Crystal structure of the 25 kDa subunit of human cleavage factor Im

Molly Coseno1, Georges Martin2, Christopher Berger3, Gregory Gilmartin1, Walter Keller2 and Sylvie Doublie1,*

1Department of Microbiology and Department of Molecular Genetics, University of Vermont, Burlington, VT 05405, USA, 2Department of Cell Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland and 3Molecular Physiology and Biophysics, University of Vermont, Burlington, VT 05405, USA

Received November 2, 2007; Revised February 8, 2008; Accepted February 9, 2008

ABSTRACT

Cleavage factor Im is an essential component of the pre-messenger RNA 3’-end processing machinery in higher eukaryotes, participating in both the polyadenylation and cleavage steps. Cleavage factor Im is an oligomer composed of a small 25 kDa subunit (CF Im25) and a variable larger subunit of either 59, 68 or 72 kDa. The small subunit also interacts with RNA, poly(A) polymerase, and the nuclear poly(A)-binding protein. These protein–protein interactions are thought to be facilitated by the Nudix domain of CF Im25, a hydrolase motif with a characteristic α/β/α fold and a conserved catalytic sequence or Nudix box. We present here the crystal structures of human CF Im25 in its free and diadenosine tetraphosphate (Ap4A) bound forms at 1.85 and 1.80 Å, respectively. CF Im25 crystallizes as a dimer and presents the classical Nudix fold. Results from crystallographic and biochemical experiments suggest that CF Im25 makes use of its Nudix fold to bind but not hydrolyze ATP and Ap4A. The complex and apo protein structures provide insight into the active oligomeric state of CF Im and suggest a possible role of nucleotide binding in either the polyadenylation and/or cleavage steps of pre-messenger RNA 3’-end processing.

INTRODUCTION

Pre-messenger RNA 3’-end processing in eukaryotes is a two-step reaction consisting of endonucleolytic cleavage of the pre-mRNA followed by addition of a poly(A) tail at the 3’ end of the upstream cleavage product (1–3). The coupling of these processing reactions relies on multiple protein–protein and protein–RNA interactions. The factors that are necessary and sufficient to reconstitute cleavage and polyadenylation in a mammalian in vitro system are poly(A) polymerase (PAP), cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor Im (CF Im), cleavage factor II m (CF II m) and the nuclear poly(A)-binding protein 1 (PABPN1). The recruitment of the mammalian polyadenylation machinery to the pre-mRNA relies on the recognition of conserved sequence elements, such as the highly conserved hexamer AAUAAA recognized by CPSF and a U-rich sequence downstream of the cleavage site recognized by CstF. Additionally a set of UGUA elements, a third sequence element found upstream of the cleavage site and not as universally conserved, is recognized by CF Im (4,5). Upon completion of the processing steps the mRNA can be efficiently transported out of the nucleus into the cytoplasm. Homologs of CF I subunits are also found in cDNA databases of Caenorhabditis elegans, Drosophila melanogaster and plants.

CF Im is an oligomer composed of a small 25 kDa subunit (CF Im25, also referred to as CPSF5 or NUDT21) and a larger subunit of either 59, 68 or 72 kDa (6). The three larger subunits share substantial sequence homology. They are encoded on two different genes and the 72 kDa subunit is a splice variant of the 68 kDa polypeptide. CF Im can be reconstituted in vitro from the 25- and 68 kDa subunits (6). CF Im, in addition to having a role in regulation of poly(A) site recognition, is also involved in stabilizing CPSF at the conserved hexamer, enhances the rate of poly(A) site cleavage in vitro, and has been shown to interact with PAP and PABPN1 via its 25 kDa subunit (7). Furthermore CF Im also interacts with splicing factors, indicating a role in communicating between different RNA-processing complexes (8).

Much of what we know about complex formation of processing factors on the pre-mRNA and its regulation...
Figure 1. Sequence alignment of CF Im25 with Nudix proteins. ClustalW sequence alignment of CF Im25 and Nudix proteins of known structure (44). PDB ID codes are shown to the right of the enzyme names followed by the residue number of the first amino acid. The position of the Nudix box is indicated below the alignment as a grey/black bar where the black part marks the position of helix α2. Residues are on a light blue background if over 70% conserved and are on yellow background if invariant. The two catalytic glutamates conserved in most Nudix enzymes are displayed in red font on light orange background. L124 and I128, which are found in place of the conserved glutamates, are boxed in the CF Im25 sequence. The abbreviations are defined as: PPHase for pyrophosphohydrolase and PPase for pyrophosphatase.

has been investigated biochemically. Structural characterization of individual and multicomponent processing factors will elucidate the domain interactions important for the 3′ pre-mRNA processing mechanism that cannot be deciphered by biochemical means. The structure of the 25 kDa subunit of CF Im (CF Im25) will help further define the domains important for substrate and protein interactions.

The structure of CF Im25 will also allow us to investigate the function of a Nudix domain present in this protein. The first Nudix protein to be characterized enzymatically and structurally was Escherichia coli MutT (9). Nudix proteins are generally characterized as housekeeping enzymes due to their role in hydrolysis of substrates described as nucleoside diphosphate linked to another moiety X, many of which are potentially toxic molecules (10,11). Nudix proteins have a conserved Nudix fold consisting of an α/β/α sandwich. Within the Nudix fold the consensus sequence of the Nudix box is GX5EX7REUXEEXGU, where U is a hydrophobic residue and X is any residue, and folds into a loop–helix–loop structure (9). Interestingly, the Nudix box in CF Im25 lacks two of the four conserved glutamate residues, of which were shown to be important for catalysis (Figure 1) (12).

In this study we present the crystal structure of human CF Im25 alone and in complex with diadenosine tetraphosphate (Ap4A). CF Im25 crystallizes as a dimer, which is also the oligomeric state of the protein in solution. The CF Im25 structure presents the classic Nudix α/β/α fold and harbors residues outside of the Nudix core that could poteniate ligand binding. Structural and biochemical evidence suggests that CF Im25 binds, but does not hydrolyze, mono and di-adenosine nucleotides.

**MATERIALS AND METHODS**

**Protein purification**

The construction of the plasmid expressing the 25 kDa subunit of human cleavage factor Im (CF Im) with an N-terminal 6xHis tag was previously described (7). The cDNA of CF Im25 was subsequently cloned into a Gateway® vector with a dual 6xHis-maltose binding protein (MBP) affinity tag provided by Dr David S. Waugh (National Cancer Institute, Frederick, MD) (13). The following primers were used in the polymerase chain reaction (PCR): 5′-GAG AAC CTG TAC TTC CAG GGT ATG TCT GTG GTA CCG CCC-3′ and 5′-GGG GAC CAC TTT GTA CAA GGA AAC TGG GTT ATT AGT TGT AAA TAA AAT TGA A-3′. The cDNA was then expressed in Rosetta (DE3) pLysS cells (Novagen) and grown in LB medium for 24 h at 25°C following induction with 0.4 mM IPTG. The protein was purified on Nickel NTA beads (Qiagen), followed by a Tobacco etch virus (Tev) protease cleavage step, and cation exchange on a Resource S column (GE Healthcare). Cells were lysed at 4°C by sonication in a buffer containing 20 mM Tris–HCl pH 8.0, 200 mM NaCl, and a protease inhibitor tablet (Roche). The lysate was centrifuged at 12 000 × g and incubated with Ni-NTA beads (Qiagen) for 1 h at 4°C. Protein was eluted with lysis buffer containing 100–500 mM imidazole. Pooled fractions were dialyzed into 10% (v/v) glycerol, 20 mM Tris–HCl pH 7.5 and 50 mM KCl. CF Im25 was cleaved from the MBP tag by the addition of equimolar amounts of Tev protease. MBP and Tev were then separated from CF Im25 by elution with a 50 mM–1 M KCl gradient at pH 7.5.

**Crystallization**

CF Im25 crystals were initially obtained with the sitting drop method in a 96-well tray format. Sitting drops were set up with a 925 PC Workstation (Gilson) by mixing 1 μl of 25 μM stock CF Im25 in 0.1 M Tris–HCl pH 80, 0.8 M ammonium sulfate, 1 mM DTT, and 2 μl of 40% (v/v) reservoir solution containing 30% (v/v) PEG 3350, 100 mM Tris–HCl pH 7.5, 20% (v/v) glycerol, and 10% (v/v) Ficoll 400. Crystals were soaked in 10% (v/v) glycerol for 2 h at 4°C and flash frozen in liquid nitrogen.
0.6 µl of protein (16 mg/ml) with 0.6 µl reservoir solution (25% w/v PEG 3350, 0.2 M MgCl₂, 0.1 M Tris–HCl pH 8.5) (Hampton Research Index Screen, condition 85) and equilibrated against 160 µl reservoir buffer. Subsequently, crystals were obtained by streak seeding hanging drops with a protein concentration of 6 mg/ml. Hanging drops were set up by mixing 1 µl of protein with 1 µl of reservoir solution under the same crystallization conditions. Trigonal crystals grew to a maximum length of 200 µm in space group P3₁21 (P3₂21) with unit cell parameters a = b = 80.11 Å, c = 72.21 Å and γ = 120°. There is one molecule per asymmetric unit with an estimated solvent content of 52%. Crystals of the complex with Ap₄A were obtained with the hanging drop method. The hanging drops were set up manually by mixing 1 µl of protein with 1 µl of reservoir solution (25% w/v PEG 3350, 0.025 M MgCl₂, 0.1 M Tris–HCl pH 7.5). The drops were streak seeded after a 24 h incubation period. When the crystals reached at least 100 µm, Ap₄A at a final concentration of 44 mM was added directly to the drop. After 6 h the soaked Ap₄A crystals were cryoprotected by the addition of 1 µl of 25% (w/v) PEG 3350 and 50% (v/v) glycerol to the 2 µl hanging drop prior to flash cooling in liquid nitrogen.

### Crystallographic data collection

Multiple wavelength anomalous diffraction (MAD) data were collected at beamline 23-ID-D (Advanced Photon Source at Argonne National Laboratory) on a MAR m300 CCD detector. One complete selenomethionyl MAD dataset was collected on one crystal at the peak, inflection and high-energy remote wavelengths to a maximum resolution of 1.85 Å. Data were collected at 1.80 Å resolution on the Ap₄A complex at beamline X12B (National Synchrotron Light Source) on a Quantum-4 CCD (ADSC) detector. The data from three Ap₄A-soaked crystals were merged to increase redundancy. Diffraction data were processed and scaled with DENZO and SCALEPACK (15). Data collection statistics are summarized in Table 1.

### Structure determination and refinement

The program SOLVE (16) identified three of the four selenium sites (the N-terminal methionine is disordered or...
missing). AutoSHARP (17) was then used for refinement of the selenium parameters. The space group was judged to be P3₁2₁ and not the enantiomorphic P3₂2₁, based on the map quality and continuity. The phasing information was then used in RESOLVE (18) for density modification and iterative model building. Seventy percent of the model was built by RESOLVE. The remaining residues were built manually using the program COOT (19). Residues 1–20 and 132–135 were omitted from the model because of poorly defined density.

Iterative rounds of refinement including simulated annealing, energy minimization, and B-factor refinement were done with CNS (20). Each refinement round was followed by rebuilding in COOT. A composite simulated annealing omit map was generated in CNS to validate the model and build the remaining side chains. Water molecules were added with CNS and COOT. The quality of the model was evaluated with PROCHECK (21). All non-glycine residues fall within either the most favored or additionally allowed regions of the Ramachandran plot. The refined model of the unliganded protein provided phases to calculate an isomorphous difference Fourier (Fo–Fo) map between the unliganded protein and the complex with Ap₄A (22). The R_crooss on amplitudes between the two datasets is 0.149, indicating good isomorphism between the two crystals. The resulting map showed clear density for one adenine base and three phosphates of Ap₄A. The complex with Ap₄A was refined with CNS (20). The refinement statistics for both structures are reported in Table 1. Figures were drawn with PyMOL (23).

**Oligomeric state determination**

Size exclusion chromatography was performed with a Superdex 75 column (GE Healthcare). The protein sample or molecular mass standards were applied to the Superdex 75 column and eluted with 10% glycerol, 20 mM Tris–HCl, pH 8.5, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, 20 μM Ap₄A, and 1 U calf intestinal phosphatase (New England Biolabs) or 1 U S. cerevisiae pyrophosphatase (Sigma), and 2 or 4 mM substrate in a total volume of 50 μl. The putative substrates tested for CF Im₂₅ hydrolyase activity include ATP, Ap₄A (Sigma) and m⁷G(₅)ppp(₅)G cap structure analog (New England Biolabs). The reactants were combined at 4°C then incubated at 25°C for 30 and 60 min. The reaction was terminated by the addition of 250 μl of 20 mM EDTA and the liberated orthophosphate was determined by the colorimetric assay of Ames and Dubin (9,25). The limit of detection of this assay is 5 μM of orthophosphate.

**Hydrolase activity assay**

The standard reaction mixture contained 50 mM Tris–HCl, pH 7.5 or 8.5, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, 20 μM CF Im₂₅, 2 U calf intestinal phosphatase (New England Biolabs) or 1 U S. cerevisiae pyrophosphatase (Sigma), and 2 or 4 mM substrate in a total volume of 50 μl. The putative substrates tested for CF Im₂₅ hydrolyase activity include ATP, Ap₄A (Sigma) and m⁷G(₅)ppp(₅)G cap structure analog (New England Biolabs). The reactants were combined at 4°C then incubated at 25°C for 30 and 60 min. The reaction was terminated by the addition of 250 μl of 20 mM EDTA and the liberated orthophosphate was determined by the colorimetric assay of Ames and Dubin (9,25). The limit of detection of this assay is 5 μM of orthophosphate.

**Fluorescence measurements**

Steady-state tryptophan fluorescence was measured with a Quantamaster fluorimeter (Photon Technology International, South Brunswick, NJ) as described (26) with a WG320 cut-off emission filter. CF Im₂₅ contains four tryptophans and only three are built in the structure: Trp148 and Trp149 are within 10–15 Å of the active site and Trp139 is located within 20 Å of the active site. The fourth tryptophan, Trp13, is located in the disordered portion of the amino terminus. The tryptophan emission spectrum was measured by excitation of the samples at 295 nm and collecting the emitted fluorescence at 90° to the incident light over the range 300–400 nm. The slit widths were set at a resolution of 1 nm for excitation and 4 nm for emission. Fluorescence measurements of all protein samples were performed using a microcuvette with a magnetic stir bar in 20 mM Tris–HCl pH 7.5, 50 mM KCl and 25 mM MgCl₂ at 25°C for the protein alone and in the presence of increasing amounts of nucleotide. All fluorescence measurements were corrected for Raman scatter and background fluorescence and represent experiments performed in triplicate and then normalized and averaged. The ATP fluorescence data were fit with a single hyperbola \[ y = ax/(b + x) \] with a \( K_d \) of 1.53 ± 0.18 μM (1.17–1.89 μM at 95% confidence interval). The Ap₄A data were fitted with a single hyperbola \[ y = ax/(b + x) \] with a \( K_d \) of 2.44 ± 0.49 μM (1.46–3.43 μM at 95% confidence interval).

**RESULTS**

**Crystal structure of human CF Im₂₅**

The original N-terminal His-tagged plasmid of human CF Im₂₅ (7) did not express to high enough levels for structural studies. We therefore inserted the coding sequence of CF Im₂₅ into a dual HisMBP vector (13). With the resulting expression vector, 1.5 mg of protein could be purified from 1 l of culture. The structure was solved to a resolution of 1.85 Å by multiple wavelength anomalous diffraction of the selenomethionyl protein variant. Residues 21–131 and 136–227 are visible in the electron density map. A complex with Ap₄A was also obtained and refined to a resolution of 1.80 Å.

**Description of the structure**

Human CF Im₂₅ is composed of 227 residues, with a calculated molecular weight of 26 kDa. CF Im₂₅ elutes as a dimer in gel exclusion chromatography with an apparent molecular weight of ~53 kDa (Supplementary Figure). The dimeric state of CF Im₂₅ has also been confirmed by dynamic light scattering. In the crystal structure, dimer formation relates two monomers by a 2-fold crystallographic axis. The Nudix domain is located in the middle of the protein and spans residues 77–202 (Figure 2). The Nudix fold comprises two mixed β sheets flanked by two helices (α₂ and α₃). The Nudix box is located in helix α₂.
Figure 2. Domain organization of CF I_m25. Ribbon diagram of the CF I_m25 dimer comprising residues 21–131 and 136–227. The ribbon color scheme corresponds to that of the domain architecture shown below. The second monomer on the left is shown in light grey. The color scheme corresponds to that of the domain architecture shown below. The second monomer on the left is shown in light grey. The secondary structure numbering is based on a DSSP analysis (45).

Metal binding

CF I_m25 is missing the second and fourth conserved glutamates of the Nudix box (residues 124 and 128), which are replaced by a leucine and isoleucine, respectively. These glutamate residues are very often involved in metal binding (12). We therefore set out to investigate whether CF I_m25 was still capable of binding metal via the two remaining carboxylates. Although the CF I_m25 crystals were obtained in the presence of 200 mM MgCl_2, there was no identifiable Mg^{2+} bound in the electron density map. Magnesium is a light atom, which is usually not easily identifiable in electron density maps and can often be mistaken for a water molecule. We therefore used metals that are more electron dense such as MnCl_2 and GdCl_3 to identify putative metal site(s) (33). Mn^{2+} and Gd^{3+} present the added advantage that they are anomalous scattering atoms, which should allow unambiguous identification of metal sites. We were unable to identify binding of either metal, regardless of whether the metal was co-crystallized or soaked into pre-existing crystals. Since Nudix enzymes require a divalent cation for catalysis (9), these experiments suggested that it is unlikely that CF I_m25 functions as a hydrolase. This finding called for further investigation of CF I_m25’s binding to potential substrates (see below).

Enzymatic assays and substrate binding

We next set out to investigate whether CF I_m25 binds a substrate. We used three different methods: a colorimetric
A colorimetric assay to measure the release of inorganic phosphate, and thus the potential hydrolytic activity of CF I\textsubscript{m}25 on nucleotides, was performed (25). The following putative substrates were tested: ATP, Ap\textsubscript{4}A and phosphate, and thus the potential hydrolytic activity of CF I\textsubscript{m}25 bind Ap\textsubscript{4}A (12,30). ATP and the cap analog were tested because of their prominent role in RNA processing. The assay indicated that, within the limits of detection of the assay, none of the putative substrates tested were hydrolyzed.

Although CF I\textsubscript{m}25 does not seem to possess a hydrolytic activity, it could still bind nucleotides. Several putative ligands were either co-crystallized or soaked into the crystals: ATP, GTP, ADP, GDP, diadenosine triphosphate (Ap\textsubscript{4}A), Ap\textsubscript{4}A, AMP, 7mG(5\textsuperscript{0})ppp(5\textsuperscript{0})G cap analog, NAD\textsuperscript{+} and GDP-mannose. In addition to testing these nucleotides, which are known substrates for Nudix enzymes, we also attempted to co-crystallize CF I\textsubscript{m}25 with a 21mer RNA derived from the PAP\(\gamma\) cDNA sequence (4,5). Of all the putative ligands tested, only Ap\textsubscript{4}A bound to the crystal (see later). The original crystallization conditions contained 200 mM MgCl\textsubscript{2} and 25% (w/v) polyethylene glycol (PEG) 3350. Because nucleotides can precipitate in the presence of high concentrations of PEG and magnesium (34), care was taken to modify the crystallization conditions so as to decrease or completely eliminate MgCl\textsubscript{2} and therefore lessen the risk of the nucleotide precipitating out of solution. Conditions with NaCl or even no salt could be used in lieu of MgCl\textsubscript{2} to grow unliganded crystals. Even when the divalent cation was omitted from the crystallization solution crystals did not form with any of the ligands tested (with the notable exception of Ap\textsubscript{4}A), demonstrating that the lack of binding was not due to the nucleotide falling out of solution.

Binding studies were then performed employing the intrinsic tryptophan fluorescence properties of CF I\textsubscript{m}25. The following Nudix ligands were tested: ATP, GTP, Ap\textsubscript{4}A, inositol hexaphosphate (IP\textsubscript{6}), 7mG(5\textsuperscript{0})ppp(5\textsuperscript{0})G cap analog, ADP and AMP. Dissociation constants (\(K_d\)) measured by titrating the ligand concentration were estimated for ATP and Ap\textsubscript{4}A (Figure 3A and B, respectively). The ATP data were fit with a single hyperbola curve, indicating one binding affinity with a \(K_d\) value of 1.53 \pm 0.18 mM. The Ap\textsubscript{4}A-binding data were fit to a single hyperbola curve and represent one binding site per monomer, resulting in a \(K_d\) of 2.44 \pm 0.49 \mu M.

**Figure 3.** CF I\textsubscript{m}25 steady-state tryptophan fluorescence experiments with ATP and Ap\textsubscript{4}A. All data are represented as normalized and averaged experiments done in triplicate. Error bars that are not represented lie within the symbol. (A) ATP steady-state tryptophan fluorescence data are fit with a single hyperbola, with \(K_d = 1.53\) mM. (B) Ap\textsubscript{4}A steady-state tryptophan fluorescence data are fit with a single hyperbola, with \(K_d = 2.44\) mM.

The adenine bases and three of the four phosphates of Ap\textsubscript{4}A, in the cavity of CF I\textsubscript{m}25 (Figure 4B). The rmsd between the bound and apo structures is low (0.41 Å), indicating that only very small changes take place upon substrate binding (20). The binding site residues Arg63, Arg150, Gln157 and Lys172 are found outside of the nucleotide moiety, and are involved in coordinating the triphosphate moiety (Figure 4B). The binding site is composed of residues from helix \(\alpha 1\), beta strand \(\beta 6\), the loop linking \(\beta 6\) and \(\beta 7\) and beta strand \(\beta 7\). In CF I\textsubscript{m}25, Arg63 and Arg150 are highly conserved across species while Gln157 and Lys172 are moderately conserved. The majority of hydrogen-bonding interactions with Ap\textsubscript{4}A involve the oxygens of the \(\beta\) and \(\gamma\) phosphates. The ordered adenine base of the Ap\textsubscript{4}A molecule stacks with Phe103, a residue contained within the Nudix domain. We note that the position of the \(\gamma\) phosphate of Ap\textsubscript{4}A coincides with the position of a sulfate ion reported for CF I\textsubscript{m}25 (Structural Genomics Consortium, Karolinska Institute; PDB ID code 2J8Q).

CF I\textsubscript{m}25 has a core structure similar to that of the *D. radiodurans* Nudix protein DR1025 (12) (PDB ID code 1SU2) (Figure 5). Variations between these two structures arise from an extension in the loop linking \(\beta 6\) and \(\beta 7\), a shortening of the loop following \(\beta 4\) and an additional \(\alpha\)-helix (helix \(\alpha 1\)) in the CF I\textsubscript{m}25 structure. The additional \(\alpha\)-helix in CF I\textsubscript{m}25 plays a role in sequestering the...
substrate binding pocket from solvent exposure. When comparing the positions of the bound nucleotides between CF I_m25 and DR1025 we also see variations in substrate fit. Superposition of the ATP bound in DR1025 on to the CF I_m25 structure shows that both the base and phosphate tail would clash with protein residues in helix a1 and beta strand b5 in the CF I_m25 structure (Figure 6).

Commonly, variations in the nucleotide-binding region of Nudix proteins occur due to differences in the side chains and motifs and contribute to the substrate specificity of the Nudix protein. Interestingly, both DR1025 and CF I_m25 stabilize the adenine base via stacking interactions with a phenylalanine residue found outside of the Nudix box region. Base-stacking interactions via a Tyr or Phe located 17 amino acids downstream of the Nudix box is commonly found in all of the Nudix Ap4A hydrolases, including DR1025 (11). Phe103 stabilizes the adenine base in the Ap4A bound CF I_m25 structure but is located six residues upstream of the Nudix box region.

DISCUSSION

In our crystal structure CF I_m25 is a homodimer. A dimeric state for CF I_m25 is consistent with dynamic light scattering (DLS) and gel filtration experiments performed with the 25 kDa subunit both unliganded and with ATP or Ap4A. The dimeric structure of CF I_m25 suggests that the active form of CF I_m may be a heterotrimer composed of a CF I_m25 homodimer and either the 59, 68 or 72 kDa subunit. This is consistent with interaction studies in C. elegans (DIP interaction database accession DIP:25083N) where the CF I_m25 homolog (Uniprot accession Q93716) was
found to interact with itself and with the CF Im25 homolog (Uniprot accession Q18937). Alternatively, the complex with the larger subunit of CF Im could be a heterotetramer. A third possibility is that CF Im25 monomerizes upon binding the larger subunit.

The Nudix box of CF Im25 lacks two of the four glutamates important for catalytic function and metal binding. Our structural and biochemical characterization suggests that CF Im25 is able to bind but not hydrolyze nucleotide substrates. In the dinucleotide bound CF Im25 structure, the dinucleotide is found outside of the Nudix box and lies deeper within the active site compared to ATP in DR1025. The fluorescence-binding data for CF Im25, indicate a binding affinity in the low micromolar range for Ap4A indicative of a potential role as a signaling molecule and a weak, albeit physiologically relevant, binding affinity for ATP. Even though CF Im25 binds both Ap4A and ATP, we have so far only been able to obtain a co-crystal complex with Ap4A. It is entirely possible that the crystallization conditions we have explored to date for the protein–ATP complex are not compatible with the formation of a crystal lattice and that further exploration of the crystallization space might yield the desired conditions. We note that pre-formed crystals of CF Im25 dissolve upon addition of ATP, indicating a possible conformation change upon binding of the nucleotide.

To our knowledge, CF Im25 is the first example of a Nudix protein binding and not hydrolyzing a nucleotide substrate. This loss of function/gain of a regulatory role is not unprecedented in evolution. A similar loss of function was reported for the *Lactobacillus lactis* ATP phosphoribosyl transferase (ATP-PRT) regulatory subunit, His Z (36). ATP-PRT functions to initiate the biosynthesis of histidine and requires both the HisZ subunit and HisG, the catalytic subunit, for activity. The regulatory subunit, HisZ, resembles the catalytic domain of functional histidyl-tRNA synthetases (HisRS) and utilizes its fold, not for catalysis, but for binding of histidine to monitor histidine levels. Although there were no metals evident in either the free or bound structures of CF Im25 and no identifiable hydrolytic activity in our colorimetric assay it remains possible that the Nudix box of CF Im25 could potentiate hydrolysis of Ap4A and ATP upon interaction of CF Im25 with one of the larger CF Im subunits or additional binding partners.

Ap4A belongs to the family of diadenosine oligophosphates, ApnA, which were first discovered 40 years ago (37,38). Ap4A is composed of two adenosines and four phosphates linked in 5′ to 5′-phosphodiester linkages. The role of ApnAs in higher eukaryotes has remained elusive. Recently, ApnAs have been suggested to play a role as putative extra- and intracellular signaling molecules. ApA itself is involved in the cellular stress response, inhibition of KATP channels, stimulation of DNA replication and repair, as well as influencing other essential cellular processes in eukaryotes (37,38). Presently, the only link between ApA and 3′-end processing was described in yeast where stimulation of primer independent synthesis by yeast poly(A) polymerase was observed in the presence of dinucleoside polyphosphates, including ApA (39). There is no CF Im25 homolog in yeast but this observation still suggests a potential role for dinucleoside polyphosphates as signaling molecules during RNA-processing events. The concentration of ATP and ApA can fluctuate within the cell in response to cellular stress or growth.
Preliminary experiments on the role of Ap4A during the polyadenylation step of 3'-end processing resulted in a non-competitive inhibition of polyadenylation in a poly (A) extension assay with mammalian PAP (results not shown). This suggests that under conditions of high concentrations of Ap4A such as stress, Ap4A can bind to a site in PAP, other than the ATP binding site, to inhibit polyadenylation.

The residues 81–160 of the Nudix domain play a dual role, binding RNA and stabilizing protein–protein interactions. Additionally, the amino terminus (1–76) participates in RNA binding (Figure 2) (7). A stable interaction between CF Im25 and CF Im68, unlike that involving PAP or PABPN1, requires the entire CF Im25 protein. CF Im68 must contact regions outside of the known RNA-binding region of CF Im25 to promote complex formation and stimulate pre-mRNA 3'-end processing. Interestingly, CF Im25 has a patch of negatively charged surface residues that runs the length of the dimer interface (Figure 7A). This charged region is composed of residues primarily from beta strand β and helix α4 and could potentiate protein–protein interactions with CF Im25's other binding partners, possibly CF Im68. Also intriguing is the observation that a binding interaction between the substrate RNA and the 25 kDa subunit occurs in the absence of a putative RNA-binding domain. This suggests another mechanism of RNA recognition, possibly through homodimer formation. A CF Im25 homodimer may enhance the binding potential to the RNA substrate compared to a monomer interaction by increased surface area interactions. A definite answer regarding the oligomeric state of the 25 kDa subunit and the RNA-binding mechanism will have to await a CF Im25 structure with RNA bound.

The results reported here suggest that CF Im25 is unable to hydrolyze nucleotides or dinucleotides even though it harbors a classic Nudix fold. The Nudix domain of CF Im25 may instead facilitate protein–protein interactions, as suggested by the large distribution of charged residues in the electrostatic surface representation of the Nudix domain. This charge distribution of the Nudix domain correlates well with results from pull-down experiments with PAP and PABPN1 (7). The interaction of CF Im with the RNA substrate and with PAP stimulates the rate of polyadenylate tail synthesis. This may be facilitated by the binding of ATP to CF Im25. Preliminary data (S. Dettwiler and W. Keller, unpublished) showed that CF Im25 interacts with hClp1, a protein shown to bind ATP which is involved in 3'–pre-mRNA processing and tRNA splicing (40,41). This observation, in conjunction with the fact that hClp1 has recently been shown to function as an siRNA kinase and a kinase that phosphorylates the 5'-end of the 3'-splicing product in human tRNA splicing (42), suggests a possible link between CF Im25's binding of ATP and protein–protein cross talk. Recently, CF Im25 has also been shown to be associated in a large RNP complex with Rae1, an mRNA export protein, in the nucleation and stabilization of microtubules during spindle assembly (43). This interaction is via CF Im25 association with an RNA component of the RNP complex and suggests a role for CF Im25 in mRNA export via direct RNA association that may be influenced by the concentration of intracellular nucleotides or dinucleotides.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

This work was supported by National Institutes of Health grant GM62239 to SD and a Department of Energy Experimental program to Stimulate Competitive Research predoctoral fellowship to M.C. Work in the laboratory of W.K. was supported by the University of Basel and the Swiss National Science Foundation. The beamline at Advanced Photon Source at Argonne National Laboratory (GM/CA-CAT) has been funded in whole or in part with federal funds from National Cancer Institute Grant Y1-CO-1020 and National Institute of General Medical Science Grant Y1-GM-1104. The beamline at the National Synchrotron Light Source, Brookhaven National Laboratory, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Sciences, under Contract No. DE-AC02-98CH10886. We are grateful to Dr David Waugh (NIH) for his generous gift of the dual His-MBP vector and the Tey protease plasmid. We thank Justin Meyette for helpful discussions concerning protein purification and Dr Mark A. Rould, Dr Pierre Aller, Dylan Murray and Karl Zahn for help with data collection and refinement, and Dr Mark A. Rould for help with troubleshooting the fluorescence experiments. Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank with entry codes 3BAP and 3BHO for the unliganded and Ap4A bound structures, respectively. Funding to pay the Open Access publication charges for this article was provided by NIH.

**Conflict of interest statement.** None declared.

**REFERENCES**

1. Zhao, J., Hyman, L. and Moore, C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, 63, 405–445.

2. Wahle, E. and Rüegsegger, U. (1999) 3' End processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.*, 23, 277–295.

3. Minville-Sebastia, L. and Keller, W. (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.*, 11, 352–357.

4. Brown, K.M. and Gilmartin, G.M. (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol. Cell.*, 12, 1467–1476.

5. Venkataraman, K., Brown, K.M. and Gilmartin, G.M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.*, 19, 1315–1327.

6. Rüegsegger, U., Beyer, K. and Keller, W. (1996) Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J. Biol. Chem.*, 271, 6107–6113.

7. Dettwiler, S., Arighieri, C., Cardinale, S., Keller, W. and Barabino, S.M. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding.
protein-protein interactions, and subcellular localization. J. Biol. Chem., 279, 35788–35797.
8. Millevoi,S., Loulergue,C., Dettwiler,S., Karaa,S.Z., Keller,W., Antoniou,M. and Vagner,S. (2006) An interaction between U2AF 65 and CF (m) links the splicing and 3' end processing machineries. EMBO J., 25, 4854–4864.
9. Bessman,M.J., Frick,D.N. and O’Handley,S.F. (1996) The MutT proteins or “Nudix” hydrolases, a family of versatile, widely distributed, “housecleaning” enzymes. J. Biol. Chem., 271, 25059–25062.
10. Mildvan,A.S., Xia,Z., Azurmendi,H.F., Saraswat,V., Legler,P.M., McLennan,A.G. (2006) The Nudix hydrolase superfamily. Cell. Mol. Life Sci., 63, 123–143.
11. McLennan,A.G. (2006) The Nudix hydrolase superfamily. Acta Crystallogr. D Biol. Crystallogr., 60, 2126–2132.
12. Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr., 55, 849–861.
13. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) In Doublé,S. (ed.), Macromolecular Crystallography Protocols. Humana Press, Totowa, Vol. 1, pp. 1–20.
14. Terwilliger,T.C. (2000) Maximum-likelihood density modification. Acta Crystallogr. D Biol. Crystallogr., 56, 965–972.
15. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) In Doublé,S. (ed.), Macromolecular Crystallography Protocols. Humana Press, Totowa, Vol. 1, pp. 215–230.
16. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr., 60, 2126–2132.
17. Logon,D.T., Andersson,J., Sjoberg,B.M. and Nordlund,P. (1999) Diadenosine oligophosphates (Ap(n)A), a novel class of signalling molecules. FEBS Lett., 427, 157–163.
18. Rouxd,M.D. and Vishwanatha,J.K. (1995) Diadenosine polyphosphates: their biological and pharmacological significance. J. Pharmacol. Toxicol. Methods, 33, 121–128.
19. Sillero,M.A., De Diego,A., Osorio,H. and Sillero,A. (2002) Dinucleoside polyphosphates stimulate the primer independent synthesis of poly(A) catalyzed by yeast poly(A) polymerase. Eur. J. Biochem., 268, 3329–3332.
20. J. Mol. Biol., 280, 4096–4104.
21. Acta Crystallogr. D Biol. Crystallogr., 50, 760–763.
22. Ames,B.N. and Dubin,D.T. (1960) The role of polyamines in the rigidity relay loop of smooth muscle myosin during the MgATPase cycle. J. Biol. Chem., 275, 25481–25487.
23. Blower,M.D., Nachury,M., Heald,R. and Weis,K. (2005) A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. Cell, 121, 223–234.
24. Nature, 447, 472–476.
25. Bessman,M.J., Chu,E., Guan,Y. and Chasin,L.A. (2005) The role of CTD phosphorylation in mitotic progression. Nature, 436, 968–972.
26. Yengo,C.M., Chrin,L.R., Rovner,A.S. and Berger,C.L. (2000) Tryp2phan 512 is sensitive to conformational changes in the rigid relay loop of smooth muscle myosin during the MgATPase cycle. J. Biol. Chem., 275, 25481–25487.
27. Shimazu,T., Horinouchi,S. and Yoshida,M. (2007) Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3' end processing. J. Biol. Chem., 282, 4470–4478.
28. Acta Crystallogr. D Biol. Crystallogr., 60, 889–892.
29. Nat. Struct. Mol. Biol., 13, 63–70.
30. Acta Crystallogr. D Biol. Crystallogr., 59, 265–274.
31. Logon,D.T., Andersson,J., Sjoberg,B.M. and Nordlund,P. (1999) Diadenosine oligophosphates (Ap(n)A), a novel class of signalling molecules. FEBS Lett., 427, 157–163.
32. Acta Crystallogr. D Biol. Crystallogr., 60, 259–266.
33. Acta Crystallogr. D Biol. Crystallogr., 56, 905–921.
34. Acta Crystallogr. D Biol. Crystallogr., 54, 139–156.
35. Acta Crystallogr. D Biol. Crystallogr., 59, 905–921.
36. Acta Crystallogr. D Biol. Crystallogr., 54, 905–921.
37. Nature, 447, 222–226.
38. Proc. Natl Acad. Sci. U.S.A., 85, 2444–2448.