Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Co-crystallization of the human nuclear cap-binding complex with a m^{7}GpppG cap analogue using protein engineering

Catherine Mazza,[a] Alexandra Segref,[b] Iain W. Mattaj[b] and Stephen Cusack[a]*

[a]European Molecular Biology Laboratory, Grenoble Outstation, c/o ILL, BP 181, F-38042 Grenoble CEDEX 9, France, and [b]European Molecular Biology Laboratory, Gene Expression Programme, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Correspondence e-mail: cusack@embl-grenoble.fr

Received 24 June 2002
Accepted 23 August 2002

The nuclear cap-binding complex (CBC) binds the 7-methyl-G(5′)ppp(5′)N cap structure at the 5′ end of pre-messenger and uracil-rich small nuclear RNAs in the nucleus. It mediates interaction of these capped RNAs with various nuclear machineries involved in RNA maturation and is co-exported with them to the cytoplasm. The structure of human CBC, which comprises the subunits CBP20 and CBP80, has previously been determined in a mildly trypsinated form which can no longer bind the cap. Here, the engineering and crystallization of two variant CBCs with deletions in CBP80 which do not affect function are described. A complex with a small N-terminal deletion in CBP80 was crystallized in space group $C2$ with one complex per asymmetric unit. The crystals diffract to 2 Å resolution and give the first structure of intact but cap-free CBC. An additional internal deletion in CBP80 of a prominent solvent-exposed coiled coil gives rise to a more compact complex. This was co-crystallized with the cap analogue m^{7}GpppG in two different crystal forms which could grow in the same drop. Form 1 belongs to space group $P3_121$ with one complex per asymmetric unit and diffracts to 2.15 Å resolution. Form 2 belongs to space group $P2_12_12_1$ with two complexes per asymmetric unit and diffracts to 2.3 Å resolution. In both forms, strong extra electron density is observed for the cap analogue and for the N- and C-terminal extensions of CBP20 which was absent or disordered in all previous structures.

1. Introduction

The nuclear cap-binding complex (CBC) is a conserved eukaryotic heterodimeric protein complex which binds tightly to the 5′ cap structure [7-methyl-G(5′)ppp(5′)N or m^{7}GpppN, where N is any nucleotide] of nascent RNA polymerase II transcripts such as pre-messenger RNA and uracil-rich small nuclear RNA (U snRNA). It remains bound to these RNAs until their export into the cytoplasm and acts as a protein-identification tag which mediates recognition of capped RNAs by various components of the nuclear RNA-processing machinery (Lewis & Izaurralde, 1997). CBC enhances pre-mRNA processing by increasing the efficiency of the interaction between the U1 snRNP and the cap-proximal 5′ splice site during the initial steps of splicing (Izaurralde *et al.*, 1994) and also stabilizes the interaction of pre-mRNA with the 3′-end processing machinery, thus promoting the poly(A) site cleavage (Flaherty *et al.*, 1997). In multicellular organisms, where U snRNP assembly is partly cytoplasmic, CBC, together with an adaptor protein called PHAX (phosphorylated adaptor for RNA export; Izaurralde *et al.*, 1995; Ohno *et al.*, 2000), is essential for the nuclear export of U snRNA (Izaurralde *et al.*, 1995). PHAX binds to and stabilizes the CBC–U snRNA complex and when phosphorylated is also able to bind, *via* its nuclear-export sequence, to the nuclear-export receptor Crm1-RanGTP (Izaurralde *et al.*, 1995; Ohno *et al.*, 2000). We are interested in elucidating the structural basis of cap recognition by CBC as well as how CBC interacts with its different partners during the various steps of RNA maturation.

CBC is a heterodimer comprising CBP80 (790 residues) and CBP20, which must be associated to bind capped RNAs. CBP20, which is phylogenetically highly conserved, contains a central RNP domain, a common RNA-binding motif (Izaurralde *et al.*, 1995). The large subunit, CBP80, is far less conserved in evolution. We have purified human CBP20 (156 residues) and CBP80 (790 residues) and reconstituted the complex (hCBC). As our attempts to crystallize intact hCBC resulted only in very poor quality crystals, we performed mild trypsination, obtaining a stable particle that contains almost all CBP80, but in two pieces (residues 20–670 and 685–790), and the middle section of CBP20 (residues 22–76 and 79–120), which includes just the RNP domain (Mazza *et al.*, 2001). Even though this

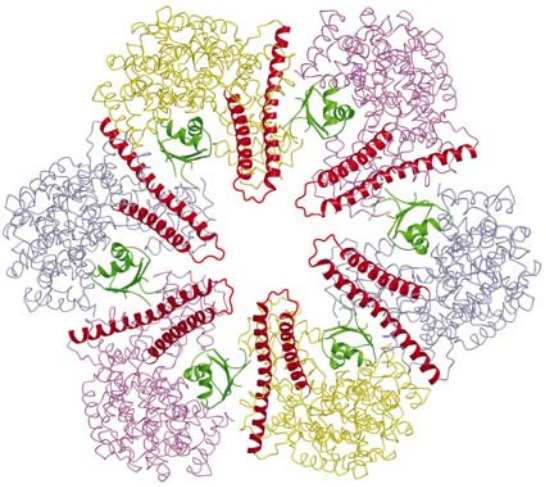

Figure 1
Crystal packing of the trypsinated CBC (Mazza *et al.*, 2001). The three CBP80 subunits from one asymmetric unit are depicted in pink, yellow and blue as well as their symmetry-related molecule. The small CBP20 subunit consisting of the RNP core (residues 38–118) is represented in green. The long protruding coiled coil (residues 635–731) from each CBP80 subunit is in red. The extremity of the coiled coil is composed of a 24-residue disordered loop which is cleaved by the trypsin treatment and is represented by a red loop in the figure. Figure generated with *BOBSCRIPT* (Esnouf, 1999) and *RENDER* (Merritt & Murphy, 1994).

complex is incapable of cap binding, it gave very good quality crystals diffracting to 2 Å resolution which allowed structure determination (Mazza *et al.*, 2001). We found that CBP80 consists of three structurally similar domains, each composed of five or six helical hairpin repeats and each resembling the so-called MIF4G domain, which occurs in the translation-initiation factor eIF4G and other proteins involved in RNA maturation (Marcotrigiano *et al.*, 2001; Ponting, 2000). The helical side of the RNP domain of CBP20 is tightly bound to the second and third domains of CBP80, leaving its β-sheet surface exposed to the solvent. By studying the cap-binding properties of a series of CBP20 alanine point mutants in intact CBC, we found that four distinct residues in the RNP β-sheet surface, Tyr43, Phe83, Phe85 and Asp116, were crucial for capped RNA binding (Mazza *et al.*, 2001). In two other cap-binding proteins where the structure of the complex with cap is known, the eukaryotic initiation factor eIF4E (Marcotrigiano *et al.*, 1997) and the vaccinia virus methyl-transferase VP39 (Hodel *et al.*, 1997, 1998), high specificity for the methylated base is achieved by sandwiching it between two aromatic residues. For CBC, we proposed Tyr43 as the bottom of a potential sandwich, but were not able to identify the other potential aromatic partner. If it exists it is likely to be on either the N- or C-terminal extensions to the currently visible part of CBP20. Extensive attempts to co-crystallize intact CBC with a cap analogue were unsuccessful. Using the known structure of trypsinated CBC, we therefore decided to engineer CBP80 in order to obtain a more compact form that might be easier to crystallize.

Here, we report the expression, purification and crystallization of two new hCBC complexes. In the CBCΔNLS complex, 19 N-terminal residues which include the nuclear localization sequence (NLS) have been removed from the CBP80 protein. The reconstituted complex has been crystallized, providing the first structure of intact cap-free hCBC. In the second complex, denoted CBCΔNLSΔCC, an additional truncation has been made of the long solvent-exposed coiled coil from the third MIF4G domain of CBP80 (Fig. 1). This complex has been co-crystallized with a cap analogue, m^7GpppG, in two crystal forms, giving the structure of cap-bound hCBC.

2. Experimental

2.1. CBP80 recloning, expression and purification

Two CBP80 deletion mutants were used in this study. In the first one, the first 19 N-terminus residues, which include the NLS sequence, were removed by PCR, resulting in the CBP80ΔNLS mutant. For the second one, a deletion of the amino acids 653–701, corresponding to the solvent-exposed part of the coiled coil, was performed using the Quick Change site-directed mutagenesis kit (Stratagene) with the oligonucleotides CAC GTG TTT GTT CAT CTT ACG AAT TGT AGA G and CAC GTG GGC GCT CAG AGT GAA CAA AAG AAT CTT. This resulted in the introduction of a glycine at position 653 followed by alanine 702 (mutant CBP80Δ653–701). A ΔNLS variant of this protein was then made by PCR, resulting in CBP80ΔNLSΔCC. The *Bam*HI–*Hin*dIII fragments containing the CBP80ΔNLS and CBP80ΔNLSΔCC mutants of hCBP80 were cloned into the corresponding sites of pFASTBac1 and recombinant baculoviruses were prepared according to the manufacturer's instructions (Life Technologies). Expression and purification of the two variant CBP80 proteins was performed as previously described for the full-length CBP80 (Mazza *et al.*, 2001), except that the concentration of NaCl was reduced to 120 m*M* in buffer 5 for the purification of the CBP80ΔNLS mutant on the heparin column.

2.2. CBP20 recloning, expression and purification

CBP20 was recloned between the *Nde*I and the *Hin*dIII sites of the pRSET A plasmid to remove the N-terminal His tag in the original construct and to allow expression of the wild-type protein sequence. This new construct was used to transform *Escherichia coli* BL21pLysS and the protein was expressed as described previously (Mazza *et al.*, 2001). The cleared lysate in buffer 1 [50 m*M* Tris pH 8.0, 50 m*M* NaCl, 10% glycerol and 5 m*M* β-mercaptoethanol and Complete protease-inhibitor cocktail (Roche Diagnostic)] was applied to a cation-exchange column (SP Sepharose fast flow, Pharmacia) and eluted with a linear gradient of 50–800 m*M* NaCl in buffer 2 (50 m*M* Tris pH 8.0, 10% glycerol, 5 m*M* DTT). The salt concentration was brought back to 50 m*M* and the protein was further purified on a heparin column (Pharmacia) and eluted with a linear gradient of 50–800 m*M* NaCl in buffer 2.

2.3. Cap-binding complex formation

CBCΔNLS and CBCΔNLSΔCC complexes were formed on ice by mixing the subunits in a 2:1 ratio of CBP20 to CBP80ΔNLS or CBP80ΔNLSΔCC mutant and the NaCl concentration was adjusted to 150 m*M*. After 30 min, the complexes were loaded onto a monoQ column (Pharmacia). Excess CBP20 was eliminated by washing the column with buffer 3 [50 m*M* HEPES pH 7.5, 150 m*M* NaCl, 10 m*M* dithiothreitol (DTT)]. The complexes were eluted, dialysed and concentrated to 10 mg ml^{-1} as previously described (Mazza *et al.*, 2001). 0.5 m*M* of cap analogue m^7GpppG was added to the CBCΔNLSΔCC complex and incubated for a minimum period of 30 min before performing crystallization trials.

2.4. Crystallization and data collection

Crystallization trials of the two complexes were performed using the hanging-drop vapour-diffusion method. A unique crystal of the CBCΔNLS complex was obtained at 277 K in a two-year-old drop containing equal volumes of the complex solution and of reservoir solution consisting of 10% polyethylene glycol 6000 and 100 m*M* MES

pH 6 (Fig. 2*a*). Analysis by SDS–PAGE showed that the drop containing this crystal still contained full-length CBP20. For data collection, the crystal was transferred into a cryoprotectant solution consisting of 100 m*M* NaCl, 20% glycerol, 15% polyethylene glycol 6000, 100 m*M* MES pH 6.0 and flash-frozen in liquid nitrogen. Native diffraction data were collected on an ADSC Quantum 4 CCD detector and processed with *XDS* (Kabsch, 1993).

Crystallization conditions for the CBCΔNLSΔCC/cap-analogue complex were screened employing sparse-matrix kits (Hampton Research, USA). Promising conditions were observed in condition 25 of the Natrix screen, which consists of 50 m*M* sodium cacodylate, 80 m*M* magnesium acetate and 30% PEG 4000. After optimization of the crystallization conditions, two crystal forms could be obtained at 277 K in drops containing equal volumes of the complex solution and of a reservoir solution consisting of 0.25–1% PEG 4000, 100 m*M* MES pH 6 and 75–100 m*M* magnesium formate. After one week, crystals reached a maximum of 0.1 × 0.1 × 0.1 mm for crystal form 1 and 0.4 × 0.03 × 0.03 mm for form 2 (Figs. 2*b* and 2*c*, respectively). As shown in Fig. 2(*b*), the two crystal forms could be observed in the same drop, but the proportion and the crystal size for each of the forms strongly depended on the PEG and magnesium formate concentrations. For data collection, crystals were transferred into a cryoprotectant solution consisting of 3% PEG 4000, 100 m*M* MES pH 6, 100 m*M* magnesium formate and 30% glycerol and flash-frozen in liquid nitrogen. Native diffraction data were collected on an ADSC Quantum 4 CCD detector and processed with *MOSFLM* (Leslie, 1999) and *SCALA* (Table 1).

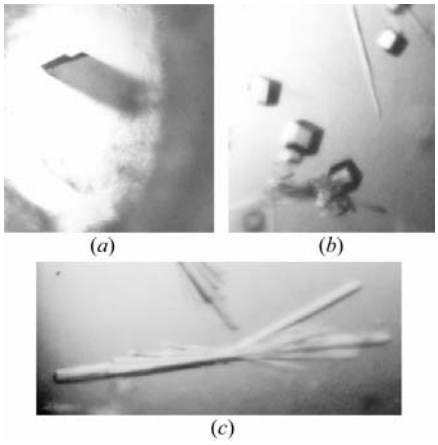

Figure 2
Crystals of CBC and its complex with the cap analogue. (*a*) CBCΔNLS, (*b*) CBCΔNLSΔCC with m^7GpppG form 1, (*c*) CBCΔNLSΔCC with m^7GpppG form 2.

Table 1
Diffraction data-collection statistics.

Values in parentheses are for reflections in the highest resolution bin.

		CBCΔNLSΔCC co-crystallized with m^7GpppG	
	CBCΔNLS	Crystal form 1	Crystal form 2
Wavelength (Å)	0.933	0.9393	0.9393
Space group	*C*2	$P3_121$	$P2_12_12_1$
Unit-cell parameters (Å, °)	*a* = 264.14, *b* = 59.61, *c* = 75.43, β = 99.52	*a* = *b* = 112.78, *c* = 158.31, γ = 120	*a* = 111.84, *b* = 125.72, *c* = 188.76
Mosaicity (°)	0.6	0.2	0.26
Oscillation range (°)	1	0.5	0.5
Complexes per asymmetric unit	1	1	2
Matthews coefficient ($Å^3$ Da^{-1})	2.7	2.9	3.3
Solvent content (%)	53.7	57.4	62.6
ESRF beamline	ID14eh2	ID14eh4	ID14eh4
Resolution	2.0	2.15	2.3
Total reflections	357588	446140	333168
Unique reflections	62416	63766	105598
Average redundancy	5.7	7.0	3.2
Completeness (%)	79.5 (58.8)	99.8 (99.8)	89.1 (65.4)
R_{merge}† (%)	5.3(47.9)	7.9 (39.8)	11.4 (45.2)

† $R_{sym} = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i (h)$, where $I_i(h)$ and $I(h)$ are the *i*th and mean measurements of the intensity of reflection *h*, respectively.

3. Results

Our attempts to crystallize intact hCBC in its cap-free form as well as in its cap-bound form remained unsuccessful for a long time. Only very thin plates of apo-CBC could be obtained, which diffracted to only 10 Å resolution. The use of limited proteolysis allowed the isolation of a more compact CBC which gave crystals of considerably better quality. After having solved the trypsinated structure of CBC at 2 Å resolution, we finally obtained one non-reproducible crystal of the intact CBCΔNLS complex. This crystal appeared after two years from a precipitate and diffracted to 2 Å resolution. The protein crystallized in space group *C*2 (unit-cell parameters *a* = 264.14, *b* = 59.61, *c* = 75.43 Å, β = 99.52°) with one complex per asymmetric unit (Table 1). The data are 91% complete to 2.3 Å resolution and 79.5% to 2.0 Å resolution. Data completeness was limited partially owing to the need to eliminate strong ice rings.

From the trypsinated CBC structure we could see that the C-terminal MIF4G-like domain of CBP80 contained a prominent solvent-exposed coiled coil which is orientated into a large solvent cavity in the crystal (Fig. 1). We hypothesized that this long projection restricts the packing of the protein in a crystal lattice. An internal deletion of CBP80 was thus made in order to remove 49 residues of the protruding part of this coiled coil. Residues Leu653–Ser701 were replaced by a glycine residue, resulting in the primary sequence Val652–Gly–Ala702. Elsewhere, we show that this deletion does not effect the cap-binding properties of CBC (Mazza *et al.*, 2002). Removal of this internal domain allowed the co-crystallization of CBC with the cap analogue in two different crystal forms. Crystal form 1 belonged to space group $P3_121$ (unit-cell parameters *a* = 112.8, *b* = 112.8, *c* = 158.3 Å, γ = 120°) with one complex per asymmetric unit and diffracted to 2.15 Å resolution. Crystal form 2 belonged to space group $P2_12_12_1$ (unit-cell parameters *a* = 111.8, *b* = 125.7, *c* = 188.8 Å, confirmed by molecular replacement) with two complexes per asymmetric unit and diffracted to 2.3 Å. Data-collection statistics are shown in Table 1.

These three structures were readily solved by molecular replacement with the program *MOLREP* (Collaborative Computational Project, Number 4, 1994) using the previously published trypsinated CBC structure as a search model (PDB code 1h6k; Mazza *et al.*, 2001). The correct solution in each case corresponded to the top peaks in both the rotation and the translation functions.

4. Discussion

In the CBCΔNLS complex, some additional positive density can be observed at both the N- and C-terminus of the RNP domain present in the structure of the trypsinated complex. In addition, the internal loop of the RNP domain (residues 73–81), previously cleaved by trypsin and leading to a distor-

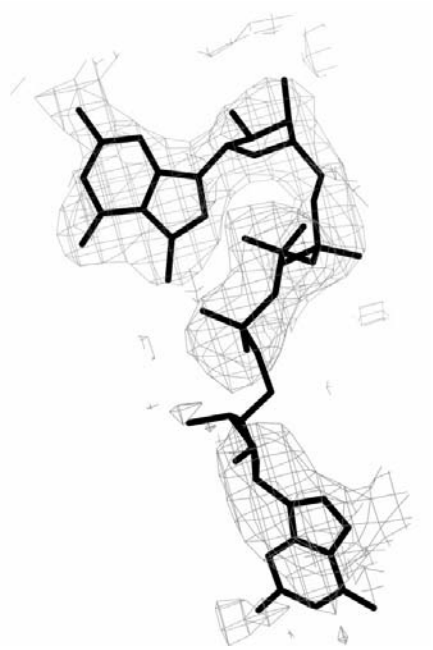

Figure 3
$2F_o - F_c$ experimental electron density for the cap analogue contoured at 1σ. The cap analogue, m^7GpppG, is placed in unbiased extra density calculated for crystal form 2 after molecular replacement. Figure generated with *BOBSCRIPT* (Esnouf, 1999).

tion of the β-sheet surface (Mazza *et al.*, 2001), is now intact. However, it is clear that there is not enough extra density to allow building of the majority of the 77 residues that are missing from the CBP20 subunit in the previously published structure. This implies that the N- and C-terminal domains of CBP20 remain largely disordered in the cap-free complex.

In contrast, in the CBCΔCC/cap complex the initial difference map shows extensive extra density above 3σ. Some of this could easily be attributed to the cap analogue (Fig. 3), whereas the remaining extra density, located around the cap analogue, should allow building the majority of the N- and C-terminal domains of CBP20. A full description of the structures will be given elsewhere (Mazza *et al.*, 2002)

Recognition of capped RNA by the nuclear CBC is one of the most crucial steps in mRNA and U snRNA maturation. Determination of the structure of the two complexes described here will help to elucidate the binding mechanism of capped RNAs by CBC. By comparison to the two other known cap-binding proteins, eIF4E and VP39, we hope to significantly advance the understanding of specific recognition of the methylated cap. More generally, the structures are a further step towards understanding the role of CBC in diverse aspects of RNA metabolism.

We thank members of the EMBL–ESRF Joint Structural Biology Group for access to ESRF beamline ID14 and to Kornelius Zeth for assistance in data processing.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W. & Gilmartin, G. M. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 11893–11898.
Hodel, A. E., Gershon, P. D. & Quiocho, F. A. (1998). *Mol. Cell*, **1**, 443–447.
Hodel, A. E., Gershon, P. D., Shi, X., Wang, S. M. & Quiocho, F. A. (1997). *Nature Struct. Biol.* **4**, 350–354.
Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C. & Mattaj, I. W. (1995). *Nature (London)*, **376**, 709–712.
Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. & Mattaj, I. W. (1994). *Cell*, **78**, 657–668.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
Lewis, J. D. & Izaurralde, E. (1997). *Eur. J. Biochem.* **247**, 461–469.
Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. (1997). *Cell*, **89**, 951–961.
Marcotrigiano, J., Lomakin, I. B., Sonenberg, N., Pestova, T. V., Hellen, C. U. & Burley, S. K. (2001). *Mol. Cell*, **7**, 193–203.
Mazza, C., Ohno, M., Segref, A., Mattaj, I. W. & Cusack, S. (2001). *Mol. Cell*, **8**, 383–396.
Mazza, C., Segref, A., Mattaj, I. W. & Cusack, S. (2002). In the press.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I. W. (2000). *Cell*, **101**, 187–198.
Ponting, C. P. (2000). *Trends Biochem. Sci.* **25**, 423–426.