Title	Water-Bridge Mediates Recognition of mRNA Cap in eIF4E
Author(s)	Lama, Dilraj; Pradhan, Mohan R.; Brown, Christopher J.; Eapen, Rohan S.; Joseph, Thomas L.; Kwoh, Chee-Keong; Lane, David P.; Verma, Chandra Shekhar
Citation	Lama, D., Pradhan, M. R., Brown, C. J., Eapen, R. S., Joseph, T. L., Kwoh, C. -K., et al. (2017). Water-Bridge Mediates Recognition of mRNA Cap in eIF4E. Structure, 25(1), 188-194.
Date	2017
URL	http://hdl.handle.net/10220/42071
Rights	© 2016 Elsevier Ltd. This is the author created version of a work that has been peer reviewed and accepted for publication by Structure, Elsevier. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [http://dx.doi.org/10.1016/j.str.2016.11.006].

Water-Bridge Mediates Recognition of mRNA Cap in eIF4E

Dilraj Lama[1#*], Mohan R Pradhan[1,2#], Christopher J Brown[3], Rohan S Eapen[4], Thomas L Joseph[1], Chee-Keong Kwoh[2], David P Lane[3] and Chandra S Verma[1, 5,6*]

[1]Bioinformatics Institute, A*STAR (Agency for Science, Technology and Research), 30 Biopolis Street, #07-01 Matrix, Singapore 138671

[2]School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

[3]p53 Laboratory, A*STAR (Agency for Science, Technology and Research), 8A Biomedical Grove, #06-04/05, Neuros/Immunos, Singapore 138648

[4]School of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, United Kingdom

[5]Department of Biological sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

[6]School of Biological sciences, Nanyang Technological University, 50 Nanyang Drive, Singapore 637551

[#]Equal authorship

Address correspondence to:

Dilraj Lama

Bioinformatics Institute, A*STAR (Agency for Science, Technology and Research),

30 Biopolis Street, #07-01 Matrix, Singapore 138671

Tel: (65) 6478 8499; Fax: (65) 6478 9047;

E-mail: dilrajl@bii.a-star.edu.sg

Chandra S Verma

Bioinformatics Institute, A*STAR (Agency for Science, Technology and Research),

30 Biopolis Street, #07-01 Matrix, Singapore 138671

Tel: (65) 6478 8273; Fax: (65) 6478 9048;

E-mail: chandra@bii.a-star.edu.sg

Summary

Ligand binding pockets in proteins contain water molecules, which play important roles in modulating protein-ligand interactions. Available crystallographic data for the 5' mRNA cap-binding pocket of the translation initiation factor protein eIF4E shows several structurally conserved waters, which also persist in molecular dynamics simulations. These waters engage an intricate hydrogen-bond network between the cap and protein. Two crystallographic waters in the cleft of the pocket show a high degree of conservation and bridge two residues, which are part of an evolutionarily conserved scaffold. This appears to be a preformed recognition module for the cap with the two structural waters facilitating an efficient interaction. This is also recapitulated in a new crystal structure of the apo protein. These findings open new windows for the design and screening of compounds targeting eIF4E.

Introduction

The eukaryotic translation initiation factor protein eIF4E is a cap-binding protein that recognizes mRNA molecules through the cap structure (7-methylguanisine triphosphate) present at the 5' end of the mRNA (Brown et al., 2007). This complex forms higher order states through interactions with other protein partners, eventually culminating in the ribosome complex, resulting in translation (Jia et al., 2012). It is therefore an essential component of the cap-dependent protein translation pathway (Merrick, 2004). Three-dimensional structures of eIF4E in complex with different cap mimics have provided a detailed understanding of their interactions (Figure S1) (Brown et al., 2009). The residues that mediate the interactions with cap include hydrophobic (W56, W102 and W166) and hydrophilic (E103, R112, R157 and K162) amino acids. The tryptophans and E103 are involved in recognition of the N7 substituted guanosine base whereas the basic residues interact with the phosphate group (Figure S1).

There are reports in the literature about explicit water molecules in the cap-binding interface mediating the interactions between eIF4E and the mRNA cap ligands (Brown et al., 2007; Marcotrigiano et al., 1997; Niedzwiecka et al., 2002; Tomoo et al., 2003); however a detailed and comprehensive study on their precise role is lacking. This knowledge becomes more critical because eIF4E is an oncogene which is known to be overexpressed in various malignancies (Carroll and Borden, 2013) and hence its inhibition using oligonucleotides (Thumma et al., 2015), small organic molecules (Chen et al., 2012; Ghosh et al., 2005) and peptides (Brown et al., 2011; Lama et al., 2013) is being pursued as an anticancer therapeutic opportunity (Borden, 2011; Hsieh and Ruggero, 2010). To date, the development of promising lead compounds targeting the cap-binding interface have not included the interfacial water molecules.

In this study, we combined structural analysis with molecular dynamics (MD) simulations to find a set of conserved water positions that bridge interactions between the protein and cap analogues. We have established the presence of two spatially conserved water positions which together with some residues, form a structural module in facilitating efficient recognition between eIF4E and the cap analogues. This is also seen in a new crystal structure of apo eIF4E. These observations strongly suggest that this water mediated recognition module is involved in the fundamental mechanism of eIF4E-mRNA cap recognition and could be exploited in the design of lead compounds against this oncogenic protein.

Results

Characterization of structural water molecules at the cap-binding interface

Analysis of waters around the cap-binding interface in the cap-bound state of the protein led to the identification of 14 significant conserved waters which occupy distinct spatial positions around the cap. These positions are termed W1 to W14 (Tables S1 and S2, Figure 1A). Of the 14 conserved waters identified, W1 and W2 show maximum conservation, being conserved in more than 90% of the structures. We also looked for conserved waters in the cap-free states of the protein and found 6 distinct groups (W1 to W4, W13 and W14) (Tables S1 and S3, Figure 1B) which are a subset of the groups observed in the cap-bound states. The elevated numbers of conserved waters in the cap-bound states indicate that water retention around the binding interface is higher in the presence of the cap; this may reflect a cap dependent ordering within the cap-binding interface. Interestingly, we observe that W1 and W2 have high conservation of more than 90% in the cap-free states of the protein also. The average crystallographic B-factors of the different conserved waters were computed from the crystal structures (Figure 1C) for both the cap-bound and cap-free states. The corresponding waters positions in the cap-bound states are more ordered as compared to the

cap-free state. Further, W1 and W2 waters have the lowest B-factor values when compared to other waters within the two different states which suggest that the waters at these positions are structured and relatively less mobile.

Solvation features around the cap-binding interface were further investigated by analysis of ensembles of structures generated from the MD simulations in terms of the water occupancy profiles (Figures 1A, 1B and S2). The occupancy map was investigated at three different threshold cutoff values (Figure S2). Most of the conserved waters that were observed in the crystal structures are also seen to be occupied in the simulations at lower threshold. Although as we increase the threshold, only the waters at W1 and W2 seem to be distinctly occupied, which indicates their relatively higher degree of ordered character. This corroborates our understanding from the crystal structure analysis that these two water positions are less dynamic and undergo relatively less exchange with the bulk.

Water mediated interaction between the protein and cap

This conservation of waters at specific locations indicates their possible role in influencing the binding of the mRNA cap. We therefore analyzed the interactions formed by these waters (Figure 1D). Potential hydrogen-bond networks bridge the nitrogen in the indole group of W166, waters at position W1 and W2 and the side-chain carbonyl oxygen of N155. The water at W1 also interacts with the nitrogen in the guanidinium group of R112 and with the water at W3. Such a favorable interaction environment could be a primary reason for the stability and high conservation of the water at W1. Basic residues R112 and K162 are involved in water-mediated interactions with the negatively charged phosphate moiety of the cap via water molecules at positions W4 to W7. This is the electrostatic component of molecular recognition that had earlier been identified as important for steering the ligands towards the cap-binding interface (Blachut-Okrasinska and Antosiewicz, 2007; Niedzwiecka

et al., 2002). The strategic ordering of water positions W4 to W7 likely ensures an efficient hydrophilic interaction after ligand-complexation associated desolvation of the region. Waters at positions W8 and W9 form a network between the indole of W102, which is one of the two tryptophan residues that sandwiches the guanosine ring, and the side-chains of H200 and T203. Structurally conserved water at position W3 bridges waters W10, W11 and W12, which together with waters W13 W14, and the hydroxyl of S92, form a closed caged network. The water at position W12 links the cap and protein via the side-chain of D90, phosphate of the cap and the side-chain indole of W56 which is the other tryptophan residue that interacts with the guanosine.

Thus, we observe that the spatial arrangement of the structural waters from W1 to W14 form an intricate interaction network that facilitates binding between the protein and the cap. A similar analysis of the cap-free state and its comparison with the cap-bound form shows that the network involving the subset of conserved waters at position W1 to W4, W13 and W14 is identical (Figure 1E). The presence of this sub-network suggests that it could act as a scaffold facilitating the formation of a higher order network as observed in the cap-bound state. Water at W1 is involved in the formation of similar interactions as seen in the bound states involving waters at W2 and W3 which could be important as it is observed to link most of the other waters in the complex. Comparatively, it appears that the set of connections involving residue W166, waters at positions W1, W2 and residue N155 lie at the core of this network and the waters at W1 and W2 act as linkers in bridging the residues.

W1 and W2 waters are held by interactions with residues W166 and N155

The presence of a subset of spatially similar structural waters and their interaction network in the cap-free state of the protein suggests that they are an integral part of the binding interface and hence likely influence the recognition of the ligand in the cap-binding

interface. This seems to be especially true for waters at W1 and W2. We analyzed the hydration properties of waters at these positions in terms of their solvation profiles and total residence times (Table S4). Both the water positions are very well hydrated throughout the simulation which is in accord with their high occupancies in the crystal structures (Tables S2 and S3). However, the total residence times of both the waters varies between the cap-bound (86 ns and 95 ns) and cap-free states (65 ns and 37 ns). The relatively lower value in the cap-free state suggests that they are more dynamic and undergo exchange with the bulk solvent in the absence of the cap. The average crystallographic B-factors of both these waters are also lower in the cap-bound state as compared to the cap-free state (Figure 1C). This analysis indicates that waters at positions W1 and W2 are retained in their spatial location for significant periods of time and which keeps the region solvated between the two different states of the protein (Figure 2A). We also examined the influence of residues N155 and W166 on the stability of these waters, by generating *in-silico* mutants, and evaluating their hydration profiles from MD simulations (Table S4). N155 was mutated to leucine and W166 to phenylalanine, both mutations designed to abrogate the side-chain hydrogen-bonds with the respective waters, without significantly affecting the van der Walls contacts with the surroundings. For the double point mutant N155L/W166F, the water properties at W1 and W2 were adversely affected, with the region getting dehydrated (Figure 2B). This *in silico* mutational experiment indicates that tryptophan and asparagine residues are both important for retaining the two respective water molecules at specific locations and thereby form the water-bridge connecting their side-chains. This bridge could be the basic structural motif that would act as a scaffold for the formation of the water mediated interaction between the mRNA cap and eIF4E.

Preformed docking site at the cap-binding interface

We next examined if the spatial location of N155 and W166 themselves will have an influence on the retention of the water molecules at W1 and W2 positions. For this, we first looked at the dynamic property of residues in the vicinity of the cap-binding interface by comparing their fluctuations in the MD simulated structures both in the cap-bound and cap-free states (Figure 2C). Interestingly, a clear distinction was observed for a subset of residues in which D90, F94, R112, N155, R157, K162, W166 and H200 show similar degree of flexibility between the two states whereas W56, W102, and E103 have reasonably high fluctuations. Crystal structure comparisons between the two structural states also highlight this difference (Figure 2D). Based on the observations, these residues in the binding interface can be categorized into two sub-groups comprising of those that are dynamic and others which are relatively rigid, irrespective of whether the ligand is bound to the protein or not. We note that the amino acids which have a preformed spatial arrangement are those which are part of the water network with W1 and W2. N155 makes favorable electrostatic interactions with D90 and R157 whereas W166 is involved in hydrophobic stacking contacts with F94 and H200. This group of residues form two separate hubs comprising of "D90, R157, N155" and "W166, H200, F94" which appear to act as 'pillars' in supporting the water bridge network (Figure 2D). Further, as stated previously, the critically important residues in the cap binding interface of eIF4E are W56, W102, E103, R112, K162 and R157 (Figure S1). This set of residues, along with those forming the solvent bridge and the pillars are also evolutionarily conserved or at-least their chemical potential is preserved (Figures 2E and S3). Hence this could likely compose the basic structural scaffold involved during the docking of the mRNA cap and hence form the nucleus of the recognition process.

The influence of the residues (D90, F94, R157 and H200) that support W166 and N155 in forming the stable solvent bridge was also evaluated by *in silico* Ala mutations and MD simulations. We observe that the side-chain orientation of N155 is dramatically affected

by D90A/R157A double point mutations (Figure 3A). This change in the spatial orientation of N155 results in the loss of water at W2 without affecting the presence of water at W1 (Figure 3B). Thus, the hydrophilic interaction network formed between residues D90-N155-R157 appears to be critical in order to maintain the distinct positioning of N155 and its interactions with water at W2. Conversely, the side-chain orientation of W166 does not change appreciably with F94A, H200A double mutations and consequently we detect waters at both W1 and W2 in these mutants (Figures 3C and 3D). This suggests that the hydrophobic packing formed by residues F94-W166-H200 is not primarily important to preserve the structural positioning of W166. However there are variations in the fluctuations of W166 between the wild and mutant structures, which indicate that F94 and H200 do effect the optimal arrangement of the residue and hence appear to be part of a wider scaffold for supporting W166. Thus the supporting residues (D90, F94, R157 and H200) are important in order to anchor N155 and W166 in their positions albeit to different degrees; consequently, they result in the retention of the water bridge and thereby maintenance of the local structural framework. Further, the evolutionary conservation of this water mediated building block across different organisms indicates its significance in the structure-activity property of eIF4E.

Crystal structure supports the presence of the water-bridge recognition module

Apo eIF4E was successfully crystallized in the space group C2221 with one copy in the asymmetric unit in the absence of cap-analogue at an atomic resolution of 2.0 Å (Table 1). Besides the structure reported here, there are only two additional apo structures (structures without peptide or m^7GTP cap bound) of eIF4E (PDB ID's: 3TF2 and 2IDR, Table S1) known, which is striking considering the relatively higher number of binary and ternary complex state structures of the protein (Table S1). The overall fold is consistent with other existing structures of eIF4E comprising of eight-stranded anti-parallel β sheet and four α

helices (Figure 4A). Comparative analysis among the three apo structures shows that the regions connecting β strands 1-2, 3-4 and 7-8 show considerable variations (Figure 4B). These regions correspond to the cap-binding interface of the protein and include W56, W102 and E103 residues which are critical for cap binding. A considerable amount of resolved water molecules that are well defined in the 2Fo-Fc electron density map is seen in the structure and more significantly, the cap-binding interface has a total of 21 distinct crystal waters. We analyzed the hydration profile at the interface region and it was exciting to observe clear electron density for both the waters at W1 and W2 (Figure 4C). They are involved in the formation of the hydrogen-bond network with W166 and N155 and hence structure the water-bridge.

The 2Fo-Fc electron density map also revealed that residues W56 and W102, which sandwich the guanosine moiety of the m^7GTP cap analogue in the cap-bound state, are significantly displaced. Hence they do not show a well-defined density profile in the map. W102 rotates out of the cap-binding interface and the loop bearing W56 swings out of the binding pocket. However, residues D90, R157 and N155 which form the hydrophilic triad and residues H200, F94 and W166 which pack against each other along with R112 are observed to have a distinct density map which suggests a more structured nature of these set of residues. Thus, the crystallization data reconfirms the presence of a water-mediated recognition module at the cap-binding region of eIF4E formed by protein residues (D90, R157, N155, H200, F94 and W166) and two specific water molecules and this further substantiates our findings from comparative structural and MD analysis.

Discussion

The role of water in mediating the interactions between the 5' cap of mRNA molecules and eIF4E has long been discussed and several crystal structures show a clear

presence of the water molecules at the interface (Brown et al., 2007; Marcotrigiano et al., 1997; Niedzwiecka et al., 2002; Tomoo et al., 2003). We observed that some of these waters are spatially conserved across the different structures and are engaged in the formation of an elaborate hydrogen-bond network between the cap and the protein. Interestingly, a subset of these waters is structurally preserved at the interface even in the absence of the ligand. This is especially true for two bound waters which are present in a crevice located in the central portion of the cap-binding interface (Figure 4D). These waters are less mobile and more ordered as compared to other structurally conserved waters in the interface. They form a water-mediated interaction connecting the side-chains of residues N155 and W166. Indeed, in the ensemble from the solution state structure of cap-free eIF4E (Volpon et al., 2006), we observe that these two residues adopt multiple conformations which are similar as seen in the ensemble from the cap-free crystal and MD simulated structures (Figure S4A). This allows us to speculate that the water-mediated interactions between N155, W166 and the two waters should be largely preserved when the protein is in the more natural aqueous medium. Both side-chains seem crucial in maintaining local hydration which is lost if these residues are mutated. Besides, the spatial arrangement of these two residues also appears to be an important factor that would conserve the specific waters in the cleft. This conformational restraint on N155 and W166 is imposed by sub-networks formed around each of the two residues involving hydrophobic (F94-W166-H200) and hydrophilic interactions (R157-N155-D90). It is also interesting to see that this group of residues exhibits a low degree of fluctuation and is relatively restrained in the cap-free state of eIF4E as compared to other residues, which are involved in the recognition of the cap. These residues along with the two water molecules form a local network in the cap-binding interface of eIF4E that provides a structural scaffold for recognition of the cap (Figure 4D). Further, these sets of residues are conserved across different species suggesting that the water-bridge recognition module has been evolutionarily preserved.

Electrostatic interactions are one of the major driving forces for the attraction of the cap analogue towards the eIF4E binding interface (Blachut-Okrasinska and Antosiewicz, 2007; Niedzwiecka et al., 2002). Biophysical and kinetic studies on the binding of the mRNA cap with eIF4E suggest a probable mechanism of binding (Blachut-Okrasinska et al., 2007; Niedzwiecka et al., 2002). The phosphate groups in the cap could act as an initial anchor followed by the formation of the fundamentally important cation-π stacking and hydrogen-bond interactions between W56, W102, E103 residues of eIF4E and the m^7G guanosine moiety of the cap (Niedzwiecka et al., 2002). It is highly probable that the electrostatically steered mRNA cap encounters the water-bridge network identified in this work as an initial docking structural motif for the ligand. It is reported previously (Niedzwiecka et al., 2002) and also seen from our comparative structure and simulation data, that the cap-binding event results in considerable uptake of water molecules in the binding interface. We suggest that the two structurally conserved waters in the bridge recruit other water molecules thereby, mobilizing the formation of an efficient interaction between the cap and the protein. They act as a strong framework on which to form the higher order interaction network. As such, we can postulate that these waters are part of the other conserved residues in the interface that will be critical in the selectivity of the mRNA cap analogues. We also observe the presence of these specific waters and the network in a new crystal structure of the apo state of the protein, which lends further credibility to our proposition and the possible role of these waters in the process of recognition.

The cap binding interface on eIF4E has been a target for the design and development of efficient inhibitors against the protein (Chen et al., 2012; Ghosh et al., 2005). A recent study reported on a structure-guided approach to design a modified cap analogue that binds eIF4E with more potency (Chen et al., 2012). However the nature of the hydration profile had not been considered explicitly in the design and development of these inhibitors, despite the

general understanding that they have a significant role in the recognition process. Thus, our current findings on the spatial conservation of specific water molecules in the cap-binding interface and their involvement in the formation of a local structural scaffold could serve as an important starting point in the rational design of more efficient inhibitors against eIF4E. Additionally, the cap-analogue scaffold could be modified accordingly to see if displacement of certain specific water positions in the binding interface could be exploited for a favorable binding with the protein. Overall, it is interesting to observe that the cap binding interface of eIF4E which is primarily formed by flexible loop regions of the protein preserves a preformed scaffold involving structurally conserved waters that could be essential for the initial recognition of the ligand.

Experimental Procedures

Analysis of available crystallographic data on eIF4E

Crystal structures of eIF4E were retrieved from the Protein Data Bank (PDB), classified into cap-bound and cap-free states and analyzed for hydration in the cap-binding interface as detailed in the supplemental experimental procedures.

Structural alignment and detection of conserved waters

The structurally conserved waters in the cap-binding interface of the cap-bound and cap-free crystal structures of eIF4E were identified as detailed in the supplemental experimental procedures.

System preparation for simulation

Deriving the force field parameters for m^7GTP, selection of representative structure for the cap-bound and cap-free structural states of eIF4E and preparation of the system for MD simulation was done as described in the supplemental experimental procedures.

Molecular dynamics simulations

MD simulations were performed according to the protocol and parameters as detailed in the supplemental experimental procedures.

Water occupancy map and total residence time

The occupancy map of water in the MD simulated trajectories and total residence time at a particular site was computed as described in the supplemental experimental procedures.

Crystallization of apo eIF4E structure

Details on the design, expression, purification and crystallization of the surface entropically reduced construct of the truncated apo human eIF4E (sER-eIF4E23-217) are described in the supplemental experimental procedures.

Protein Data Bank accession number

Coordinates and structure factors of the crystal structure have been deposited in the Protein Data Bank with accession number 5GW6.

Author Contributions

DL, MRP and CSV conceived and designed the project. DL and MRP performed the computational analysis. CJB and RSE did the X-ray and other associated experimental work. DL, CSV, CJB, TLJ, CK and DPL wrote the manuscript. All Authors read and approved the manuscript.

Acknowledgments

Computational resource for the project was supported by A*PRE*COT (PRE-COT 2010-14) grant, Singapore. Mohan R. Pradhan was funded by the Research Scholarship Award by Bioinformatics Institute-NTU/SCE Joint PhD Program. Dilraj Lama was partly funded by Ipsen Pharmaceuticals.We would also like to thank A*STAR's Biomedical Sciences Institute (BMSI) for support and all the members of the Chandra group for helpful discussions and suggestions.

References

Blachut-Okrasinska, E., and Antosiewicz, J.M. (2007). Brownian dynamics simulations of binding mRNA cap analogues to eIF4E protein. J Phys Chem B *111*, 13107-13115.

Blachut-Okrasinska, E., Bojarska, E., Stepinski, J., and Antosiewicz, J.M. (2007). Kinetics of binding the mRNA cap analogues to the translation initiation factor eIF4E under second-order reaction conditions. Biophys Chem *129*, 289-297.

Borden, K.L. (2011). Targeting the oncogene eIF4E in cancer: From the bench to clinical trials. Clin Invest Med *34*, E315.

Brown, C.J., Lim, J.J., Leonard, T., Lim, H.C., Chia, C.S., Verma, C.S., and Lane, D.P. (2011). Stabilizing the eIF4G1 alpha-helix increases its binding affinity with eIF4E: implications for peptidomimetic design strategies. J Mol Biol *405*, 736-753.

Brown, C.J., McNae, I., Fischer, P.M., and Walkinshaw, M.D. (2007). Crystallographic and mass spectrometric characterisation of eIF4E with N7-alkylated cap derivatives. J Mol Biol *372*, 7-15.

Brown, C.J., Verma, C.S., Walkinshaw, M.D., and Lane, D.P. (2009). Crystallization of eIF4E complexed with eIF4GI peptide and glycerol reveals distinct structural differences around the cap-binding site. Cell Cycle *8*, 1905-1911.

Carroll, M., and Borden, K.L. (2013). The oncogene eIF4E: using biochemical insights to target cancer. J Interferon Cytokine Res *33*, 227-238.

Chen, X., Kopecky, D.J., Mihalic, J., Jeffries, S., Min, X., Heath, J., Deignan, J., Lai, S., Fu, Z., Guimaraes, C., *et al.* (2012). Structure-guided design, synthesis, and evaluation of guanine-derived inhibitors of the eIF4E mRNA-cap interaction. J Med Chem *55*, 3837-3851.

Ghosh, P., Park, C., Peterson, M.S., Bitterman, P.B., Polunovsky, V.A., and Wagner, C.R. (2005). Synthesis and evaluation of potential inhibitors of eIF4E cap binding to 7-methyl GTP. Bioorg Med Chem Lett *15*, 2177-2180.

Hsieh, A.C., and Ruggero, D. (2010). Targeting eukaryotic translation initiation factor 4E (eIF4E) in cancer. Clin Cancer Res *16*, 4914-4920.

Jia, Y., Polunovsky, V., Bitterman, P.B., and Wagner, C.R. (2012). Cap-dependent translation initiation factor eIF4E: an emerging anticancer drug target. Med Res Rev *32*, 786-814.

Konagurthu, A.S., Whisstock, J.C., Stuckey, P.J., and Lesk, A.M. (2006). MUSTANG: a multiple structural alignment algorithm. Proteins *64*, 559-574.

Lama, D., Quah, S.T., Verma, C.S., Lakshminarayanan, R., Beuerman, R.W., Lane, D.P., and Brown, C.J. (2013). Rational optimization of conformational effects induced by hydrocarbon staples in peptides and their binding interfaces. Sci Rep *3*, 3451.

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

Merrick, W.C. (2004). Cap-dependent and cap-independent translation in eukaryotic systems. Gene *332*, 1-11.

Niedzwiecka, A., Marcotrigiano, J., Stepinski, J., Jankowska-Anyszka, M., Wyslouch-Cieszynska, A., Dadlez, M., Gingras, A.C., Mak, P., Darzynkiewicz, E., Sonenberg, N., *et al.* (2002). 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 *319*, 615-635.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Thumma, S.C., Jacobson, B.A., Patel, M.R., Konicek, B.W., Franklin, M.J., Jay-Dixon, J., Sadiq, A., De, A., Graff, J.R., and Kratzke, R.A. (2015). Antisense oligonucleotide targeting

eukaryotic translation initiation factor 4E reduces growth and enhances chemosensitivity of non-small-cell lung cancer cells. Cancer Gene Ther *22*, 396-401.
Tomoo, K., Shen, X., Okabe, K., Nozoe, Y., Fukuhara, S., Morino, S., Sasaki, M., Taniguchi, T., Miyagawa, H., Kitamura, K., *et al.* (2003). Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations. J Mol Biol *328*, 365-383.
Volpon, L., Osborne, M.J., Topisirovic, I., Siddiqui, N., Borden, K.L. (2006). Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands. EMBO J. 25, 5138-5149.

Figure and Table Legends:

Figure 1. Hydration of the cap-binding interface. (A, B) Conserved waters from comparative structure analysis and water occupancy maps from MD simulations in the cap-binding interface for the cap-bound and cap-free states of the protein. The water occupancy map is shown at an isocutoff value of 3.5 times the peak value (the mode of the isovalue distribution). **(C)** Average B-factors of the conserved waters present in the cap-bound and cap-free states. (**D, E**) Potential hydrogen-bond network in the cap-binding interface for cap-bound and cap-free states. UCSF chimera visualization software package (Pettersen et al., 2004) was used. The hydrogen-bond was calculated between the heavy atoms of the protein, m^7GTP cap and the water oxygens. Residues involved in the interaction network between the water molecules and ligand are shown as sticks. See also Figure S2.

Figure 2. Water-bridge structural motif at the cap-binding interface. (A, B) Occupancy map highlighting the waters at positions W1 and W2 from MD simulations of the protein in the cap-bound and cap-free states for wild and N155L/W166F mutant. The occupancy map is shown at an isocutoff value of 3.5 times the peak value (the mode of the isovalue distribution). The representative structures are averages from the respective MD simulations. For clarity, only m^7GMP is shown to represent the cap-bound state. **(C)** Root Mean Square Fluctuation **(**RMSF) of all heavy atoms from MD simulations for the selected set of residues in the cap-binding interface compared between the two structural states. **(D)** End state structure comparison of the selected residues between cap-bound (PDB ID: 2W97, Chain A) and cap-free states (PDB ID: 2W97, Chain B). The potential hydrogen-bond interaction network involving residues N155, W166 and the waters at position W1 and W2 is indicated. The interacting hubs formed by H200, W166, F94 and D90, R157 and N155 are also shown. **(E)** The non-redundant set of structures of eIF4E proteins from different organisms are aligned using MUSTANG (Konagurthu et al., 2006). Residue number shown at the top of the

alignment refers to human eIF4E (Uniprot ID: P06730). The structurally equivalent residues which are involved in the recognition of the cap analogue are highlighted in red and those identified in this study as part of the water bridge structural motif are shown in green. See Figure S3 for complete alignment.

Figure 3. Stability of the water-bridge. (**A, C**) Population distribution for residues N155 and W166 was analyzed in terms of the $\chi 2$ side-chain torsion angle (measured along CA, CB, CG and CD) in MD simulations of both wild and mutant eIF4E. (**B, D**) Water occupancy map highlighting the presence/absence of waters at position W1 and W2 of D90A/R157A and F94A/H200A mutant proteins. The map is displayed at an isocutoff value of 3.5 times the peak value (the mode of the isovalue distribution). The representative structures are averages from the respective MD simulations.

Figure 4. Crystal structure of apo eIF4E and water-bridge recognition module at the cap-binding interface (A) Cartoon view of the structure with secondary structural elements labeled. **(B)** Structural superimposition of the three known Apo structures of eIF4E (PDB IDs: 5GW6, 3TF2 and 2IDR). The loop regions showing maximum variation among the three structures are shown in orange and labeled. The three critical cap binding residues which are part of these regions are also highlighted. **(C)** Electron density of selected residues and water molecules present in the cap-binding interface. **(D)** Surface representation of the protein with the two structurally conserved waters at positions W1 and W2 shown in magenta. The interaction involving residue W166, waters W1, W2 and residue N155 form the water bridge, and together with the hydrophilic (D90, N155 and R157) and hydrophobic (H200, W166 and F94) residues constitute the preformed recognition module is represented. The other critical residues which are recognized to be involved in the binding of the cap are indicated. The cap-binding interface is shown with an arrow.

Table 1. Crystallographic data collection and refinement statistics[a]. [a]Experimental details are provided in the supplemental experimental procedure.

Research highlights

1. Discovered a water-bridge structural module at the mRNA cap binding region of eIF4E.
2. The module facilitates efficient cap-protein interaction and is evolutionarily conserved.
3. X-ray structure of an apo eIF4E supports the presence of this recognition module.
4. Critical insights into water's role in mediating mRNAcap-eIF4E interaction.

eTOC Blurb

Lama et.al. have employed comparative structural analysis, molecular dynamics simulation and X-ray crystallography to identify and establish the presence of an evolutionarily conserved water-bridge structural module which is an integral component of the recognition of the 5' mRNA cap by eIF4E during the process of translation initiation.