The mechanism of RNA 5’ capping with NAD⁺, NADH and desphospho-CoA

Jeremy G. Bird¹,2*, Yu Zhang², Yuan Tian³, Natalya Panova⁴, Ivan Barvšek⁵, Landon Greene⁶, Min Liu⁷, Brian Buckley⁷, Libor Krásný⁴, Jeehiun K. Lee⁵, Craig D. Kaplan⁸, Richard H. Ebright² & Bryce E. Nickels¹

The chemical nature of the 5’ end of RNA is a key determinant of RNA stability, processing, localization and translation efficiency¹,², and has been proposed to provide a layer of ‘epitranscriptomic’ gene regulation³. Recently it has been shown that some bacterial RNA species carry a 5’-end structure reminiscent of the 5’-7-methylguanylate ‘cap’ in eukaryotic RNA. In particular, RNA species containing a 5’-end nicotinamide adenine dinucleotide (NAD⁺) or 3’-desphospho-coenzyme A (dpCoA) have been identified in both Gram-negative and Gram-positive bacteria⁴–⁶. It has been proposed that NAD⁺, reduced NAD⁺ (NADH) and dpCoA caps are added to RNA after transcription initiation, in a manner analogous to the addition of 7-methylguanylate caps⁶–⁸. Here we show instead that NAD⁺, NADH and dpCoA are incorporated into RNA during transcription initiation, by serving as non-canonical initiating nucleotides (NCINs) for de novo transcription initiation by cellular RNA polymerase (RNAP). We further show that both bacterial RNAP and eukaryotic RNAP II incorporate NCIN caps, that promoter DNA sequences at and upstream of the transcription start site determine the efficiency of NCIN capping, that NCIN capping occurs in vivo, and that NCIN capping has functional consequences. We report crystal structures of transcription initiation complexes containing NCIN-capped RNA products. Our results define the mechanism and structural basis of NCIN capping, and suggest that NCIN-mediated ‘ab initio capping’ may occur in all organisms.

NAD⁺, NADH, and dpCoA are nucleotides that share an adenosine-diphosphate substructure with ATP (Fig. 1a, red). Therefore, in principle, NAD⁺, NADH, and dpCoA could serve in place of ATP as initiating nucleotides for de novo transcription initiation by a cellular RNAP, and thereby be incorporated ab initio into RNA (see Supplementary Discussion). To assess whether NAD⁺, NADH, and dpCoA serve as initiating nucleotides for a cellular RNAP, we performed in vitro transcription experiments with Escherichia coli RNAP and DNA templates containing promoters for RNAs identified as the most-highly NAD⁺-capped in vivo⁹ (Pₐgal and Pₐlac; Fig. 1b; Extended Data Fig. 1b). We performed reactions using ATP, NAD⁺, NADH, or dpCoA as initiating nucleotide and [α³²P]-CTP as extending nucleotide. In each case, we observed efficient formation of an initial RNA product (Fig. 1b; Extended Data Fig. 1b). RppH, which processes 5’-triphosphate and 5’-diphosphate RNAs to 5’-monophosphate RNAs⁹, processed the product obtained with ATP, but had no effect on products obtained with NAD⁺, NADH, and dpCoA (Fig. 1b; Extended Data Fig. 1b). Conversely, NudC, which processes 5’-capped RNAs to 5’-monophosphate RNAs⁹, had no effect on the product formed with ATP but processed products obtained with NAD⁺, NADH, and dpCoA (Fig. 1b; Extended Data Fig. 1b). Liquid chromatography/tandem mass spectrometry analysis confirmed the detection and structural assignment of the product of NAD⁺-mediated transcription initiation (Extended Data Fig. 2).

To assess whether initial RNA products formed with NCINs can be extended to full-length RNA products, we performed transcription reactions with ATP, NAD⁺, NADH, or dpCoA as initiating nucleotide and [α³²P]-UTP, CTP, and GTP as extending nucleotides on templates containing PrnaI linked to a 100-nucleotide cassette that did not contain any adenine bases. We observed production of full-length

---

¹Department of Genetics and Waksman Institute, Rutgers University, Piscataway, New Jersey 08854, USA. ²Department of Chemistry and Chemical Biology and Waksman Institute, Rutgers University, Piscataway, New Jersey 08854, USA. ³Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA. ⁴Institute of Microbiology, Czech Academy of Sciences, v.v.i., Vídeňská 1083, 142 20 Prague 4, Czech Republic. ⁵Institute of Physics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 5, 121 16 Prague 2, Czech Republic. ⁶Analytical and Bioanalytical Development, Bristol-Myers Squibb Company, One Squibb Drive, New Brunswick, New Jersey 08903, USA. ⁷Environmental and Occupational Health Sciences Institute, Rutgers University, Piscataway, New Jersey 08854, USA. ⁸Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843, USA. *These authors contributed equally to this work.
Figure 2 | Promoter-sequence effects on efficiency of NCIN-mediated transcription initiation. a, NCIN capping requires A₂₁. Top, promoters of NAD⁺-capped RNAs (promoter elements and start sites in grey). Bottom, initial RNA products of in vitro transcription reactions with ATP, NAD⁺, NADH, or dpCoA as initiating nucleotide and [α-3²P]-CTP as extending nucleotide (E. coli RNAP; WT (+1A), P₅₃₇ (upper) or P₅₃₇₋₁ (lower); Mut. (+1G), +1 G derivative of P₅₃₇ (upper) or P₅₃₇₋₁ (lower)). b, Promoter sequence determinants in addition to A₂₁ affect NCIN capping. Top, control +1A promoters. Bottom left, dependence of NAD⁺ capping on [NAD⁺]/[(ATP) ratio (mean ± s.e.m. of four determinations). Bottom right, relative efficiencies of NAD⁺ capping. c, Promoter position −1 affects NCIN capping. Top, P₅₃₇₋₁ (−1 in black). Other features as in b (mean ± s.e.m. of three determinations). For gel source data, see Supplementary Fig. 1.

products in all cases, and observed that full-length products with NAD⁺, NADH, and dpCoA contained NudC-sensitive, capped 5′-ends (Fig. 1c). Equivalent results were obtained using [²²P]-NAD⁺, rather than [α-3²P]-UTP, to detect products (Fig. 1d; Extended Data Fig. 3).

Because NAD⁺, NADH, and dpCoA have the same Watson–Crick base pairing preferences as ATP (Fig. 1a), NCIN-mediated initiation should occur only for RNAs transcribed from promoters that contain A:T at the transcription start site +1 (1A promoters). Consistent with this inference, we find that NCIN-mediated initiation occurs only with +1A promoters, and not with corresponding +1G promoter derivatives (Fig. 2a).

In principle, other aspects of promoter sequence, in addition to +1A at the transcription start site, could affect NCIN-mediated initiation. Consistent with this inference, we find that relative efficiencies of NAD⁺-mediated initiation and ATP-mediated initiation, (Kcat/Km, NAD⁺)/(Kcat/Km, ATP), differ over one to two orders of magnitude at different +1A promoters, with P₅₃₇ and P₅₃₇₋₁ exhibiting high relative efficiencies and P₅₂₅ and P₇₇₄₁ exhibiting low relative efficiencies (Figs 2a, b; Extended Data Figs 4 and 5). Substitution of the P₅₇₄₁ promoter position immediately upstream of the transcription start site, −1, results in an order-of-magnitude reduction in relative efficiency (Fig. 2c), indicating that position −1 is a key determinant of NAD⁺-mediated initiation.

Having obtained results establishing that NAD⁺, NADH, and dpCoA function as NCINs for transcription initiation in vitro and that promoter sequence determines the efficiency of NCIN-mediated initiation in vitro, we hypothesized that NCIN-mediated initiation also may occur in vivo and be responsible for the generation of NAD⁺-, NADH-, and dpCoA-capped RNAs in vivo. Consistent with this hypothesis, the promoters for full-length RNAs identified as enriched for NAD⁺-capping in vivo all are +1A promoters*, the class of promoters competent for NAD⁺-mediated initiation in vitro (Fig. 2a). Further consistent with this hypothesis, the promoters identified as most-highest enriched for NAD⁺-capping in vivo are P₅₃₇ and P₅₃₇₋₁, the two promoters with highest efficiencies of NAD⁺-mediated initiation in vitro (Fig. 2b). To show directly that NCIN-mediated initiation occurs in vivo, we assessed whether the promoter sequence dependence for NCIN-mediated initiation observed in vitro is observed in vivo. We fused promoters observed in vitro to have high (P₅₃₇) and low (P₇₇₄₁, P₅₃₇₋₁)
To define the structural basis of transcription initiation with ATP, we performed in vitro transcription experiments with Saccharomyces cerevisiae RNAP II using ATP, NAD$^+$, or NADH as initiating nucleotide and $\alpha$-UTP as extending nucleotide (Extended Data Fig. 6). The results matched those for bacterial RNAP: initiation occurred with NAD$^+$ and NADH, and we suggest that ab initio capping with NAD$^+$ and NADH also may occur in eukaryotes.

To establish whether NCIN capping has functional consequences in vivo, we measured stabilities of NCIN capped and uncapped RNA in vivo with a ΔNudC strain to eliminate NudC processing. We observe that NCIN capping results in a large, approximately three- to fourfold, increase in RNA stability (Fig. 3c), demonstrating a functional consequence of NCIN capping in vivo. We further observe that levels of NCIN capping are approximately twofold higher in stationary phase versus exponential phase (Fig. 3c), demonstrating a growth-phase dependence in NCIN capping.

All cellular RNAPs are members of a protein family having conserved structures and mechanisms.$^{10-12}$ To assess whether eukaryotic RNAP II, like bacterial RNAP, can use NCINs as initiating nucleotides, we performed in vitro transcription experiments with $\Delta$NudC protein and NudC-sensitive products in vivo (P$_{\text{nuc}}$, Fig. 3b), whereas the promoters with low efficiency of NAD$^+$-mediated initiation in vitro yielded low, undetectable, levels of NudC-sensitive products in vivo (P$_{\text{nuc}}$, Fig. 3b). We conclude that NCIN-mediated initiation occurs in vivo, and that most, or all, NCIN-capped RNA products in vivo arise from NCIN-mediated initiation.

Figure 4 | Structural basis of NCIN-mediated transcription initiation. a–c, Crystal structures of RPo-pppApC, RPo-NAD$^+$pC, and RPo-dpCoApC. Left, electron density and atomic model for initial RNA product. Green mesh, $F_o - F_c$ omit map (contoured at 2.5σ in a, b and 2.2σ in c); red, DNA; pink, RNA product and diphosphate in ‘E site’ (see refs 15–17); violet, spheres, Mg$^{2+}$ (I) and Mg$^{2+}$ (II); grey, RNAP bridge helix (BH). Right, contacts between RPo and initial RNA product. Green and orange, carbon and phosphorus atoms derived from initiating nucleotide; pink, atoms derived from extending nucleotide; red, DNA atoms and non-DNA oxygen atoms; blue, nitrogen atoms; grey sticks, RNAP carbon atoms; purple sphere, Mg$^{2+}$ (I); grey ribbon, RNAP bridge helix.
ATP and CTP, NAD\(^+\) and CTP, and dpCoA and CTP (Fig. 4; Extended Data Fig. 7; Extended Data Table 1). The results indicate that, in each case, the initiating entity—ATP, NAD\(^+\), or dpCoA—and an extending nucleotide (CTP) were able to react in *crystallo* to form an initial RNA product—pppApC, NAD\(^+\)pC, or dpCoApC, respectively—and that RNAP was able to translocate in *crystallo* by one base pair relative to the nucleic-acid scaffold, yielding a complex in a post-translocated state poised for RNA extension (Fig. 4). The results for ATP provide the first structural description of an initial product complex for a cellular RNAP. The results for NAD\(^+\) and dpCoA show graphically that NAD\(^+\) and dpCoA serve as initiating nucleotides in transcription initiation and define the interactions that the initial RNA products make with DNA and RNAP (see Supplementary Discussion).

Our results establish that NAD\(^+\), NADH, and dpCoA function as non-canonical initiating nucleotides (NCINs) for *de novo* transcription initiation and demonstrate the occurrence of *ab initio*, as opposed to post-transcription-initiation, capping of cellular RNAs.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 8 February; accepted 1 June 2016.**

**Published online 6 July 2016.**
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Proteins. E. coli RNA polymerase core was prepared from BL21(DE3) cells transformed with plasmid pV510 (gift from I. Artsimovich) as described in ref. 18. σ70 was prepared from BL21(DE3) cells transformed with plasmid pRc-70 (gift from J. Roberts) as described in refs 19, 20. RppH was purchased from NEB. NudC was purified from BL21(DE3) cells pET NudC-His, which was prepared as described previously21. Briefly, nudC was amplified by PCR from E. coli strain MG1655 using oligonucleotides J221 and J222. The PCR product was digested with XbaI and NotI and inserted into pET28c digested with XbaI and NotI. T. thermophilus RNA polymerase holozyme was prepared as described previously13. Saccharomyces cerevisiae RNA polymerase II was prepared as described previously21. MazF-mt3 protein, prepared as described previously22, was a gift from N. Woychik.

In vitro transcription assays with E. coli RNA polymerase. Linear transcription templates were synthesized by PCR using Phusion HF Polymerase master mix (Thermo) and oligonucleotides listed in Supplementary Table 1. PCR reactions contained 5 mM of the indicated forward and reverse primers. PCR products were purified using Qiagen PCR-clean up columns before use in transcription reactions.

Sequence of gadY (−65 to +35) template used in Figs 2a and 2b, and Extended Data Figs 4a and 3 and was prepared as described previously21. Briefly, nudC was amplified by PCR from E. coli strain MG1655 using oligonucleotides J221 and J222. The PCR product was digested with XbaI and NotI and inserted into pET28c digested with XbaI and NotI. T. thermophilus RNA polymerase holozyme was prepared as described previously13. Saccharomyces cerevisiae RNA polymerase II was prepared as described previously21. MazF-mt3 protein, prepared as described previously22, was a gift from N. Woychik.

In vitro transcription assays with E. coli RNA polymerase. Linear transcription templates were synthesized by PCR using Phusion HF Polymerase master mix (Thermo) and oligonucleotides listed in Supplementary Table 1. PCR reactions contained 5 mM of the indicated forward and reverse primers. PCR products were purified using Qiagen PCR-clean up columns before use in transcription reactions.

Sequence of gadY (−65 to +35) template used in Figs 2a and 2b, and Extended Data Figs 4a and 3 and was prepared as described previously21. Briefly, nudC was amplified by PCR from E. coli strain MG1655 using oligonucleotides J221 and J222. The PCR product was digested with XbaI and NotI and inserted into pET28c digested with XbaI and NotI. T. thermophilus RNA polymerase holozyme was prepared as described previously13. Saccharomyces cerevisiae RNA polymerase II was prepared as described previously21. MazF-mt3 protein, prepared as described previously22, was a gift from N. Woychik.

In vitro transcription assays with E. coli RNA polymerase. Linear transcription templates were synthesized by PCR using Phusion HF Polymerase master mix (Thermo) and oligonucleotides listed in Supplementary Table 1. PCR reactions contained 5 mM of the indicated forward and reverse primers. PCR products were purified using Qiagen PCR-clean up columns before use in transcription reactions.
Promoter—35 elements and 10 elements are in bold and underlined. DNA that directs synthesis of the reference RNA is coloured blue. Recognition site for MazF-mt3 (UCCUU; ref. 22) is underlined. Reference RNA includes sequence of the rT2 terminator.

For in vitro reactions shown in Fig. 3b, open complexes were formed using 10 mM of plasmid (pJB99, pJB91, or pJB95) and 50 nM core RNAP and 250 nM α32P-UTP in 30 μl transcription buffer (see above). Transcription was carried out for 10 min at 37 °C in 1 mM NAD+, 2.5 mM ATP, UTf, CTP and GTP. Reactions were stopped by addition of 120 μl of stop solution and samples were extracted with acid phenol:chloroform and ethanol precipitated. The resulting RNAs were resuspended in 10 mM Tris pH 8.0, supplemented with 2 μg of RNA isolated from wild-type MG1655 cells and treated with 2 μM MazF-mt3 for 1 h at 37 °C. Reactions were divided into two, 9 μl aliquots and 1 μl transcription buffer was added to each. 1 μl NudC was added to one of these aliquots and NudC storage buffer was added to the other. Reactions were incubated at 37 °C for 90 min and stopped by addition of an equal volume of gel loading buffer.

**Analysis of RNA products generated in vitro.** For the experiments of Fig. 3b, E. coli MG1655 cells containing plasmids pJB99, pJB91, or pJB95 were shaken at 220 c.p.m. at 37 °C in 25 ml LB (10 g Bacto-tryptone, 5 g Bacto-yeast extract and 10 g NaCl per litre) containing 25 μg ml−1 chloramphenicol in 125 ml flasks (Bellco). When cell density reached an OD600 ~ 0.6, 2 ml of the cell suspension was centrifuged (30 s, 10,000g at room temperature) to collect cells, supernatants were removed and cell pellets were rapidly frozen on dry ice and stored at −80 °C.

For the RNA stability analysis of Fig. 3c, strain MG1655 ΔnuD::C-Kan was constructed by P1 transduction of the ΔnuD::C-Kan cassette from JW5348 into MG1655. Strain MG1655 ΔnuD::C-Kan cells containing plasmids pJB99, pJB91, or pJB95 were grown as described above to an OD600 ~ 3.5 (stationary phase). 2 ml of the cell suspension was collected as described above (the RNA derived from these cells was used as the 0 min time point). Rifampicin was added to a final concentration of 2 μg ml−1 to the remainder of the culture. 2 ml aliquots were taken at the indicated times and cells were collected as described above. Frozen pellets were resuspended in 1 ml of TRI Reagent solution (Molecular Research Center). Samples were incubated at 70 °C for 10 min and then centrifuged (10 min, 21,000g, 4 °C) to remove insoluble material. The supernatant was transferred to a fresh tube and 200 μl of chloroform was added, samples were mixed by vortexing, and then centrifuged (10 min, 21,000g, 4 °C). The aqueous phase was transferred to a fresh tube, extracted with acid phenol:chloroform, and RNA transcripts were recovered by ethanol precipitation and resuspended in 10 mM Tris, pH 8.0.

5 μg of isolated RNA was treated with 2 μl MazF-mt3 in 10 mM Tris, pH 8.0 (total volume, 25 μl) and incubated for 60 min at 37 °C. Reactions were stopped by addition of an equal volume of gel loading buffer. In cases where RNA products were treated with NudC, reactions were divided into two aliquots and 10 μl transcription buffer was added to each to a final concentration of 1 × 1 μl NudC was added to one of these aliquots and NudC storage buffer was added to the other. Reactions were incubated at 37 °C for 45 min and stopped by addition of an equal volume of gel loading buffer.

**Detection of RNA products generated in vitro or in vivo by hybridization.** In Fig. 3b, c, RNA products were detected by a procedure comprising electrophoresis on 10% 7.5 M urea slab gels (equilibrated and run in 1 x TBE, transfer to and UV cross-linking to a membrane (Nytran supercharge nylon membrane, GE Healthcare Life Sciences), hybridization with a 32P-labelled ‘locked-nucleic-acid’ probe complementary to positions +31 to +42 of the reference RNA sequence, agCaaAttAaaC; LNA bases capitalized; purchased from Exiqon; 32P-labelled using T4 polynucleotide kinase), high stringency washing (procedure as described in ref. 25), and storage-phosphor imaging. Bands were quantified using ImageQuant software. The percentage of RNA products in the upper band was determined for each sample as: 100 × [upper band] / [upper band + lower band]. Values of the per cent of NudC-sensitive products represent the reduction in the per cent of RNA products in the upper band upon treatment with NudC. Values shown represent the average of three technical replicates (in vitro) or three biological replicates (in vivo).

For the experiments in Fig. 3c, values of half-life were determined by fitting data to a single-exponential decay function (mean ± s.e.m. of three biological replicates for exponential phase and five biological replicates for stationary phase).

**In vitro transcription assays with E. coli RNAP; liquid chromatography/tandem mass spectrometry (LC/MS/MS) detection of RNA·PC.** Open complexes formed as described above were mixed with 2 mM NAD+ and 10 μM CTP, incubated at 37 °C for 1 h, passed through a Nanosep Centrifugal Device, and the flow-through was analysed by LC/MS/MS. (Control samples in which NAD+, CTP, or RNA polymer were not present in the reactions were also prepared.) LC/MS/MS analyses were performed in negative ion mode with a Finnigan LTQ mass spectrometer equipped with an electrospray ionization (ESI) interface coupled with a Finnigan Surveyor HPLC system. The flow rate from the Finnigan Surveyor pump was 0.3 ml min−1. The autosampler temperature was 5 °C. The liquid chromatography method used a YMC ODS-A 5 μm particle size 120 Å pore size column, 3.0 mm × 100 mm. The samples were separated using a gradient mobile phase consisting of 5 mM ammonium formiate buffer to pH 7.9 in water (A) and methanol (B). The gradient condition was: 0–5 min, 100% A; 5–15 min, 100–30% A; 15–23 min, 30–0% A; 23–28 min, 0% A; 28–40 min, 100% A. Column temperature was 25 °C. The temperature of the heated capillary was 200 °C. Fragmentation was activated by collision-induced dissociation (CID) with a collision energy of 20–25%. The search for NAD+·PC was conducted by isolating its ion ((m−2H)+ at m/z 967) at isolation width 3 using full scan mode, and analysing its fragmentation spectrum (collision energy 20%) from m/z 210 to m/z = 1000. The characteristic fragment m/z = 845 (loss of nicotinamide) was used to identify NAD+·PC. The instrument control, data acquisition, and data analysis were performed by Xcalibur software. Ammonium formiate (99%), water (HPLC grade) and methanol (HPLC grade) were used in these LC/MS/MS experiments were purchased from Sigma-Aldrich.

**In vitro transcription assays with Saccharomyces cerevisiae RNAP II.** Bacterial templates for initiation were used following the approach in ref. 26 by mixing 2.5 mM of oligonucleotide CKO1639 and 2.5 mM of oligonucleotide CKO1621 (see Supplementary Table 1) in 100 μl of annealing buffer (40 mM Tris, pH 8.0, 50 mM NaCl, 1 mM EDTA), incubating at 95 °C for 3 min, and slowly cooling to 23 °C (1 °C min−1) to allow annealing. The oligonucleotide mixture was run on a 2.5% agarose gel (Apex, Genesee) equilibrated and run in 1 x TBE (90 mM Tris base (BioRad), 90 mM boric acid (Calbiochem), 2 mM EDTA (JT Baker)), the gel was stained with ethidium bromide and the band corresponding to the bubble template was excised from the gel, crushed, and incubated with 500 μl elution buffer (500 mM NH4OAc, 10 mM Mg(OAc), 1 mM EDTA, 0.1% SDS). Gel debris was removed using a Spin-X column (Costar) and nucleic acids were washed following magnetic capture twice with 200 μl 10 mM Tris, pH 8.0, supplemented with 2.5 μM MazF-mt3 in 10 mM Tris, pH 8.0. The samples were separated using a gradient mobile phase consisting of 5 mM ammonium formiate buffer to pH 7.9 in water (A) and 50% acetonitrile (B) at a flow rate of 0.1 ml min−1. The characteristic fragment m/z 845 (loss of nicotinamide) was used to identify NAD+·PC. The instrument control, data acquisition, and data analysis were performed by Xcalibur software. Ammonium formiate (99%), water (HPLC grade) and methanol (HPLC grade) were used in these LC/MS/MS experiments were purchased from Sigma-Aldrich.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
ATP, 2 mM CTP, and 17.5% (v/v) (2R, 3R)-(−)-2,3-butanediol (Sigma-Aldrich) and were flash-cooled with liquid nitrogen.

Diffraction data were collected at BNL beamline X29A (temperature, 100 K; wavelength, 1.08 Å), processed and scaled using HKL200028. Structure factors were converted using the French-Wilson algorithm in Phenix29 and were subjected to anisotropy correction using the UCLA MBI Diffraction Anisotropy server30 (http://services.mbi.ucla.edu/anisoscale/). The structure was solved by molecular replacement with Molrep31 using one RNA molecule from the structure of T. thermophilus RPo (PDB 4G7H)13 as the search model. Early-stage refinement included rigid-body refinement of each RNA molecule, followed by rigid-body refinement of each subunit of RNA molecules. Cycles of iterative model building with Coot32 and refinement with Phenix33 were performed. Atomic models of the DNA non-template strand, the DNA template strand, and RNA were built into mF0 – DF omit maps, and subsequent cycles of refinement and model building were performed. The final crystallographic model of RPo-pppApC, refined to Ramachandran statistics of 97.81% favoured, 2.15% allowed, and 0.04% outliers, has been deposited in the PDB with accession code 5D4D.

Structure determination: RPo-NAD+′pc. NAD+ (Sigma-Aldrich) and CTP were soaked into RPo crystals (prepared as described above) by addition of 0.2 μl 30 mM NAD+ and 40 mM CTP in 50% (v/v) RB to the crystallization drop, and incubated for 2 h at 22 °C. Crystals were transferred into reservoir solutions containing 1.5 mM NAD+, 2 mM CTP, and 17.5% (v/v) (2R, 3R)-(−)-2,3-butanediol and were flash-cooled with liquid nitrogen.

Diffraction data were collected at APS beamline 19-ID-D (temperature, 100 K; wavelength, 0.98 Å), processed and scaled using HKL200028, and subjected to anisotropic correction using the UCLA MBI Diffraction Anisotropy server30 (http://services.mbi.ucla.edu/anisoscale/). The structure was solved and refined using procedures analogous to those described above for RPo-pppApC. The final crystallographic model of RPo-NAD+′pc, refined to Ramachandran statistics of 97.81% favoured, 2.15% allowed, and 0.04% outliers, has been deposited in the PDB with accession code 5D4D.

Structure determination: RPo-dpCoApC. dpCoA (Sigma-Aldrich) and CTP were soaked into RPo crystals (prepared as described above) by addition of 0.2 μl 60 mM 3′-desphospho-CoA and 40 mM CTP in 50% (v/v) RB to the crystallization drop, and incubated for 2 h at 22 °C. Crystals were transferred into reservoir solutions containing 3 mM 3′-desphospho-CoA, 2 mM CTP, and 17.5% (v/v) (2R, 3R)-(−)-2,3-butanediol and were flash-cooled with liquid nitrogen.

Diffraction data were collected at APS beamline 19-ID-D (temperature, 100 K; wavelength, 0.98 Å), processed and scaled using HKL200028, and subjected to anisotropic correction using the UCLA MBI Diffraction Anisotropy server30 (http://services.mbi.ucla.edu/anisoscale/). The structure was solved and refined using procedures analogous to those described above for RPo-pppApC. The final crystallographic model of dpCoApC, refined to Ramachandran statistics of 97.54% favoured, 2.43% allowed, and 0.03% outliers, has been deposited in the PDB with accession code 5D4E.

Analysis of dpCoA by liquid chromatography (LC)-UV. AMP (Sigma-Aldrich) and dpCoA were diluted to 1 mM in water (HPLC grade) before injection. LC-UV analyses were performed using a Waters 2960 HPLC system coupled with a Waters 2996 photodiode array detector. The flow rate from the Finnigan Surveyor pump was 0.3 ml min⁻¹. The autosampler temperature was 5 °C. The LC method used an YMC ODS-A 5μm particle size 120 Å pore size column, 3.0 mm x 100 mm.

The samples were separated using a gradient mobile phase consisting of 25 mM triethylammonium formate buffer to pH 3.0 in water (A) and methanol (B). The gradient condition was: 0–5 min, 100% A; 5–15 min, 100–30% A; 15–23 min, 30–0% A; 23–28 min, 0% A; 28–40 min, 100% A. Column temperature was 25°C. Detector wave length was 260 nm. The instrument control, data acquisition, and data analysis were performed by MassLynx software. Triethylammonium acetate buffer (2.0 M) and formic acid (>95%) were purchased from Sigma-Aldrich.

18. Artsimovitch, I., Svetlov, V., Murakami, K. S., & Landick, R. Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. J. Biol. Chem. 278, 12344–12355 (2003).
19. Marr, M. T. & Roberts, J. W. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science 276, 1258–1260 (1997).
20. Perdue, S. A. & Roberts, J. W. A backtrack-inducing sequence is an essential component of Escherichia coli σ70-dependent promoter-proximal pausing. Mol. Microbiol. 78, 636–650 (2010).
21. Kaplan, C. D., Larsson, K. M. & Kornberg, R. D. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by α-amanitin. Mol. Cell 30, 547–556 (2008).
22. Schifano, J. M. et al. An RNA-seq method for defining endonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. Nat. Commun. 5, 3538 (2014).
23. Vvedenskaya, I. O. et al. Massively systematic transcript end readout, “MASTER”: transcription start site selection, transcriptional slippage, and transcript yields. Mol. Cell 60, 953–965 (2015).
24. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 0008 (2006).
25. Goldman, S. R., Nair, N. U., Wells, C. D., Nickels, B. E. & Hochschild, A. The primary + factor in Escherichia coli can access the transcription elongation complex from solution in vivo. eLife 4, 4 (2015).
26. Čabart, P., Jin, H., Li, L. & Kaplan, C. D. Activation and reactivation of the RNA polymerase II trigger loop for intrinsic RNA cleavage and catalysis. Transactions 5, e28869 (2014).
27. Murakami, K. S. Structural biology of bacterial RNA polymerase. Biomolecules 5, 848–864 (2015).
28. Olwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
29. French, S. & Wilson, K. On the treatment of negative intensity observations. Acta Crystallogr. A 34, 517–525 (1978).
30. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPPE protein complex from Mycobacterium tuberculosis. Proc. Natl Acad. Sci. USA 103, 8060–8065 (2006).
31. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. J. Appl. Crystallogr. 30, 1022–1025 (1997).
32. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010).
33. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).
Extended Data Figure 1 | De novo transcription initiation by ATP and NCINs. a, Structures of ATP, NAD\(^+\), NADH, and dpCoA. Red, identical atoms. b, Initial RNA products of in vitro transcription reactions with ATP, NAD\(^+\), NADH, or dpCoA as initiating nucleotide and \([\alpha^32P]\)-CTP as extending nucleotide (E. coli RNAP, \(P_{\text{rnlI}}\); see analogous data for \(P_{\text{gadY}}\) in Fig. 1b). Products were treated with RppH (processes 5\(^\prime\)-triphosphate RNA to 5\(^\prime\)-monophosphate RNA and 5\(^\prime\)-NTP to 5\(^\prime\)-NDP/5\(^\prime\)-NMP\(^9,14\)) or NudC (processes 5\(^\prime\)-NAD\(^+\)/NADH-capped RNA to 5\(^\prime\)-monophosphate RNA\(^9\)) as indicated. For gel source data, see Supplementary Fig. 1.
Extended Data Figure 2 | LC/MS/MS analysis of initial RNA products of in vitro transcription reactions with NAD$^+$ as initiating nucleotide and CTP as an extending nucleotide. **a**, Structure of NAD$^+$pC (red, atoms corresponding to CID-generated fragment ion). **b**, Extracted ion chromatogram (signal derived from detection of parent ion of $m/z = 967$ and CID fragment of $m/z = 845$ corresponding to NAD$^+$pC minus nicotinamide). Reactions contained the indicated components. **c**, Mass spectrum of CID fragment.
Extended Data Figure 3 | Sensitivity of full-length RNA products to alkaline phosphatase treatment. Full-length RNA products of *in vitro* transcription reactions with $[^{32}\text{P}]-\text{ATP}$ or $[^{32}\text{P}]-\text{NAD}^+$ as initiating nucleotide and CTP, GTP, and UTP as extending nucleotides (*E. coli* RNAP; $P_{\text{RNAI}}$ fused to an A-less cassette). Products were treated with alkaline phosphatase (AP; processes 5′ phosphates) or NudC (processes 5′-NAD$^+$/NADH-capped RNA to 5′-monophosphate RNA) as indicated. Results indicate that full-length RNA products generated in reactions with $[^{32}\text{P}]-\text{NAD}^+$ as initiating nucleotides are not sensitive to AP until they are processed by NudC. M, 100-nucleotide marker. For gel source data, see Supplementary Fig. 1.
Extended Data Figure 4 | Promoter-sequence effects on efficiency of NCIN-mediated transcription initiation: NAD\(^+\). a. Templates having \(rnia\), \(gadY\), \(N25\), and \(T7A1\) promoters used in the assays. b. Representative raw data from experiments of Fig. 2b. Initial RNA products of \textit{in vitro} transcription reactions performed in the presence of 50 \(\mu\)M ATP and 1 mM NAD\(^+\) as initiating nucleotides and \([\alpha^{32}\text{P}]-\text{CTP}\) as extending nucleotide (\textit{E. coli} RNAP; \(P_{rnia}\), \(P_{gadY}\), \(P_{N25}\), or \(P_{T7A1}\)). (We note that contaminating AMP in the NAD\(^+\) stock results in production of pAp\(^+\).) For gel source data, see Supplementary Fig. 1.
Extended Data Figure 5 | Promoter-sequence effects on efficiency of NCIN-mediated transcription initiation: NADH and dpCoA.  

a, Left, dependence of NADH capping on [NADH]/[ATP] ratio (mean ± s.e.m. of three determinations). Right, relative efficiencies of NADH capping. (E. coli RNAP; P$_{\text{mal}}$, P$_{\text{gadY}}$, P$_{\text{N25}}$, or P$_{\text{T7A1}}$).  

b, Left, dependence of dpCoA-capping on [dpCoA]/[ATP] ratio (mean ± s.e.m. of three determinations). Right, relative efficiencies of dpCoA capping. (E. coli RNAP; P$_{\text{mal}}$, P$_{\text{gadY}}$, P$_{\text{N25}}$, or P$_{\text{T7A1}}$).
Extended Data Figure 6  |  NCIN-mediated de novo transcription initiation by eukaryotic RNAP II. Initial RNA products of \textit{in vitro} transcription reactions with ATP, NAD\(^+\), or NADH as initiating nucleotide and \([\alpha^{32}\text{P}]-\text{UTP}\) as extending nucleotide. Reactions were performed with yeast RNAP II and an artificial bubble transcription initiation template. Products were treated with RppH or NudC as indicated. For gel source data, see Supplementary Fig. 1.
Extended Data Figure 7 | Structural basis of NCIN-mediated transcription initiation: stereoviews. a–c, Crystal structures of RPo-pppApC, RPo-NAD$^+$pC, and RPo-dpCoApC. Stereo views of density and fit for initial RNA product. Green mesh, $F_o - F_c$ omit map (contoured at 2.5$\sigma$ in a, b and 2.2$\sigma$ in c); red, DNA; pink, RNA product and diphosphate in 'E site' (see refs 15–17); violet spheres, Mg$^{2+}$ (I) and Mg$^{2+}$ (II); grey, RNAP bridge helix.
Extended Data Figure 8 | AMP content of dpCoA stock. HPLC chromatogram of dpCoA stock (Sigma-Aldrich, lot SLBJ2886V; 50 nmol). Green, HPLC chromatogram of AMP (20 nmol). Comparison of chromatograms indicates that the dpCoA stock contains ~2% AMP. The observation that the dpCoA stock contains ~2% AMP in the dpCoA stock accounts for the formation of pApC in reactions performed with dpCoA (Fig. 1b).
## Extended Data Table 1 | Data collection and refinement statistics

|                       | RPo-pppApC | RPo-NAD\textsuperscript{+}pC | RPo-dpCoApC |
|-----------------------|------------|-------------------------------|-------------|
| **Data collection**   |            |                               |             |
| Space group           | P2(1)      | P2(1)                         | P2(1)       |
| Cell dimensions       |            |                               |             |
| $a$, $b$, $c$ (Å)     | 186.0, 103.6, 297.4 | 185.0, 103.5, 296.3 | 185.8, 103.9, 297.1 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90.0, 98.3, 90.0 | 90.0, 98.3, 90.0 | 90.0, 98.5, 90.0 |
| Resolution (Å)        | 50.00-3.30(3.36-3.30) * | 50.00-3.00(3.05-3.00) * | 40.00-3.10(3.15-3.10) * |
| $R_{\text{sym}}$ or $R_{\text{merge}}$ | 0.185(0.954) | 0.119(0.920) | 0.120(0.840) |
| $I/\sigma(I)$         | 8.7(1.7)   | 10.6(1.5)                    | 10.4(1.4)   |
| Completeness (%)      | 0.961(0.875) | 0.979(0.934) | 0.989(0.960) |
| Redundancy            | 5.4(5.0)   | 4.0(3.7)                     | 3.8(3.5)    |
| **Refinement**        |            |                               |             |
| Resolution (Å)        | 50.00-3.30 | 50.00-3.00                    | 40.00-3.10  |
| No. reflections       | 154105     | 218773                       | 187101      |
| $R_{\text{work}}$ / $R_{\text{free}}$ | 0.211/0.258 | 0.201/0.254 | 0.210/0.247 |
| No. atoms             |            |                               |             |
| Protein               | 55006      | 54790                        | 55297       |
| Ligand/ion            | 1536       | 1664                         | 1651        |
| Water                 | 58         | 895                          | 775         |
| **B-factors**         |            |                               |             |
| Protein               | 48.0       | 31.2                         | 49.1        |
| Ligand/ion            | 68.0       | 65.3                         | 80.4        |
| Water                 | 12.9       | 11.7                         | 27.4        |
| **R.m.s deviations**  |            |                               |             |
| Bond lengths (Å)      | 0.004      | 0.008                        | 0.004       |
| Bond angles (°)       | 0.79       | 0.95                         | 0.69        |

*Highest resolution shell is shown in parentheses.