Crystal structure of the 25kDa 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 I m25 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 UGUΑ 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 25kDa 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...
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 been conserved Nudix fold consisting of an α/β/α sandwich. Within the Nudix fold the consensus sequence of the Nudix box is GX2EX3-REUXEXGU, where U is a hydrophobic residue and X is any residue, and folds into a loop–α helix–loop structure (9). Interestingly, the Nudix box in CF Im25 lacks two of the four conserved glutamate residues, three 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 potentiate 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 CAG ATG TCT GTG GTA CCG CCC-3' and 5'-GGG CAC TTT GGA AAA AGC TGG GTT ATT AGT TGT AAA TAA AAT TGA A-3'.

**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 structure. The position of the 25 kDa subunit of CF Im (CF Im25) will help further define the domain interactions important for substrate and protein interactions.

**Figure 2.** Alignment of human CF Im25 with 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 structure. The position of the 25 kDa subunit of CF Im (CF Im25) will help further define the domain interactions important for substrate and protein interactions.

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 50 μl of the protein solution and 50 μl of the reservoir solution, which was 0.1 M Na acetate pH 5.0 and 15% (v/v) PEG 3350. The protein solution contained 20 mg/ml of CF Im25 and 0.1 M Na acetate pH 5.0. The reservoir solution was filtered through a 0.22 μm filter (Millipore Amicon Ultra-15), flash frozen and stored at −80°C.

**Table 1.** Conditions for crystallization of CF Im25. Plates were set up with 925 PC Workstation (Gilson) by mixing 50 μl of the protein solution and 50 μl of the reservoir solution, which was 0.1 M Na acetate pH 5.0 and 15% (v/v) PEG 3350. The protein solution contained 20 mg/ml of CF Im25 and 0.1 M Na acetate pH 5.0. The reservoir solution was filtered through a 0.22 μm filter (Millipore Amicon Ultra-15), flash frozen and stored at −80°C.
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 (P3₂₁) 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 3500 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

---

Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Se-Peak</th>
<th>Se-Edge</th>
<th>Se-Remote</th>
<th>Ap₄A complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.97923</td>
<td>0.97939</td>
<td>0.97166</td>
<td>0.9</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30–1.82 (1.89–1.82)</td>
<td>30–1.85 (1.92–1.85)</td>
<td>30–1.90 (1.97–1.90)</td>
<td>15–1.80 (1.86–1.80)</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>Measured</td>
<td>Unique</td>
<td>Completeness (%)</td>
<td>Rₘₐᵣₑₑ ( %)²</td>
</tr>
<tr>
<td></td>
<td>157128</td>
<td>23262</td>
<td>97.7 (85.9)</td>
<td>8.6 (51.7)</td>
</tr>
<tr>
<td></td>
<td>148933</td>
<td>22468</td>
<td>96.5 (80.6)</td>
<td>7.3 (57.7)</td>
</tr>
<tr>
<td></td>
<td>136117</td>
<td>20608</td>
<td>95.7 (79.9)</td>
<td>7.0 (52.8)</td>
</tr>
<tr>
<td></td>
<td>543233</td>
<td>25257</td>
<td>99.8 (98.2)</td>
<td>13.8 (26.0)</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>Measured</td>
<td>Unique</td>
<td>Completeness (%)</td>
<td>Rₘₐᵣₑₑ ( %)²</td>
</tr>
<tr>
<td></td>
<td>157128</td>
<td>23262</td>
<td>97.7 (85.9)</td>
<td>8.6 (51.7)</td>
</tr>
<tr>
<td></td>
<td>148933</td>
<td>22468</td>
<td>96.5 (80.6)</td>
<td>7.3 (57.7)</td>
</tr>
<tr>
<td></td>
<td>136117</td>
<td>20608</td>
<td>95.7 (79.9)</td>
<td>7.0 (52.8)</td>
</tr>
<tr>
<td></td>
<td>543233</td>
<td>25257</td>
<td>99.8 (98.2)</td>
<td>13.8 (26.0)</td>
</tr>
</tbody>
</table>

---

**Notes:****

²Rmerge = \frac{\sum |I – \langle I⟩|/\sum I}{\langle I⟩}, where \langle I⟩ is the average intensity from multiple observations of symmetry-related reflections.

²Rcal⁻ᵣ = \left[ \frac{\sum |F_o - F_c|}{\sum F_o} \right]^{1/2}, where LOC is the lack-of-closure error.

²Before and after density modification.

²Rwork and R_free = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}, where F_o and F_c are the observed and calculated structure factor amplitudes. R_free was calculated with 10% of the reflections not used in refinement.
missing). AutoSHARP (17) was then used for refinement of the selenium parameters. The space group was judged to be $P_2_1_2_1$ and not the enantiomorphic $P_2_1_2_1$, 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 $R_{cross}$ 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 7.5 or 8.5, 50 mM KCl, 5 mM MgCl$_2$, 1 mM DTT, 20 mM CF Im$_{25}$, 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 Im$_{25}$ hydrolase activity include ATP, Ap$_4$A (Sigma) and m$^7$G(5)ppp(5)G cap structure analog (New England Biolabs). The putative substrates were combined at 4°C and 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 Im$_{25}$ 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 MgCl$_2$ 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 mM (1.17–1.89 mM at 95% confidence interval). The Ap$_4$A 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 Im$_{25}$

The original N-terminal His-tagged plasmid of human CF Im$_{25}$ (7) did not express to high enough levels for structural studies. We therefore inserted the coding sequence of CF Im$_{25}$ 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 Ap$_4$A was also obtained and refined to a resolution of 1.80 Å.

Description of the structure

Human CF Im$_{25}$ is composed of 227 residues, with a calculated molecular weight of 26 kDa. CF Im$_{25}$ elutes as a dimer in gel exclusion chromatography with an apparent molecular weight of ~53 kDa (Supplementary Figure). The dimeric state of CF Im$_{25}$ 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 box is located in helix α2

Hydrolase activity assay

The standard reaction mixture contained 50 mM Tris–HCl, pH 7.5 or 8.5, 50 mM KCl, 5 mM MgCl$_2$, 1 mM DTT, 20 µM CF Im$_{25}$, 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 Im$_{25}$ hydrolase activity include ATP, Ap$_4$A (Sigma) and m$^7$G(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 Im$_{25}$ 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 MgCl$_2$ 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 mM (1.17–1.89 mM at 95% confidence interval). The Ap$_4$A 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).
Dimer interactions

Dimer formation buries an interface of about 2700 Å² (20,24). The buried surface is extensive and represents 13% of the surface of the homodimer. The dimer interface is maintained by approximately 20 residues per monomer, which participate in both hydrogen bonding and non-polar contacts. The monomer interactions are facilitated by the amino terminal extension (residues 21–32) and by helix α5, the loop preceding helix α5, helix α6, beta strand β6, the loop linking β6 and β7 and beta strand β10.

The main hydrogen-bonding contributions are between the amino terminal extension, beta strand β6 and beta strand β10. These involve residues Thr32 of chain A and Asp142 of chain B, Ser220 of chain A and Asn147 of chain B as well as the reverse monomer interactions. The loop preceding helix α5 and the loop between β6 and β7 contribute to dimer formation through hydrophobic contacts involving residues Pro159 and Tyr160 from chain A and Tyr202 and Phe199 from chain B. Additional interactions are provided by stacking helix α5 from one monomer on to strand β6 of the opposing monomer in a tail-to-tail fashion allowing hydrogen bonding between residues in the loop following helix α5. These hydrogen-bonding interactions include Arg221 and its symmetry mate and the carboxyl group of Ser220 with Asn147 from the opposing monomers.

Metal binding

CF Im25 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 Im25 was still capable of binding metal via the two remaining carboxylates. Although the CF Im25 crystals were obtained in the presence of 200 mM MgCl₂, there was no identifiable Mg²⁺ 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 MnCl₂ and GdCl₃ to identify putative metal site(s) (33). Mn²⁺ and Gd³⁺ 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 Im25 functions as a hydrolase. This finding called for further investigation of CF Im25’s binding to potential substrates (see below).

Enzymatic assays and substrate binding

We next set out to investigate whether CF Im25 binds a substrate. We used three different methods: a colorimetric
assay to measure the release of inorganic phosphate (25), co-crystallization/crystal soaking with putative substrates and steady-state tryptophan fluorescence experiments.

A colorimetric assay to measure the release of inorganic phosphate, and thus the potential hydrolytic activity of CF Im25 on nucleotides, was performed (25). The following putative substrates were tested: ATP, Ap4A and the 7mG(5′)ppp(5′)G cap analog. Ap4A was chosen because the two enzymes found to be most structurally similar to CF Im25 bind Ap4A (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 Im25 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 (Ap4A), Ap4A, AMP, 7mG(5′)ppp(5′)G cap analog, NAD+ and GDP-mannose. In addition to testing these nucleotides, which are known substrates for Nudix enzymes, we also attempted to co-crystallize CF Im25 with a 21mer RNA derived from the PAP γ cDNA sequence (4,5). Of all the putative ligands tested, only Ap4A bound to the crystal (see later). The original crystallization conditions contained 200 mM MgCl2 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 MgCl2 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 MgCl2 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 Ap4A), 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 Im25. The following Nudix ligands were tested: ATP, GTP, Ap4A, inositol hexaphosphate (IP6), 7mG(5′)ppp(5′)G cap analog, ADP and AMP. Dissociation constants (Kd) measured by titrating the ligand concentration were estimated for ATP and Ap4A (Figure 3A and B, respectively). The ATP data were fit with a single hyperbola curve, indicating one binding affinity with a Kd value of 1.53 ± 0.18 mM. The Ap4A-binding data were fit to a single hyperbola curve and represent one binding site per monomer, resulting in a Kd of 2.44 ± 0.49 μM.

**Complex of CF Im25 with Ap4A**

Co-crystallization experiments were performed in order to resolve the ligand-binding interactions of CF Im25. Of all the molecules screened only Ap4A was captured in a co-crystal (Figure 4A). The Ap4A-co-crystal structure was determined by using the phases from the unliganded structure. The resulting unbiased isomorphous difference Fourier (Fo–Fo) map revealed distinct density for one of the adenine bases and three of the four phosphates of Ap4A, in the cavity of CF Im25 (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 Nudix box region, within 3 Å of the triphosphate moiety, and are involved in coordinating the triphosphate moiety (Figure 4B). The binding site is composed of residues from helix α1, beta strand β6, the loop linking β6 and β7 and beta strand β7. In CF Im25, Arg63 and Arg150 are highly conserved across species while Gln157 and Lys172 are moderately conserved. The majority of hydrogen-bonding interactions with Ap4A involve the oxygens of the β- and γ phosphates. The ordered adenine base of the Ap4A molecule stacks with Phe103, a residue contained within the Nudix domain. We note that the position of the γ phosphate of Ap4A coincides with the position of a sulfate ion reported for CF Im25 (Structural Genomics Consortium, Karolinska Institute; PDB ID code 2J8Q).

CF Im25 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 β6 and β7, a shortening of the loop following β4 and an additional α-helix (helix α1) in the CF Im25 structure. The additional α-helix in CF Im25 plays a role in sequestering the
substrate binding pocket from solvent exposure. When comparing the positions of the bound nucleotides between CF I\textsubscript{m}25 and DR1025 we also see variations in substrate fit. Superposition of the ATP bound in DR1025 on to the CF I\textsubscript{m}25 structure shows that both the base and phosphate tail would clash with protein residues in helix α1 and β strand β5 in the CF I\textsubscript{m}25 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\textsubscript{m}25 stabilize the adenine base via stacking interactions with Phe103. Arg63, Arg150, Gln157 and Lys172 contact the oxygens of the β and γ phosphates.

Electrostatic surface and putative RNA and protein-binding regions

The electrostatic surface potential of CF I\textsubscript{m}25 was calculated with GRASP (Figure 7A) (35). The amino terminal region and residues 78–160 of the Nudix domain were shown to participate in RNA binding (7). We observe a good correlation between the location of the putative RNA binding residues and that of the positively charged residues on the surface of the protein (Figure 7B).

DISCUSSION

In our crystal structure CF I\textsubscript{m}25 is a homodimer. A dimeric state for CF I\textsubscript{m}25 is consistent with dynamic light scattering (DLS) and gel filtration experiments performed with the 25 kDa subunit both unliganded and with ATP or Ap\textsubscript{4}A. The dimeric structure of CF I\textsubscript{m}25 stabilizes 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 Ap\textsubscript{4}A hydrolases, including DR1025 (11). Phe103 stabilizes the adenine base in the Ap\textsubscript{4}A bound CF I\textsubscript{m}25 structure but is located six residues upstream of the Nudix box region.
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, HisZ (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.

**Figure 6.** Positions of DR1025 ATP and Ap4A in the CF Im25 structure. Positions of the DR1025 ATP shown in red and CF Im25 Ap4A colored in orange (phosphates) and blue/green (adenine base) within the CF Im25 structure. Ligand position was determined by superposition of the Nudix region over residues 12–145 in DR1025 (PDB ID code 1SU2) and 77–202 in CF Im25. Only one CF Im25 monomer is shown.

**Figure 7.** Surface representation of CF Im25. (A) Electrostatic potential surface representation of dimeric CF Im25 calculated by the program GRASP (32). The surface is shown in the same orientation as in Figure 2 (left) and rotated 90° (right). The surface on the right was made semi-transparent to show the bound Ap4A (shown with arrow). The areas of negative charge are depicted in red, whereas the positively charged regions are colored in blue. (B) Ribbon diagram of dimeric CF Im25 in the same orientation as in A, with residues 21–160 corresponding to the RNA-binding region highlighted in green.

Ap4A belongs to the family of diadenosine oligophosphates, ApnA, which were first discovered 40 years ago (37,38). ApnA 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. Ap4A itself is involved in the cellular stress response, inhibition of K<sub>ATP</sub> channels, stimulation of DNA replication and repair, as well as influencing other essential cellular processes in eukaryotes (37,38). Presently, the only link between Ap4A 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 Ap4A (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 Ap4A can fluctuate within the cell in response to cellular stress or growth.

found to interact with itself and with the CF Im68 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.
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 β3 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 sub-strate 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 Rael1, 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 Tev protease plasmid. We thank Justin Mettey 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**


