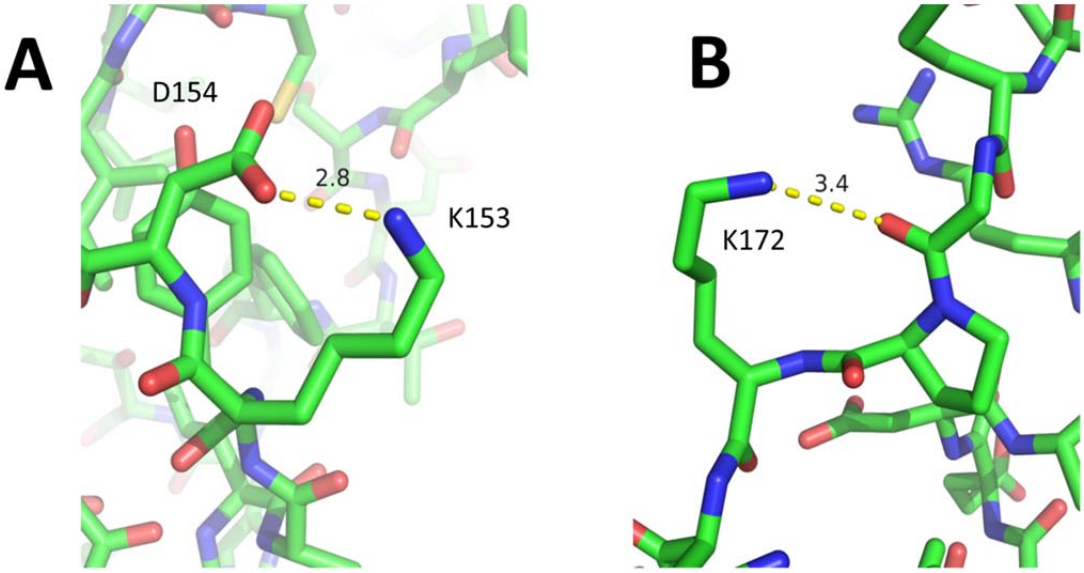
Supplemental Figure 1. Structure of NN amine-reactive chemical probe.



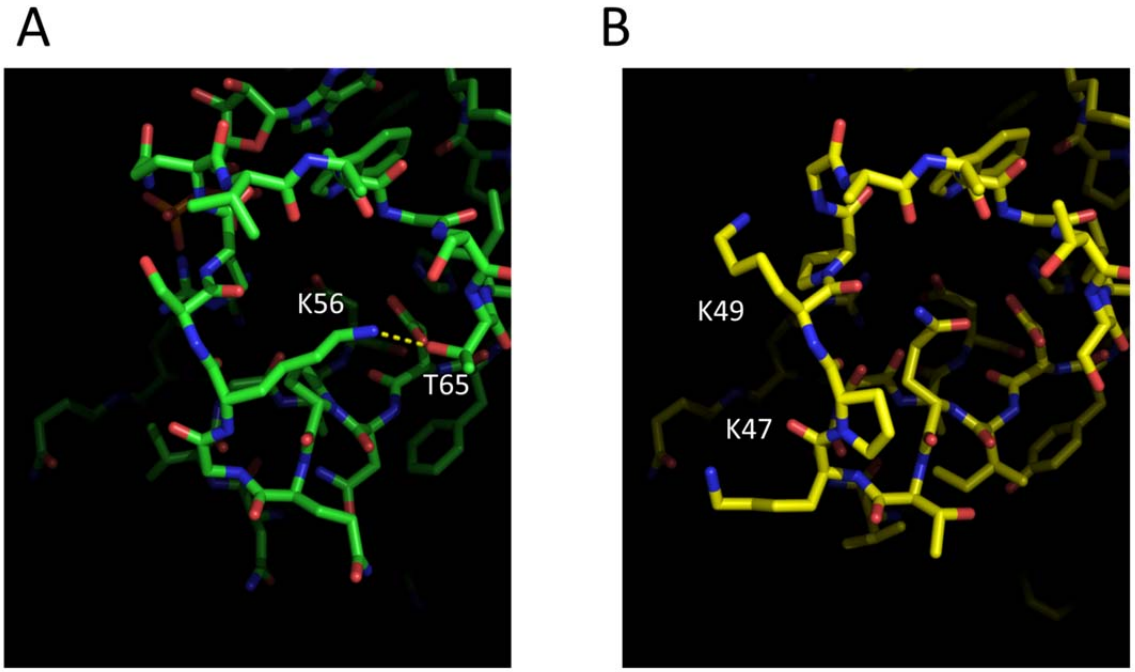
Supplemental Figure 2. Histograms of the observed NN reactivities (for the data shown in **Table 1**) for eIF4E and eIF4F (top) and eIFiso4E (bottom). Legend keys for each of the corresponding proteins are shown below the graphs. Overlapping conserved residues between eIF4E and eIFiso4E are shown in alignment of the two histograms. The predicted solvent accessibilities are shown as gold bars.



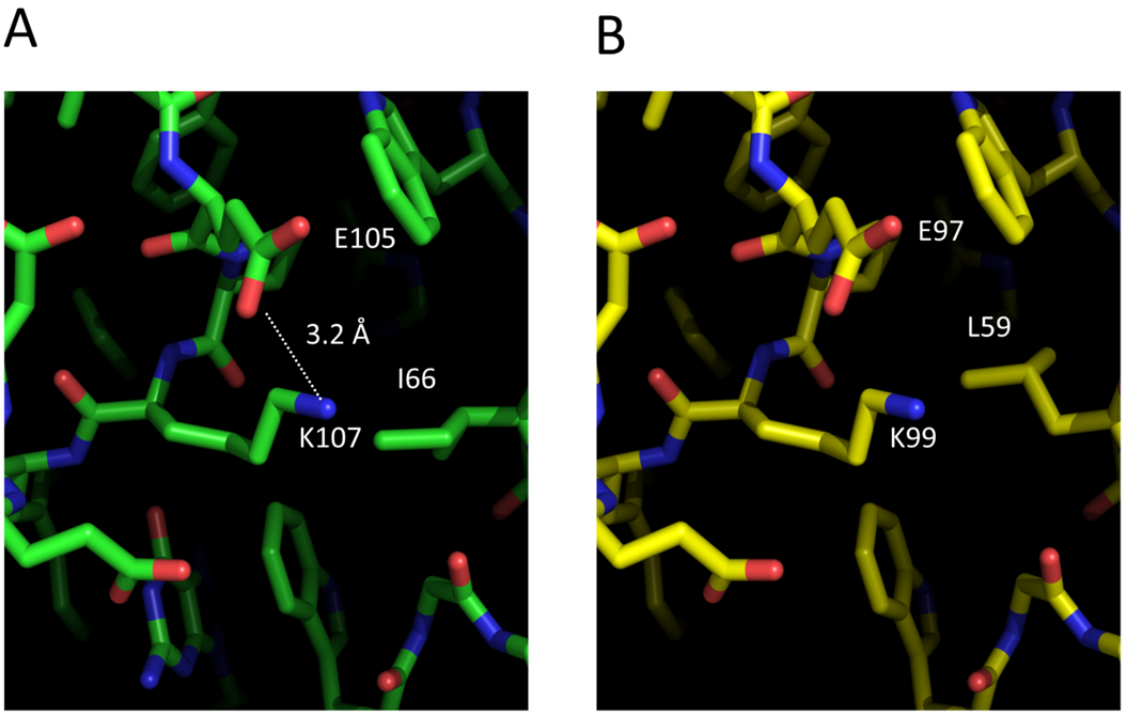
Supplemental Figure 3. Structural homology models of eIFiso4E (**A**) in the absence of m^7GTP and (**B**) with m^7GTP bound.



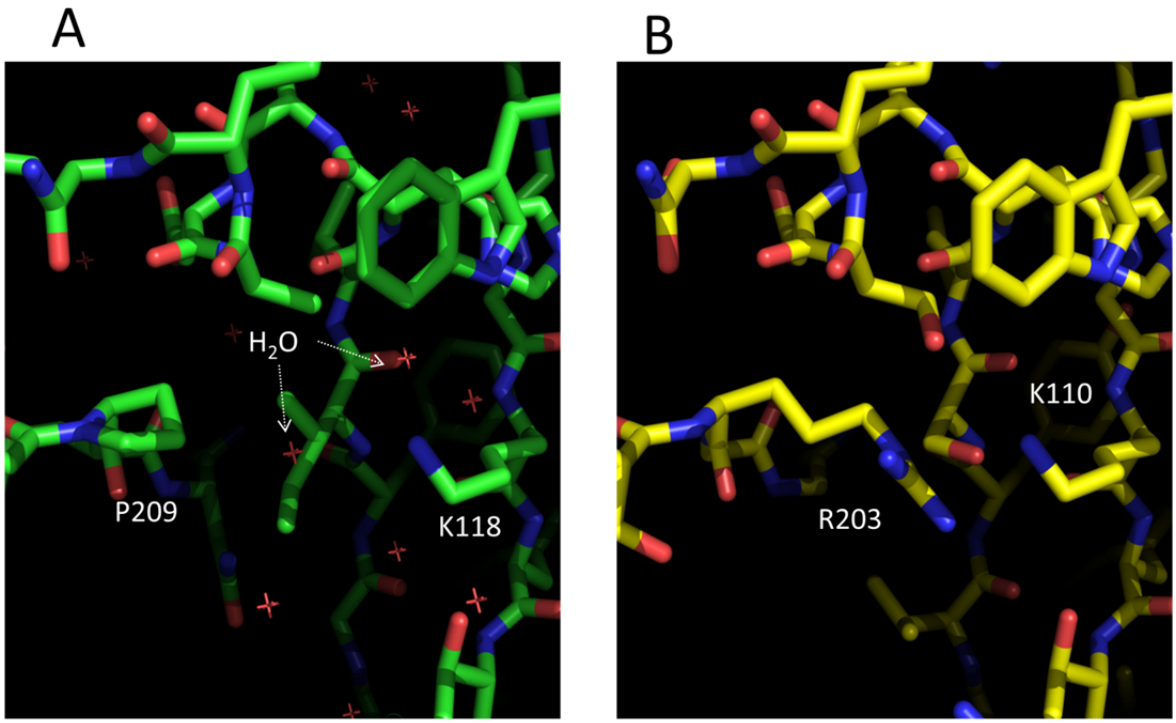
Supplemental Figure 4. Images of truncated eIF4E showing the polar contacts between (A) the ε-amine of lysine 153 and the carboxylate group of aspartic acid 154 and (B) the ε-amine of lysine 172 and the amide backbone of eIF4E. The distances between these sites are shown in angstroms (Å).



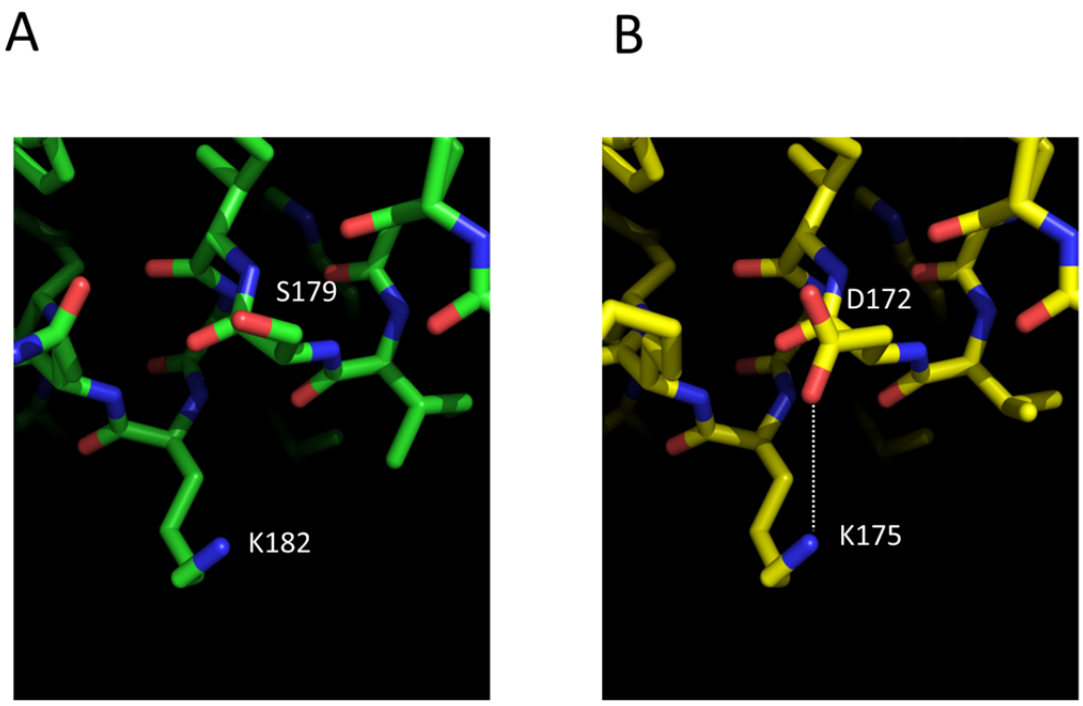
Supplemental Figure 5. Side-by-side structural images showing the (A) polar contacts between eIF4E K56 and T65 (PDB 2IDV) and (B) location of K47 and K49 in the model of eIFiso4E.



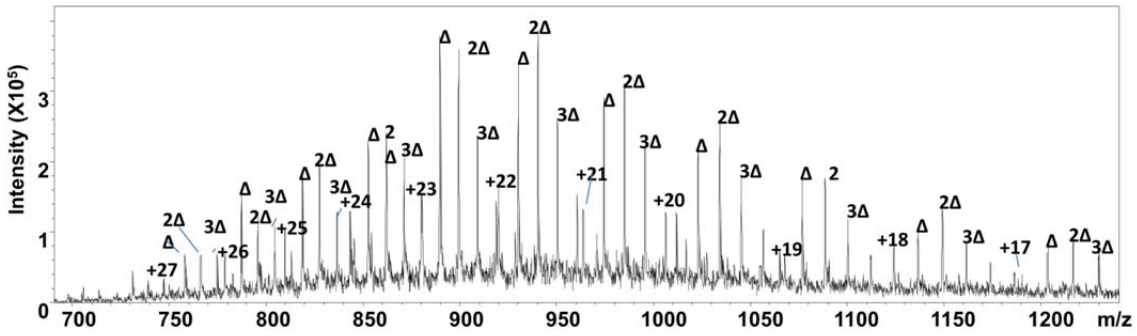
Supplemental Figure 6. Side-by-side structures showing the localized region of (A) eIF4E K107 (PDB 2IDV) and (B) eIFiso4E K99.



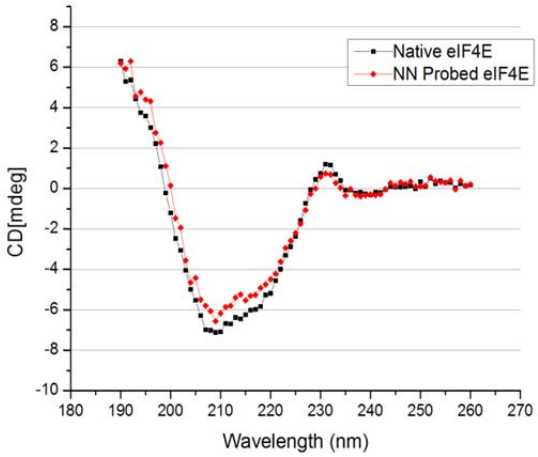
Supplemental Figure 7. Side-by-side structures showing the localized region of (A) K118 of eIF4E (PDB 2IDV) and (B) K110 of eIFiso4E. Water molecules are shown as red stars.



Supplemental Figure 8. Side-by-side structures showing the localized region of (A) K182 of eIF4E (PDB 2IDV) and (B) K175 of eIFiso4E.



Supplemental Figure 9. ESI mass spectrum of eIF4E reacted with NN using a NN:protein molar ratio of 20:1. Δ symbolizes an addition of NN.



Supplemental Figure 10. Circular dichroism spectra of truncated eIF4E and eIF4E reacted with NN using a NN:protein molar ratio of 20:1.

Supplemental Tables.

Peptide Sequence	EIC Area	Modified Residue
AHPLENAWTFWFDNPQGK	1.33E+09	
AHPLENAWTFWFDNPQG**K**	6.41E+08	K18
NKIEPK	1.40E+07	
N**K**IEPK	1.35E+08	K65
IEPK	8.31E+07	
WEDPICANGGK	1.82E+08	
WTISCGRGK	9.60E+06	
Q**K**QER	6.27E+07	K122
QER	5.67E+06	
VAIWTK	4.86E+08	
VAIWT**K**NAANEAAQISIGK	1.67E+07	K131
NAANEAAQISIGK	7.52E+08	
NAANEAAQISIG**K**QWK	1.70E+09	K144
QWK	5.46E+06	
EFLDYK	4.04E+08	
DSIGFIVHEDAK	1.90E+09	
DSIGFIVHEDA**K**R	6.64E+08	K165
SD**K**GPK	6.67E+08	K169
SDKGPKNRYTV	9.22E+06	

Supplemental Table 1. List of identified peptides and their corresponding ion abundances from a tryptic digest of eIF4E. Each NN-modified lysine is designated in bold font. The total sequence coverage was 64%.

Peptide Sequence	EIC Area	Modified Residue
AHPLENAWTFWFDNPQGK	3.29E+08	
AHPLENAWTFWFDNPQG**K**	1.48E+08	K18
AHPLENAWTFWFDNPQG**K**SR	9.98E+06	K18
LNVGADFHCFK	3.87E+06	
NKIEPK	8.90E+06	
N**K**IEPK	5.31E+07	K65
IEPK	3.37E+07	
WEDPICANGGK	1.27E+07	
WTISCGRGK	7.49E+06	
Q**K**QER	2.13E+07	K122
QER	3.05E+06	
VAIWTK	3.07E+08	
NAANEAAQISIGK	2.65E+08	
NAANEAAQISIG**K**QWK	3.70E+08	K144
QWK	2.35E+07	
EFLDYK	5.22E+08	
DSIGFIVHEDAK	1.31E+09	
DSIGFIVHEDA**K**R	5.45E+07	K165
SD**K**GPK	3.66E+07	K169
SDKGPKNRYTV	1.60E+06	

Supplemental Table 2. List of identified peptides and their corresponding ion abundances from a tryptic digest of eIF4E in a reduced environment. Each NN-modified lysine is designated in bold font. The total sequence coverage was 67%.

Peptide Sequence	EIC Area	Modified Residue
AHPLENAWTFWFDNPQGK	4.67E+09	
AHPLENAWTFWFDNPQG**K**	2.26E+09	K18
LNVGADFHCFK	2.88E+09	
NKIEPK	3.04E+07	
N**K**IEPK	2.09E+08	K65
IEPK	1.36E+08	
WEDPICANGGK	7.25E+09	
WTISCGRGK	4.72E+07	
Q**K**QER	2.71E+08	K122
QER	7.09E+07	
VAIWTK	1.17E+09	
VAIWT**K**NAANEAAQISIGK	9.67E+07	K131
NAANEAAQISIGK	1.75E+09	
NAANEAAQISIG**K**QWK	5.97E+09	K144
QWK	1.27E+08	
EFLDYK	1.92E+09	
DSIGFIVHEDAK	5.70E+09	
DSIGFIVHEDA**K**R	1.13E+08	K165
SD**K**GPK	8.78E+08	K169
SD**K**GPKNRYTV	2.19E+08	K169

Supplemental Table 3. List of peptides and their corresponding ion abundances from a tryptic digest of truncated eIF4E while complexed to eIF4G (mini-complex) for one of three replicates. Each NN-modified lysine is designated in bold font. The total sequence coverage was 66%.

Peptide Sequence	EIC Area	Modified Residue
AEVEAALPVAATETPEVAAEGDAGAAEAK	3.25E+08	
AEVEAALPVAATETPEVAAEGDAGAAEAK	3.00E+07	A1
AEVEAALPVAATETPEVAAEGDAGAAEA**K**GPHK	4.07E+06	K28
GPH**K**LQR	2.32E+06	K33
QWTFWYDIQTKPKPGAAWGTSLK	1.22E+06	
QWTFWYDIQT**K***P**K***PGAAWGTSLK	2.69E+06	K47/K49
LVGSADFHLFK	1.15E+08	
LVGSADFHLF**K**AGVEPK	2.64E+06	K93
AGVEPK	8.10E+06	
AGVEP**K**WEDPECANGGK	2.02E+07	K99
WEDPECANGGK	1.91E+08	
WEDPECANGG**K**	5.64E+06	K110
WEDPECANGGKWTVISSR	4.73E+06	
WEDPECANGG**K**WTVISSR	9.64E+07	L110
QRQD**K**LSLWTK	3.16E+06	K156
QD**K**LSLWTK	7.59E+07	K156
LSLWTK	2.24E+08	
LSLWT**K**TASNEAVQVDIGK	4.44E+06	K162
LSLWTKTASNEAVQVDIGKK	8.14E+07	
QDKLSLWTK	4.69E+07	
TASNEAVQVDIGK	2.86E+08	
TASNEAVQVDIGKK	1.46E+08	
TASNEAVQVDIG**K**K	5.70E+07	K175
KW**K**EVIDYNDK	4.06E+06	K176 and K178
TASNEAVQVDIG**KK**WK	2.87E+07	K175 and K176
WKEVIDYNDK	5.37E+07	
W**K**EVIDYNDK	1.97E+07	K178
WKEVIDYNDKMVYSFHDDSR	1.35E+06	
EVIDYNDK	2.62E+08	
EVIDYNDKMVYSFHDDSR	6.37E+07	K186
SQKPSR	1.08E+07	
SQKPSR	2.01E+07	K199

Supplemental Table 4. List of peptides and their corresponding ion abundances from a tryptic digest of eIFiso4E. Each NN-modified lysine is designated in bold font. The total sequence coverage was 97%.

Peptide Sequence	EIC Area	Modified Residue
AEVEAALPVAATETPEVAAEGDAGAAEAK	1.90E+08	
AEVEAALPVAATETPEVAAEGDAGAAEAK	3.10E+06	A1
AEVEAALPVAATETPEVAAEGDAGAAEA**K**GPHK	4.25E+05	K28
GPH**K**LQR	4.00E+05	K33
QWTFWYDIQTKPKPGAAWGTSLK	3.12E+05	
QWTFWYDIQT**K**PKPGAAWGTSLK or QWTFWYDIQTKP**K**PGAAWGTSLK	1.61E+06	K47/K49
LVGSADFHLFK	7.02E+07	
LVGSADFHLF**K**AGVEPK	7.20E+05	K93
AGVEPK	3.30E+06	
AGVEP**K**WEDPECANGGK	4.46E+06	K99
WEDPECANGGK	9.12E+07	
WEDPECANGG**K**	1.93E+06	K110
WEDPECANGGKWTVISSR	1.49E+06	
WEDPECANGG**K**WTVISSR	2.42E+06	L110
QD**K**LSLWTK	1.83E+06	K156
LSLWTK	1.36E+08	
LSLWT**K**TASNEAVQVDIGK	1.38E+06	K162
LSLWTKTASNEAVQVDIGKK	5.06E+07	
QDKLSLWTK	5.65E+06	
TASNEAVQVDIGK	1.27E+08	
TASNEAVQVDIGKK	1.07E+08	
TASNEAVQVDIG**K**K	6.93E+06	K175
KW**K**EVIDYNDK	8.96E+05	K176 and K178
TASNEAVQVDIG**K**KWK	2.71E+06	K175 and K176
WKEVIDYNDK	5.01E+07	
W**K**EVIDYNDK	6.34E+06	K178
WKEVIDYNDKMVYSFHDDSR	3.38E+06	
EVIDYNDK	1.28E+08	
EVIDYNDKMVYSFHDDSR	7.24E+06	K186
SQKPSR	4.93E+06	
SQ**K**PSR	4.91E+06	K199

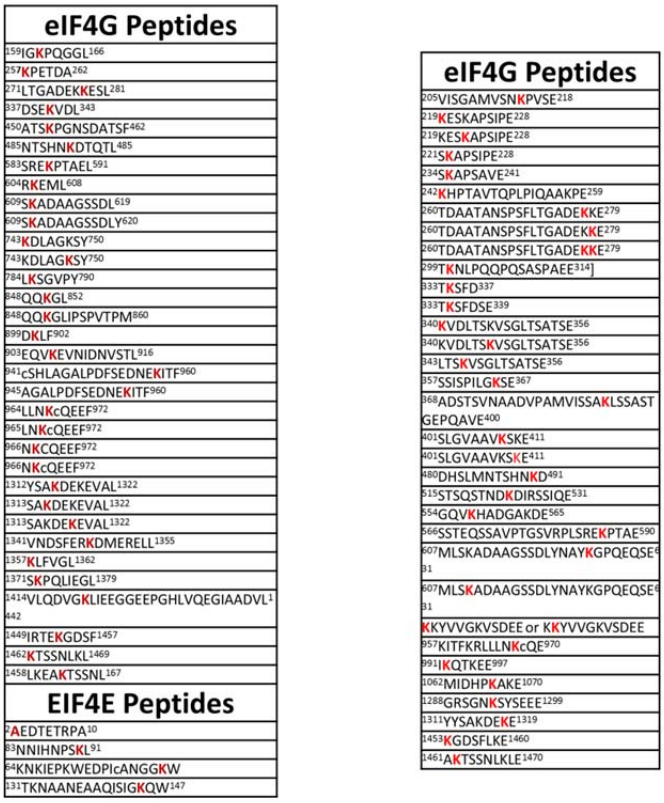
Supplemental Table 5. List of peptides and their corresponding ion abundances from a tryptic digest of eIFiso4E. Each NN-modified lysine is designated in bold font. The total sequence coverage was 97%.

Peptide Sequence	EIC Area	Modified Residue
AEDTETR	2.14E+06	
AEDTETR	2.48E+06	A1
AEDTETRPASAGAEER	4.86E+07	
AEDTETRPASAGAEER	8.33E+06	A1
ITAHPLENAWTFWFDNPQGK	1.96E+08	
ITAHPLENAWTFWFDNPQG**K**SR	1.94E+06	K56
NKIEPK	3.54E+06	
N**K**IEPK	8.40E+06	K103
WEDPICANGGK	5.74E+07	
WEDPICANGG**K**WTISCGR	7.85E+06	K118
Q**K**QER	8.67E+06	K160
NAANEAAQISIGK	4.40E+08	
NAANEAAQISIG**K**QWK	7.93E+07	K182
EFLDYK	1.60E+08	
EFLDYKDSIGFIVHEDAK	1.57E+06	
EFLDY**K**DSIGFIVHEDAK	2.92E+06	K191
DSIGFIVHEDAK	3.73E+08	
DSIGFIVHEDA**K**R	1.97E+06	K203
SD**K**GPK	6.65E+06	K207
GP**K**NR	4.18E+06	K210

Supplemental Table 6. List of peptides and their corresponding ion abundances from a tryptic digest of full eIF4E while complexed to the eIF4G. Each NN-modified lysine is designated in bold font. The total sequence coverage was 79%.

Eukaryotic translation initiation factor 4E	
2AEDTETRPASAGAEER17	170NAANEAAQISIGKQWK185
37ITAHPLENAWTFWFDNPQGKSR58	186EFLDYKDSIGFIVHEDAK203
91LNVGADFHcFK101	192DSIGFIVHEDAKR204
102NKIEPK107	205SDKGPK210
108WEDPIcANGGKWTIScGR125	205SDKGPKN212
159QKQER163	208GPKNR212
Eukaryotic translation initiation factor 4G	
200QKDNEVISGAMVSNKPVSEK219	869KYVVGKVSDEEQAK882
202DNEVISGAMVSNKPVSEK219	870YVVGKVSDEEQAK882
202DNEVISGAM$_{ox}$VSNKPVSEK219	875VSDEEQAKQR884
223APSIPEKHSK232	885QLKAILNK892
233ESKAPSAVEK242	888AILNK892
278KESLPMTDSLK288	888AILNKLTPQNFDK900
278KESLPMTDSLKDNK291	963LLLNKcQEEFER974
279ESLPMTDSLKDNK291	1015LIGELYKK1022
292KNATR296	1029IMHECIK1035
297NDTKNLPQQPQSASPAEELK316	1029IMHEcIKK1036
317GQTSVKLGDDVVGHMETK334	1036KLLGNYQNPDEENIEALcK1054
323LGDDVVGHMETKSFDSEK340	1037LLGNYQNPDEENIEALcK1054
335SFDSEKVDLTSK346	1055LMSTIGEMIDHPKAK1069
341VDLTSKVSGLTSATSESSISPILGK365	1068AKEHMDAYFDR1078
366SEADSTSVNAADVPAMVISSAKLSSASTGEPQAVESLGVAAVK408	1105NKWQQR1110
388LSSASTGEPQAVESLGVAAVKSK410	1112KVDGPK1117
585EKPTAELAR593	1118VDGPKK1118
594TKSTAGR600	1118KIDEVHR1124
604RKEMLSK610	1200SVKDEAITLGPQGGLAR1216
606EMLSKADAAGSSDLYNAYK624	1290SGNKSYSEEELR1301
658EVMcEDDGKK667	1302EKSIATIR1309
668KVEPDDWEDAADMSTPK684	1348KDMER1352
734MDTVTSTLFKDLAGK748	1407FVVEKILVLQDVGK1420
744DLAGKSYVIDR754	1451TEKGDSFLK1459
774RGPAMDDDKWLK785	1454GDSFLKEAK1462
775GPAMDDDKWLK785	1454GDSFLKEAKTSSNLK1468
775GPAMDDDKWLKSGVPYSPNR794	1460EAKTSSNLK1468
874KYVVGK869	1469LEDFRPQHLKR1479
869KYVVGKVSDEEQAK882	1480SKLDAFMLT1488

Supplemental Table 7. NN-modified tryptic peptides identified from the full eIF4F complex using LC-UVPD-MS



eIF4G Peptides
159IGKPQGGL166
257KPETDA262
271LTGADEKKESL281
337DSEKVDL343
450ATSKPGNSDATSF462
485NTSHNKDTQTL485
583SREKPTAEL591
604RKEML608
609SKADAAGSSDL619
609SKADAAGSSDLY620
743KDLAGKSY750
743KDLAGKSY750
784LKSGVPY790
848QQKGL852
848QQKGLIPSPVTPM860
899DKLF902
903EQVKEVNIDNVSTL916
941cSHLAGALPDFSEDNEKITF960
945AGALPDFSEDNEKITF960
964LLNKcQEEF972
965LNKcQEEF972
966NKCQEEF972
966NKcQEEF972
1312YSAKDEKEVAL1322
1313SAKDEKEVAL1322
1313SAKDEKEVAL1322
1341VNDSFERKDMERELL1355
1357KLFVGL1362
1371SKPQLIEGL1379
1414VLQDVGKLIEEGGEEPGHLVQEGIAADVL1442
1449IRTEKGDSF1457
1462KTSSNLKL1469
1458LKEAKTSSNL167

EIF4E Peptides
2AEDTETRPA10
83NNIHNPSKL91
64KNKIEPKWEDPIcANGGKW
131TKNAANEAAQISIGKQW147

eIF4G Peptides
205VISGAMVSNKPVSE218
219KESKAPSIPE228
219KESKAPSIPE228
221SKAPSIPE228
234SKAPSAVE241
242KHPTAVTQPLPIQAAKPE259
260TDAATANSPSFLTGADEKKE279
260TDAATANSPSFLTGADEKKE279
260TDAATANSPSFLTGADEKKE279
299TKNLPQQPQSASPAEE314]
333TKSFD337
333TKSFDSE339
340KVDLTSKVSGLTSATSE356
340KVDLTSKVSGLTSATSE356
343LTSKVSGLTSATSE356
357SSISPILGKSE367
368ADSTSVNAADVPAMVISSAKLSSASTGEPQAVE400
401SLGVAAVKSKE411
401SLGVAAVKSKE411
480DHSLMNTSHNKD491
515STSQSTNDKDIRSSIQE531
554GQVKHADGAKDE565
566SSTEQSSAVPTGSVRPLSREKPTAE590
607MLSKADAAGSSDLYNAYKGPQEQSE631
607MLSKADAAGSSDLYNAYKGPQEQSE631
KKYVVGKVSDEE or KKYVVGKVSDEE
957KITFKRLLLNKcQE970
991IKQTKEE997
1062MIDHPKAKE1070
1288GRSGNKSYSEEE1299
1311YYSAKDEKE1319
1453KGDSFLKE1460
1461AKTSSNLKLE1470

Supplemental Table 8. Chymotrypsin and GluC peptides identified using LC-UVPD-MS for NN-modified eIF4F (a complex comprised of eIF4E and eIF4G).

Residue	SEA (Å^2)	SA Fraction: SEA (Å^2) / 211 Å^2	Solvent Accessibilities (%)
Lys 33	141.290	0.67	67
Lys 47	45.983	0.22	22
Lys49	185.412	0.87	87
Lys 59	59.830	0.28	28
Lys 60	95.455	0.45	45
Lys 82	151.305	0.71	71
Lys 93	18.022	0.09	9
Lys 99	159.958	0.75	75
Lys 110	68.192	0.32	32
Lys 118	158.868	0.75	75
Lys 156	76.392	0.36	36
Lys 162	58.182	0.27	27
Lys 175	124.184	0.59	59
Lys 176	52.743	0.25	25
Lys 178	34.112	0.16	16
Lys 186	164.165	0.77	77
Lys 199	167.214	0.79	79

Supplemental Table 9. Results of PyMOL calculations for solvent exposed areas (SEA) and solvent accessibilities (SA) of lysine residues for eIFiso4E. The SEA was calculated in PyMOL using the get_area command for each lysine residue. The SA fraction for each residue was obtained by dividing the SEA by the surface area of an exposed lysine residue (211 Å^2 for a lysine that is considered 100% solvent exposed as in the tripeptide G-K-G). Finally the SA was calculated by multiplying the SA fraction by 100%.