AD_____

Award Number: DAMD17-99-1-9157

TITLE: Structural Studies of a New Nuclear Target for EGF Receptor Tyrosine Kinase

PRINCIPAL INVESTIGATOR: Guillermo A. Calero, M.D., Ph.D.
Rick Cerione, Ph.D.
Jon Clardy, Ph.D.

CONTRACTING ORGANIZATION: Cornell University
Ithaca, New York 14853-2801

REPORT DATE: August 2002

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2002	Annual Summary (15 Jul 99 - 14 Jul 02)

4. TITLE AND SUBTITLE
Structural Studies of a New Nuclear Target for EGF Receptor Tyrosine Kinase

5. FUNDING NUMBERS
DAMD17-99-1-9157

6. AUTHOR(S)
Guillermo A. Calero, M.D., Ph.D.
Rick Cerione, Ph.D.
Jon Clardy, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cornell University
Ithaca, New York 14853-2801

E-Mail:gac9@cornell.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. Abstract ***(Maximum 200 Words)*** *(abstract should contain no proprietary or confidential information)*

This project involves structural studies of a nuclear target for the EGF receptor, and the elated Neu/ErbB2 tyrosine kinase, named the CBC for RNA-capped binding protein complex. The CBC consists of two subunits, CBP20 (Mr 18 kDa) and CBP80 (Mr 90 kda), and undergoes a growth factor (EGF, heregulin)-dependent binding of RNAs transcribed by the RNA polymerase II at a 5′ cap structure that consists of a guanosine residue methylated at the N7 position. This represents a first key step in the cap-dependent splicing of precursor messenger RNA (mRNA) and in the mucleocytoplasmic transport of U snRNAs which are necessary for the formation of the spliceosome complexes. While, EGF stimulates CBC activity, it is most strongly stimulated by heregulin, an activator of the Neu/ErbB2 tyrosine kinase, and appears to be constitutive in breast cancer where Neu/ErbB2 expression is high. Thus, we believe that the CBC represents an exciting nuclear target for receptor tyrosine kinases, linking growth factor-dependent gene expression to RNA processing. We have solved the atomic structure of the CBC alone and in complex with the cap structure analog m7GpppG at 2.1 Å. The atomic structure of this triple complex represents the second eukaryotic cap binding structure solved to date and reveals interesting aspects of capped RNA binding and regulation.

14. SUBJECT TERMS	15. NUMBER OF PAGES
breast cancer, nuclear target, EGF receptor	29
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

Normal and malignant human mammary epithelial cells are able to synthesize and to respond to various different, locally acting growth factors trough specific receptors. Among these are the type 1 family of growth factor receptors, which consist of the epidermal growth factor receptor, ErbB-2/Neu, ErbB-3, and ErbB-4 [1-4]. They are required for normal mammary development and lactation and are aberrantly expressed in approximately 40% of breast carcinomas. Indeed, in human breast cancer cases the prognosis of a patient is inversely correlated with the over expression and/or amplification of this receptor family. The physiological ligand for these receptors has been shown to be heregulin [5,6]. Interaction of heregulin with the ErbB3 induces a heterodimerization between ErbB2 and ErbB3, which results in the transphosphorylation and activation of the ErbB2 receptor. Phosphorylation of this receptor initiates signaling cascades, which in turn can impact upon cell function, growth and division.

Wilson *et al.* [7] have identified a novel nuclear target for heregulin signaling which responds to the growth factor treatment of cells with an increase ability to be labeled with GTP. They identified this target as the 20-kDa subunit of the nuclear cap binding complex (CBC) and demonstrated that the CBC is stimulated to bind to capped RNAs in response to heregulin. Based on these observations Wilson *et al* suggested that heregulin could impact upon cell growth by modulating gene expression at the level of RNA processing via the CBC. They further suggested that in a situation where the heregulin signal is constitutive, the active CBC could affect gene expression by amplifying the rate of RNA processing, and thus contribute to unregulated cell growth and division.

Eucaryotic RNA polymerase II transcripts (mRNAs and U snRNAs) are modified with the addition of a ^{7}methyl guanosine cap in a 5'-5' triphosphate linkage. The CBC binds cotranscriptionally to the monomethylated cap-structure and is required for the efficient (cap-dependent) splicing of precursor mRNA[8,9,10], most likely by promoting the assembly of the commitment complex for spliceosome formation. The CBC also plays an essential role in the nuclear export of U snRNAs[11,12], which is prerequisite for U snRNP assembly in the cytosol, as well as ensures efficient 3'-end processing[13]. Recently, a role for the CBC mediating a pioneer round of translation for mRNAs subject to nonsense-mediated decay, has been described[14]. Ultimately, mature and error-proofed transcripts are relayed from the CBC to eIF-4E, the

cytosolic cap-binding component of the translation initiation complex, eIF-4F, to undergo several rounds of translation in the cytosol. Both mitogenic and stress signals activate eIF-4E in order to stimulate protein synthesis [15]. We have demonstrated that the CBC is subject to similar types of regulation[7,16], thus raising the likelihood of a regulated, temporal processing and exchange of capped mRNAs between the CBC and eIF-4E. Given the essential role for cap-binding in a number of aspects of RNA processing, there has been a great deal of interest in understanding the molecular basis by which the cap-structure associates with these proteins. X-ray crystallographic structures have been reported for m^7GDP[17], m^7GTP, and m^7GpppG[18] bound to eIF-4E providing detailed information regarding cap recognition by this cytosolic cap-binding protein. An X-ray structure for a proteolyticially-derived CBC has also been described[19], yet structural information detailing interactions between the nuclear cap-binding complex and its cap-ligand has remained elusive. Thus, the fundamental questions of how the CBP20 subunit specifically recognizes the methylated base and how CBP80 promotes high affinity interactions with the cap, remain to be addressed.

BODY

1. Experimental Procedures

Protein purification and crystallization experiments.

Purification of the m^7GpppG bound complex: Recombinant CBC was prepared from insect cell co-expressed CBP20 and CBP80. Cells were lysed using 0.4% CHAPS in a solution containing 150 mM NaCl, 30 mM Tris (pH 8.0), 2 mM DTT and 1 mM sodium azide (Buffer A). Lysates were centrifuged at 40,000 rpm for 50 minutes, and the supernatant was then loaded onto a Q-sepharose resin (Pharmacia) and eluted with Buffer A using a NaCl gradient from 150–450 mM. The fraction containing CBC was then bound to a m^7GTP sepharose (Pharmacia) column and eluted with 200 μM m^7GpppG (New England Biolabs) in Buffer A. The m^7GpppG-bound CBC was applied to a Q-resource column (Pharmacia) and further purified with a 250-450 mM NaCl gradient in Buffer A using an AKTA system (Pharmacia). As a final step, the protein was loaded onto a Superdex 200 (26/60 Pharmacia) gel filtration column in Buffer A without NaCl. The final CBC-m^7GpppG complex was assessed to be >98% pure. Purification of the unbound complex was achieved by substituting the m^7GTP-sepharose resin with a butyl-sepharose resin (Pharmacia). CBC crystals were obtained at 4°C by sitting drop vapor diffusion. Two microliters of protein solution (12 mg/ml) were mixed

with the same volume of a solution containing 22% PEG 400, 100 mM Tris-HCl (pH 7.25), 125 mM $MgCl_2$, 10% glycerol, 2 mM DTT, 10 mM glycine and 3% ethylene glycol to produce crystals diffracting up to 2.1 Å. Crystallization of the unliganded CBC was accomplished using the same conditions as above except that several rounds of seeding and dehydration were performed in order to obtain diffracting crystals. For cryocrystallography, the crystals were pre-soaked in a stabilization solution containing 40% PEG 400 and 20% glycerol, mounted in nylon loops and flash-frozen in liquid nitrogen.

Derivatization of crystals with krypton.

Crystals pre-soaked in cryoprotectant solution were mounted in nylon loops and placed in a Hampton Research Xenon chamber adapted to withstand 1000 psi of pressure. Crystals were subject to 800 psi of pressure using different protocols for 6, 12, 24 or 48 hours, with longer exposures giving the best results. After derivatization, crystals were immediately flash-frozen in liquid nitrogen.

Diffraction experiments and data processing.

All diffraction experiments were collected using synchrotron radiation at the Cornell High Energy Synchrotron Source (CHESS, Ithaca NY) and at the Argonne National Laboratory, IMCA-CAT (Argonne, Ill). An oscillation step of 0.5° was used throughout, and the crystal to detector distance varied from 200 to 225 mm. Raw reflection intensities were reduced with Denzo and Scalepack[27]. The space group was determined to be P3(2)21 for both structures using reciprocal lattice characteristics. The unit cell dimensions for the free and the m^7GpppG-bound crystals were a=b=111.68 Å, c=177.58 Å, and a=b=112.01 Å, c=175.59 Å, respectively. Both crystals contained one molecule in the asymmetric unit. The structure of CBC was solved by the molecular replacement method (MOLREP[28]) using CBP80 (PDPID code 1H6K, [19]) as a search model and the phases from a multiwavelength anomalous diffraction experiment using krypton. Krypton sites were found using SOLVE[29] and refined using SHARP[30]. After rigid body refinement CNS[31], the model was subject to several cycles of simulated annealing and model building using O[32]. The crystallographic *R*work and *R*free values for the CBC in complex with m^7GpppG are 22.3 and 24.5, respectively (40–2.1 Å data). The *R*work and *R*free values for the substrate free complex are 24.5 and 29.1, respectively (40–2.75 Å data). Other

data-collection and refinement statistics are shown in the supplementary information (Table 3). Figures were generated with Bobscript [33] and SPOCK [34] and rendered using RASTER3D [35].

2. Results

We have determined the three dimensional structure of the full-length CBC bound to m^7GpppG by molecular replacement, using CBP80 as an initial search model[19] and the phases obtained from a krypton multiple-wavelength anomalous dispersion (MAD) experiment (Figure 1A). The atomic structure of CBP20 includes residues 6-153 of the total 156, and conforms to the classical ribonucleotide binding domain (RNP) fold containing four anti-parallel β-strands arranged in the order β4-β1-β3-β2, packed against the αA and αB helices. The RNP-fold is comprised of two conserved motifs designated RNP1 (β2, residues 81-88) and RNP2 (β1, residues 41-46) (Figure 1B; grey). Additions to the RNP-fold through N-terminal (helices α1, α2, and α3; magenta) and C-terminal (the β4-αC loop and helix αC; green) insertions contribute substantially to the binding of the ^{7}methyl guanosine cap. These extensions enable the CBC to exhibit functional specificity relative to other RNP-containing proteins like the U1A protein[20,21], the Sex–lethal protein[22], and the poly-(A)-binding protein[23] which all bind primary and secondary elements of RNA in distinct ways but are incapable of binding the cap-structure.

CBP80 is a super-helical structure consisting of three domains (domain 1: residues 23-244, helices 1-11; domain 2: residues 309-478, helices 15-24; and domain 3: residues 498-790, helices 25-37) connected by two linkers (Figure 1A). The N-terminal domain 1 is structurally similar to the middle domain of eIF-4G (MIF4G) which plays a key role in cap-dependent mRNA translation[24]. The atomic structure of CBP80 includes residues 23-526, 538-667 and 687-790 (residues 1-22, 527-537 and 668-686 have no visible electron density). CBP80 does not directly bind to nor cover the cap-binding site, despite the well-documented requirement of this subunit for the high affinity binding of the ^{7}methyl guanosine cap to CBP20[8]. Nonetheless, there are extensive interactions between CBP80 and the cap-bound CBP20 subunit which bury a total of 3176 A^2 of surface. Helices A and B of CBP20, which are arranged orthogonal to one another and form an underlying layer to the β-sheet core of the RNP domains, engage domains 1 and 2 of CBP80.

Additionally, there are interactions between the N-termini of CBP20 and CBP80 that contribute to cap-binding and will be detailed below.

Figure 2A illustrates the electron density map for the m^7GpppG cap-analog and Figure 2B shows a surface representation of the cap-binding site within CBP20. The methylated cap-analog fits within a deep cavity of ~180 $Å^3$, with the electrostatic potential switching from negative to positive in the clockwise direction, thus counterbalancing charges contributed by the cap-structure. Recognition and binding of the cap analog is achieved in two ways. First, π-π stacking interactions occur between the electron deficient ^{7}methyl guanosine ring and two electron rich aromatic residues, Tyr 20 (loop α2-α3) and Tyr 43 (RNP2). The stacking residues and the methylated base are oriented along three planes parallel to the β-platform. The upper plane is formed by Tyr 20 and is 3.25 Å away from the methylated base (which constitutes the middle plane) (Figure 2C, left) and overlaps partially with the guanine ring (Figure 2C, right). Given that the average van der Waals distance for π-π stacking interactions is 3.4 Å (ref), the interaction between Tyr 20 and the methylated guanosine ring is likely to be strong. The lower plane is formed by Tyr 43 which is separated from the methylated base by 3.4 Å (Figure 2C, left) and its degree of overlap with the methylated base is larger than that of Tyr 20 (Figure 2C, right). Theoretical data and experimental studies using site directed mutagenesis suggest that the area of overlap between the stacking residues and the methylated base contribute significantly to the binding affinity[25-27]. Methylation of the guanine ring draws electron density away from the imidazole ring, allowing strong interactions with the side chain of the aromatic residues and thereby enhancing the π-π stacking interactions. In addition to the π-π stacking interactions, there are a number of specific interactions between the cap-structure and residues in the β4-αC loop (Figure 2D). These include hydrogen bonds with the ^{7}methyl guanosine that mimic Watson-Crick pairing between guanosine and cytosine, and hydrogen bonds and van der Waals contacts with the ribose and phosphate oxygen atoms. Additionally, there is a stacking interaction between Tyr 138 and the non-methylated guanosine moiety (inter-planar distance = 3.6 Å) conferring the specificity for a 5'-5' triphosphate linkage. Figure 2E shows that the residues responsible for binding the cap-structure are highly conserved among different species.

Our structural data explain the results from previous mutagenesis studies[19]. For example, mutation of the π-π stacking residue Tyr 43 to alanine in CBP20 reduces the affinity of the CBC for m^7G-capped U1 snRNA by more than 100 fold. Likewise, mutations of Asp 116 (from the β4-α6 loop) and Phe 83 (from the β3 strand) to alanine residues cause greater than 100 fold reductions in the affinity for the cap-structure. Our

structural data predict that mutating Asp 116 to Ala disrupts Watson-Crick pairing between Asp 116(OD2) and the exocyclic NH_2, as well as potentially alters the geometry of the β4-α6 loop, as Asp 116 hydrogen bonds with Gly 118 and Arg 123. The effect of changing Phe 83 to alanine is less clear. Given that Phe 83 participates in C-H-O bonds with atom O4 from the ribose, the disruption of these interactions could decrease the binding affinity.

The interactions between the CBC and the methylated base described above support the observed high affinity interaction between methylated cap structures and the CBC (approximate Kd = 5 nM [7]. The cavity where the methylated base binds is larger in the case for CBP20 than for eIF-4E, and interplaner distances between the methylated base and the residues participating in the π-π stacking interactions are shorter. Although the two proteins bind the methylated base by satisfying Watson-Crick base pairing, the overall number of hydrogen bonds between CBP20 and the methylated base, the ribose, and the phosphate oxygen is greater than that for eIF-4E. Thus, it is not surprising that the CBC binds cap structures with significantly higher affinity than does eIF-4E (Kd = 200 nM)[28]. The differences between the CBC and eIF-4E in their binding affinities for capped RNA speaks to their divergent roles in RNA processing. It is attractive to envision that the higher affinity of CBP20 for the cap makes the CBC more suitable for "chaperoning" capped RNAs through sequential RNA processing events during the life span of RNA molecules in the nucleus.

A key question concerns how access to the cap-binding site of the CBC is regulated. Such regulation is necessary so that each mature mRNA molecule can dissociate from the CBC, enabling the processed mRNA to bind to eIF-4E and to be translated by the eIF-4F complex, as well as ultimately allowing a precursor mRNA molecule to bind to the CBC in place of the mature mRNA. Comparisons of the structures for the cap-bound CBC with that for the free, unbound protein provide some insights into the molecular basis of this regulation. Figures 3A and 3B show the electrostatic surface representations of CBP20 in the unbound and bound states, respectively, and demonstrate that a significant conformational change is associated with the binding of the cap-analog. The blue arrows in Figure 3A show the location of the empty cap-binding slot. The transition from the open (unbound) to the closed (bound) state buries a previously accessible surface area of 667 $Å^2$. This conformational change results in the hinge-like movement of the amino-terminal residues 16-30 from the α2-α3 loop of CBP20 toward the β-sheets comprising the RNP-fold. In the open state, the conformationally-sensitive α2-α3 loop is stabilized by intra-chain interactions, including Asn 19(O)-Arg 21(N), Arg 21(O)-Asn 23(N), and Asn 23(O)-Phe 25(N), whereas the closed state is maintained mainly

through hydrogen bond interactions between Ser 18 and Asp 114, and salt-bridges between Arg 21 and Glu 33, and between Asp 22 and Arg 127 (see Figure 3C). The net effect of the conformational change accompanying cap-binding is to rotate the α2-α3 loop approximately 55 degrees, forming a 'roof' for the cap-binding site and possibly the RNA-binding groove (Figure 3B; the black arrows point to a possible RNA-binding groove extending from the cap-binding site). This movement enables Tyr 20 to enter into a stacking position together with Tyr 43, thus sandwiching the ^{7}methyl guanosine moiety of the cap (Figure 3C). Arginine 26 moves nearly 11 Å from its position in the unbound CBP20 subunit, such that Tyr 20 now partially occupies the position formerly held by the arginine residue. The conformational change also impacts the C-terminal residues of CBP20, as residues 128-156 within this region show no detectable electron density in the unbound or open state, suggesting that they are disordered or flexible, whereas a clear electron density is observed for this region in the closed (bound) state. Given that there is very little change within the RNP-fold, for either the cap-bound or unbound states, an attractive possibility is that this region serves as the initial binding site for the cap-structure, with the ensuing conformational change then bringing Tyr 20 and the carboxyl-terminal residues into position to further contribute to the binding interaction.

The necessity for CBP80 in achieving a high affinity interaction between the cap structure and the CBC can now clearly be appreciated. Two 3-10 helices, α1 and α2 from CBP20, which are connected by a short loop and form the N-terminal extension to the RNP domain (see Figure 4), fit into a binding groove (shown in cyan) within CBP80. This groove in effect serves as 'fulcrum' for the hinged motion of the α2-α3 loop that accompanies the cap-binding interaction (see Figure 4, inset). The N-terminal region of CBP20 forms no other contacts with CBP80, aside from those within the groove. Thus, it appears that this contact between CBP80 and CBP20 is responsible for maintaining the N-terminus of CBP20 in a position that allows Tyr 20 to participate in π-π stacking with the methylated base.

Given what is known about the growth factor-dependent regulation of cap binding to eIF-4E, comparisons between the two major cap-binding protein complexes offer some plausible mechanisms regarding the regulation of CBC activation and capped RNA binding imparted by extracellular stimuli. Specifically, the structurally similar CBP80 and eIF-4G proteins act in essence as regulatory factors to ensure the high affinity binding of the cap-structure to CBP20 and eIF-4E, respectively. Both CBP80 and eIF-4G make important contacts with the β-sheet core and portions of the N-terminal extensions of their respective

binding partners. In the case of eIF-4E, the growth factor-stimulated phosphorylation of 4E-BP1 [29] [30], a regulatory protein that binds to eIF-4E and significantly weakens its affinity for capped RNA, results in the dissociation of 4E-BP1 and the binding of eIF-4G in its place [31] [32]. For the case of the CBC, the CBP80 subunit, while serving a function analogous to eIF-4G, differs from it by remaining bound to the cap-binding subunit (CBP20) in both its high affinity (cap-bound) and low affinity (cap-free) states. However, the N-terminal portion of CBP80, which constitutes part of the binding groove for the conformationally-sensitive loop of CBP20, contains consensus phosphorylation sites for the ribosomal p70 S6 kinase, which is under growth factor control. Thus, CBP80 may combine features of 4E-BP1 which serves as a growth factor/phosphorylation-sensor, and eIF-4G which helps stabilize the conformational state for the high affinity cap-binding interaction. It is worth noting that the N-terminal portion of CBP80 also contains a binding site for α̃importin[12]. We have observed that the binding of α-importin to CBP80 is translated into an increased binding affinity for capped RNA by the CBC, whereas the binding of β-importin to α-importin may weaken the binding of the capped RNA (K.W. Cerione and G. Calero, unpublished data). Subtle alterations in the position of the N-terminal portion of CBP80, as influenced by whether α-importin and/or β-importin (or other as yet to be identified CBC-binding partners) are bound to that region, could alter the hinge-like movement of the conformationally-sensitive loop of CBP20. Such alterations, induced by a growth factor- or stress response-induced phosphorylation of CBP80, could in turn be translated into an opening of the cap-binding site, thus facilitating the exchange of a newly processed capped mRNA for another precursor mRNA molecule and thereby providing a molecular basis for the growth factor- (or stress-) dependent activation of the CBC.

KEY RESEARCH ACCOMPLISHMENTS

Expression of CBC in SF9 cells.

Purification and crystallization of CBC.

Solution of the atomic structure of CBC in complex with m7GpppG at 2.1 Å resolution.

Refinement of the CBC structure with m7GpppG with a final R_{free}/R_{factor} of 24.7/22.2 (Table 1)

Solution of the atomic structure of the CBC without the substrate m7GpppG at 2.7 Å resolution.

Refinement of the CBC without m7GpppG with a final R_{free}/R_{factor} of 29.1/24.7 (Table 1).

REPORTABLE OUTCOMES

Submitted manuscript:

ATOMIC STRUCTURE OF THE NUCLEAR CAP BINDING PROTEIN (CBP20) IN COMPLEX WITH CAP BINDING PROTEIN 80 (CBP80) AND THE CAP ANALOG m7GpppG.

G.A. Calero, K.F. Wilson, J.L. Rios, T.K. Ly, R.A. Cerione and J.C.Clardy.

REFERENCES

1. Kraus, M. H., Issing, W., Miki, T., Popescu, N. C. & Aaronson, S. A. Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors. *Proc Natl Acad Sci U S A* **86**, 9193-7. (1989).
2. Slamon, D. J. et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-82. (1987).
3. Slamon, D. J. et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**, 707-12. (1989).
4. van de Vijver, M. et al. Amplification of the neu (c-erbB-2) oncogene in human mammmary tumors is relatively frequent and is often accompanied by amplification of the linked c-erbA oncogene. *Mol Cell Biol* **7**, 2019-23. (1987).
5. Plowman, G. D. et al. Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal growth factor receptor family. *Proc Natl Acad Sci U S A* **90**, 1746-50. (1993).
6. Sliwkowski, M. X. et al. Coexpression of erbB2 and erbB3 proteins reconstitutes a high affinity receptor for heregulin. *J Biol Chem* **269**, 14661-5. (1994).
7. Wilson, K. F. et al. The nuclear cap-binding complex is a novel target of growth factor receptor-coupled signal transduction. *J Biol Chem* **274**, 4166-73. (1999).
8. Izaurralde, E. et al. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657-68. (1994).
9. Lewis, J. D., Gorlich, D. & Mattaj, I. W. A yeast cap binding protein complex (yCBC) acts at an early step in pre- mRNA splicing. *Nucleic Acids Res* **24**, 3332-6. (1996).
10. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, I. W. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* **10**, 1683-98. (1996).
11. Izaurralde, E. et al. A cap-binding protein complex mediating U snRNA export. *Nature* **376**, 709-12. (1995).

12. Gorlich, D. et al. Importin provides a link between nuclear protein import and U snRNA export. *Cell* **87**, 21-32. (1996).

13. Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W. & Gilmartin, G. M. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc Natl Acad Sci U S A* **94**, 11893-8. (1997).

14. Ishigaki, Y., Li, X., Serin, G. & Maquat, L. E. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**, 607-17. (2001).

15. Rhoads, R. E. Regulation of eukaryotic protein synthesis by initiation factors. *J Biol Chem* **268**, 3017-20. (1993).

16. Wilson, K. F., Wu, W. J. & Cerione, R. A. Cdc42 stimulates RNA splicing via the S6 kinase and a novel S6 kinase target, the nuclear cap-binding complex. *J Biol Chem* **275**, 37307-10. (2000).

17. Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. X-ray studies of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Nucleic Acids Symp Ser* **36**, 8-11 (1997).

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* **362**, 539-44. (2002).

19. Mazza, C., Ohno, M., Segref, A., Mattaj, I. W. & Cusack, S. Crystal structure of the human nuclear cap binding complex. *Mol Cell* **8**, 383-96. (2001).

20. Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515-20. (1990).

21. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-8. (1994).

22. Handa, N. et al. Structural basis for recognition of the tra mRNA precursor by the Sex- lethal protein. *Nature* **398**, 579-85. (1999).

23. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835-45. (1999).

24. Marcotrigiano, J. et al. A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell* **7**, 193-203. (2001).

25. Ishida, T., Shibata, M., Fujii, K. & Inoue, M. Inter- and intramolecular stacking interaction between indole and adeninium rings. *Biochemistry* **22**, 3571-81. (1983).

26. Ishida, T., Ueda, H., Segawa, K., Doi, M. & Inoue, M. Prominent stacking interaction with aromatic amino acid by N- quarternization of nucleic acid base: X-ray crystallographic characteristics and biological implications. *Arch Biochem Biophys* **278**, 217-27. (1990).

27. Ishida, T. et al. Cooperative face-to-face and edge-to-face aromatic interactions of tryptophan indole ring with N7-quarternized guanine and neutral cytosine bases. *FEBS Lett* **333**, 214-6. (1993).

28. Scheper, G. C. et al. Phosphorylation of eukaryotic initiation factor 4E markedly reduces its affinity for capped mRNA. *J Biol Chem* **277**, 3303-9. (2002).

29. Pause, A. et al. Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* **371**, 762-7. (1994).

30. Lin, T. A. et al. PHAS-I as a link between mitogen-activated protein kinase and translation initiation. *Science* **266**, 653-6. (1994).

31. Haghighat, A. & Sonenberg, N. eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. *J Biol Chem* **272**, 21677-80. (1997).

32. von Der Haar, T., Ball, P. D. & McCarthy, J. E. Stabilization of eukaryotic initiation factor 4E binding to the mRNA 5'- Cap by domains of eIF4G. *J Biol Chem* **275**, 30551-5. (2000).

33. Otwinowsky, Z & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326 (1997).

34. Collaborative computational project, number 4. The CCP4 Suite: Programs for X-ray Crystallography. *Acta Crystallogr. D.* **50**, 760-763 (1994).

35. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**, 849-61. (1999).

36. De la Fortelle, E. & Bricogne, G. Maximum likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494 (1997).

37. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21. (1998).

38. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47**, 110-9. (1991).

39. Esnouf, R.M. An extensive modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 133-138 (1997).

40. Christopher, J. A, and Baldwin, T.O *J. Mol. Graph Model.* **16**, 285 (1998).

41. Merritt, Ethan A. and Bacon, David J. "Raster3D: Photorealistic Molecular Graphics" Methods in Enzymology 277, 505-524 (1997).

42. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-90. (1988).

43. Barton, G.J. ALSCRIPT - A Tool to Format Multiple Sequence Alignments, *Prot. Eng.*, **6**, .37-40. (1993).

APPENDIX

Figure 1A

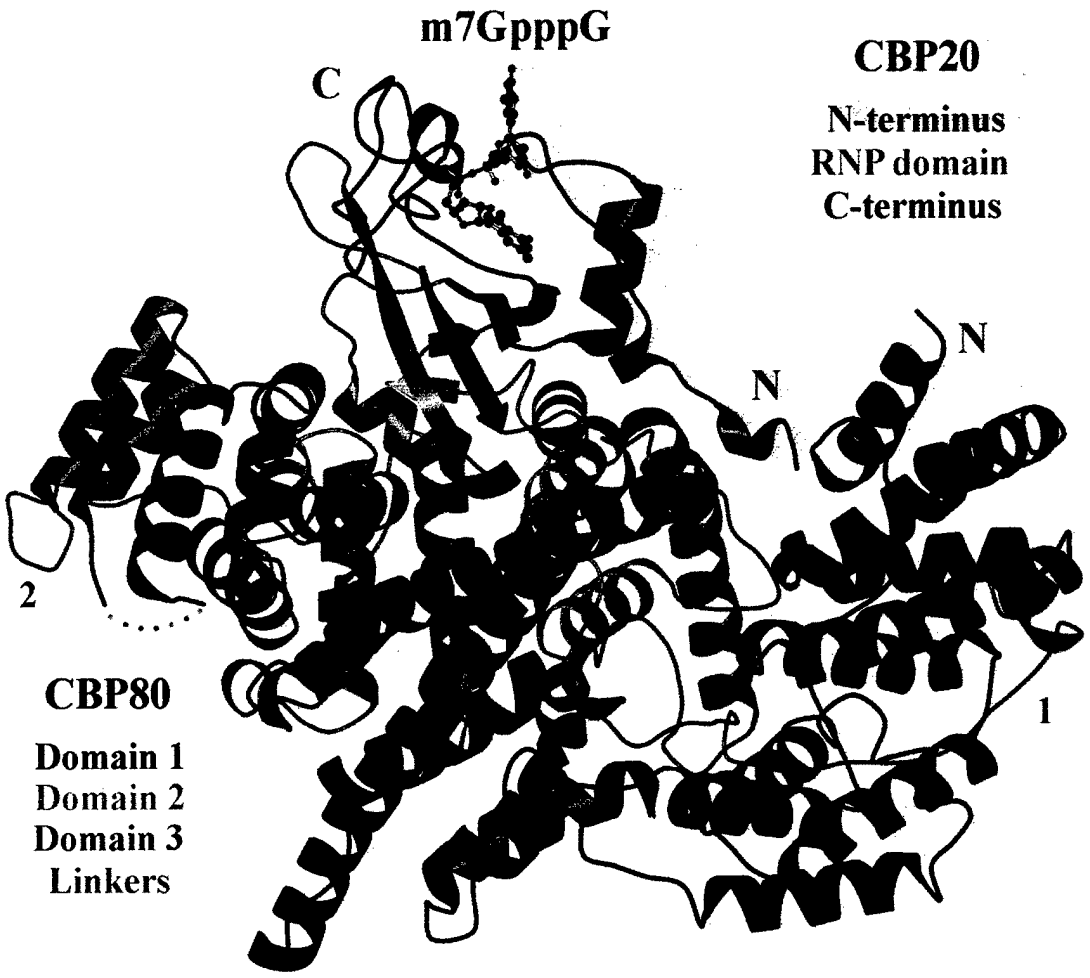


Figure 1B

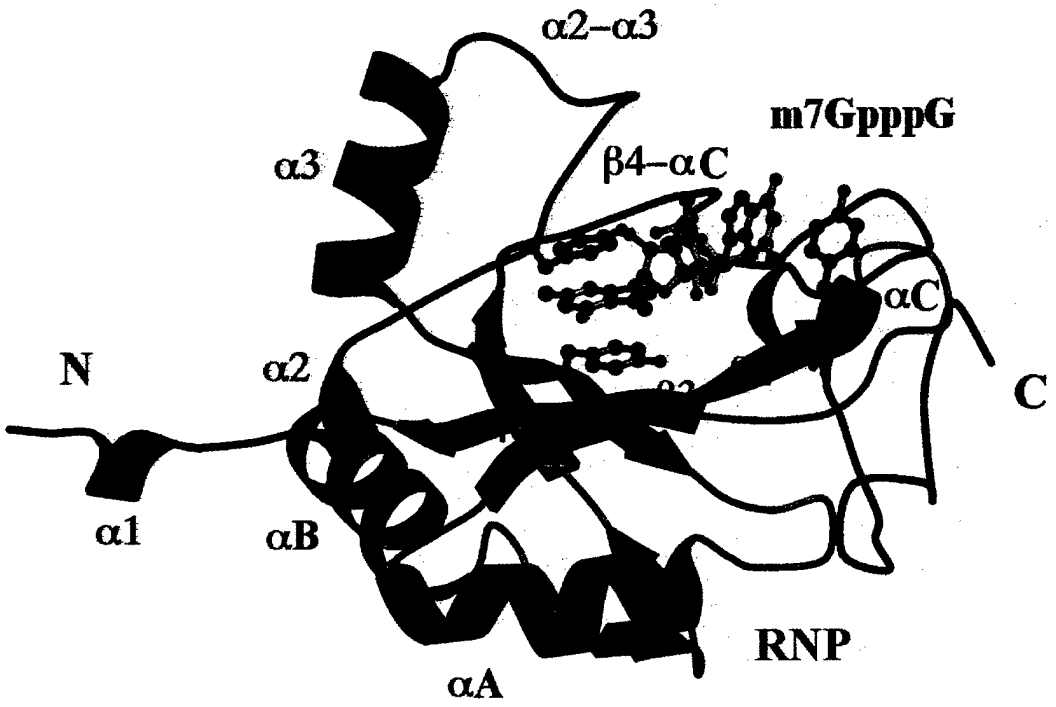


Figure 2A

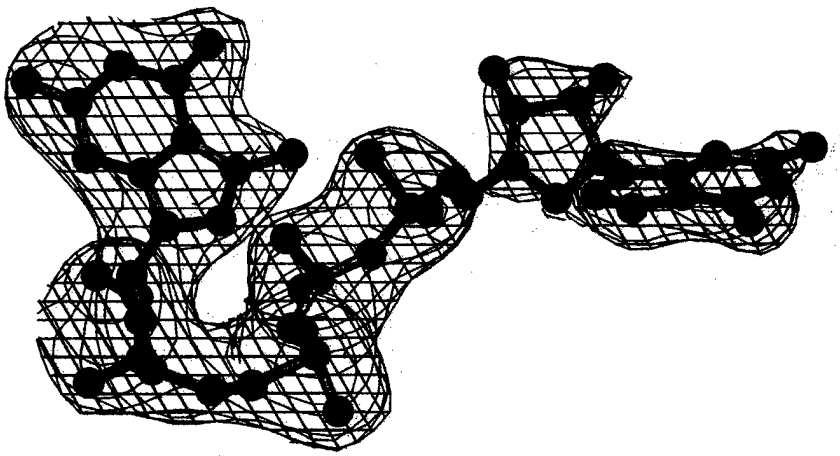
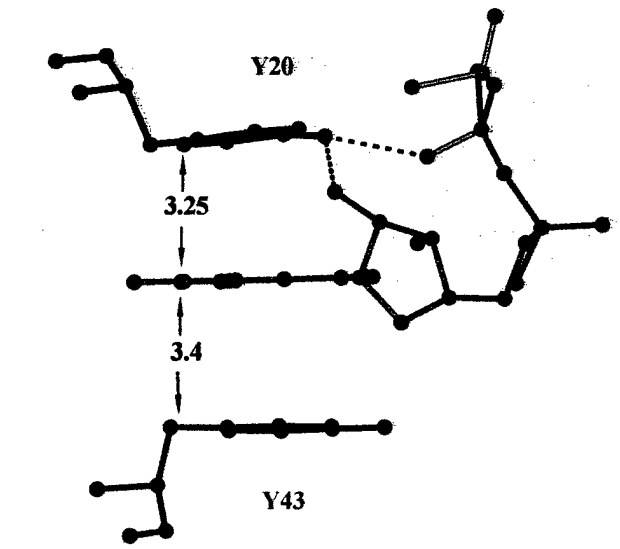


Figure 2B

Figure2c



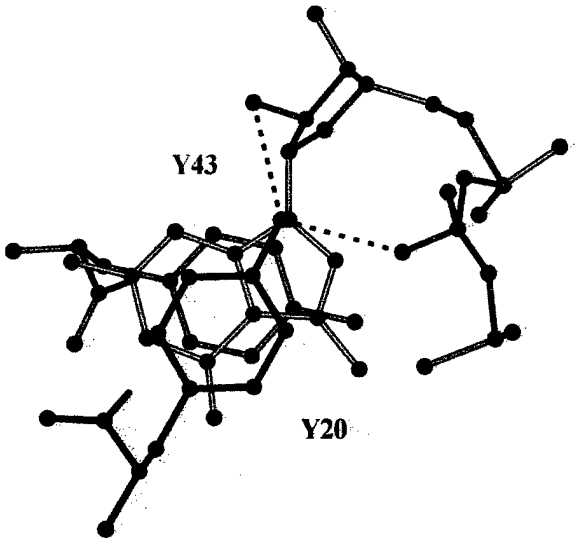


Figure 2D

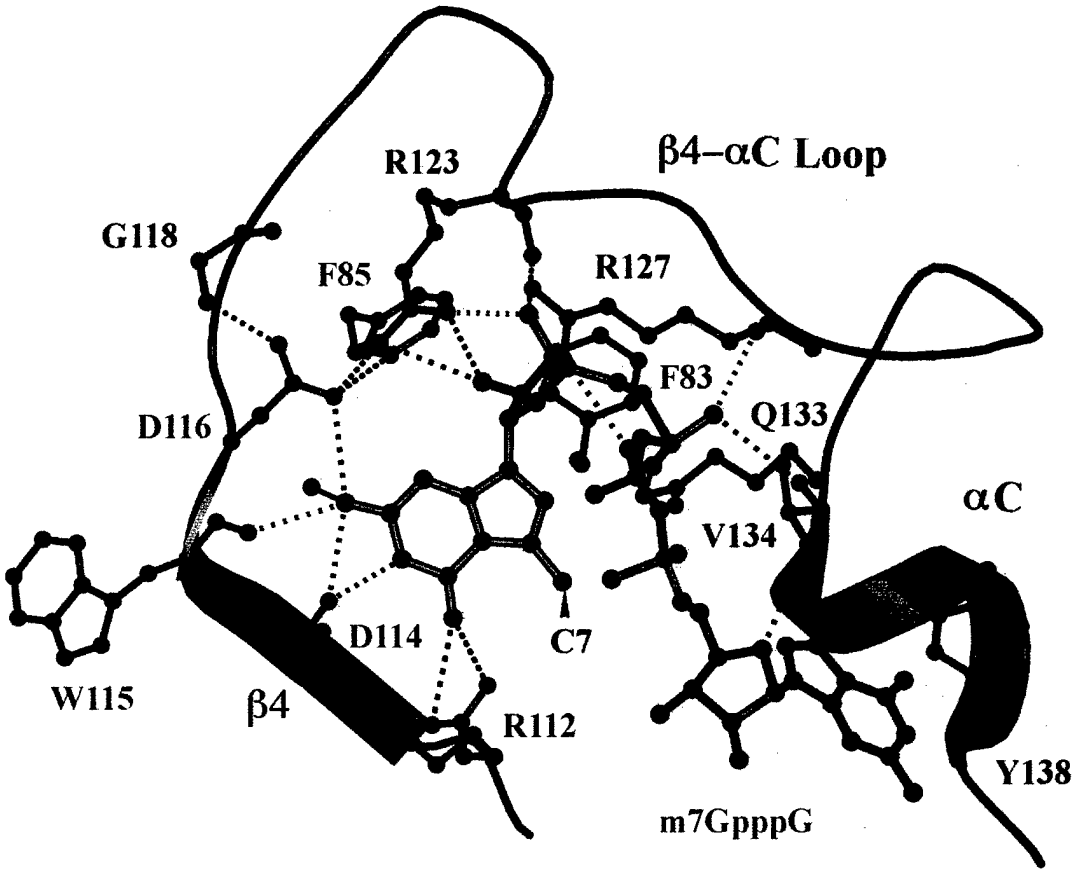

R123
β4–αC Loop
G118
F85
R127
F83
Q133
D116
αC
V134
D114
C7
W115
β4
R112
Y138
m7GpppG

Figure 2E

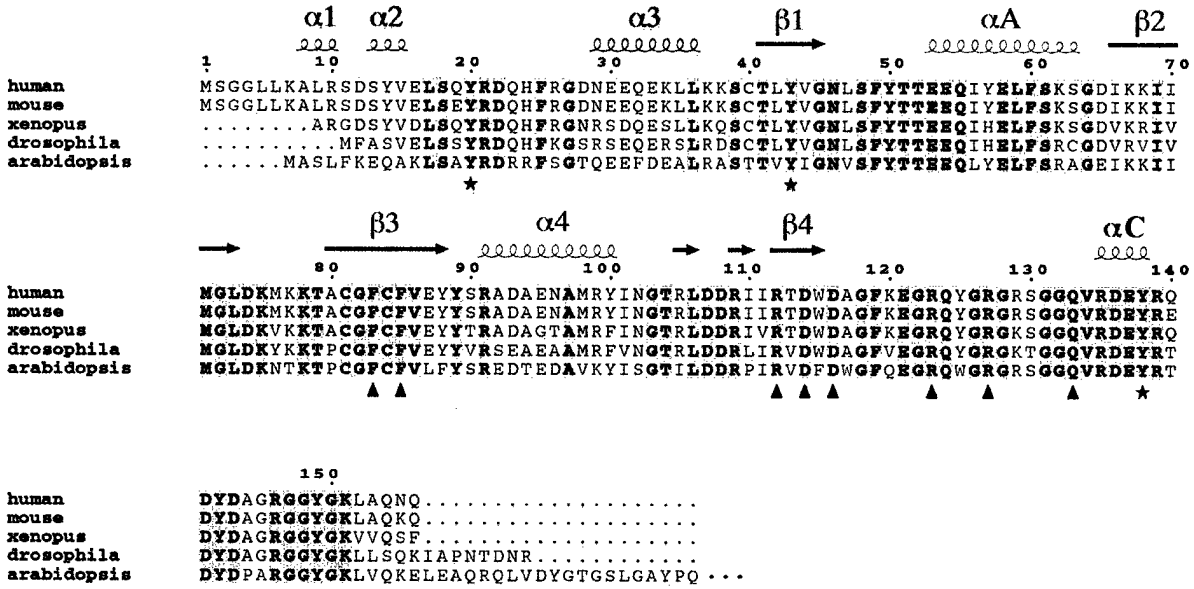

α1 α2 α3 β1 αA β2
1 10 20 30 40 50 60 70
human MSGGLLKALRSDSYVELSQYRDQHFRGDNEEQEKLLKKSCTLYVGNLSFYTTEEQIYELFSKSGDIKKII
mouse MSGGLLKALRSDSYVELSEYRDQHFRGDNEEQEKLLKKSCTLYVGNLSFYTTEEQIYELFSKSGDIKKII
xenopus .....ARGDSYVDLSQYRDQHFRGNRSDQESLLKQSCTLYVGNLSFYTTEEQIHELFSKSGDVKRIV
drosophila .....MFASVELSSYRDQHFKGSRSEQERSLRDSCTLYVGNLSFYTTEEQIHELFSRCGDVRVIV
arabidopsis .....MASLFKEQAKLSAYRDRRFSGTQEEFDEALRASTTVYIGNVSFYTTEEQLYELFSRAGEIKKII
β3 β4 αC
80 90 100 110 120 130 140
human MGLDKMKKTACGFCFVEYYSRADAENAMRYINGTRLDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYRQ
mouse MGLDKMKKTACGFCFVEYYSRADAENAMRYINGTRLDDRIIRTDWDAGFKEGRQYGRGRSGGQVRDEYRE
xenopus MGLDKVKKTACGFCFVEYYTRADAGTAMRFINGTRLDDRIVRTDWDAGFKEGRQYGRGKSGGQVRDEYRQ
drosophila MGLDKYKKTPCGFCFVEYYVRSEAEAAMRFVNGTRLDDRLIRVDWDAGFVEGRQYGRGKTGGQVRDEYRT
arabidopsis MGLDKNTKTPCGFCFVLFYSREDTEDAVKYISGTILDDRPIRVDFDWGFQEGRQWGRGRSGGQVRDEYRT
150
human DYDAGRGGYGKLAQNQ.....
mouse DYDAGRGGYGKLAQKQ.....
xenopus DYDAGRGGYGKVVQSF.....
drosophila DYDAGRGGYGKLLSQKIAPNTDNR.....
arabidopsis DYDPARGGYGKLVQKELEAQRQLVDYGTGSLGAYPQ...

Figure3A

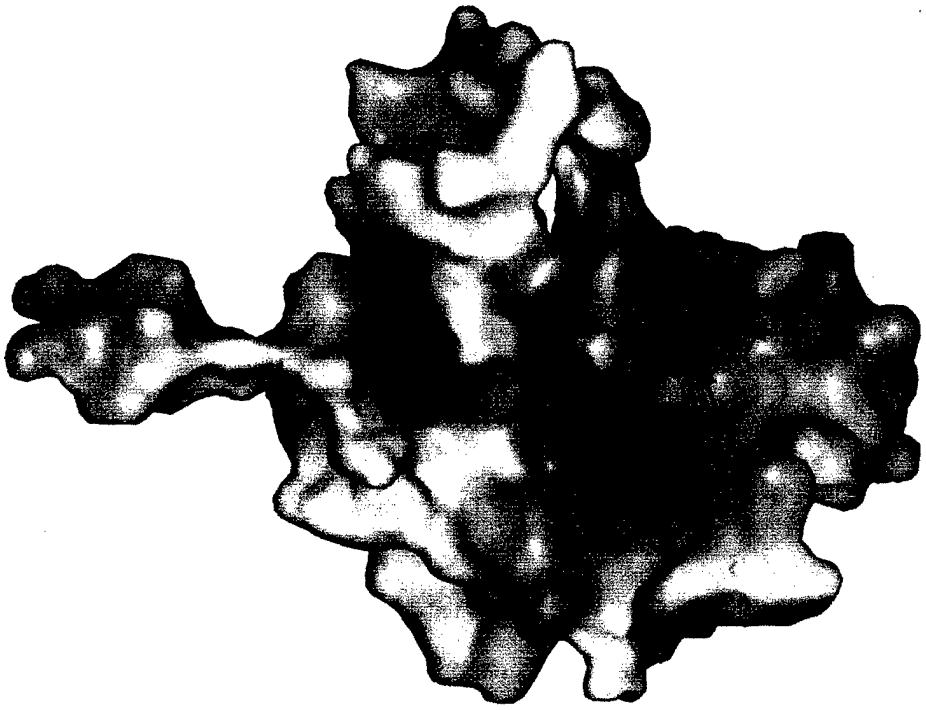


Figure3B

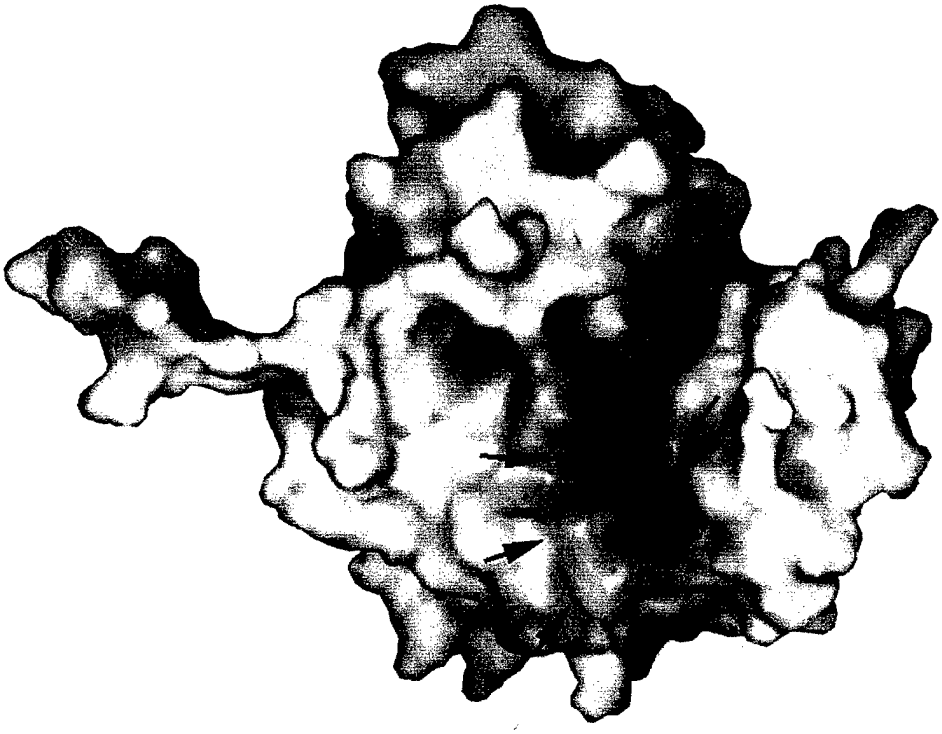


Figure 3C

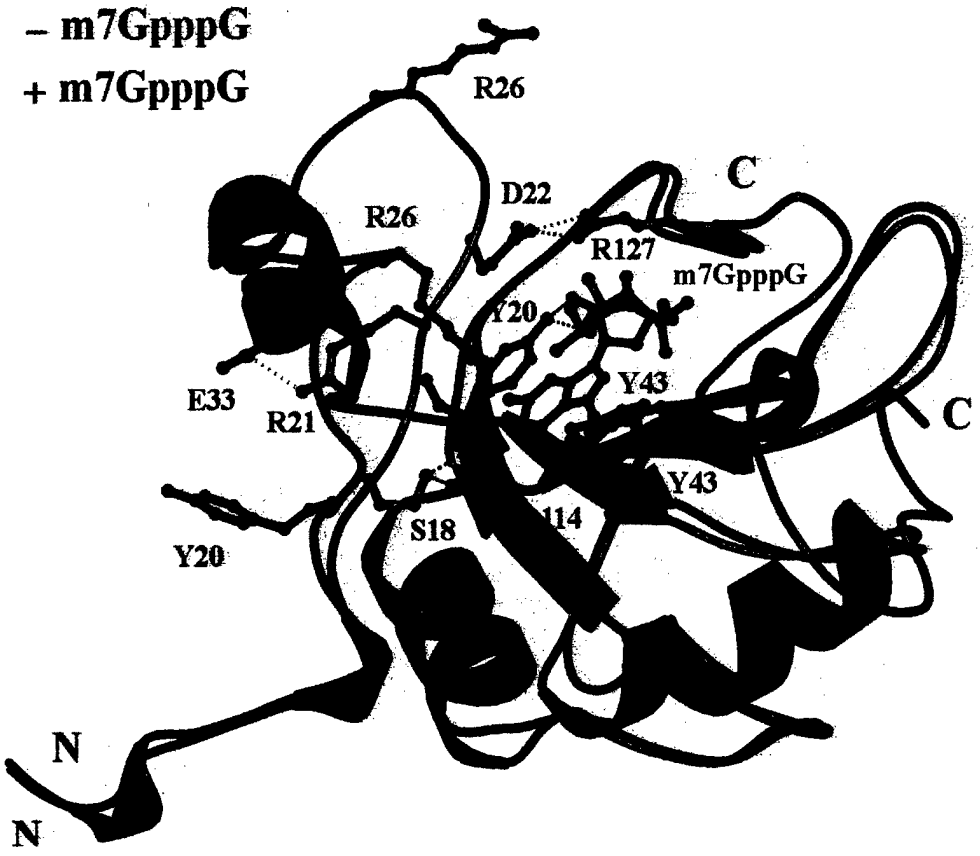

– m7GpppG
+ m7GpppG
R26
D22
C
R26
R127
m7GpppG
Y20
Y43
E33
R21
C
Y43
S18
I14
Y20
N
N

Figure4d

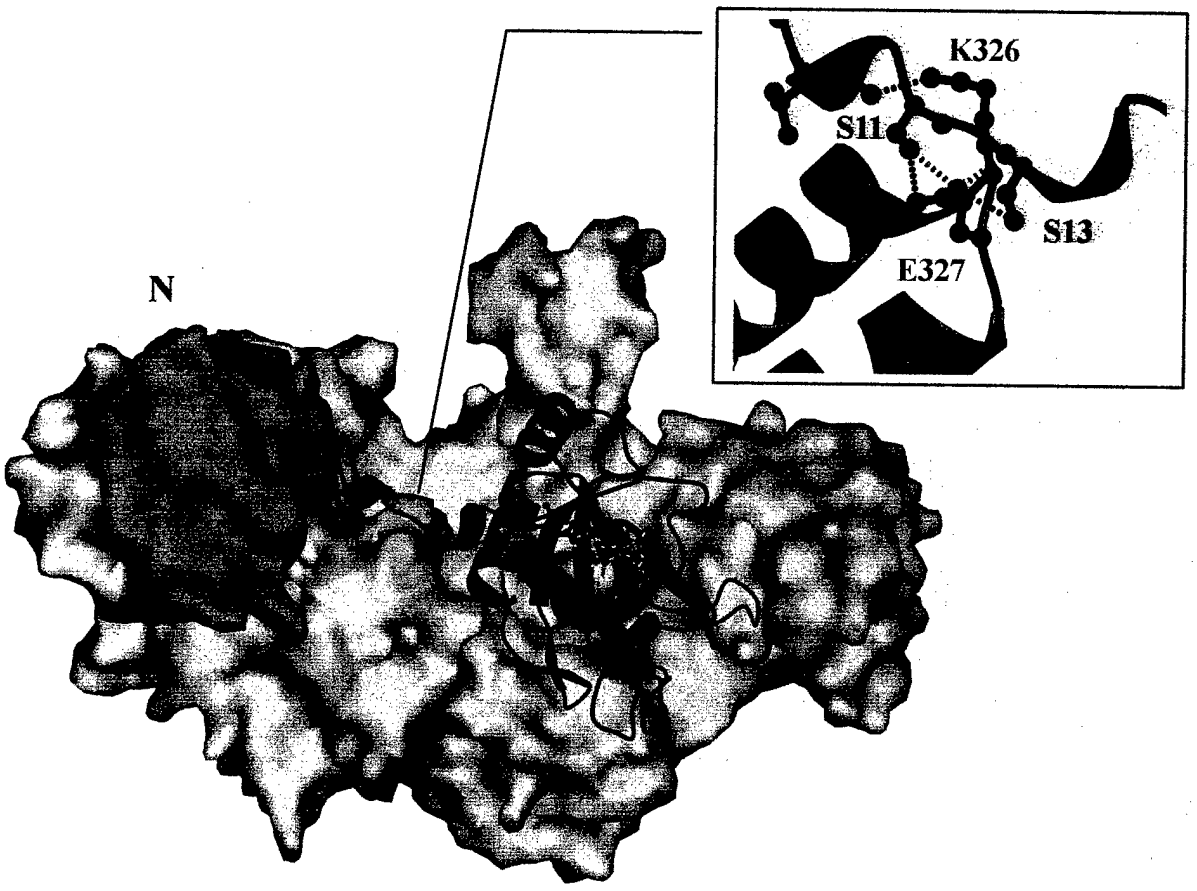


Figure Legends

Figure 1. (A) Ribbon diagrams for the CBC in complex with the cap structural analog m^7GpppG. The three structural domains of CBP80 and the three main regions of CBP20 are shown as is the m^7GpppG (ball-and-stick). (B) The fold of CBP20 conforms to the RNP fold (shown in gray color) with the following topology α1-α2-α3-β**1**-α**A**-β**2**-β**3**-α**B**-βA-βB-β**4**-αC (bold letters denote canonical elements). The strands of the β-sheet are anti-parallel, twisted and arranged in the order β**4**-β**1**-β**3**-β**2**. The two middle strands, β**1** (residues 41-46) and β**3** (residues 81-88), form the RNP2 and RNP1 motifs, respectively. Alpha helices α**A** and α**B** (which interact with CBP80) are orthogonal to each other and are located almost horizontally below the plane of the β sheet creating a layered core. N-terminal insertions (shown in magenta) to the RNP fold comprise two short 3_{10} helices (α1 and α2 that interact with CBP80), loop α2-α3 that contains Y20 and helix α3. C-terminal insertions (shown in green) participate in cap binding and consist of a semi-circular loop β**4**-αC followed by a short helix (αC) and a C-terminal loop. Also shown is the cap analog m^7GpppG and the residues which participate in π-π stacking interactions with it, Y20 (magenta) and Y43 (gray).

Figure 2. Interactions of CBP20 with m^7GpppG. (A) Electron density map (2Fo-Fc) for the methylated dinucleotide m7GpppG calculated at 1.2 sigma. (B) Surface electrostatic representation calculated with SPOCK[32] for the m^7GpppG-binding cavity. The cavity is formed from residues Y20 and L36 (top), L42 and Y43 (bottom), and F85, R112, D114, W115, D116, A117 R123 and R127 (semicircular wall). (C) Ball and stick representation of π-π stacking and degree of overlap between Y20, m7GpppG and Y43 (see text) (D) Ball and Stick representation of the hydrogen bond network stabilizing m7GpppG. Hydrogen bonds with 7-methyl guanine include: R112(NH1)-O6=3.3Å, R112(NH2)-O6=3.4Å, D114(OD2)-N1=2.49Å, D114(OD2)-N2=2.96Å, W115(O)-N2=2.96Å and D116(OD2)-N2=3.13Å. 2). Hydrogen bonds with the ribose and phosphate oxygen atoms comprise: R123(NH1)-O2=2.7Å, R123(NH2)-O2=3.26Å, R123(NH1)-O3=2.37Å, R123(O)-O3=2.78Å, R127(N)-O2A=2.71Å, G128(N)-O2A=3.72Å, N133(NE2)-O1A=2.69Å, V134(N)-O2A=3.26Å, Y20(OH)-O1B=2.81Å and F83(CE1)-O4=3.5Å, F83(CE1)-O4=3Å and D116(OD2)-F85(CE2)=3.02Å 3) Intra-chain interactions include hydrogen bonds D116(OD1)-G118(N)=2.64Å and Y43(OH)-N133(NE2)=2.64Å and a salt bridge D116(OD2)-R123(NH2)=2.58Å. 4) Interactions with the non-methylated guanosine include stacking interactions between Y138 and the guanine ring (inter-planar distance=3.6Å) and the C-H-O bond V134(CG1)-O4=3.18Å..

(E) Amino acid sequence alignment (MultAlin[37]) using Alscript[38] for human and selected non-human CBP20s. Conserved residues include among others: 1) Most residues in RNP1 (β2) and RNP2 (β3) motifs. 2) Residues involved in stacking interactions with m7GpppG (Y20, Y43 and Y138). 3) Hinge residues (L17 and S18). 4) Residues involved in stabilization of the closed conformation (Y20, R21, D22, E33); 5) Residues in loops β2-β3 and 6) Residues involved in hydrogen bonding with the cap di-nucleotide (see text).

Figure 3. Conformational change associated with cap-binding to CBP20. (A) and (B) Electrostatic surface representation of CBP20 in the open and closed conformations respectively. Rotation of loop 1 upon cap binding forms the roof of the cap binding cavity. Blue arrows in (A) show the location of the cap binding site (see text); black arrows in (B) show a possible RNA binding surface in CBP20. (C) Analysis of the free (open) (red) versus the substrate bound (closed) (green) conformations of CBP20 (Rms deviations on Cα=0.508 Å) reveals considerable structural changes associated with the binding of the m^7GpppG cap-analog. Calculation of RMS distances on Cα and ϕ-ψ angles for the two conformations suggest that the N-terminus residues 16-29 from loop α2-α3 undergo a hinged motion towards the β-sheet. The net effect of this motion is to rotate loop α2-α3 approximately 55° so that Y20 moves about 8.3 Å on top of Y43 to engage in cap binding while Arg26 (that was occupying Y20 position) moves outward about 10.8Å. In addition, a small rotation (opposite to the direction of the hinged motion) of the β-sheet as a rigid body aligns Y43 in a plane almost parallel to the plane of Y20.

Figure 4. Surface (CBP80) and ribbons (CBP20) representation to illustrate the stabilization of the N-terminal (magenta) hinge of CBP20 by CBP80. The N-terminus of CBP20 (residues S11, D12 and S13 shown in magenta) crosses through a "saddle" formed by residues K326 and E327 in CBP80 (cyan) making multiple contacts that fix it, allowing the hinged motion of residues 16-29 in loop α2-α3. The N-terminal region of CBP80 is in blue. The insert shows in detail these interactions which include hydrogen bonds K326(NZ)-L9(O) = 3.3Å, E327(OE2)-S11(OG) = 2.93Å and E327(OE2)-S13(N) = 2.74Å).

Table 1 Data collection and refinement statistics

	Native 1 + cap	*Native 2 + cap*	*Native - cap*	*Kr λ1 (edge)*	*Kr λ2 (peak*	*Kr λ3 (remote*
Beamline	A1 CHESS	17-ID IMCA-CAT	F1 CHESS	F2 CHESS	F2 CHESS	F2 CHESS
Resolution (Å)	40-2.35	40-2.1	40-2.75	40-2.65	40-2.65	40-2.65
Wavelength (Å)	0.9347	1.00	0.9407	0.8651	0.8659	0.850
Rmerge (%)	6.0	5.8	6.1	7.5	7.6	7.7
I/σI	3.4	3.4	4.5	3.1	3.1	3.1
Reflections	54292	70263	51260	34246	34259	34160
Redundancy	5.6	6.5	6.4	12.8	12.7	12.4
Completenes	99%	99%	100%	95%	95%	95%
No of sites				1	1	1
Phasing Power				Iso Ano	Iso Ano	Iso Ano
				NA 0.98	1.24 1.23	1.39 1.04
Rworking		22	24.5			
RFree		25	29			
Monomers / AU		1	1			
Non-H protein atoms		7521	7487			
Non-H ligand atoms		242	125			
Number of waters		375	280			
Ramachandran core		90.0%	84%			
Ramachandran allowed		10%	16%			