An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA

Sarah Nainar¹, Bonnie J. Cuthbert³, Nathan M. Lim², Whitney England¹, Ke Ke¹, Kanika Sophal², Robert Quechol², David L. Mobley¹,², Celia W. Goulding¹,³, Robert C. Spitale*,¹,²,³

(1) Department of Pharmaceutical Sciences, (2) Department of Chemistry, and (3) Department of Molecular Biology and Biochemistry. University of California, Irvine. Irvine, California. 92697

*Correspondence: rspitale@uci.edu
Supplementary Figure 1. Quantification of dot blots in Figure 2c. Integrated chemiluminescence signal of each dot across a concentration titration of 5EU analog fitted with a non-linear regression, and signals relative to 5EUd incorporation (bar charts) for a. HeLa, b. K562, c. HEK293T, d. NIH3T3, and e. PCN cells.
Supplementary Figure 2. CuAAC imaging of 5EU background incorporation in human and mouse cultured cells. Cells were incubated with titrated 5EU for 5 h, followed by fixation permeabilization and click with Alexa-488 to visualize background labeling of nascent transcription. Imaging analysis also revealed that 5EU incorporation mirrors 5EUd, with strongest signal originating in the nucleolus, the location of ribosomal RNA synthesis. This experiment was performed twice across four cell lines, with similar results.
Supplementary Figure 3. HEK293T cells were incubated with titrated 4-thiouracil (4SU) and 4-thiouridine (4SUd) as a positive control for 5 h. Following incubation, RNA was extracted and subjected to MTSEA biotinylation, followed by streptavidin-HRP dot blot detection. 4SU is able to be incorporated through endogenous cellular machinery by wildtype cells. This experiment was independently repeated three times (three different cell-culture experiments), with similar results.
Supplementary Figure 4. Anti-FLAG Western blot detection of a. HsUPRT and b. HsUMPS ectopic expression in HEK293T cells. Lane 1 represents un-transfected, WT lysate, while lane 2 represents lysate expressing the proteins of interest. These experiments were performed three times with similar results.
**Supplementary Figure 5.** Effects of HsUPRT and HsUMPS overexpression and knockdown on 4SU incorporation in HEK293T cells. 

**a.** Dot blot demonstrating that in the presence of overexpressed HsUPRT or HsUMPS, there is increased incorporation of 4SU into HEK293T cellular RNA. Bar chart shows mean integrated chemiluminescence signal of each dot in each treatment, normalized to the WT signal. Samples were obtained from two independent cell culture experiments (biological replicates), as shown. 

**b.** Dot blot showing the effects of 4SU RNA incorporation when cells are treated with 25 nM non-targeting scramble, HsUPRT or HsUMPS siRNA pools. Bar chart shows mean integrated chemiluminescence signal of each dot in each treatment, normalized to the scramble signal. Samples were obtained from two independent cell culture experiments (biological replicates), as shown. 

**c.** RT-qPCR analysis confirming successful siRNA knockdown of endogenous HsUMPS or HsUPRT mRNA (mean relative gene expression) in samples treated with 4SU, normalized to GAPDH control. Samples were obtained from two independent experiments, and run in technical triplicates.
Supplementary Figure 6. Human UPRT and UMPS facilitate 5EU incorporation into cellular RNA. **a.** Direct comparison blot of HsUPRT and HsUMPS overexpression and 5EU incorporation in HEK293T cells. This blot was performed once, however overexpression of HsUPRT and HsUMPS was independently repeated three times (biological cell culture experiments) as described in Figure 2), with similar results. **b.** Structure of human uridine monophosphate synthase PDB: 2WNS, demonstrates a spacious pocket around C5 of the substrate orotidine monophosphate. This region may allow for C5 functionalized pyrimidine substrates to bind. **c.** Comparison of 5EU labeling signals produced upon expression of Toxoplasma gondii UPRT or Homo sapiens UPRT homologs suggest similar activity levels. Similar labeling signals may indicate that the *Toxplasma gondii* and *Homo sapiens* UPRT enzymes have similar activities with the substrate 5EU. This experiment was independently repeated twice, with similar results.
Supplementary Figure 7. Primary sequence alignments of a. UPRT and b. UMPS homologs amongst various representative organisms suggests there may be similar levels of activity.
Supplementary Figure 7. Primary sequence alignments of a. UPRT and b. UMPS homologs amongst various representative organisms suggests there may be similar levels of activity.
Supplementary Figure 8. Additional replicate dot blots showing the effects of 5EU RNA incorporation when cells are treated with 25 nM non-targeting scramble and a. HsUMPS or b. HsUPRT siRNA pools. Integrated signals from these blots were used in integrated signal bar chart in Figure 2. In total, knockdown of HsUMPS and HsUPRT was independently repeated four times, all with similar results. c. RT-qPCR analysis confirming successful siRNA knockdown of endogenous HsUPRT or HsUMPS mRNA (mean relative gene expression) in samples treated with 5EU, normalized to GAPDH control. Samples were obtained from four independent experiments, and run in technical triplicates where *** p = 0.0005, **** p = 0.000065 (95% CI) as determined by a two-tailed Student’s t-test. Error bars represent standard deviation.
Supplementary Figure 9. Western blot analysis of metabolic enzyme knockdown. **a.** Anti-UMPS probing of lysate extracted from cells transfected with 25 nM scramble siRNA (left) or 25 nM UMPS siRNA (right) for 72 h. The expected size of functional UMPS isoform A is ~52 kDa, which has reduced upon successful knockdown. Other bands are likely non-specific bands, as has been demonstrated previously in the literature with antibodies against this protein. This experiment was performed once. **b.** Anti-UPRT probing of lysate extracted from cells transfected with 25 nM scramble siRNA (left) or 25 nM UPRT siRNA (right) for 72 h. The expected size of functional UPRT is ~33 kDa, which was not detected in either treatment. This experiment was independently repeated twice, with similar results.
Supplementary Figure 10. Incorporation of 2’azido-nucleosides into nascent RNA. HEK293T cells were treated with 1 mM 2’azidoadenosine, -uridine or –cytidine for 12 h, followed by RNA extraction, biotinylation and streptavidin HRP dot blot. Cells were able to incorporate 2’AzAd and 2’AzCd, while 2’AzUd was not detected. This experiment was independently repeated twice, with similar results.
Site directed mutagenesis was performed on a vector encoding UCK2 fused to TurboGFP. Following cloning and sequence confirmation, we had two sets of 13 mutants for screening against 5mAzUd and 2’AzUd. Mutants were transfected into HEK293T cells and expression was monitored via the GFP reporter. Images were taken at different confluencies. We note that 24 of the 26 mutants were robustly expressed in cells, with the exception of R166H which significantly affected cell proliferation and Y65D, which appeared to cause severe inclusion body formation. Expression of UCK2 or variants was confirmed via fluorescent imaging of the GFP reporter prior to all experiments using each construct.
Supplementary Figure 12. Quantification of dot blots represented in Figure 3. Integrated chemiluminescence signal of each dot of the in-cell screening of UCK2 mutants with a. 5mAzUd and b. 2’AzUd. Signals were normalized to the positive control in each blot. These experiments were repeated twice, with cell-culture duplicates, with similar results.
Supplementary Figure 13. HEK293T cells transfected with WT UCK2 followed by treatment with 5mAzUd. This titration experiment and endpoint cell imaging was carried out after all labeling experiments.
Supplementary Figure 14. HEK293T cells transfected with Y112A UCK2 followed by treatment with 5mAzUd. This titration experiment and endpoint cell imaging was carried out after all labeling experiments.
Supplementary Figure 15. HEK293T cells transfected with F114S UCK2 followed by treatment with 5mAzUd. This titration experiment and endpoint cell imaging was carried out after all labeling experiments.
Supplementary Figure 16. HEK293T cells transfected with F83A; F114S UCK2 followed by treatment with 5mAzUd. This titration experiment and endpoint cell imaging was carried out after all labeling experiments.
**Supplementary Figure 17.** HEK293T cells transfected with Y112S; F114S UCK2 followed by treatment with 5mAzUd. This titration experiment and endpoint cell imaging was carried out after all labeling experiments.
Supplementary Figure 18. WT HEK293T cells treated with 5mAzUd followed by treatment with 5mAzUd. This titration experiment and endpoint cell imaging was carried out after all labeling experiments.
Supplementary Figure 19. HEK293T cells transfected with WT UCK2 followed by treatment with 2’AzUd. This titration experiment was performed twice, with similar results in addition to repeated endpoint cell imaging after all labeling experiments.
Supplementary Figure 20. HEK293T cells transfected with R166G UCK2 followed by treatment with 2’AzUd. This titration experiment was performed twice, with similar results in addition to repeated endpoint cell imaging after all labeling experiments.
Supplementary Figure 21. HEK293T cells transfected with R166A UCK2 followed by treatment with 2’AzUd. This titration experiment was performed twice, with similar results in addition to repeated endpoint cell imaging after all labeling experiments.
**Supplementary Figure 22.** WT HEK293T cells following treatment with 2’AzUd. This titration experiment was performed twice, with similar results in addition to repeated endpoint cell imaging after all labeling experiments.
Supplementary Figure 23. HEK293T cells were transduced with lentivirus to create a stably-expressing UCK2 cell line. a. Design of lentivirus vector encoding UCK2 upstream of a read-through linker attached to an EGFP reporter. b. WT or UCK2-expressing cells were treated with 1 mM 2’AzUd for 5 h, followed by streptavidin dot blot to demonstrate robust UCK2-dependent labeling. This experiment was independently repeated three times, with similar results. c. Western blot analysis of UCK2 expression for lentivirus encoded UCK2. Anti-UCK2 probing of lysate extracted from WT cells (left) or cells expressing UCK2 (right) demonstrates presence of UCK2 protein only in stable cell line. This experiment was performed once. d. Imaging of EGFP reporter to identify UCK2 positive cells. Imaging of GFP reporter was monitored each time cells were used for experiments.
Supplementary Figure 24. Cell proliferation measured via the ability of metabolically active cells to reduce resazurin to resorufin. WT or HEK293T cells stably expressing UCK2 were grown in standard media and treated with titrated a. & b. 2'AzUd or c. & d. 5mAzUd for 5 – 72 h. Mean percent cell proliferation was normalized to cells treated with the DMSO control only for each respective time point (0.5% final). This experiment was with replicate cell cultures of four for each treatment. Error bars represent standard deviation.
Supplementary Figure 25. Effects of nucleosides on nucleophosmin (NPM1) localization, an indicator of nucleolar stress.² HEK293T cells were treated 500 μM nucleoside for 5 h. After staining with anti-NPM1 followed by anti-rabbit Alexa 647, the localization of NPM1 was assessed. Cells treated with DMSO showed normal staining, with NPM1 properly localized to the nucleoli, which maintained their spherical morphology. Cells treated with 4SUd showed some dispersed staining outside of the nucleoli, while the nucleoli were less pronounced and spherical. The appearance of this stress phenotype was observed in ~13% of cells. Cells treated with both 2'AzUd and 5mAzUd (+/- UCK2) showed normal staining, with NPM1 properly localized to the nucleoli, which maintained their spherical morphology. There were no indications of nucleolar
stress based on the translocation of NPM1. This observation suggests that when incorporated, 2'AzUd does not interfere with nucleolar processing of RNA. Furthermore, despite the observed toxicity of 5mAzuUd, these results support that the mode of toxicity is not through RNA incorporation. Rather, it may be through some intermediary, phosphorylated 5mAzuUd species and other metabolic pathways. This experiment was independently repeated twice with similar results.
Supplementary Figure 26. Assessment of 2’ modification on the decay rate of cellular mRNAs. HEK293T cells treated with DMSO or HEK293T cells expressing UCK2 and treated with 1 mM 2’AzUd were incubated for 5 h, followed by Actinomycin D transcriptional inhibitor for 0 – 6 h. RNA was isolated from the samples followed by RT-qPCR analysis of the MYC, EEF1A and MZT1 (a – c) transcript abundance. These results are from two independent cell culture experiments, with technical triplicates. d. The decay rates of several mRNAs was then determined between the two treatments.
Supplementary Figure 27. Imaging of HEK293T cells following incubation with 1 mM 4SUd, 5EUd, 2’AzCd or DMSO control for 5 – 72 h. Cells treated with 4SUd followed by 5EUd showed the greatest changes in cell morphology and detachment. This experiment was performed twice with similar results.
**Supplementary Figure 28.** Assessment of the effects of non-specific nucleosides on cell viability via trypan blue exclusion and Alamar Blue analysis.  

**a.** Cells treated with 1 mM 4SUd, 5EUd, 2’AzCd or DMSO for 5 – 72 h were assessed for their ability to exclude Trypan blue, indicating sustained membrane integrity. This experiment was performed in duplicate cell cultures for each treatment.  

**For b-d,** Alamar blue analysis measured the ability of metabolically active cells to reduce resazurin to resorufin. HEK293T cells were grown in standard media and treated with b. 5EUd, c. 4SUd, or d. 2’AzCd for 5 – 72 h. Mean percent cell proliferation was normalized to cells treated with the DMSO control only for each respective time point (0.5% final). This experiment was performed in replicate cell cultures of six for each treatment. Error bars represent standard deviation.
Supplementary Figure 29. Aromatic residues in a. UCK1 (PDB: 2JEO) and b. UCK2 (PDB: 1UEJ) used in fluorescence quenching experiments to determine binding affinity of canonical and 2'azido analogs.
Supplementary Figure 30. Kinetic curves of uridine, cytidine, 2’AzUd and 2’AzCd phosphorylation by purified UCK1 and UCK2. The data was fitted to the Michaelis-Menten equation using a non-linear fit. Each data point is the mean of three experiments. Error bars represent standard deviation.
Supplementary Figure 31. Comparison of UCK1 and UCK2 sequence and structure. a. UCK1 PDB: 2JEO and UCK2 PDB: 1UEJ overlay similarly and have 67% sequence similarity. b. UCK1 has a shorter helix of the lid domain, and less structure (circled in blue) compared to UCK2. Arginine residues critical for induced fit substrate binding are differentially oriented. ARG175 of UCK1 faces the exterior of the enzyme, making it less likely to interact with the base. ARG174 and ARG176 face the interior, toward the binding pocket (circled in red).
**Supplementary Figure 32.** Crystal structures of UCK2•2’AzUd representing the flipped binding mode. 

- **a.** Crystal structure of UCK2•2’AzUd (orange) taken from subunit D and the bound cytidine molecule (gray) from PDB: 1UEJ.
- **b.** UCK2•2’AzUd subunit D (above) and subunit E (below) and
- **c.** 2’AzUd from subunit D (above) and subunit F (below) demonstrate a wide degree of flexibility in the base orientation.
Supplementary Figure 33. Representations of the canonical binding mode. a. Crystal structure of UCK2-2’AzUMP (orange) taken from subunit B and the crystallized cytidine molecule (gray) from PDB ID: 1UEJ. b. Overlay of crystal structures of UCK2-2’AzUMP taken from subunit A and subunit B, representing the canonical binding mode. Subunit A is represented by the molecule in which the oxygen atoms are oriented upwards. Subunit B is represented by the molecule in which the oxygen atoms are oriented downward.
Supplementary Figure 34. Frequency of hydrogen bond contacts from the five 100 ns MD simulations of 2’AzUd in the flipped binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.
Supplementary Figure 35. Frequency of hydrogen bond contacts from five 100 ns MD simulations of 2’AzUd in the canonical binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.
Supplementary Figure 36. TPM (transcripts per million count) expression of UCK1 and UCK2 from tissues and two immortalized cell lines to be profiled.
Supplementary Figure 37. Quantification of dot blots in Figure 5. a. & b. Integrated chemiluminescence signal of each dot relative to the DMSO controls. c. & d. Integrated chemiluminescence signal of each dot plotted as time and concentration titration curves of 2’AzUd + UCK2, fitted with a non-linear regression.
Supplementary Figure 38. UCK2-mediated incorporation of 2’AzUd into HEK293T cell RNA.

a. LC-MS/MS traces of uridine and 2’AzUd detection in WT or UCK2-expressing cells.

b. Concentrations of 2’AzUd and uridine were determined through a simultaneously-run standard curve, and percent RNA incorporation of 2’AzUd into cells expressing UCK2 was determined by the ratio of 2’AzUd to uridine within each sample. Samples were obtained from two independent labeling experiments and analyzed in technical duplicates, ± SD. n.d. = not determined.
Supplementary Figure 39. UCK2-GFP and mCherry co-culture enrichment RT-qPCR analysis. 

a. Co-culture image of HEK293T cells separately transfected with UCK2-TurboGFP fusion or mCherry. 2’AzUd was then added for 0 or 5 h followed by RNA extraction, biotinylation, reverse transcription streptavidin bead enrichment of cDNA:hybrids. This was performed twice, with independent co-cultures generated from different cell passages. 

b. qPCR was used to assess enrichment of TurboGFP or mCherry, where the average difference between the input and enriched Cq values is shown for each time point. A difference below zero indicates the transcript was de-enriched in the sample. Analysis was performed on samples obtained from two independent experiments, and technical triplicates. 

c. Overall mean fold enrichment of both genes in samples (from b) labeled with 2’AzUd for 5 h relative to samples with no 2’AzUd labeling (0 h). TurboGFP was shown to be more enriched at 5 h than mCherry.
References

1 Griffith, M. et al. Novel mRNA isoforms and mutations of uridine monophosphate synthetase and 5-fluorouracil resistance in colorectal cancer. *The Pharmacogenomics Journal* **13**, 148, doi:10.1038/tpj.2011.65
https://www.nature.com/articles/tpj201165#supplementary-information (2012).

2 Burger, K. et al. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol.* **10**, 1623-1630, doi:10.4161/rna.26214 (2013).
Supplementary Note 1: Molecular Dynamics Supporting Information and Figures

An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA

Sarah Nainar¹, Bonnie Cuthbert², Nathan M. Lim², Whitney England¹, Kanika Sophal², Robert Quechol², David L. Mobley¹,², Celia Goulding², Robert C. Spitale*¹,²,³

(1) Department of Pharmaceutical Sciences, (2) Department of Chemistry, and (3) Department of Molecular Biology and Biochemistry. University of California, Irvine, Irvine, California. 92697

*Correspondence: rspitale@uci.edu
1.0 Computational Methods

The starting structure used for the molecular dynamics (MD) simulations was taken from the crystal structure of human uridine-cytidine kinase 2 complex with the cytidine substrate (PDB: 1UEJ). In this study four different MD simulations were carried out, using two different binding modes for each of the respective ligands 2'AzUd and 2'AzCd. Here, we differentiate between the two binding modes by the orientation of the ribose moiety of each molecule. The "canonical" binding mode refers to the pose in which the azide (substituted at the 2' position) is pointed inwards into the binding site (Fig. 1a; main text Fig. 4a). We refer to this as the canonical binding mode as this pose is similar to the pose found in the crystal structure with the bound cytidine substrate (PDB: 1UEJ). The "flipped" binding mode refers to the pose in which the azide is pointed outwards from the binding site, (Fig. S1B; main text Fig. 4b).

A total of 500 ns of MD simulation time was conducted for each binding mode by running 5 simulations each, where each copy of the simulation began from the same starting structure. The metastable binding modes sampled during our MD simulations were defined by constructing a Markov State Model (MSM) from our pool of MD simulation data and clustering with perron-cluster cluster analysis (PCCA). We then compared these metastable binding modes with X-ray crystal structures for 2'AzUd and 2'AzCd by computing the root-mean-square deviation (RMSD) between the ligand heavy atoms. We would like to note that the x-ray crystal structures of 2'AzUd and 2'AzCd were not used for setup or seen prior to conducting our MD simulations. Our simulation results, along with the x-ray crystal structures, support the hypothesis that the ligands must adopt the "flipped" binding mode for catalytic turnover.

1.1 Setup

1.1.1 Docking

After equilibration of the UCK2-cytidine complex, we used HYBRID docking from OpenEye Toolkits (v2019.4b2) to dock our nucleoside analogs (2'AzUd and 2'AzCd) into the binding site. HYBRID differs from the standard docking approach, such that the software will use the co-crystallized ligand as a reference point and attempt to fit the nucleoside analogs within the binding site by overlaying the docked ligands with the crystallographic ligand. From HYBRID docking, we generated up to 50 different poses for each nucleoside analog and then followed the same equilibration protocol described previously. After equilibration, we found by visual inspection, that the poses began to resemble two separate groups: 1 pose which resembled the "canonical" binding mode and another which had the ribose moiety "flipped". We then selected 1 conformer from each representative group and carried these forward for our NPT 100 ns MD simulations (no restraints).

1.1.2 Protein preparation

To prepare the protein system (PDB: 1UEJ) for MD simulations, we used PDBFixer to model in missing residues, add missing hydrogen atoms, protonate sidechains, and solvate our system. Sidechains were protonated in accordance with the pH = 8.0 environment of the enzyme assay experiments and as described in other computational studies. Following protein preparation, with the cytidine molecule bound (using the pose from PDB: 1UEJ), we energy minimized the protein-ligand complex for a maximum of 30,000 steps and followed with an equilibration protocol as follows. The equilibration protocol occurs in four 10 picoseconds (ps) stages, whereby a progressively declining restraining force was used to help the protein-ligand
system gradually relax. First, we apply a restraining force of 2.0 kcal/mol* Å² to the heavy atoms of the protein-ligand complex, simulate for 10 ps using constant volume (NVT), and follow-up with 10 ps at constant pressure (NPT). Next, we decrease the restraining force to 0.5 kcal/mol* Å² and then conduct an NPT simulation for 10 ps. Last, we use a 0.1 kcal/mol* Å² restraining force on the alpha-carbons (protein backbone) and the ligand heavy atoms and then simulate in the NPT ensemble simulate for 10 ps.

1.1.3 MD simulation parameters

All simulations in this study were conducted using OpenMM v7.1.1 at T = 300K and P=1atm. Bonds with hydrogens were constrained. Here, we use timesteps of 4 fs by employing the hydrogen mass repartitioning (HMR) scheme. This scheme allows us to take larger timesteps by slowing down the fastest motions (i.e. hydrogen bond stretching) in our MD simulations. The HMR scheme re-allocates mass from the connected heavy atom to the hydrogens. The protein-ligand systems were placed in a periodic box with explicit TIP3P water molecules using a 10 Å solvent padding distance and counter ions (Na+ Cl-) were added using a concentration of 150 mM. A 10 Å cut-off distance was used for the particle-mesh Ewald method for computing long-range (e.g. electrostatic) interactions. Protein atoms were parameterized using the ‘amber99sbildn’ forcefield and the ligands were parameterized using GAFF2 in which atomic charges were assigned using the AM1-BCC charge model. Both charges and GAFF2 parameters were assigned using openmoltools v0.8.1 and OpenEye Toolkits v2019.4b2.

1.2 Analysis

For each nucleoside analog (2’AzCd and 2’AzUd), we ran five 100 nanoseconds (ns) MD simulations starting from each of the two binding modes (canonical and flipped). They are denoted as 2’AzCd<sub>canc</sub>, 2’AzCd<sub>flip</sub>, 2’AzUd<sub>canc</sub>, and 2’AzUd<sub>flip</sub>. Collectively, over the course of the 100ns of simulation time, the root-mean-square deviation (RMSD) for the ligand atoms began to stabilize after 25 ns. Thus, we discard trajectory frames from 0 – 25 ns as additional equilibration time and only perform further analysis from the 25 ns – 100 ns time frame Fig.2a. For construction of our Markov State models (MSM), we use the PyEMMA v2.5.5 toolkit. To analyze hydrogen bonding contacts we use the MDTraj v1.9.1 and VMD v1.9.3 toolkits.
Supplementary Note 1; Figure S1. Comparison of the canonical binding mode and the flipped binding mode. The canonical binding mode is represented by the cytidine molecule (gray) from the crystallized UCK2-cytidine complex (PDB: 1UEJ), wherein a. one binding mode of simulated 2’AzCd (cyan) similarly overlays in the canonical binding mode (2’AzCd_{canc}) and b. a second binding mode is flipped, wherein the sugars are oppositely oriented (2’AzCd_{flip}).
**Supplementary Note 1; Table 1.** Calculated RMSDs (Å) of the ligand heavy atoms from each MD simulation for the 25-100 ns time frame.

| Ligand | RMSD  | RMSD  | RMSD  | RMSD  | RMSD  | AVG  |
|--------|-------|-------|-------|-------|-------|------|
| 2’AzCd_{canc} | 6.66  | 1.36  | 1.14  | 4.12  | 2.06  | 3.07 |
| 2’AzUd_{canc} | 0.94  | 1.17  | 2.89  | 1.31  | 1.41  | 1.54 |
| 2’AzCd_{flip} | 2.24  | 3.72  | 0.99  | 1.88  | 1.90  | 2.15 |
| 2’AzUd_{flip} | 3.19  | 1.36  | 1.39  | 1.37  | 6.39  | 2.74 |

The average RMSD of each ligand binding mode is shown in Table 1. They are 3.07 Å, 1.54 Å, 2.94 Å, and 2.54 Å for 2’AzCd_{canc}, 2’AzUd_{canc}, 2’AzCd_{flip}, and 2’AzUd_{flip}, respectively.

**Supplementary Note 1; Figure S2A.** Calculated RMSDs (Å) for the ligand heavy atoms for 2’AzCd in the canonical binding mode from 25 – 100 ns. Each colored line represents one of the five MD simulations.
Supplementary Note 1; Figure S2B. Calculated RMSDs (Å) over time for the ligand heavy atoms for 2’AzUd in the canonical binding mode from 25 – 100 ns. Each colored line represents one of the five MD simulations.
Supplementary Note 1; Figure S2C. Calculated RMSDs (Å) over time for the ligand heavy atoms for 2’AzCd in the flipped binding mode from 25 – 100 ns. Each colored line represents one of the five MD simulations.

Supplementary Note 1; Figure S2D. Calculated RMSDs (Å) over time for the ligand heavy atoms for 2’AzUd in the flipped binding mode from 25 – 100 ns. Each colored line represents each of the five MD simulations.

1.2.1 Defining the metastable binding modes

Supplementary Note 1; Figure S3. The projected TICA components for 2’AzCd in the canonical binding mode (left) and the discretized trajectory frames after cluster assignment by PCCA (right). Each color indicates a different metastable binding mode. For 2’AzCd in the canonical binding mode, we find 4 different metastable binding modes from our MD simulations. The dominant binding mode (red) corresponds to the representation we show in Figure S26 a.
Supplementary Note 1; Figure S4. The projected TICA components for 2’AzUd in the canonical binding mode (left) and the discretized trajectory frames after cluster assignment by PCCA (right). Each color indicates a different metastable binding mode. For 2’AzUd in the canonical binding mode, we find 2 different metastable binding modes from our MD simulations. The dominant binding mode (pink) corresponds to the representation we show in Figure 4 a.

Supplementary Note 1; Figure S5. The projected TICA components for 2’AzCd in the flipped binding mode (left) and the discretized trajectory frames after cluster assignment by PCCA (right). Each color indicates a different metastable binding mode. For 2’AzCd in the flipped binding mode, we find 3 different metastable binding modes from our MD simulations. The dominant binding mode (pink) corresponds to the representation we show in Figure S26 b.

Supplementary Note 1; Figure S6. The projected TICA components for 2’AzUd in the flipped binding mode (left) and the discretized trajectory frames after cluster assignment by PCCA (right). Each color indicates a different metastable binding mode. For 2’AzUd in the flipped binding mode, we find 2 different metastable binding modes from our MD simulations. The dominant binding mode (green) corresponds to the representation shown in Figure 4 b.
Supplementary Note 1; Figure S7A. Implied timescales at varying lag times for 2’AzCd in the canonical binding mode

Supplementary Note 1; Figure S7B. Implied timescales at varying lag times for 2’AzCd in the flipped binding mode
Supplementary Note 1; Figure S7C. Implied timescales at varying lag times for 2’AzUd in the flipped binding mode

Supplementary Note 1; Figure S7D. Implied timescales at varying lag times for 2’AzUd in the canonical binding mode.

In order to define our metastable binding modes and visualize their structures, we construct a Markov State model (MSM)\(^\text{14}\) from our five separate MD simulations. Our features for constructing the MSMs consists of the distance between the closest heavy atoms on the ligand and the following 9 residues: ASP62, PHE83, ASP84, TYR112, PHE114, HIS117, ILE137, ARG166, and ARG176. These residues were selected as they have been noted in the literature to play important roles in binding\(^\text{9}\). From this feature space, we apply the time-lagged independent component analysis (TICA) method using a lag time of 1 ns. In Figure S7, we show the implied timescales as a function of lag time and find that 1 ns provides good discretization or separation of states for each of our MD simulation sets. TICA transforms our 9 dimensional feature space to a new set of reaction coordinates which maximizes the autocorrelation of the transformed coordinates\(^\text{15}\). In other words, TICA allows us to extract the slow order parameters and project them into a lower dimensional space; here, we use the first two TICA coordinates (Fig. S3-S6).
Then, we apply k-means clustering to discretize our trajectory frames into discrete microstates and project them into TICA space (denoted by individual Xs). Following, we use perron-cluster cluster analysis (PCCA)\textsuperscript{16} to assign each microstate to a metastable macrostate (denoted by color in Fig. S3-S6). From each of our assigned macrostates, we randomly sample 100 representative structures and then visualize the structure which minimizes the RMSD to the crystallographic ligand.

**Supplementary Note 1; 1.2.2 Comparisons to X-ray Crystal Structures**

Here, we compared our metastable binding modes from our MD simulations against each subunit found in the experimental x-ray crystal structures and list the values for the randomly sampled frame against the subunit which minimizes the computed RMSD. To compare the metastable binding modes from our MD simulations with experimental x-ray crystal structures, we first align the two structures using the protein backbone. Particularly, we align by the protein backbone using residues 19 to 229 but exclude residues 48 to 52 as these were missing in the x-ray crystal structures. Once aligned by the protein backbone, we then compute the RMSD between crystallized ligand and the matching ligand heavy atoms from our defined metastable binding modes.

**Supplementary Note 1; Table 2A.** Calculated RMSDs (in Angstroms) of the matching ligand heavy atoms of each metastable binding mode for 2’AzCd against the crystallized ligand (within each subunit) from our UCK2-2’AzCd-ATP structure.

| Metastable Binding Mode | Subunit D | Subunit E | Subunit F |
|-------------------------|-----------|-----------|-----------|
| 2’AzUd\textsubscript{flip} (cluster 0) | 3.17 | 5.02 | 4.93 |
| 2’AzUd\textsubscript{flip} (cluster 1) | 2.88 | 4.30 | 4.02 |
| 2’AzUd\textsubscript{canc} (cluster 0) | 5.28 | 4.54 | 4.14 |
| 2’AzUd\textsubscript{canc} (cluster 1) | 5.15 | 4.27 | 3.93 |

**Supplementary Note 1; Table 2B.** Calculated RMSDs (in Angstroms) of the matching ligand heavy atoms of each metastable binding mode for 2’AzUd against the crystallized ligand (within each subunit) from our UCK2-2’AzUMP structure.

| Metastable Binding Mode | Subunit A | Subunit B |
|-------------------------|-----------|-----------|
| 2’AzUd\textsubscript{flip} (cluster 0) | 4.76 | 4.70 |
| 2’AzUd\textsubscript{flip} (cluster 1) | 4.45 | 4.17 |
| 2’AzUd\textsubscript{canc} (cluster 0) | 3.61 | 2.71 |
| 2’AzUd\textsubscript{canc} (cluster 1) | 3.34 | 2.44 |

**Supplementary Note 1; Table 2C.** Calculated RMSDs (in Angstroms) of the matching ligand heavy atoms of each metastable binding mode for 2’AzUd against the crystallized ligand (within each subunit) from our UCK2-2’AzUd structure.
|                  | Subunit A | Subunit B |
|------------------|-----------|-----------|
| $2'\text{AzCd}_{\text{flip}}$ (cluster 0) | 3.89      | 4.05      |
| $2'\text{AzCd}_{\text{flip}}$ (cluster 1) | 2.95      | 2.90      |
| $2'\text{AzCd}_{\text{flip}}$ (cluster 2) | 3.10      | 2.89      |
| $2'\text{AzCd}_{\text{canc}}$ (cluster 0) | 0.76      | 1.22      |
| $2'\text{AzCd}_{\text{canc}}$ (cluster 1) | 0.76      | 0.98      |
| $2'\text{AzCd}_{\text{canc}}$ (cluster 2) | 0.84      | 1.17      |
| $2'\text{AzCd}_{\text{canc}}$ (cluster 3) | 0.72      | 1.12      |

**Supplementary Note 1; Figure S8.** Post-catalytic $2'\text{AzCMP}$ represents the canonical binding mode. Shown here are the structures of our crystal structure from UCK2-$2'\text{AzCMP}$ (subunit A, orange), MD model of $2'\text{AzCd}$ (cyan) in the canonical binding mode, and cytidine (PDB: 1UEJ, gray).
1.2.3 Hydrogen bond contacts
Using the ‘compute_contacts’ tool from MDTraj, we compute the distance between the closest heavy atoms in the ligand and nearby surrounding residues for hydrogen bonding. Here, we calculate the frequency in which the distance between the ligand and the residues are less than or equal to 3.0 Å, which we choose to define as the minimum distance needed to form a hydrogen bond. The frequency of hydrogen bond contact is calculated by taking the number of frames, in which the minimum distance is met, over the total number of frames from each of our 100 ns MD simulations. Contact frequencies from each MD simulation is represented by each individual colored bar.

Supplementary Note 1; Figure S9A. Frequency of hydrogen bond contacts from five 100 ns MD simulations of 2’AzCd in the flipped binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.
Figure S9B. Frequency of hydrogen bond contacts from five 100 ns MD simulations of 2′AzCd in the canonical binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.

PDB Files: [https://drive.google.com/file/d/1-SqBdbT27jehi42dxGw8BIJTFkqlc_sm/view?usp=sharing](https://drive.google.com/file/d/1-SqBdbT27jehi42dxGw8BIJTFkqlc_sm/view?usp=sharing)

References
1. Suzuki, N. N., Koizumi, K., Fukushima, M., Matsuda, A. & Inagaki, F. Structural basis for the specificity, catalysis, and regulation of human uridine-cytidine kinase. Structure 12, 751–764 (2004).
2. Eastman, P. et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. PLoS computational biology 13, e1005659 (2017).
3. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. J. chemical theory computation 11, 1864–1874 (2015).
4. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the amber ff99sb protein force field. Proteins: Struct. Funct. Bioinforma. 78, 1950–1958 (2010).
5. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. J. computational chemistry 25, 1157–1174 (2004).
6. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. am1-bcc model: ii. parameterization and validation. J. computational chemistry 23, 1623–1641 (2002).
7. Eastman, P. Pdbfixer version 1.5. https://github.com/pandegroup/pdbfixer (2018).
8. Tomoike, F., Nakagawa, N., Kuramitsu, S. & Masui, R. A single amino acid limits the substrate specificity of thermus thermophilus uridine-cytidine kinase to cytidine. Biochemistry 50, 4597–4607, DOI: 10.1021/bi102054n (2011). PMID: 21539325, https://doi.org/10.1021/bi102054n.
9. Tanaka, W. et al. Molecular mechanisms of substrate specificities of uridine-cytidine kinase. Biophys. physicobiology 13, 77–84 (2016).
10. McGann, M. Fred and hybrid docking performance on standardized datasets. J. computer-aided molecular design 26, 897–906 (2012).
11. Scherer, M. K. et al. Pyemma 2: A software package for estimation, validation, and analysis of markov models. J. chemical theory computation 11, 5525–5542 (2015).
12. McGibbon, R. T. et al. Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. Biophys. journal 109, 1528–1532 (2015).
13. Humphrey, W., Dalke, A. & Schulten, K. Vmd: visual molecular dynamics. J. molecular graphics 14, 33–38 (1996).
14. Prinz, J.-H. et al. Markov models of molecular kinetics: Generation and validation. The J. chemical physics 134, 174105 (2011).
15. Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for markov model construction. The J. chemical physics 139, 07B604_1 (2013).
16. Röblitz, S. & Weber, M. Fuzzy spectral clustering by pcca+: application to markov state models and data classification. Adv. Data Analysis Classif. 7, 147–179 (2013).
Supplementary Note 2: X-Ray Crystallography Supporting Information and Figures

An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA

Sarah Nainar¹, Bonnie Cuthbert², Nathan M. Lim², Whitney England¹, Kanika Sophal², Robert Quechol², David L. Mobley¹,², Celia Goulding², Robert C. Spitale*,¹,²,³

(1) Department of Pharmaceutical Sciences, (2) Department of Chemistry, and (3) Department of Molecular Biology and Biochemistry. University of California, Irvine. Irvine, California. 92697

*Correspondence: rspitale@uci.edu
Supporting Information – Crystallographic Studies

UCK2 crystallization experiments.
UCK2 crystals were grown by hanging drop vapor-diffusion. UCK2 (15 mg/mL) was screened for crystallization with 100 mM 2'-azidouridine (2'AzUd); or 50 mM 2'-AzUd, 5 mM ATP, 10 mM MgCl₂ or 50 mM 2'-azidocytidine (2'AzCd), 5 mM ATP and 10 mM MgCl₂. All resulting crystals were mounted and collected under cryoconditions at 70K with the addition of 10-25% (vol/vol) glycerol as a cryoprotectant to the reservoir condition.

Datasets were indexed, integrated and scaled in HKL2000 (Otwinowski and Minor 1997), solved by molecular replacement using Phaser and refined using Phenix.refine in the Phenix suite (Table S1) (Adams et al. 2010). Molecular replacement searches were performed with UCK2 structures PDB: 1UJ2 (UCK2•CMP) and 1UEI (UCK2•UTP) (Suzuki et al. 2004).

Structure of UCK2 complexed with 2'AzUd.
UCK2 with 2'AzUd crystallized in 50 mM BIS-TRIS pH 6.5, 25 mM ammonium sulfate, and 32% pentaerythritol ethoxylate and were cryo-cooled after 12 days. The dataset was collected at ALS 8.2.1, resulting in a 100% complete dataset up to 3.1 Å (Table S1). A molecular replacement solution of the processed dataset was found with a search for eight molecules with PDB: 1UJ2. Refinement was carried out in subsequent rounds using Phenix.refine (Adams et al. 2010) and Coot (Emsley and Cowtan 2004) to a final R-work/R-free of 21.1/25.8. Data collection and refinement statistics are presented in Table 1.

Structures of UCK2 complexed with ATP, MgCl₂ and 2'AzCd or 2'AzUd.
UCK2 was screened for crystallization with 2'AzCd, MgCl₂ and ATP. The best diffracting crystals were found in the PACT suite condition 58: 0.2 M sodium/potassium phosphate, 20% PEG 3350. Crystals were optimized and grown in this condition for 10 days before cryo-cooling for data collection. To obtain a structure of UCK2 with ATP and 2'AzUd, UCK2 was screened for crystallization with 2'AzUd, ATP, and MgCl₂. However, the best diffracting crystals of UCK2 with ATP and 2'AzUd were observed in the UCK2•2'AzCMP crystallization condition. Following crystallization optimization an UCK2•2'AzUMP dataset was collected from crystals grown for 3 days in 0.2 M sodium/potassium phosphate, 22% PEG 3350.

The UCK2•2'AzCMP dataset was collected at ALS 8.2.1 with diffraction to 2.4 Å and was processed as described above (Table 1). Molecular replacement was carried out using a search for two molecules of PDB: 1UEI, and the solution was refined to a final R-work/R-free of 18.9/22.7, as described above. The UCK2•2'AzUMP dataset was collected at
SSRL 9-2 with diffraction to 2.7 Å. Data processing and molecular replacement were performed as per the UCK2•2′AzCMP dataset. Refinement resulted in final R-work/R-free of 19.6/24.4.

**Supplementary Note 2; Table S1.** Data collection and refinement statistics for UCK2 datasets with 2′azidouridine and cytidine.

| Data collection | 2′-azidouridine | 2′-azidouridine, ATP, MgCl₂ | 2′-azidocytidine, ATP, MgCl₂ |
|-----------------|----------------|----------------------------|----------------------------|
| **Space group** | P 2₁           | F 2 2 2                    | F 2 2 2                    |
| **Cell dimensions** | 92.7, 85.8, 153.4 | 89.9, 142.4, 248.6 | 90.3, 141.3, 247.9 |
| a, b, c (Å) | 90, 95.4, 90 | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 50.0-3.1 (3.15-3.1)* | 50.0-2.7 (2.75-2.7) | 50.0-2.4 (2.44-2.4) |
| **R_{sym}** | 0.089 (0.607) | 0.078 (0.448) | 0.097 (1.033) |
| **I / I** | 22.4 (1.9) | 29.3 (1.8) | 31.9 (1.7) |
| **Completeness (%)** | 99.8 (100.0) | 98.0 (85.5) | 99.9 (99.7) |
| **Redundancy** | 3.7 (3.4) | 8.0 (5.3) | 7.0 (6.5) |

**Refinement**

| Resolution (Å) | 38.5-3.1 (3.2-3.1) | 46.8-2.7 (2.8-2.7) | 45.2-2.4 (2.5-2.4) |
| No. reflections | 43704 (3766) | 21895 (1867) | 29822 (2616) |
| R_{work} / R_{free} | 0.211 / 0.258 | 0.196 / 0.244 | 0.189 / 0.227 |
| **Protein atoms** | 12755 | 3256 | 3286 |
| **Ligands:** Nucleotide, Phosphate, Mg^{2+} | 3, 8, NA | 2, 2, 2 | 2, 2, 2 |
| **Solvent molecules:** water, glycerol | 184, 27 | 65, 5 | 145, 7 |
| **B-factors** | 76.5 | 68.6 | 59.9 |
| **Ligands:** Nucleotide, Phosphate, Mg^{2+} | 103.3, 68.1, NA | 100.0, 58.1, 91.0 | 79.2, 51.1, 79.9 |
| **Solvent:** water, glycerol | 62.8, 88.0 | 73.6, 93.2 | 63.0, 87.5 |
| Ramachandran favored, outliers (%) | 98.7, 0 | 99.8, 0 | 99.8, 0 |
| **R.M.S. deviations** | 0.002 | 0.002 | 0.003 |
| Bond lengths (Å) | 0.56 | 0.76 | 0.52 |

*Values in parentheses are for highest-resolution shell.*
Supplementary Note 2; Table S2. Distances between ligand and UCK2 residue atoms representing polar contacts and potential hydrogen bonds. Interaction distances for each chain are shown for UCK2-2’AzUd, UCK2-2’AzUMP and UCK2-2’AzCMP. Only one representative chain shown for canonical structures of UCK2-UTP (PDB: 1UEI) and UCK2-CMP (PDB: 1UJ2).

| Residue: | Thr29 | Ala30 | Lys33 | Asp62 | Tyr65 | Asp84 | Asp84 | Tyr112 | Phe114 | His117 | Glu135 | Arg166 | Arg166 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|
| Atom:    | Cα1  | N    | NZ    | O5/1O62 | OH | O1 | O2 | OH | O | NH1/N2 | O1 | NH1 | NH2 |
| UCK2-2’AzUd D | 4.1 (O2) | 2.1 (O2) | 3.4 (O4) | 3.5 (N1) | 3.3 (O4) |
| UCK2-2’AzUd E | 4.2 (O3) | 3.9 (O4) |
| UCK2-2’AzUd F | 2.8 (O4) | 3.7 (O6) |
| UCK2-2’AzUMP A | 2.7 (O2), 3.6 (O5) | 3.2 (O2), 2.8 (O2) | 2.3 (O2), 4.0 (O2) | 2.9 (O4), 3.0 (O2) | 2.7 (N2), 3.3 (N4) | 3.7 (O4), 3.8 (O3) | 2.4 (O2), 3.0 (N3), 3.1 (O2), 3.3 (O4), 3.8 (O6) |
| UCK2-2’AzUMP B | 2.2 (O2), 3.1 (O4), 3.5 (O3) | 2.6 (O2), 4.0 (O4) | 2.9 (O4) | 2.8 (N2), 3.2 (O3) | 2.8 (O4) | 3.0 (N3), 3.1 (N4), 3.9 (O3) |
| UCK2-2’AzCMP A | 2.5 (O5), 4.0 (O6) | 3.0 (O5), 3.2 (N15), 3.2 (N14), 4.0 (N15), 3.4 (O11), 4.0 (N15), 3.7 (O10), 2.5 (N25), 2.8 (N23), 2.9 (O5), 3.5 (N13), 2.9 (N16), 3.0 (N15), 3.3 (O11), 4.0 (N15), 4.0 (O11) |
| UCK2-2’AzCMP B | 2.2 (O8), 3.7 (O9), 3.4 (O7), 3.7 (O8), 3.7 (O10), 2.6 (N15), 3.2 (N14), 3.4 (O11), 4.0 (N15), 3.7 (O10), 2.3 (O11), 2.4 (O11), 3.7 (O10), 2.3 (O11), 2.9 (O2), 2.9 (N23), 3.3 (O11), 4.0 (N15), 3.0 (N15), 3.3 (O11), 4.0 (N15), 4.0 (O11) |

We determined the structures of UCK2 complexes with 2’azidouridine (2’AzUd) in the presence and absence of ATP, and with 2’-azidocytidine (2’AzCd) and ATP. We compare our structures throughout this section with the structures determined by Suzuki et al. of apo, cytidine monophosphate/adenosine diphosphate (CMP/ADP) bound and uridine triphosphate (UTP) bound structures of UCK2 (PDB: 1UFQ, 1UJ2, and 1UEI) (Suzuki et al. 2004). All figures and table mentioned herein correspond to those found in this section, unless otherwise noted.

Supplementary notes and figures on the structure of UCK2 bound to 2’AzUd

UCK2 complexed with 2’AzUd (UCK2-2’AzUd) crystallized with eight protein subunits in the asymmetric unit (ASU). When comparing the protein subunits with previously determined UCK2 protein complex structures, the subunits align best to the apo-UCK2 structure with an average root mean square deviation (RMSD) of 0.42 ± 0.06
Å, and more poorly to CMP/ADP and UTP bound structures with average RMSD of 0.50 ± 0.07 and 0.90 ± 0.11 Å, respectively, suggesting that the eight protein subunits are all in an “open” conformation. Of the eight protein subunits, only three (D, E, and F) show reasonable electron density for a bound 2’AzUd. Interestingly, all subunits show bound phosphate in a highly coordinated binding pocket that corresponds to the ADP/ATP binding site seen in the CMP/ADP bound structure (Fig. S1).

Only a few residues provide potential hydrogen-bond (H-bond) contacts with the bound 2’AzUd molecules (Table S2). In all three subunits, 2’AzUd H-bonds with Tyr65, whereas additional H-bonds are observed in subunit D by Asp62, Asp84, Arg166, and Gln184; in subunit E by Asp62, Arg174, and Arg176; and in subunit F site by only Asp 84, and Arg174 and a water molecule (Table S2). Within the UCK2•UTP structure, UTP H-bonds to some of these same residues—Asp62, Asp84, Arg166, Arg174, Arg176, and Gln184, however the uridine portion of UTP is coordinated by H-bonds from 13 residues, while 2’AzUd is coordinated by 6-8 H-bonds (Table S2). Notably, the azido-group is coordinated by Arg174 in two of the three occupied subunits.

The three bound subunits show three different ligand conformations for 2’AzUd (Fig. 4 c) Of these conformations, two (subunit E and F) could not bind in the closed conformation due to potential clashes with Arg174 and Arg176 (Table S2), which both coordinate with 2’-AzUd. Thus one may speculate that Arg174 and Arg176 residues are important in guiding the ligand into and locking it in the binding site as the protein transitions into its closed conformation. Further, we postulate that the ligand conformations observed in subunits E and F represent the ligand entering or exiting the binding site. The high B factors for the bound nucleosides in this structure reflects the highly mobile nature of 2’-AU binding, where the nucleosides have an average B factor of 103.3 Å² compared to 76.5 Å² for the protein, and are modelled with an average occupancy of 94%.

Supplementary Note 2; Figure S1. Coordinated phosphate in the ASU of UCK2•AzUd.
Supplementary notes and figures on the structure of UCK2 bound to 2’AzUMP

UCK2 complexed with 2’AzUd, ATP and MgCl₂ results in crystals with two protein subunits in the ASU. Notably, UCK2 carried out a phosphorylation reaction in situ, which resulted in bound 2’-azidouridine monophosphate (2’AzUMP) and an Mg²⁺ ion in each subunit (Fig. S2). The byproduct of the phosphorylation reaction, ADP, is not observed in the structure, whereas the site that should be occupied by the ADP terminal phosphate is occupied by a phosphate, as observed for the UCK2•AU structure.

While the phosphate of 2’AzUMP is in a similar location in both subunits (phosphate shifted 1.1 Å), the orientation of the rest of the nucleotide differs. Within the two subunits the uridine base is roughly planar, however in one subunit the O2-carbonyl oxygen is flipped reflecting a ~170° rotation; additionally, one of the O4-carbonyl oxygens is shifted 2.5 Å between the two subunits (Fig. 4d). In native-ligand containing UCK2 complex structures, the nucleoside base π-stacks with Phe83. The base of 2’AzUMP also retains this π-stacking interaction; however, the base does not lie parallel to Phe83, likely weakening the interaction (Fig. S3). Interestingly, the azido-group in both subunits is posed to form H-bonds with Asp84 and Arg166, possibly stabilizing the charge on the azido-group.

Overall, the A and B subunit have very similar coordinating interactions with bound 2’AzUMP as each molecule is coordinated by residues: Thr29, Ala30, Lys33, Asp62, Tyr65, Asp84, His117, Arg166, and Arg169 (Table S2, Fig. S3). Of these, all but Tyr65 were observed interacting with UTP in the UCK2•UTP structure (Suzuki et al. 2004). Interestingly, the participation of Tyr65 in this binding site appears to sterically block the nucleobase from optimal positioning for base stacking interactions with Phe83. Within the A subunit, 2’AzUd also coordinates Glu135, while this interaction is prevented in the B subunit by a slight shift in the terminal phosphate – however both subunits show Glu135 coordinating the Mg²⁺ ion. Overall, the majority of the known interactions between UCK2 and phosphorylated uridine are also observed with 2’AzUMP.

The bound 2’AzUMP molecules are not modelled with full occupancy, where in the A subunit 2’AzUMP is modelled at 78% and in the B subunit 2’AzUMP is at 68% occupancy. Additionally, the B factors of the azido-UMP molecules at 100.0 Å² are high relative to the rest of the protein structure at 71.7 Å², indicating that the bound ligands are mobile. Thus, though 2’AzUMP appears much more tightly coordinated than 2’AzUd, there is still mobility within 2’AzUMP binding of UCK2.
Supplementary Note 2; Figure S2. Density of subunit A and B illustrating the product of phosphoryl

Supplementary Note 2; Figure S3. 2’AzUMP interactions within the binding pocket largely coordinate to the same residues found in PDB: 1UEI.
Supplementary notes and figures on the structure of UCK2 bound to 2’AzCMP

As with UCK2•2’AzUMP, UCK2 complexed with 2’AzCd, ATP and MgCl₂ crystallized with two protein subunits in the ASU, both of which contain the product of the phosphorylation reaction, 2’-azidocytidine monophosphate (2’AzCMP) and an Mg²⁺ ion (Fig. S4a). Like UCK2•2’AzUMP, ADP was not detected in the UCK2•azido-CMP structure, and instead we observe a phosphate molecule seen in both UCK2•2’AzUd and UCK2•2’AzUMP structures.

The bound 2’AzCMP molecule superimposes well with the CMP molecule from the UCK2•CMP structure (Suzuki et al. 2004). Additionally, the binding of 2’AzCMP in both A and B subunits is quite similar (Fig. S4b). Overall, the monophosphate and ribose are similarly positioned for CMP and azido-CMP, wherein one subunit the monophosphate and ribose atoms are shifted by approximately 0.7-0.8 Å (as measured at the phosphate, O3’ and O4’ atoms). The cytidine base in both subunits is in a similar orientation to the UCK2•CMP structure, however, both bases in the 2’AzCMP molecules are slightly tilted (20 °) in comparison to UCK2•CMP structure.

While the UCK2 bound 2’AzCMP molecule orientation is similar to CMP, the average B factors for 2’AzCMP are somewhat higher than the rest of the protein: 81.2 and 77.2 Å² for A and B, respectively, indicating that the bound ligand has some mobility. Additionally, the occupancy for the 2’AzCMP molecules was refined to 86% and 71%, again, indicating some propensity for the molecule to be absent from the binding pocket as was seen with 2’AzUMP.

Overall the H-bond contacts that stabilize the native ligand are conserved with azido-modified cytidine ligand. In UCK2•CMP, the CMP is coordinated by 14 residues, and 10 of these contacts are conserved in the UCK2•2’AzCMP structure. The bound 2’AzCMP is coordinated by Thr29, Ala30, Lys33, Asp62, Tyr65, Asp84, Phe114, His117, Arg166 and Arg169 residues in both subunits (Table S2, Fig. S5). Additionally, in subunit B we see an additional contact provided by the carbonyl of Asp170. Notably, Gln184, which provides contacts to the O2’ atom of CMP, is displaced in both subunits upon 2’AzCMP binding, potentially due to the introduction of the azido group. Both His117 and Tyr112, which make contacts to the N4 atom of CMP normally, are on β-strands (β-strand 3 and 4) that have shifted in the 2’AzCMP-bound structure resulting in a loss of Tyr112 contact with the ligand. Similarly, Arg174, another coordinating residue in the CMP bound structure, is displaced in the UCK2•2’AzCMP structure: Arg174 is facing away from bound ligand in subunit A, but was left unmodeled in subunit B as it lacks clear electron density.
Supplementary Note 2; Figure S4. UCK2 complexed with 2'AzCd, ATP and MgCl₂ crystallized with two protein subunits in the ASU, both of which contain the product of the phosphorylation reaction, 2'-azidocytidine monophosphate (2'AzCMP) and a Mg²⁺ ion. a. Density of A and B subunits showing the phosphorylation product 2'AzCMP. b. Superimposed structures of 2'AzCMP and CMP from PDB: 1UJ2 display agreement in binding configurations.
Supplementary Note 2; Figure S5. Structure of UCK2•2’AzCMP demonstrating conserved contacts.

Supplementary notes and figures on the open and closed conformations of UCK2

A previously published structures bound to a native substrate (PDB: 1UJ2, Fig. S6 gray) shows a closed conformation: the bound ligand is fully enclosed in the binding pocket where the base is coordinated by an aromatic environment from one direction and a series of arginine residues coordinate the other face of the molecule (Suzuki et al. 2004). Conversely, the apo-UCK2 structure (PDB: 1UFQ, Fig. S6, light blue) shows a fully open conformation (Suzuki et al. 2004). Strikingly, within the azido-modified nucleoside bound structures we have captured a series of conformations between the apo and fully closed conformation. With the UCK2•2’AzUd structure, none of the subunits were in the closed conformation, reflecting the transitory positions of the bound 2’AzUd molecules (Fig. S6, green). Indeed, while one molecule occupies a position in or near canonical binding, the other two appear to be in a state of entry or exit. In the UCK2•AzUMP structure, the subunits are in a conformation closer to the closed conformation (Fig. S6, yellow). This is reflected in an increased commonality between the residues coordinating 2’AzUMP to residues previously observed coordinating CMP and UTP (Table S2). Finally,
the structure of the UCK2•2’AzCMP complex best approximates the closed conformation (Fig. S6, magenta).

**Supplementary Note 2; Figure S6.** Overlay of UCK2•CMP (PDB: 1UJ2) in gray, UCK2 apo (PDB: 1UFQ) in light blue, UCK2•2’AzUd in green, UCK2•2’AzUMP in yellow and UCK2•2’AzCMP in magenta illustrates the open and closed conformations found in the azido analog structures.

**Additional notes for this section that were taken from the main text.**

[[UCK2•2’AzUd]] Simulated hydrogen bond (H-bond) frequencies of 2’AzUd$_{flip}$ reveal contact for key residues ASP62 (~7%), ASP84 (~30%), and ARG174 (lid region; ~60%), which are likewise found in our UCK2•2’AzUd structures (**Supplementary Fig. 33; Supplementary Note 1, Table S2**). As such, these residues likely play a significant role in forming the flipped binding conformation.

[[UCK2•2’AzUMP]] Observing H-bond contact frequencies for 2’AzUd$_{canc}$ across all five simulations, we saw almost no contact with the catalytic residue ASP62 (**Supplementary Fig.**
We also noticed a drastic increase in contact with TYR112 (~80%) and HIS117 (~40%), which supports the observed changes in the UCK2•2'AzUMP structure, illustrating these residues have shifted toward the binding pocket (yellow, Fig. 4e). The H-bonds of HIS117 (loop residue) and ARG166 (lid residue) predicted by the MD model are only observed the post-catalytic state captured in our UCK2•2'AzUMP structures (Supplementary Fig. 35; Supplementary Note 2, Table S2).

References

Adams, Paul D, Pavel V Afonine, Gabor Bunkoczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, et al. 2010. “PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution.” Acta Crystallogr. D Biol. Crystallogr. 66: 213–21. doi:10.1107/s0907444909052925.

Emsley, P, and K Cowtan. 2004. “Coot: Model-Building Tools for Molecular Graphics.” Acta Crystallogr. D Biol. Crystallogr. 60: 2126–32. doi:10.1107/s0907444904019158.

Otwinowski, Zbyszek, and Wladek Minor. 1997. “Processing of X-Ray Diffraction Data Collected in Oscillation Mode.” Methods 276 (January 1993): 306–15. doi:10.1016/S0076-6879(97)76066-X.

Suzuki, Nobuo N., Katsuhisa Koizumi, Masakazu Fukushima, Akira Matsuda, and Fuyuhiko Inagaki. 2003. “Crystallization and Preliminary X-Ray Analysis of Human Uridine-Cytidine Kinase 2.” Acta Crystallographica - Section D Biological Crystallography 59 (8). International Union of Crystallography: 1477–78. doi:10.1107/S0907444903011533.

Suzuki, Nobuo N., Katsuhisa Koizumi, Masanori Fukushima, Akira Matsuda, and Fuyuhiko Inagaki. 2004. “Structural Basis for the Specificity, Catalysis, and Regulation of Human Uridine-Cytidine Kinase.” Structure 12 (5): 751–64. doi:10.1016/j.str.2004.02.038.