Published in final edited form as:

Nat Struct Mol Biol. 2012 August ; 19(8): 782–787. doi:10.1038/nsmb.2329.

Structural Basis for the Activity of a Cytoplasmic RNA Terminal U-transferase

Luke A. Yates^{#1}, Sophie Fleurdépine^{#2}, Olivia S. Rissland^{2,3}, Luigi De Colibus¹, Karl Harlos¹, Chris J. Norbury², and Robert J. C. Gilbert¹

¹Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.

²Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK.

[#] These authors contributed equally to this work.

Abstract

Cytoplasmic terminal uridylyltransferases (TUTases) comprise a conserved family of enzymes that negatively regulate the stability or biological activity of a variety of eukaryotic RNAs, including mRNAs and tumor suppressor let-7 miRNAs. Here we describe crystal structures of the *Schizosaccharomyces pombe* TUTase Cid1 in two Apo conformers and bound to UTP. We demonstrate that a single histidine residue, conserved in mammalian Cid1 orthologs, is responsible for discrimination between UTP and ATP. We also describe a novel high-affinity RNA substrate binding mechanism of Cid1, which is essential for its enzymatic activity and is mediated by three basic patches across the surface of the enzyme. Overall, our structures provide a basis for understanding the activity of Cid1 and a mechanism of UTP selectivity conserved in its human orthologs, with potential implications for anti-cancer drug design.

The post-transcriptional addition of uridyl ribonucleotides to cytoplasmic RNA 3' ends has recently been implicated in several key aspects of eukaryotic RNA biology, including mRNA turnover and regulation of the biogenesis and activity of microRNAs (miRNAs) ¹⁻⁷. In let-7 tumour suppressor miRNA biogenesis, the cytoplasmic terminal U-transferase (TUTase) ZCCHC11 (also known as TUT4 or PUP-2) catalyses the 3' uridylation of cytoplasmic let-7 precursors (pre-miRNAs), which targets them for destruction ^{1,2,4}. Additionally, ZCCHC11-dependent uridylation is important in the regulation of mature miRNAs ³ and replication-dependent histone mRNAs in human cells ⁶. In the fission yeast

Correspondence should be addressed to R.J.C.G. or C.J.N. (gilbert@strubi.ox.ac.uk, chris.norbury@path.ox.ac.uk).

³Current address: Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA.

Author Contributions The project was initiated by O.S.R., L.A.Y., C.J.N. and R.J.C.G.; C.J.N. and R.J.C.G. designed and supervised the project. O.S.R. designed the tCid1 expression construct. S.F. characterised the RNA binding capacity of Cid1, designed and purified the tCid1 mutants, designed and performed EMSA, TUTase assays and Surface Plasmon Resonance. L.A.Y. purified, crystallized tCid1 and optimised crystals for data collection. K.H. soaked and handled crystals for data collection. L.A.Y. collected, processed and analysed the X-ray diffraction data. L.D.C. solved the initial structure; L.A.Y. solved subsequent structures and built and refined all tCid1 structures. L.A.Y. and R.J.C.G. performed the structural and phylogenetic analysis. L.A.Y., S.F., C.J.N. and R.J.C.G. wrote the manuscript. All authors read and approved the manuscript.

Accession codes The structures reported in this manuscript have all been deposited in the RCSB PDB, accession codes 4e7x (crystal form I, Apo I conformer), 4e80 (UTP-bound) and 4e8f (crystal form II containing Apo II conformer).

Schizosaccharomyces pombe the orthologous enzyme, Cid1 (caffeine-induced death suppressor protein 1), uridylylates polyadenylated mRNAs and stimulates their decay^{8,9}. These enzymes belong to the same family as the nuclear poly(A) polymerase^{10,11}, but the structural basis for their RNA binding and UTP selectivity has not yet been described.

Cid1 is a 46 kDa protein containing two recognisable sequence motifs: a nucleotidyl transferase motif common to all members of the DNA polymerase β (Pol β) superfamily and a poly(A) polymerase (PAP)-associated motif. By contrast, ZCCHC11 is much larger (185 kDa), containing a duplication of both motifs found in Cid1, as well as three CCHC-type zinc knuckle motifs (Supplementary Fig. 1). The C-terminal Cid1-homologous region in ZCCHC11 is catalytically active³ and shows striking domain conservation (Supplementary Fig. 1b and c). ZCCHC11 interacts with the RNA-binding protein Lin28 in order to associate stably with (and uridylylate) pre-miRNAs of the let-7 family^{1,2,4} and ZCCHC11 inhibition in Lin28A-expressing cancer cells resulted in tumor regression and suppression of invasiveness¹². No equivalent RNA-binding partner has been described for Cid1, which in monomeric recombinant form efficiently uridylylates RNA substrates *in vitro*⁸. Nevertheless, we reasoned that structural studies of Cid1 might provide an insight into the mode of action of ZCCHC11, while being substantially more straightforward than the direct study of the metazoan enzyme.

Here we report the crystal structure of Cid1 in two Apo conformers and in complex with UTP. The architecture of the active site of Cid1 identifies a histidine residue crucial for the UTP selectivity of the enzyme, and a dramatic remodelling of the active site on transition between the two conformers suggests a mechanism for the clearance of uridylated product. Cid1 lacks a canonical RNA recognition motif, and yet efficiently binds RNA. The crystal structure identifies three basic patches of amino acid side chains on the surface that are non-contiguous in the primary sequence, and which together form a high affinity RNA-binding stripe across the surface of the enzyme. This first structure of a cytoplasmic RNA uridylyltransferase opens the way to novel avenues in anti-cancer drug design.

RESULTS

Structures of Apo and UTP-bound Cid1

We crystallised a functional but slightly truncated variant of Cid1 (residues 33-405; termed tCid1; ref. 8) in two crystal forms (Apo I and Apo II, space groups C2 and P2₁2₁2₁). In addition, a UTP-bound structure was obtained by soaking crystals of form I (space group C2) in UTP. The structures were determined by molecular replacement and refined to 3.2Å, 2.6Å and 3.0Å for Apo structures and UTP-bound structure, respectively (Table 1).

The structure of Cid1 reveals two globular domains (overall dimensions 60Å × 40Å × 40Å) bordering a large catalytic cleft of approximately 15Å × 32Å × 18Å, resulting in limited interfacial contacts (Fig. 1a, b). The N-terminal domain (NTD) is comprised of a mixed five-stranded twisted β -sheet (β 1- β 5) and four α -helices (α A- α D) sandwiched together to give a Pol β 'palm' domain topology¹³ and possesses the magnesium ion co-ordinating catalytic triad of aspartate residues found in all members of the DNA polymerase β family (Fig. 1a). In our crystal structure we were unable to visualise a magnesium ion despite its presence

during the crystallisation process; this probably reflects its sequestration by the mildly chelating citrate anions and DNA present in the crystallisation condition. The C-terminal domain (CTD) shares structural homology with the central domain of poly(A) polymerase (PAP)^{10,11} and C-terminal domains of the trypanosomal TUTases¹⁴ and is comprised of α -helices (α F- α O) and two β -strands (β 6- β 7). We performed a structural phylogenetic analysis^{15,16} of the tCid1 structure with respect to other members of the Pol β superfamily (Fig. 1c, d; see Online Methods and Supplementary Note in Supplementary Information). Of all the polymerase structures compared, Cid1 is most closely related to the last common ancestor, and is placed on an evolutionary branch between the Pol β family and the remaining uridylyltransferases. The most notable difference between the Apo I and UTP-bound states is the deployment of a Cid1-specific two-stranded antiparallel β -sheet we denote the 'trap door' (Fig. 1a, b) that extends over the catalytic cleft on one side, but is too flexible to be seen in either the Apo I or II crystal structure (Fig. 1e, f). No analogous loop is present in Cid1 orthologs from other species, however (Supplementary Fig. 1 online). The oscillatory motion of the NTD, especially the movement of strands β 1 and β 3 (Fig. 1f), determines whether or not the trap door is deployed. In the Apo structures, and in two of the four molecules of the asymmetric unit (ASU) of the UTP-bound crystal structure, the conformation of the β 1 and β 3 strands is such that the β -trap door is sterically excluded. However, in two molecules of the ASU of the UTP-bound structure the β 1 and β 3 strands are positioned so that the β -trap door can traverse the two domains and is held in place via contacts between the β 1- β 3 loop. Additionally, we observe a large sweeping motion of the NTD with respect to the CTD in the second crystal form of Apo tCid1 (Fig. 1f), which is discussed later.

Structural basis of UTP binding

On the basis of multiple sequence comparisons, we initially suspected that the uridine base selectivity in Cid1 would be analogous to that seen in trypanosomal TUTases¹⁴. The active site architecture reveals similar interactions that allow the selection of UTP at the sugar-base edge, most notably the simultaneous interaction between Asn171 and the pyrimidine carbonyl and the ribose 2' OH (Fig. 2a); Asn171 is one of the only amino acid side-chains in the active site to alter conformation on transition from the Apo to the UTP-bound state. This interaction maintains the π -stacking interactions between the uracil base and the Tyr212 aromatic ring and allows the selection of ribose moieties over deoxyribose. Additionally, Phe88 is within van der Waals contact of the uridine ribose sugar moiety and is homologous to Phe100 of PAP¹¹. As in other members of the Pol β family, the triphosphate moiety is coordinated by direct hydrogen bonding with serines and lysines (Fig. 2a). However, despite these similarities in the modality of UTP binding, we find that Cid1 employs a different mechanism for uridine base selectivity from that used by the trypanosomal TUTases. In the latter an aspartate and a glutamate residue cluster around the uracil N3 amine group and interact with uracil *via* a water molecule^{14,17} and both residues are positionally conserved in Cid1, as Asp330 and Glu333. However, structural comparisons of the Cid1 and trypanosomal TUTase active sites (Supplementary Fig. 2 online) reveals an additional histidine residue with no equivalent in the trypanosomal enzymes (His336) that is clearly involved in uridine selection by Cid1, being positioned in close proximity to and contacting the pyrimidine ring. To test this hypothesis we mutated His336 to alanine and found,

remarkably, that it transforms the TUTase into PAP activity (Fig. 2b). Sequence alignments of the nucleotide recognition motif (NRM)¹⁸ reveal conservation of this histidine residue in all three known human TUTases, including ZCCHC6/11 (Fig. 2c), suggesting this mechanism of uracil identification is conserved among *S. pombe* and mammals although it is absent from trypanosomal TUTases. Assessment of the UTP-bound crystal structure by MOLPROBITY¹⁹ suggests that His336 adopts two conformations, between which the imidazole ring is flipped by 180°, each being equally represented in the asymmetric unit. In the Apo-like conformer the distance between the His336 imidazole ring and the uracil amine suggests that a single water molecule, structurally conserved in other TUTases (Fig. 2d) is simultaneously coordinated by the N81 and the uracil N3 amine. In the alternative conformer, the His336 imidazole amine (Nε2) is positioned towards the uracil carbonyl oxygen (O4) at a distance of 3.1 Å. The hydrogen bonding pattern observed in either mechanism is capable of discriminating uracil over other pyrimidines and other nucleotides. This redundancy suggests a ‘failsafe’ mechanism for UTP discrimination. Thus, the acquisition of His336 during evolution may have been sufficient to convert an ancestral PAP into the common ancestor of Cid1 and its human orthologs. The observation that the His336Ala mutant has PAP activity itself strongly supports this contention. The measured distances between His336, Asp330, the uridine N3 amine and a proposed modelled structural water molecule suggest ideal hydrogen bonding distances and a donor:acceptor system involving Asp330 that allows the detection of a uracil-specific amine²⁰. Mutation of Asp330 to alanine reduced but did not ablate the TUTase activity (Fig. 2b). This would be expected if Cid1 employs an additional U-selection process, using the flipped His336 conformer to detect the uracil-specific carbonyl. We therefore suggest that initial UTP binding involves Asp330, but the cross-checking state requires only His336. The deployment of the β-trap door correlated with the flipped conformation of His336 (Fig. 2a), whereas in the UTP-bound state where the β-trap door is not deployed, His336 adopts an Apo-like conformation. It is possible that the trap door acts to prevent UTP from leaving the active site when the His336 is flipped. This correlation suggests that subtle allosteric effects specific to the flipped His336 conformer cause β-trap door deployment, perhaps because it is an adaptation required for the carbonyl-detection stage of uracil decoding.

In the trypanosomal TUTases, the protrusion of an NTD loop region into the active site, specifically an arginine residue (Arg121 in TbTUT4), has been shown to be essential for substrate binding, and in the TbTUT4:UpU crystal structure contacts the 3′ RNA ribonucleotide¹⁷. Interestingly, in TbTUT4 and RET2 this arginine side chain forms a salt bridge with a glutamate residue (Glu300 and Glu424 in TbTUT4 and RET2, respectively) that has been suggested to provide structural support for the NRM²⁰. The presence of this salt bridge in the trypanosomal TUTases elevates the NRM by ~1 Å compared to that of Cid1. As a result Cid1 Glu333, despite being equivalent to Glu300 and Glu424 in the trypanosomal enzymes, is further from the uridine base and is consequently unable to function analogously. Mutation of Glu333 to alanine in Cid1 resulted in greatly diminished TUTase activity (Fig. 2b). It is possible that Glu333 contacts the 3′ RNA ribonucleotide, stabilising a triple-stacked sandwich of Tyr121-UTP-3′ ribonucleotide for efficient transfer¹⁷.

Structural basis of RNA binding

Unlike other members of the Pol β superfamily such as the non-canonical PAP Trf4 (refs. 21-23), Cid1 has not been found in a stable multi-protein complex with RNA-binding proteins and, unlike canonical PAPs, Cid1 lacks an RNA recognition motif (RRM; Supplementary Fig. 1 online). Nonetheless, electrophoretic mobility shift assays (EMSA) showed that purified recombinant Cid1 bound RNA (Fig. 3a), with no apparent sequence specificity (Supplementary Fig. 3a online). Cid1 must therefore contain a cryptic RNA-binding domain or site(s), which on the basis of the nucleic acids bound is likely to depend, in part, on sequence-independent charge interactions.

The electrostatic surface potential of tCid1 revealed several prominent positively charged patches (Fig. 3b). To investigate the involvement of these regions in RNA binding, we mutated several of the basic amino acids to alanines and examined the effect on RNA binding and TUTase activity *in vitro*. EMSA showed that double mutations within any one patch (respectively K133A R137A: mutant 1, K321A R323A: mutant 2, R277A K282A: mutant 3) still allowed RNA binding (Fig. 3c). However, combinations of double mutations in two regions were more disruptive; tCid1 doubly defective in regions 2 and 3 lacked any detectable RNA binding by EMSA (Fig. 3c; Supplementary Fig. 3b online). All the mutant proteins impaired in RNA binding displayed defective TUTase activity (Fig. 3d); in contrast the inactive tCid1DADA mutant, lacking two aspartates in the catalytic triad²⁴, displayed unimpaired RNA binding. An approximate RNA length that the putative RNA-binding surface is capable of accommodating was estimated using the ssRNA from the poly(A) binding protein crystal structure¹⁰ (PDB ID: 1cvj) and suggests that Cid1 is capable of binding to an ssRNA molecule of ~13nt. This is similar in length to the minimal oligonucleotide length (15-mer) observed in EMSA experiments (Supplementary Fig. 3c online).

The RNA-binding characteristics of tCid1 and its mutant derivatives were characterized in detail by surface plasmon resonance (SPR; Supplementary Figs 4, 5 online). Wild type tCid1 exhibited tight RNA binding with surprisingly high on and off rate constants, which complicated estimation of a K_D from kinetic measurements. Steady-state SPR analyses gave higher apparent K_D values (Supplementary Fig. 5 online) of approximately 100 nM for tCid1, 200 nM for mutant 1, 450 nM for mutant 2 and 410 nM for mutant 3. Scatchard plots for tCid1 and mutants 1 and 2 were nonlinear and concave (Fig. 3e), consistent with heterogeneous binding at more than one sub-site. The plot for mutant 3 was almost linear with a shallow curvature, indicating binding dependent on a single site. We therefore propose that RNA binding by Cid1 involves two sites, the first depending on regions 1 and 2 and the second depending on region 3.

Substrate recognition and product ejection by Cid1

Altogether, our data suggest that Cid1 has a two-stage RNA recognition mechanism, in which its general RNA binding ability is succeeded by specific binding of the 3' end alongside UTP. The bound RNA could be threaded along the basic groove on the surface of Cid1 and fed into the active site cleft (Fig. 3b) prior to uridylylation. This mode of RNA binding is distinct from that used by trypanosomal TUTases, canonical PAPs and PAPD1, a

human non-canonical PAP²⁵. The latter lacks the RRM found in canonical PAPs, but does possess an N-terminal region similar to RNP-type RNA binding domains.

The Apo II crystal structure of Cid1 obtained at higher resolution contained two molecules of Cid1 in the asymmetric unit, with radically different conformations (Fig. 1f). Domain movement analysis using DYNDOM²⁶ of the two structures revealed that the NTD (residues 31-164) rotates by 42°, about Asn166, with respect to the CTD, and also translates by 5 Å (Fig. 1f). Animation of the conformational change between Apo conformers I and II (Supplementary Movies 1 and 2 online) indicates a sweeping motion that brings the NTD down into the UTP-binding region of the active site cleft. We propose that this movement is responsible for the ejection of uridylated products prior to the resumption of an open (Apo conformer I) structure and a new round of RNA binding and uridylation.

DISCUSSION

We have described how tCid1 adopts two different Apo conformers – one primed for substrate binding, the other engaged in product ejection. The UTP-bound state that intervenes demonstrates a previously undescribed mechanism of uracil discrimination dependent on His336 (Fig. 2). Electrostatic and mutational analysis has shown how RNA substrates of tCid1 traverse the surface of the enzyme *via* a series of basic patches (Fig. 3) to enable them to be fed into the active site where 3' uridylation occurs. Given that, like Cid1, numerous other members of the DNA Polβ-nucleotidyltransferase superfamily lack recognisable RNA-binding modules, this novel mode of RNA binding may be more widespread, although to what extent will only be revealed by further structural studies. The RNA-class specificity of the human orthologs of Cid1 presumably reflects their acquisition of additional structural elements (such as zinc knuckles) as well as their interaction with substrate-specific RNA-binding proteins such as Lin28A.

Perhaps the most striking finding of our study is, however, the histidine dependence of uracil selection. This critical residue, His336, is conserved in the mammalian TUTases (Fig. 2c and Supplementary Fig. 1b) but is absent from the trypanosomal enzymes, providing further evidence for the functional and structural homology between ZCCHC6, ZCCHC11 and Cid1 (ref. 8). Inhibition of ZCCHC11 expression was recently shown to suppress cancer-related phenotypes including invasion and metastasis in human cell lines and xenografts¹². The identification of His336-dependent nucleotide selectivity in Cid1 thus provides a rationale for the identification of selective small molecule inhibitors of mammalian TUTases on the one hand and trypanosomal TUTases on the other. The insights afforded by this study into the basis of catalysis by Cid1 open up the possibility of using this structural information to target TUTase enzymes therapeutically in both cancer and trypanosomiasis.

ONLINE METHODS

Expression and Purification of tCid1

S. pombe tCid1 was cloned as described previously 8 and expressed in *E. coli* BL21(DE3) pLysS (Novagen) and purified by glutathione, then heparin, affinity chromatography. See Supplementary Note in Supplementary Information for more details.

tCid1 crystallization

Purified tCid1 (in 20mM HEPES, pH 7.0, 50mM NaCl, 1mM DTT) was crystallised in two crystal forms using nanolitre sitting drop vapour diffusion. A UTP-bound complex was obtained by soaking crystals in 10mM UTP. See Supplementary Note in Supplementary Information for more details.

Data Collection

X-ray diffraction data from the two forms of native tCid1 crystals together with the UTP soaked crystals were collected under cryo-temperature (~100K) conditions using single wavelength synchrotron radiation at Diamond Light Source (Didcot, UK). Due to the marked difference in crystal size diffraction data were collected using two different beamlines. In the case of the thin plate-like crystals the specialised microfocus beamline I24 was used for data collection of native and UTP-soaked crystals. Data collection was performed by collecting several consecutive wedges of data from adjacent positions within a single mounted crystal using a beamsize of $10\mu\text{m} \times 10\mu\text{m}$. In the case of the large crystals the standard macromolecular crystallography beamline I04 was used for data collection.

The thin plate-like crystals were radiation sensitive and therefore the sweeps of data were initially processed with XDS²⁷ and SCALA²⁸ implemented in Xia2²⁹. The subsequent merging statistics were used in order to create an optimised subset of data to be merged. Optimal datasets were comprised of a total of 90 degrees of diffraction data from three to four positions on the mounted crystal. Data from each sweep were combined, merged and scaled in Xia2²⁹ to give a final working native dataset at a resolution of 3.2Å and a UTP soaked dataset at a resolution of 3.0Å, the statistics of which are summarised in Table 1. In the case of the large crystals, several complete datasets to ~2.6Å were collected from a single frozen crystal at adjacent positions within the mounted crystal. Each dataset was indexed and integrated in MOSFLM³⁰ and merged and scaled using the CCP4 program SCALA²⁸.

The two crystal morphologies belonged to different space groups; C2 (thin plate-like crystals) and P2₁2₁2₁ (large rod-like crystals), which were denoted as crystal form I and crystal form II, respectively (summarised in Table 1).

tCid1 and UTP complex structure determination, model building and refinement

The structure of tCid1 was determined by molecular replacement using multiple search models, terminal uridylyl transferase 4 (ref. 17, PDB ID: 2q0e) and RET2 (ref. 20, PDB ID: 2b4v) without its middle domain, in Phaser³¹. Both TUTases are from *Trypanosoma brucei* and both provided marginal solutions. The PDB ID 2q0e search model solution gave an R_{free} of 49% after rigid body refinement. This relatively high R factor was due to low

sequence identity (21%) and conformational differences between Cid1 and the adopted search model. Therefore, the marginal solution was subjected to positional and B-factor refinement with autoNCS and targeted against the trimmed version of the search model, as generated by Chainsaw³² and using local similarity symmetry restraints (LSSR) in autoBUSTER-TNT³³. As a result the R_{free} decreased to 39%. The phase distribution obtained from this protocol was subjected to 4-fold NCS averaging, histogram matching and solvent flattening as implemented in Parrot³⁴. The resulting map allowed the automatic building of nearly 80% of tCid1 model by Buccaneer³⁵.

Model building of the Apo (crystal form I) tCid1 structure was performed in COOT³⁶ and a complete Apo I model was refined to 3.2Å using 4-fold NCS restraints and individual B-factor refinement in autoBUSTER³³. Final refinement using translation-libration-screw (TLS)³⁷ with a single TLS group per chain yielded a final R_{work} and R_{free} of 17.9% and 21.3%, respectively. Protein model geometry was assessed during rounds of model building and refinement using MOLPROBITY¹⁹, which scored the model in the 100th percentile against structures of equivalent resolution¹⁹. The Ramachandran plot as determined by MOLPROBITY shows 98.6% of all residues in favoured regions and 100% of all residues in allowed regions. The Apo I crystal structure was used in a rigid body refinement against the UTP-bound data set. Unbiased electron density (Fc-Fo) for a single UTP molecule per chain was clearly visible (see Supplementary Fig. 2). Coordinates for the UTP molecules were taken from HIC-UP database³⁸ and were fitted in the electron density in COOT³⁶. Refinement to 3.0Å was performed in autoBUSTER-TNT³³ analogously to the Apo crystal structures and yielded a final R_{work} and R_{free} of 17.6% and 20.1%, respectively. The Ramachandran plot for the UTP-complex was determined by MOLPROBITY and shows 98.2% of all residues in favoured regions and 100% in allowed regions. MOLPROBITY also scored this structure in the 100th percentile. The Apo II (second crystal form) was solved by Phaser³¹ using the Apo I refined structure. Positional and individual B-factor refinement of the Apo II crystal structure was performed using autoBUSTER-TNT³³ using 2-fold NCS restraints and TLS with 6 groups per chain determined by TLSMD³⁷. The model was refined to 2.6Å yielding a final R_{work} and R_{free} of 19.3% and 23.7%, respectively. Final data collection and refinement statistics are given in Table 1.

The structures encompass residues 41 – 380 of Cid1 according to the UniProt Accession code O13833. In all three structures a flexible loop in the N-terminal domain between α D and β 3 was not built due to weak electron density.

Structural Analysis

Domain motion between Apo I and Apo II structures was analysed using the DYNDOM web server (<http://fizz.cmp.uea.ac.uk/dyndom/>)²⁶. Phylogenetic analysis made use of the structure database webserver DALI³⁹ prior to superimposition using SHP^{15,16} and comparison using Chimera⁴⁰. Full details are given online in Supplementary Information, Supplementary Note and Supplementary Tables 1 and 2.

Plasmids

tCid1 mutants were generated by PCR-based site-directed mutagenesis using Pfu Turbo DNA polymerase (Stratagene) and pGEX6P-1 tCid1⁸ as template. The oligonucleotides used for mutagenesis are listed in Supplementary Table 3. The combination of mutation sites was generated by successive site-directed mutagenesis procedures on previously mutagenised templates. All constructs were verified by sequencing. See Supplementary Note in Supplementary Information for more details.

EMSA

Binding assays were carried out in 10 μ L. 10,000 - 30,000 cpm of 5' end-labeled RNA (corresponding approximately to 0.2 pmol) were boiled for 1 min and subsequently incubated for 2 min on ice. 5 $\times$ buffer was added to final concentrations of 20 mM Tris HCl (pH 8), 150 mM KCl, 2 mM MgCl₂, 2.5 mM DTT, 100 μ g/mL BSA, 5% glycerol. The indicated amounts of proteins were added in 1 μ L. Binding reactions were incubated at RT for 5 min or on ice for 15 min and then loaded on a nondenaturing Tris-glycine (25 mM-250 mM) 5% polyacrylamide gel, pre-run for 15 min. Gels were run for 80 min at 170 V (10 V cm⁻¹), dried and exposed to a storage phosphor screen. Screens were scanned with a Fuji FLA-5000 PhosphorImager and results analysed with AIDA software. Oligonucleotides used for EMSA are listed in Supplementary Table 4.

In vitro RNA nucleotidyl transferase assays

Polymerisation reactions were set up as the binding reactions except that UTP, or when appropriate CTP, GTP or ATP, was added to 0.5 mM. Reactions were incubated for 15 min at room temperature and stopped by adding 20 μ L of STOP buffer (20 mM Tris-Cl [pH 7.5], 0.1 M NaCl, 10 mM EDTA). The RNA was ethanol precipitated, pellets were resuspended in 6 μ L of 8M urea loading buffer and separated by 8 M urea 8% polyacrylamide gel electrophoresis. Gels were run for 80 min at 40 W, dried and exposed to a storage phosphor screen.

Surface plasmon resonance

SPR experiments were carried out using a Biacore 3000 (GE Healthcare). The 21 nt RNA oligonucleotide (GAUUUGACGUUGGUUUUGACG) was chemically synthesised with a 5' biotin tag. Biotinylated RNA was immobilized to CM5 sensor chip (GE Healthcare) indirectly by covalently coupled streptavidin at various levels (30 RU and 60 RU). Flow cells 1 and 3 were kept as control surfaces to account for nonspecific binding.

All experiments were performed at 22°C using a flow rate of 50 μ L/minute in HBS-EP buffer (0.01 M HEPES buffer (pH 7.4), 0.15 M NaCl, 0.005% Surfactant P20). The proteins were injected for 2 min at concentrations of 0.91, 2.74, 8.23, 24.69, 74.07, 222.22, 666.67 and 2000 nM (and 50 and 100 nM for tCid1 and K133A-R137A). Analysis at each protein concentration was repeated at least twice, and samples were run first increasing the concentrations and then decreasing them. Any protein that remained bound after a 2 min dissociation phase was removed by injecting 2 M NaCl for 60 s at 20 μ L/min, which regenerated the RNA surface completely. The data were analysed using BIAevaluation 3.0 (GE Healthcare). Background signal from a streptavidin-only reference flow cell was

subtracted from every data set, as well as an average blank buffer injection. Kinetics data were fitted to a 1:1 Langmuir interaction model with no correction for refractive index. Non-linear regression analysis in Prism (GraphPad) was used to calculate steady state apparent K_D values. Prism was also used for Scatchard transformations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Nick Proudfoot, Anton van der Merwe, Fumiko Esashi and Torben Schiffner for discussions and comments on the manuscript, Aleks Watson and Jane Endicott for their input in the early stages of this project, and Cancer Research UK (C.J.N.), The Royal Society (R.J.C.G.), the Medical Research Council (L.A.Y. and K.H.) and the EP Abraham Research Fund (C.J.N.) for financial support.

References

1. Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 2009; 16:1021–1025. [PubMed: 19713958]
2. Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, Yeom KH, Han J, Kim VN. TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell.* 2009; 138:696–708. [PubMed: 19703396]
3. Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, Ivanov AR, Wolf DA, Mizgerd JP. Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat. Cell Biol.* 2009; 11:1157–1163. [PubMed: 19701194]
4. Lehrbach NJ, Armisen J, Lightfoot HL, Murfitt KJ, Bugaut A, Balasubramanian S, Miska EA. LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 2009; 16:1016–1020. [PubMed: 19713957]
5. Mullen TE, Marzluff WF. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev.* 2008; 22:50–65. [PubMed: 18172165]
6. Schmidt M-J, West S, Norbury C. The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA.* 2011; 17:39–44. [PubMed: 21051505]
7. Shen B, Goodman HM. Uridine addition after microRNA-directed cleavage. *Science.* 2004; 306:997. [PubMed: 15528436]
8. Rissland OS, Mikulasova A, Norbury CJ. Efficient RNA polyuridylation by noncanonical poly(A) polymerases. *Mol. Cell. Biol.* 2007; 27:3612–3624. [PubMed: 17353264]
9. Rissland OS, Norbury CJ. Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat. Struct. Mol. Biol.* 2009; 16:616–623. [PubMed: 19430462]
10. Bard J, Zhelkovsky AM, Helmling S, Earnest TN, Moore CL, Bohm A. Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science.* 2000; 289:1346–1349. [PubMed: 10958780]
11. Martin G, Keller W, Doublié S. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J.* 2000; 19:4193–4203. [PubMed: 10944102]
12. Piskounova E, Polytaichou C, Thornton JE, Lapierre RJ, Pothoulakis C, Hagan JP, Iliopoulos D, Gregory RI. Lin28A and Lin28B Inhibit let-7 MicroRNA Biogenesis by Distinct Mechanisms. *Cell.* 2011; 147:1066–1079. [PubMed: 22118463]
13. Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J. Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science.* 1994; 264:1930–1935. [PubMed: 7516581]
14. Stagno J, Aphasizheva I, Rosengarth A, Luecke H, Aphasizhev R. UTP-bound and Apo structures of a minimal RNA uridylyltransferase. *J. Mol. Biol.* 2007; 366:882–899. [PubMed: 17189640]

15. Riffel N, Harlos K, Iourin O, Rao Z, Kingsman A, Stuart D, Fry E. Atomic resolution structure of Moloney murine leukemia virus matrix protein and its relationship to other retroviral matrix proteins. *Structure*. 2002; 10:1627–1636. [PubMed: 12467570]
16. Stuart DI, Levine M, Muirhead H, Stammers DK. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. *J. Mol. Biol.* 1979; 134:109–142. [PubMed: 537059]
17. Stagno J, Aphasizheva I, Aphasizhev R, Luecke H. Dual role of the RNA substrate in selectivity and catalysis by terminal uridylyl transferases. *Proc. Natl. Acad. Sci. USA*. 2007; 104:14634–14639. [PubMed: 17785418]
18. Martin G, Keller W. RNA-specific ribonucleotidyl transferases. *RNA*. 2007; 13:1834–1849. [PubMed: 17872511]
19. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 2010; 66:12–21. [PubMed: 20057044]
20. Deng J, Ernst NL, Turley S, Stuart KD, Hol WG. Structural basis for UTP specificity of RNA editing TUTases from *Trypanosoma brucei*. *EMBO J.* 2005; 24:4007–4017. [PubMed: 16281058]
21. Hamill S, Wolin SL, Reinisch KM. Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc. Natl. Acad. Sci. USA*. 2010; 107:15045–15050. [PubMed: 20696927]
22. LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*. 2005; 121:713–724. [PubMed: 15935758]
23. Vanacova S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biology*. 2005; 3:e189. [PubMed: 15828860]
24. Read RL, Martinho RG, Wang SW, Carr AM, Norbury CJ. Cytoplasmic poly(A) polymerases mediate cellular responses to S phase arrest. *Proc. Natl. Acad. Sci. USA*. 2002; 99:12079–12084. [PubMed: 12218190]
25. Bai Y, Srivastava SK, Chang JH, Manley JL, Tong L. Structural basis for dimerization and activity of human PAPD1, a noncanonical poly(A) polymerase. *Mol. Cell*. 2011; 41:311–320. [PubMed: 21292163]
26. Hayward S, Berendsen HJ. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*. 1998; 30:144–154. [PubMed: 9489922]

Online References

27. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 1993; 26:795–800.
28. Evans P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* 2006; 62:72–82. [PubMed: 16369096]
29. Winter G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* 2010; 43:186–190.
30. Leslie, AGW.; Powell, HR. Evolving methods for macromolecular crystallography. Read, RJ.; Sussman, JL., editors. Springer; 2007. p. 41–51. Vol. 245 *Nato Science Series IIMathematics, Physics and Chemistry*
31. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. *J. Appl. Cryst.* 2007; 40:658–674. [PubMed: 19461840]
32. Stein N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.* 2008; 41:641–643.
33. Blanc E, Roversi P, Vornrhein C, Flensburg C, Lea SM, Bricogne G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* 2004; 60:2210–2221. [PubMed: 15572774]
34. Zhang KY, Cowtan K, Main P. Combining constraints for electron-density modification. *Meth. Enzymol.* 1997; 277:53–64. [PubMed: 18488305]

35. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* 2006; 62:1002–1011. [PubMed: 16929101]
36. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
37. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* 2006; 62:439–450. [PubMed: 16552146]
38. Kleywegt GJ. Crystallographic refinement of ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* 2007; 63:94–100. [PubMed: 17164531]
39. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucl. Acids Res.* 2010; 38:W545–549. [PubMed: 20457744]
40. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004; 25:1605–1612. [PubMed: 15264254]

Figure 1.

The crystal structure of tCid1. (a) The N-terminal (blue) and C-terminal (green) domains border the central catalytic cleft where UTP is bound. A Cid1-specific two stranded β -sheet denoted as the trap door (red) extends from the CTD to the NTD only in the UTP-bound form. The NTD-side of the catalytic cleft houses three aspartate residues essential for catalytic activity (orange). The conserved catalytic sequence h[GS]X9-13Dh[D/E]h, (where h is a hydrophobic residue, X is any amino acid and the aspartate/glutamate residues are catalytic), which is the hallmark the DNA pol β nucleotidyltransferase family, is indicated in

cyan. **(b)** Orthogonal view labelled as in panel **a**. **(c)** Structural phylogenetic analysis of members of the DNA polymerase β superfamily. **(d)** Structures of the trypanosomal TUTases (*Tb*RET2, ref. 20; PDB 2b56 and *Tb*TUT4, ref. 14; PDB 2ikf), yeast poly(A) polymerase (*Sc*PAP, ref. 10; PDB 1q78), and DNA polymerase β (ref. 13; PDB 2bpc) superimposed using SHP¹⁵ and displayed in the same orientation as the tCid1 structure in panel **a**. **(e)** Comparison between Apo (grey) and UTP-bound tCid1 (coloured as in panel **a** with additional conformational states) illustrating domain oscillation motion (coloured as in panel **a**, and slightly transparent) showing an induced fit and the deployment of the trap door in the UTP-bound state only. **(f)** Cid1 undergoes a large conformational change whereby the NTD rotates by 42° about a pivot point resulting in the closure of the open end of the catalytic cleft together with a remodelling of the active site. The open structure (UTP-bound) is shown coloured as in panel **a**, and the second conformation (Apo) is coloured dark yellow. The extent of the domain motion is highlighted by a dashed line along helix α B in the open form (grey dashed line) and the closed form (blue dashed line).

Figure 2.

The UTP specificity of Cid1 is conferred by a conserved histidine residue. **(a)** A detailed view of the Cid1 active site highlighting the catalytic triad (Asp101, 103 and 160) and amino acid residues involved in UTP binding. **(b)** Mutation of His336 to alanine converts Cid1 from a TUTase into a poly(A) polymerase. Activity was assayed in vitro using recombinant tCid1 (100 nM) and an 5' end-labelled (A)₁₅ RNA substrate. Products were resolved by denaturing polyacrylamide gel electrophoresis. **(c)** Conservation of Cid1 His336 in human TUTases. The aspartate residue asterisked (Asp330 in Cid1) is conserved among

all the RNA nucleotidyl transferases. **(d)** Comparison of uracil selection mechanisms found in Cid1 and the high-resolution crystal structures of TbTUT4 (ref. 14; PDB: 2ikf), RET2 (ref. 20; PDB: 1b51) and MEAT1 (ref. 17; PDB: 3hj1) from *T. brucei*. Highly co-ordinated waters that form hydrogen bonds with the uridine base from the *T. brucei* structures are shown as red spheres. In Cid1 a proposed water is modelled (from 3hj1) based on structural superposition. Hydrogen bonds involved in achieving U-specificity are shown as green arrows. Invariant residues shared by the trypanosomal TUTases and Cid1 are asterisked.

Figure 3.

Cid1 binds RNA via an extended basic surface. **(a)** Electrophoretic mobility shift assay (EMSA) of RNA binding by tCid1 *in vitro*. **(b)** Space-filling representation of tCid1 structure colour-coded for surface charge (red, acidic; blue, basic; range -4 to +4 KT/e). Amino acid residues contributing to the putative RNA-binding surface are indicated. **(c, d)** RNA binding (EMSA; **c**) and TUTase activity assays (**d**) using 25 nM tCid1 (1) and mutants (2) K133A R137A, (3) K321A R323A, (4) R277A K282A, (5) K133A R137A R277A, (6) K133A R137A K282A, (7) K133A R137A R277A K282A, (8) R277A K321 R323A, (9)

K282A K321A R323A, (10) R277A K282A K321A R323A, (11) K133A R137A K321A R323A, (12) K144A, (13) K144A R277A K282A and (14) the catalytic mutant D101A D103A. Numbers circled in panels **c** and **d** denote mutation of the three basic patches as indicated in panel **b**. (**e**) Scatchard plot derived from steady-state SPR binding data for tCid1 and the mutant proteins indicated (for corresponding primary steady-state SPR data see Supplementary Fig. 5).

Table 1

Data collection and refinement statistics

	tCid1 Apo (I)	tCid1 UTP-bound	tCid1 Apo (II)
Data collection			
Space group	C2	C2	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	164.4, 78.0, 152.5	164.7, 78.0, 151.7	81.2, 101.2, 113.6
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 90, 109.5	90, 90, 109.3	90, 90, 90
Resolution (Å)	64.24–3.20 (3.28–3.20) *	77.72–3.02 (3.10–3.02)	56.8 – 2.6 (2.74 – 2.6)
<i>R</i> _{merge}	13.8 (71.2)	9.3 (47.9)	7.3 (48.1)
<i>I</i> / <i>σI</i>	8.9 (2.4)	12.8 (3.2)	24.7 (6.2)
Completeness (%)	99.5 (97.7)	99.4 (99.5)	99.6 (99.7)
Redundancy	4.1 (4.1)	3.7 (3.9)	18.0 (18.4)
Refinement			
Resolution (Å)	64.27 – 3.20	28.72 – 3.02	56.79 – 2.60
No. reflections	30013	35591	29283
<i>R</i> _{work} / <i>R</i> _{free}	17.88/21.30	17.64/20.11	19.19/23.88
No. atoms			
Protein	10262	10339	5202
Ligand/ion	26	133	14
Water	10	17	81
<i>B</i> -factors			
Protein	73.28	61.20	74.62
Ligand/ion	67.00	55.44	94.30
Water	39.21	56.25	65.95
R.m.s. deviations			
Bond lengths (Å)	0.009	0.009	0.009
Bond angles (°)	0.93	0.98	0.97

* Values in parentheses are for highest-resolution shell.