



RNA polymerase II trapped on a molecular treadmill: Structural basis of persistent transcriptional arrest by a minor groove DNA binder

Juntaek Oh^a, Tiezheng Jia^b, Jun Xu^a, Jenny Chong^a, Peter B. Dervan^{b,1}, and Dong Wang^{a,c,d,1}

^aDivision of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093; ^bDivision of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; ^cDepartment of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92093; and ^dDepartment of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093

Edited by Seth Darst, Rockefeller University; received July 29, 2021; accepted December 7, 2021 by Editorial Board Member Graham C. Walker

Elongating RNA polymerase II (Pol II) can be paused or arrested by a variety of obstacles. These obstacles include DNA lesions, DNA-binding proteins, and small molecules. Hairpin pyrrole-imidazole (Py-Im) polyamides bind to the minor groove of DNA in a sequence-specific manner and induce strong transcriptional arrest. Remarkably, this Py-Im-induced Pol II transcriptional arrest is persistent and cannot be rescued by transcription factor TFIIS. In contrast, TFIIS can effectively rescue the transcriptional arrest induced by a nucleosome barrier. The structural basis of Py-Im-induced transcriptional arrest and why TFIIS cannot rescue this arrest remain elusive. Here we determined the X-ray crystal structures of four distinct Pol II elongation complexes (Pol II ECs) in complex with hairpin Py-Im polyamides as well as of the hairpin Py-Im polyamides-dsDNA complex. We observed that the Py-Im oligomer directly interacts with RNA Pol II residues, introduces compression of the downstream DNA duplex, prevents Pol II forward translocation, and induces Pol II backtracking. These results, together with biochemical studies, provide structural insight into the molecular mechanism by which Py-Im blocks transcription. Our structural study reveals why TFIIS fails to promote Pol II bypass of Py-Im-induced transcriptional arrest.

RNA polymerase II | transcription blockage | minor groove | pyrrole-imidazole polyamide | transcription factor

Eukaryotic RNA polymerase II (Pol II) transcription elongation process is subject to pausing and arrest by various regulatory factors and obstacles, such as pause-inducing DNA sequences or secondary structures, DNA modifications, DNA lesions, DNA-binding proteins, or DNA-binding small molecules (1–14). The fate of Pol II is determined by the nature of obstacles and whether this Pol II pausing/arrest can be resolved by other transcription factors in a timely manner. For example, Pol II can bypass natural DNA-pausing sequences, small DNA lesions or modifications, and nucleosome barriers, with the help of other transcription factors, such as Spt4/5, TFIIS, or CSB (5, 7, 15–21). In contrast, Pol II gets arrested by helix distorting bulky DNA lesions (8, 16, 22–28) and certain types of DNA-binding small molecules, such as pyrrole-imidazole polyamides (Py-Im) (18, 29). Intriguingly, transcription factors TFIIS and Spt4/5 cannot rescue the prolonged Pol II arrest induced by Py-Im (18, 29).

The binding of Py-Im to its target DNA sequences has several features. The sequence selectivity of Py-Im is governed by side-by-side arrangement of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) and the corresponding functional groups on the DNA minor groove floor (30–32). For example, Im/Py recognizes the G/C base pair while Py/Py recognizes A/T or T/A base pairs (30). Hairpin Py-Im oligomers bind to target DNA sequences with high affinity (at nanomolar range) comparable to transcription factors and nucleosomes (30). Hairpin Py-Im oligomers function as a molecular wedge causing the expansion of the minor groove and compression of the major groove of DNA (33, 34).

These features of Py-Im oligomers led to several studies on the inhibition of specific gene expression in colon, cervical, and prostate cancer (35–40). For example, the Py-Im polyamide ARE-1 targets androgen receptor (AR) consensus element and directly competes with AR to decrease the occupancy of AR in the KLK3 promoter and enhancer (37). As a result, ARE-1 can greatly reduce the expression level of KLK3 in enzalutamide-resistant LNCaP cells (37). Similarly, Py-Im 1 targets estrogen response elements and reduces the expression of estrogen receptor-controlled luciferase in breast cancer tumor xenografts (36). In addition to targeting the promoter and enhancer regions, our previous results showed that Py-Im induces strong transcriptional arrest of the elongation complex (EC) in a sequence-specific manner, paving the way to target transcription of specific genes (16, 18, 29). Indeed, Py-Im can also be used to target cancer-specific coding region, such as the gene body of cancer-driving mutant E545K in PIK3CA, a hotspot mutation that occurs in 23 to 36% of cervical cancer cases (41). The alkylating polyamide P3AE5K binds to the coding region of E545K mutation and specifically reduces the expression level of PIK3CA E545K mRNA and protein, leading to apoptotic cell death (38).

Significance

Hairpin pyrrole-imidazole (Py-Im) polyamides can be programmed to bind a broad repertoire of DNA sequences. Py-Im small molecules can be used to target cancer-specific coding regions and block transcription elongation. This transcription blockage by Py-Im cannot be rescued by transcription elongation factors, such as TFIIS. The mechanism by which Py-Im blocks transcription remains elusive. To understand the structural basis of this strong transcription blockage, we solved five different structures containing an eight-ring hairpin Py-Im bound with either a Pol II elongation complex (EC) or a DNA duplex. These structures revealed that Py-Im can trap Pol II EC in a treadmill-like manner. This knowledge may pave the way for the development of small molecules that inhibit transcriptional addiction in cancer.

Author contributions: J.O., P.B.D., and D.W. designed research; J.O. and J.X. performed research; T.J. and J.X. contributed new reagents/analytic tools; J.O., P.B.D., and D.W. analyzed data; and J.O., J.C., P.B.D., and D.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. S.D. is a guest editor invited by the Editorial Board.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: dongwang@ucsd.edu or dervan@caltech.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2114065119/-DCSupplemental>.

Published January 12, 2022.

The structural mechanism of Py-Im-induced persistent transcriptional arrest remains elusive. In this study, we solved seven X-ray crystal structures, including a double-stranded DNA complexed with hairpin Py-Im 1 and four different Pol II elongation complexes bound with Py-Im 1. Our structural and functional analyses showed that hairpin Py-Im 1 traps Pol II elongation complex in the n-5 to n-3 position, which forces the elongation complex to be trapped on a futile “treadmill” where Pol II moves back and forth by repetitive extension and cleavage in the presence of TFIIS.

Results and Discussion

Py-Im Induces Strong Transcriptional Pausing at Both Full-Bubble Scaffold and Miniscaffold. We previously showed that hairpin Py-Im 1 (Fig. 1A) induces consecutive Pol II pausing/arrest using the scaffolds containing a fully matched transcription bubble (full-bubble scaffolds) (29). These featuring pausing/arrest bands are located upstream of the actual Py-Im-binding site (i.e., from the n-5 to n-2 positions, where n is the first 5'-base of the hairpin Py-Im-binding site). As the first step toward obtaining the structural insights into Py-Im-induced Pol II arrest, we tested whether we could recapitulate these Py-Im-induced pausing/arrests using the miniscaffold for structural studies. To this end, we designed a miniscaffold harboring a hairpin Py-Im 1 binding sequence (5'-TGACCA-3') (Fig. 1B). This template strand (TS)-binding orientation puts the γ -turn moiety of hairpin Py-Im 1 facing the leading edge of transcribing Pol II (Fig. 1B). We performed transcription assays with two different primers (scaffold-1; 9mer RNA, n-5 and scaffold-2; 10mer RNA, n-4) and found that Py-Im 1 induces strong transcriptional pausing/arrest at the miniscaffold in a similar manner as that with the full-bubble scaffolds (Fig. 1B) (29). Among all pausing/arrest bands, the n-3 is the strongest pausing/arrest site (Fig. 1B).

Crystal Structure of a Hairpin Py-Im-dsDNA Complex at 1.8-Å Resolution. The detailed atomic interaction between hairpin Py-Im 1 and DNA duplex is obscured due to a lack of high-resolution crystal structure of hairpin Py-Im 1 bound with dsDNA

complex. Here we obtained a high-resolution crystal structure of the hairpin Py-Im 1-dsDNA (5'-CCTGGTCAGG-3') complex at 1.8-Å resolution (Fig. 1C and *SI Appendix*, Fig. S1A). We observed an intensive hydrogen bonding network that stabilizes the hairpin Py-Im 1-dsDNA complex (*SI Appendix*, Fig. S1B). Our structure shows the conserved base recognition pattern between Py-Im and DNA base pairs: Im/Py recognizes G/C and Py/Py recognizes A/T or T/A (*SI Appendix*, Fig. S1B) (33, 34, 42, 43). The binding of hairpin Py-Im 1 leads to minor groove widening as well as major groove compression of the DNA duplex in a similar manner as cyclic Py-Im molecules (*SI Appendix*, Fig. S1C) (33, 34, 42, 43). This high-resolution structure hairpin Py-Im-dsDNA complex was used as a starting model for building the downstream Py-Im-dsDNA region of the Pol II EC-Py-Im complex and paved the way for our structural studies of Py-Im-induced Pol II-arrested complexes.

Comparison of Hairpin Py-Im-dsDNA Complex and Pol II-Py-Im Encounter Complex. To understand the molecular mechanism of Py-Im-induced transcription pausing/arrest in a stepwise manner, we determined six Pol II elongation complex structures. These structures include four distinct Py-Im-bound Pol II elongation complexes and two elongation complexes (*SI Appendix*, Table S1). We observed substantial structural rearrangements of Pol II-Py-Im complexes induced by Py-Im 1.

We first solved the Pol II-Py-Im paused complex at the n-5 position (scaffold-1), referred to as the Pol II-Py-Im encounter complex. The encounter complex represents the structural snapshot of when Pol II first senses the approaching Py-Im molecule (Fig. 2A and B). By comparing the Pol II-Py-Im encounter complex and the Py-Im-dsDNA complex, we observed substantial changes in terms of molecular interactions and structural rearrangement. First, the molecular interaction network surrounding the γ -turn moiety of hairpin Py-Im 1 is very different. In the Py-Im-dsDNA structure, we found that the hydrogen bonding network interactions between the γ -turn moiety of hairpin Py-Im 1 (α -ammonium tip) and the DNA minor groove are mediated by two ordered water molecules

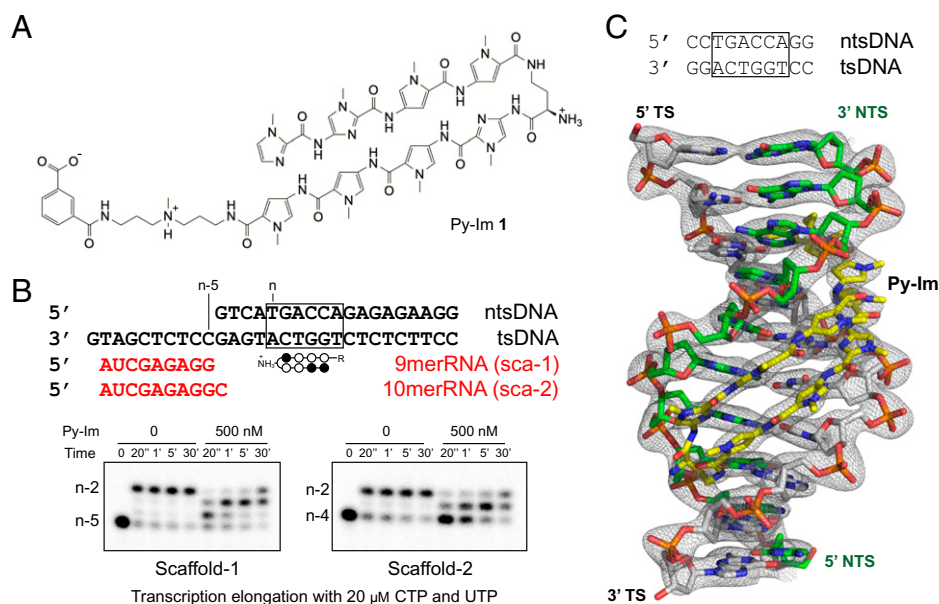


Fig. 1. Structural and functional analysis of hairpin Py-Im 1. (A) Chemical structure of hairpin Py-Im 1. (B) Py-Im blocks transcription elongation. Miniscaffolds used for the biochemical assay (Top). Transcription elongation assays were performed with scaffolds containing either a 9mer RNA (Bottom Left, scaffold-1) or a 10mer RNA (Bottom Right, scaffold-2). CTP and UTP were added for transcription elongation. Strong pausing at n-4 and n-3 positions was observed in the presence of Py-Im. (C) The crystal structure of Py-Im-DNA duplex complex at 1.8-Å resolution. The 2Fo-Fc electron density map is contoured at 1.3 σ . DNA sequence with the Py-Im-binding site highlighted in the black box (Top).

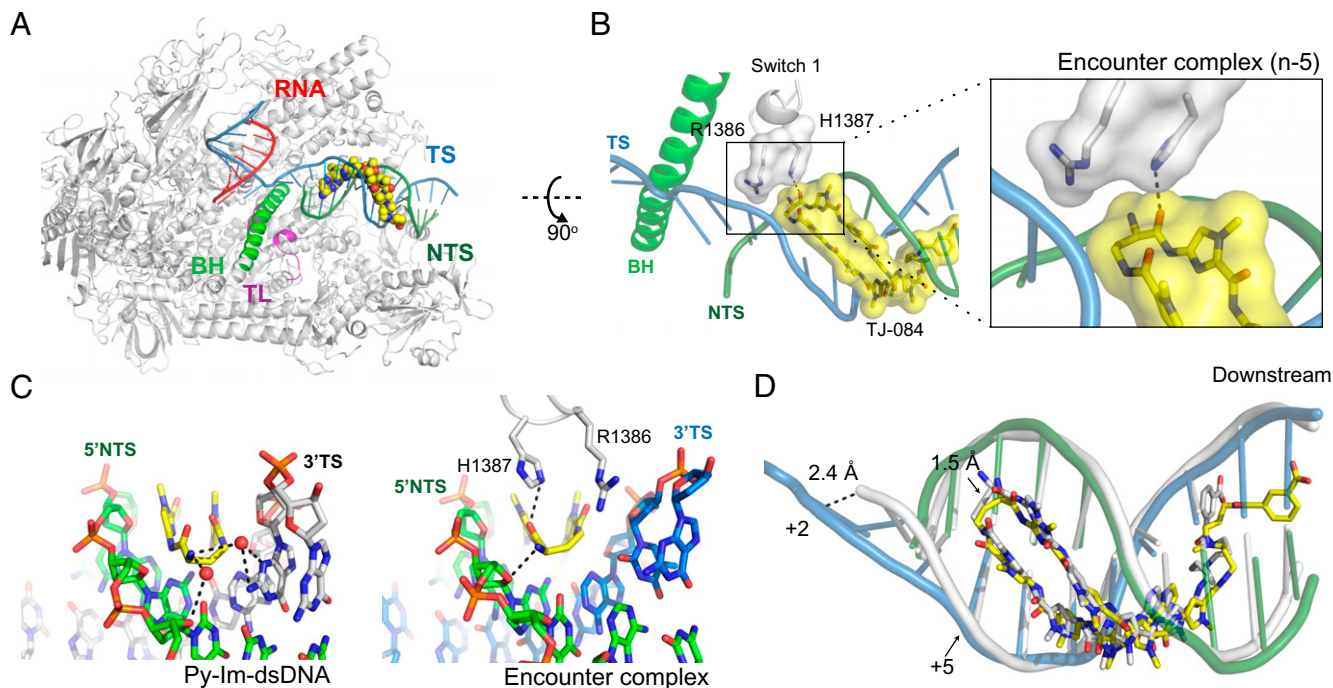


Fig. 2. Structural comparison of the Py-Im-dsDNA and Pol II-Py-Im encounter complex. (A) Overall structure of the Pol II-Py-Im encounter complex. Pol II is shown in gray (with part of Rpb2 omitted for clarity). RNA (red), template DNA strand (TS, blue), nontemplate DNA strand (NTS, green), bridge helix (BH, green helix), and trigger loop (TL, magenta) are highlighted and labeled. Py-Im molecule is shown as spheres. (B) Rotated, enlarged view showing interaction between Py-Im (yellow surface) and the switch 1 motif of Pol II (white surface). Hydrogen bond is shown in black dashed line. (C) Hydrogen bond interaction network in the Py-Im-dsDNA (Left) and Pol II-Py-Im encounter complex (Right). (Left) Water-mediated hydrogen bonding between (R)- α -amine- γ -turn tip of Py-Im and DNA minor groove in Py-Im-dsDNA. (Right) (R)- α -amine- γ -turn tip of Py-Im is sandwiched with the DNA minor groove and Pol II via direct hydrogen bonding in the Pol II-Py-Im encounter complex. (D) Structure superposition of the Py-Im-dsDNA complex and Py-Im encounter complex. Least squares fit superpose was applied using NTS sequence 5'-TGACCA-3' in COOT (52).

(Fig. 2C). By contrast, in the Pol II-Py-Im encounter complex structure, the γ -turn moiety of hairpin Py-Im 1 is sandwiched by two direct hydrogen bonds with DNA and the side chain of Pol II Rpb1 His1387 from the switch region (Fig. 2C). This structure confirms our prediction from previous functional and modeling analyses that the Rpb1 switch region is involved in the interaction with the approaching γ -turn of Py-Im (29). Second, we noticed a 2.4-Å shift of the DNA backbone of a nontemplate strand (NTS) nearby the n-5 position in the Pol II-Py-Im encounter complex in comparison with that in the Py-Im-dsDNA complex (Fig. 2D). Third, we observed an approximate 1.5- to 1.9-Å shift of the Py-Im molecule toward the downstream DNA in the Pol II-Py-Im encounter complex structure in comparison with that in the Py-Im-dsDNA structure (Fig. 2D). It is noteworthy that despite this downstream shift, all key hydrogen bonds are maintained, including those between amide NH and purine N3 and pyrimidine O2 lone pairs, which consist of DNA minor groove “floor” interactions, together with an imidazole lone pair and exocyclic amine of guanine (N2 hydrogen of G) (SI Appendix, Figs. S1B and S24). These interactions dictate the specificity of Py-Im to its target DNA. Interestingly, in addition to those conserved hydrogen bonds, amide NH adjacent to the imidazole group was in a hydrogen bonding distance with exocyclic amine of guanine (N2 of G) in the Pol II-Py-Im encounter complex.

Py-Im Induces a 1-bp Downstream DNA Compression in the Pol II-Py-Im Encounter Complex. To accommodate hairpin Py-Im 1, we found substantial rearrangement of downstream DNA in both NTSs and TSs of the encounter complex (n-5) (red arrows in Fig. 3A). Binding of hairpin Py-Im 1 to Pol II elongation complex induces minor groove widening and major groove

narrowing at the downstream DNA (Fig. 3A and SI Appendix, Fig. S2B). In contrast, binding of Py-Im does not change the structure of the Pol II active site and upstream DNA/RNA (Fig. 3A and SI Appendix, Fig. S2C). By superposing the structures of Pol II elongation complexes in the absence (apo Pol II EC) and presence of Py-Im (Pol II-Py-Im encounter complex), we found that there is a 1-bp squeezing in the downstream DNA that occurred upon Py-Im binding (Fig. 3). We found that, upon Py-Im binding, an accumulative base rise from i+2 to i+13 nt of tsDNA was decreased by 3.2 Å, from 39.0 Å to 35.8 Å (Fig. 3B). This decrease in the Py-Im-induced base rise is similar to the average rise of 1-bp in B-form DNA (3.4 Å) (44). Consistently, we found that the base pair per turn has been increased from 10.0 (apo Pol II EC) to 10.8 (Pol II-Py-Im encounter complex), which is about 1 bp, upon Py-Im binding (SI Appendix, Table S2). Taken together, these results revealed how Py-Im induces 1-bp squeezing between the active site and Rpb5 Jaw motif of Pol II.

Binding of Py-Im Prevents Pol II Forward Translocation and Induces Pol II Backtracking. To understand how Py-Im affects Pol II forward translocation from the n-5 to the n-4 position, we determined the crystal structures of Pol II elongation complexes using two different scaffolds in the absence or presence of hairpin Py-Im 1. Scaffold-1 is associated with a 9mer RNA (n-5, Fig. 4A), whereas scaffold-2 is associated with a 10mer RNA (n-4, Fig. 4B). As controls, we first solved the structures of two canonical Pol II elongation complexes (scaffold-1 and scaffold-2) in the absence of hairpin Py-Im 1. These Pol II elongation complexes are in two consecutive posttranslocation states, mimicking Pol II 1-bp forward translocation after RNA extension from a 9mer (scaffold-1)

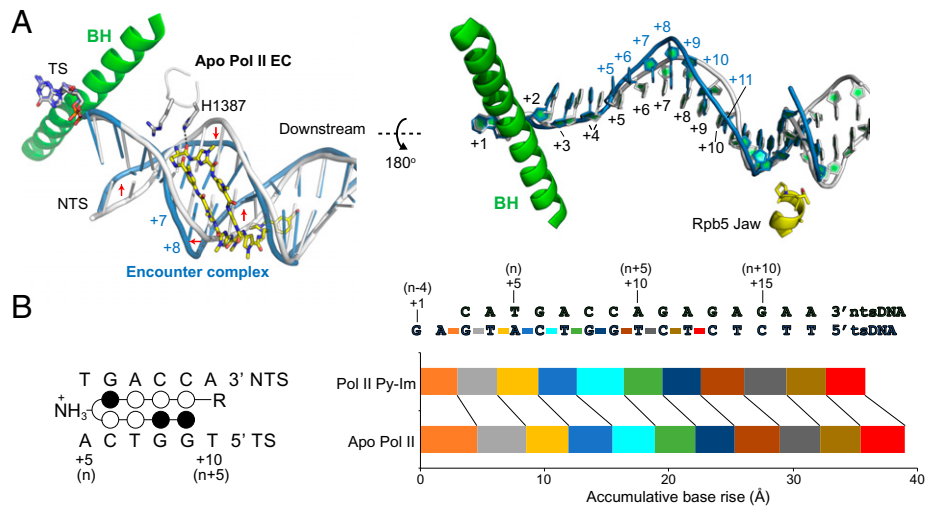


Fig. 3. Structural comparison of canonical Pol II elongation complex (EC) and Py-Im encounter complex. (A) Structure superposition of the Py-Im encounter complex and canonical Pol II elongation complex (apo, in the absence of Py-Im). The downstream DNA duplex in apo Pol II EC and Pol II-Py-Im encounter complex are shown in white and blue, respectively. Major structural changes in the DNA backbone are highlighted with red arrows. The positions of the downstream TS are labeled. For example, +1 refers to the $i+1$ template position, the template base in the Pol II active site that is read for pairing with incoming nucleotide substrate. (B) Comparison of accumulative base rise parameters of the downstream DNA in the Py-Im encounter complex and canonical Pol II elongation complex (apo). Note that the difference of parameter between apo Pol II EC and Pol II-Py-Im was 3.2 Å. (Left) The position of the Py-Im-binding site relative to the Pol II active site, where n is the first 5'-base of the hairpin Py-Im-binding site.

to a 10mer (scaffold-2) (Fig. 4 A and B). In both posttranslocation states, the 3'-RNA end is translocated into the -1 position and the Pol II active site is available for nucleotide substrate binding for a new round of incorporation (red circles in Fig. 4 A and B).

We then determined the crystal structures of Pol II complexes in the presence of hairpin Py-Im 1. Intriguingly, we found a very distinct scenario in the presence of Py-Im. As shown in Fig. 4C, the Pol II-Py-Im complex with scaffold-1 (i.e., encounter complex at $n-5$) is in a posttranslocation state.

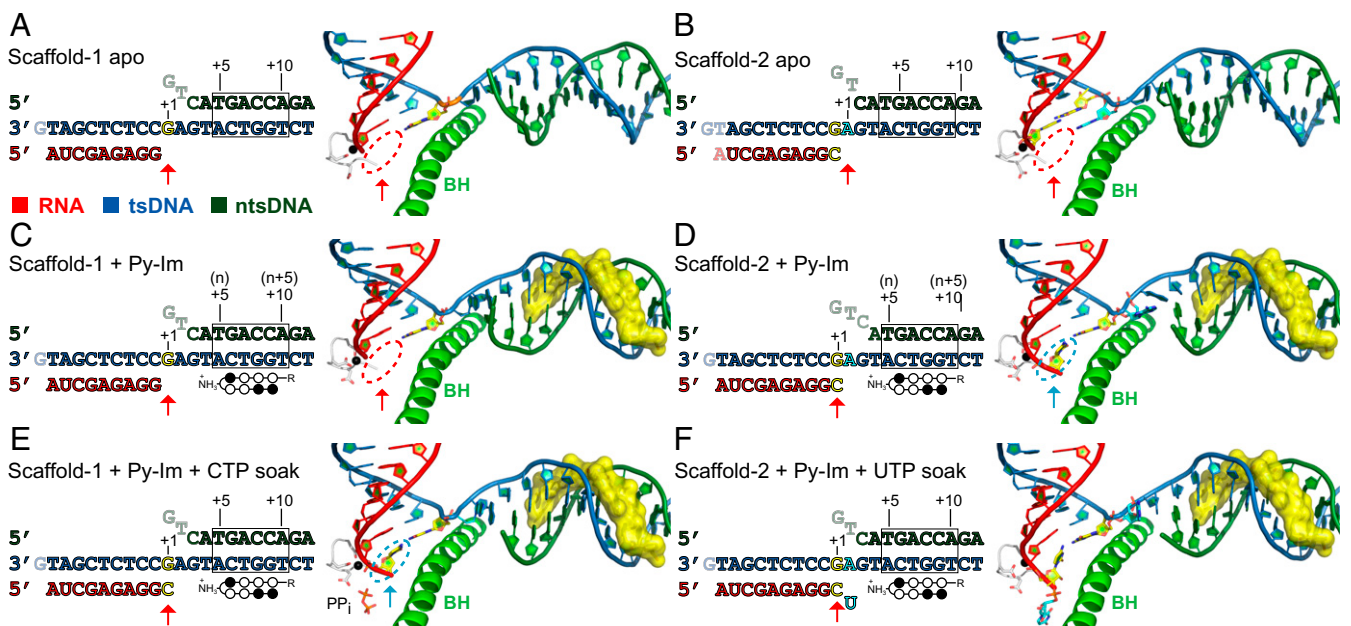


Fig. 4. Stepwise structural snapshots of Py-Im-induced transcriptional arrest revealed by six distinct structures using scaffold-1 (Left) and scaffold-2 (Right). Red and cyan arrows indicate the +1 position of each structure. Note that for the posttranslocation state, the 3'-RNA is located at the -1 position and the active site at the +1 position is empty (highlighted in red circles); for the pretranslocation state, the 3'-RNA occupies the +1 position (highlighted in cyan circles). (A and B) Canonical Pol II elongation complexes at two consecutive posttranslocation states in the absence of Py-Im (controls): Pol II elongation complex with scaffold-1 with a 9mer RNA (scaffold-1 apo, A); Pol II elongation complex with scaffold-2 with a 10mer RNA (scaffold-2 apo, B). The empty active site is shown in red circles. (C and D) Two Pol II complexes in the presence of Py-Im were determined: Py-Im encounter complex ($n-5$) (C, posttranslocation state [with red circle]) and Py-Im engaged complex ($n-4$) (D, pretranslocation state [with cyan circle]). (E) Engaged state: ppi ($n-4$) was captured in a pretranslocation state by soaking the crystal of the Py-Im encounter complex ($n-5$) with CTP. The newly incorporated CMP occupies the +1 position (shown in cyan circle). (F) Backtracked state ($n-3$) was captured by soaking the crystal of the engaged complex ($n-4$) with UTP. Color code is the same as in Fig. 2. A total of 7-bp of downstream DNA duplex was omitted for clarity.

In sharp contrast to the apo Pol II EC with scaffold-2, we found that new Pol II–Py-Im EC with scaffold-2 is in the pretranslocation state in the presence of Py-Im (Fig. 4D). In the pretranslocation state, the 3'-end of RNA transcript still occupies the Pol II active site at the +1 position (cyan circle in Fig. 4D). We refer to this Py-Im-induced pretranslocation state as the “engaged complex (n-4)” (Fig. 4D).

To test whether we can observe a nucleotide addition reaction from the Pol II–Py-Im encounter complex (n-5) in crystal, we soaked the crystal of the encounter complex (scaffold-1, n-5) with cytidine triphosphate (CTP) overnight. The structure we captured reveals a postchemistry state. In this state, the RNA transcript product is extended from a 9mer to a 10mer by Pol II via a S_N2 reaction. As a result, the matched cytidine monophosphate (CMP) is incorporated at the 3'-end of RNA transcript. Pol II is in a pretranslocation state in which the newly added CMP still occupies the active site at the +1 position (cyan circles in Fig. 4E). Interestingly, the pyrophosphate group is not yet released from the Pol II active site. We refer to this state as the n-4:ppi state (Fig. 4E). Taking together the results from the structural studies of both the engaged complex (n-4) and n-4:ppi states, we found that the pretranslocation state at the n-4 position is energetically favored in the presence of Py-Im (Fig. 4D and E). The 3'-end of RNA transcript occupies the active site. Consequently, the next round of nucleotide binding and incorporation is greatly compromised. Therefore, this structure provides a structural explanation for why the extension from the n-4 state to the n-3 state is very slow.

To gain further structural insights into how Pol II extends from the n-4 state to the n-3 state (despite being slow), we soaked the crystals of the engaged complex (n-4) with matched uridine triphosphate (UTP) overnight to test whether we could force slow extension in the presence of hairpin Py-Im 1, and if so, what the

dominant state of Pol II would be. Interestingly, we found that UTP can be incorporated into the RNA strand to form 11mer in crystal by overnight soaking. However, this newly uridine monophosphate (UMP)-incorporated Pol II–Py-Im EC is adopted in a backtracked state (Fig. 4F). This indicates even though the Pol II engaged complex (n-4) dominates at the pretranslocation state in crystal, Pol II may slowly and transiently move forward to incorporate UTP. Once UMP is incorporated, Pol II moves backward, indicating the backtracked state is energetically favored in the presence of Py-Im. We refer to this Py-Im-induced backtracked state as the “backtracked state (n-3)” (Fig. 4F). This structure also explains why the n-3 state is a much stronger arrest state (in comparison with n-5 and n-4 states) and further extension to the n-2 state would be expected to be much slower.

Pol II Gets Stuck on a Molecular Treadmill in the Presence of Py-Im and TFIIS. TFIIS promotes RNA transcript cleavage at the backtracked state or pretranslocation state to generate a new posttranslocation state (45, 46). Therefore, based on the structures of the engaged complex (n-4, pretranslocation state) and backtracked state (n-3), we can have a clear predication that the RNA transcripts in these states are readily cleaved 1 nt and 2 nt by TFIIS, respectively, to generate a new posttranslocation state at n-5. To test this, we performed a transcription assay in the presence or absence of hairpin Py-Im 1 and TFIIS. Consistent with our previous report, TFIIS cannot rescue Py-Im-induced transcriptional arrest even with prolonged incubation (up to 2 h, Fig. 5) (29). However, we noticed the pattern of pausing/arrest bands undergoes a significant change in the presence of TFIIS. While the major pausing bands correspond to n-3 and n-4, followed by the n-5 band, in the absence of TFIIS, the major pausing band shifts toward the n-5 band in the presence of TFIIS (Fig. 5). This result fully supports our prediction that the n-4 state (engaged

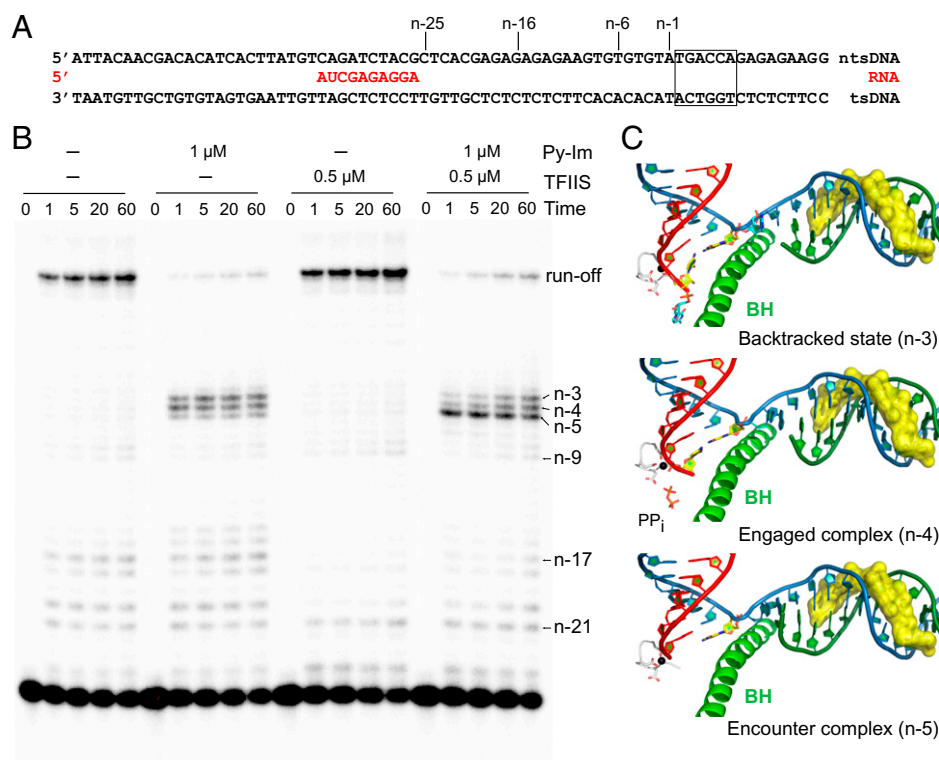


Fig. 5. TFIIS cannot rescue Py-Im-induced strong transcriptional arrest. (A) The full-bubble scaffold for transcription assay. (B) TFIIS can stimulate RNA cleavage but fails to rescue Py-Im-induced transcriptional arrest. Presence of Py-Im induces strong arrest at n-5 to n-3 positions. Addition of TFIIS to the system significantly increases the band at the n-5 position. This indicates the majority of the n-3 complex is in a backtracked state, therefore allowing a 2-nt cleavage. (C) The corresponding Py-Im-arrested Pol II structures (n-5, n-4, and n-3) are shown. Color code is the same as in Fig. 2.

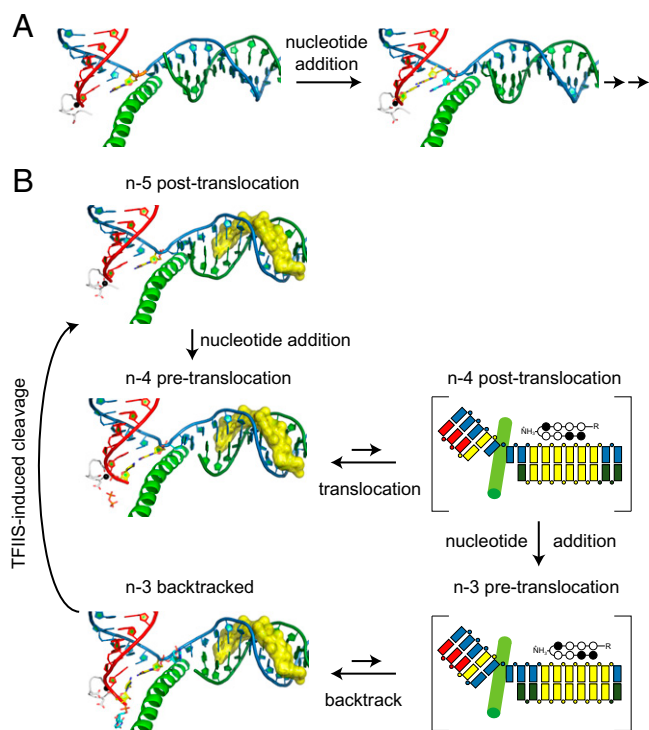


Fig. 6. Molecular mechanism of Py-Im-induced transcriptional arrest. (A) Pol II incorporates incoming nucleotide and moves forward effectively in the absence of Py-Im. (B) Pol II is trapped among three states in the presence of Py-Im. Pol II undergoes futile cycles of nucleotide addition and cleavage in the presence of Py-Im and TFIIS.

complex) and the n-3 state (backtracked state) are more prone to TFIIS cleavage, whereas the n-5 state (encounter state at the post-translocation state) is resistant to TFIIS cleavage. Taking together the results from our biochemical and structural studies, we are now able to attribute the three consecutive stalling bands to the three defined Py-Im-induced Pol II-arrested structures (Fig. 5). We also obtained structural insights into why TFIIS fails to rescue Py-Im-induced Pol II arrest. Py-Im represents a class of very stable and bulky barriers. The binding affinity of Py-Im to its target DNA is extremely high (with nanomolar scale) (47, 48). The off-rate of Py-Im from Pol II elongation complex is extremely slow (even slower than the dsDNA). Indeed, we found that Pol II cannot bypass the Py-Im barriers even after over 40 h of incubation time (29). The presence of TFIIS leads to futile cycles of slow nucleotide addition and TFIIS-stimulated RNA transcript cleavage without bypassing the Py-Im barriers. Pol II may slowly extend from the n-5 state to the n-4 or n-3 state, both of which (either in pretranslocation or backtracked state) are prone to TFIIS cleavage, and return to the n-5 state, which is resistant to further cleavage. As a result, Pol II is trapped by the Py-Im barrier, such that it is stuck on a “molecular treadmill” (Fig. 6).

In summary, we investigated the molecular mechanism of transcriptional arrest by a noncovalent DNA-binding small molecule, pyrrole-imidazole polyamide **1**. Our structural studies captured

stepwise structural snapshots at distinct states of Py-Im **1**-induced Pol II-arrested complexes, including one n-5 state (encounter complex, posttranslocation state), two n-4 states that are trapped at the pretranslocation state (engaged complex, n-4:ppi), and one n-3 state (Py-Im-induced backtracked state). Our structural and functional analyses provide important structural insights into how Py-Im **1** is bound within the Pol II elongation complex and interacts with Pol II; how Py-Im **1** leads to conformational changes of downstream DNA duplex; how bound Py-Im **1** prevents Pol II forward translocation and induces backtracking; how Py-Im **1** traps Pol II into persistent arrest; and why TFIIS fails to rescue Py-Im-induced arrest. Our study may also have implications in targeting transcriptional addition for cancer therapy with small molecules (49–51).

Materials and Methods

Detailed descriptions of synthesis of hairpin Py-Im **1** and purification of RNA Pol II are provided in *SI Appendix*. For in vitro transcription assay with a miniscaffold, 200 nM of 5'-³²P-labeled RNA (5'-AUCGAGAGG-3'), 600 nM of TS DNA (5'-CCTTCTCTCTGGTCATGAGCCTCTCGATG-3'), and 800 nM of nontemplate strand DNA (5'-GTATGACCATGAGAGAGAGG-3') was annealed in elongation buffer to prepare the miniscaffold. Py-Im was dissolved in dimethylsulfoxide, and the concentration was validated by measuring the absorbance at 310 nm using a NanoDrop. Various concentrations of Py-Im were added to the miniscaffold and incubated for 3 h at room temperature. The prepared miniscaffold was then mixed with Pol II and preincubated for 10 min at room temperature to assemble the EC. Reaction was started by mixing equal volume of EC and ribonucleoside triphosphate (rNTP) or TFIIS and was quenched by adding quench-loading buffer (90% formamide, 50 mM ethylenediaminetetraacetic acid (EDTA), 0.05% xylene cyanol, and 0.05% bromophenol blue) and analyzed by 12% urea/tris/borate/EDTA polyacrylamide gel electrophoresis (TBE PAGE). For transcription assays with a full-bubble scaffold, the scaffold was assembled by annealing TS DNA and RNA, followed by the addition of Pol II and incubated for 20 min at room temperature. After adding nontemplate strand DNA and incubating for 10 min, transcription was initiated as described above. All transcription assays were repeated three times. For structure determination of Py-Im-bound dsDNA complexes, an aliquot of 0.7 mM duplex DNA:Py-Im was mixed with an equal volume of crystallization solution containing 10 mM Tris (pH 7.5), 24% 2-methyl-2,4-pentanediol (MPD), and 35 mM calcium acetate, with 35% MPD as a reservoir. Crystals were obtained by the sitting-drop vapor diffusion method at 4 °C for 4 to 6 wk. For structural determination of Py-Im-bound Pol II elongation complex, the crystals of Py-Im-bound Pol II elongation complexes were obtained by hanging-drop methods with incubation, with crystallization buffer (390 mM ammonium phosphate [pH 6.0], 5 mM dithiothreitol [DTT], 5 mM dioxane, and 9 to 13% [wt/vol] polyethylene glycol [PEG 6,000]) at 22 °C for 7 to 14 d. A detailed description of structural determination and refinement can be found in *SI Appendix*.

Data Availability. All atomic coordinates and structure factors are deposited at the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB). All PDB codes are listed in *SI Appendix, Table S1*. Structure coordinates data have been deposited in the PDB (7R1L, 7R1Q, 7R1M, 7R1P, 7R1W, 7R1X, and 7R1Y). All other study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by grants from the NIH (R01 GM102362 to D.W. and R01 GM27681 to P.B.D.). We appreciate the beamline staff for their kind support during our remote data collection. This research used resources of the Advanced Light Source (beamlines 5.0.1, 8.2.1, and 8.2.2), which is supported by the US Department of Energy (DOE), Office of Science User Facility under Contract DE-AC02-05CH11231. We also acknowledge Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, which is supported by the DOE, Office of Science, Office of Basic Energy Sciences under Contract DE-AC02-76SF00515.

1. X. Liu, D. A. Bushnell, R. D. Kornberg, RNA polymerase II transcription: Structure and mechanism. *Biochim. Biophys. Acta* **1829**, 2–8 (2013).
2. K. Adelman, J. T. Lis, Promoter-proximal pausing of RNA polymerase II: Emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
3. A. Mayer, H. M. Landry, L. S. Churchman, Pause & go: From the discovery of RNA polymerase pausing to its functional implications. *Curr. Opin. Cell Biol.* **46**, 72–80 (2017).
4. L. S. Churchman, J. S. Weissman, Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
5. L. Wang et al., Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* **523**, 621–625 (2015).

6. F. Brueckner, U. Hennecke, T. Carell, P. Cramer, CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859–862 (2007).
7. W. Wang, C. Walmacq, J. Chong, M. Kashlev, D. Wang, Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2538–E2545 (2018).
8. J. Oh et al., RNA polymerase II stalls on oxidative DNA damage via a torsion-latch mechanism involving lone pair- π and CH- π interactions. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9338–9348 (2020).
9. L. Xu et al., Molecular basis of transcriptional fidelity and DNA lesion-induced transcriptional mutagenesis. *DNA Repair (Amst.)* **19**, 71–83 (2014).

10. L. Xu *et al.*, RNA polymerase II transcriptional fidelity control and its functional interplay with DNA modifications. *Crit. Rev. Biochem. Mol. Biol.* **50**, 503–519 (2015).
11. W. Wang, J. Xu, J. Chong, D. Wang, Structural basis of DNA lesion recognition for eukaryotic transcription-coupled nucleotide excision repair. *DNA Repair (Amst.)* **71**, 43–55 (2018).
12. J. H. Shin, L. Xu, D. Wang, RNA polymerase II acts as a selective sensor for DNA lesions and endogenous DNA modifications. *Transcription* **7**, 57–62 (2016).
13. J. H. Shin, L. Xu, D. Wang, Mechanism of transcription-coupled DNA modification recognition. *Cell Biosci.* **7**, 9 (2017).
14. J. Oh, J. Xu, J. Chong, D. Wang, Molecular basis of transcriptional pausing, stalling, and transcription-coupled repair initiation. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1864**, 194659 (2021).
15. N. Charlet-Berguerand *et al.*, RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *EMBO J.* **25**, 5481–5491 (2006).
16. J. Xu *et al.*, Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature* **551**, 653–657 (2017).
17. A. Hirtreiter *et al.*, Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
18. J. Xu *et al.*, Cockayne syndrome B protein acts as an ATP-dependent processivity factor that helps RNA polymerase II overcome nucleosome barriers. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25486–25493 (2020).
19. M. W. Kellinger *et al.*, 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).
20. W. Wang *et al.*, Epigenetic DNA modification N⁶-methyladenine causes site-specific RNA polymerase II transcriptional pausing. *J. Am. Chem. Soc.* **139**, 14436–14442 (2017).
21. L. Xu *et al.*, Mechanism of DNA alkylation-induced transcriptional stalling, lesion bypass, and mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7082–E7091 (2017).
22. A. Tufegđić Vidaković *et al.*, Regulation of the RNAPII pool is integral to the DNA damage response. *Cell* **180**, 1245–1261.e21 (2020).
23. M. Sanz-Murillo *et al.*, Structural basis of RNA polymerase I stalling at UV light-induced DNA damage. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8972–8977 (2018).
24. D. Wang, G. Zhu, X. Huang, S. J. Lippard, X-ray structure and mechanism of RNA polymerase II stalled at an antineoplastic monofunctional platinum-DNA adduct. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9584–9589 (2010).
25. J. Xu, J. Chong, D. Wang, Opposite roles of transcription elongation factors Spt4/5 and Elf1 in RNA polymerase II transcription through B-form versus non-B DNA structures. *Nucleic Acids Res.* **49**, 4944–4953 (2021).
26. M. W. Kellinger, G. Y. Park, J. Chong, S. J. Lippard, D. Wang, Effect of a monofunctional phenanthriplatin-DNA adduct on RNA polymerase II transcriptional fidelity and translesion synthesis. *J. Am. Chem. Soc.* **135**, 13054–13061 (2013).
27. C. Walmacq *et al.*, Mechanism of RNA polymerase II bypass of oxidative cyclopurine DNA lesions. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E410–E419 (2015).
28. L. Xu *et al.*, A chemical probe targets DNA 5-formylcytosine sites and inhibits TDG excision, polymerases bypass, and gene expression. *Chem. Sci. (Camb.)* **5**, 567–574 (2014).
29. L. Xu *et al.*, RNA polymerase II senses obstruction in the DNA minor groove via a conserved sensor motif. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12426–12431 (2016).
30. J. W. Trauger, E. E. Baird, P. B. Dervan, Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559–561 (1996).
31. S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468–471 (1998).
32. C. L. Kielkopf *et al.*, A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **282**, 111–115 (1998).
33. D. M. Chenoweth, P. B. Dervan, Allosteric modulation of DNA by small molecules. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13175–13179 (2009).
34. D. M. Chenoweth, P. B. Dervan, Structural basis for cyclic Py-Im polyamide allosteric inhibition of nuclear receptor binding. *J. Am. Chem. Soc.* **132**, 14521–14529 (2010).
35. A. A. Kurmis, F. Yang, T. R. Welch, N. G. Nickols, P. B. Dervan, A pyrrole-imidazole polyamide is active against enzalutamide-resistant prostate cancer. *Cancer Res.* **77**, 2207–2212 (2017).
36. N. G. Nickols *et al.*, Activity of a Py-Im polyamide targeted to the estrogen response element. *Mol. Cancer Ther.* **12**, 675–684 (2013).
37. A. A. Kurmis, P. B. Dervan, Sequence specific suppression of androgen receptor-DNA binding in vivo by a Py-Im polyamide. *Nucleic Acids Res.* **47**, 3828–3835 (2019).
38. S. Krishnamurthy *et al.*, Targeting the mutant PIK3CA gene by DNA-alkylating pyrrole-imidazole polyamide in cervical cancer. *Cancer Sci.* **112**, 1141–1149 (2021).
39. F. Yang *et al.*, Antitumor activity of a pyrrole-imidazole polyamide. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1863–1868 (2013).
40. K. Hiraoka *et al.*, Inhibition of KRAS codon 12 mutants using a novel DNA-alkylating pyrrole-imidazole polyamide conjugate. *Nat. Commun.* **6**, 6706 (2015).
41. A. A. Wright *et al.*, Oncogenic mutations in cervical cancer: Genomic differences between adenocarcinomas and squamous cell carcinomas of the cervix. *Cancer* **119**, 3776–3783 (2013).
42. C. L. Kielkopf, S. Ding, P. Kuhn, D. C. Rees, Conformational flexibility of B-DNA at 0.74 Å resolution: d(CCAGTACTGG)(2). *J. Mol. Biol.* **296**, 787–801 (2000).
43. K. Abe *et al.*, X-ray crystal structure of a cyclic-PIP-DNA complex in the reverse-binding orientation. *J. Am. Chem. Soc.* **142**, 10544–10549 (2020).
44. R. R. Sinden, *DNA Structure and Function* (Gulf Professional Publishing, 1994).
45. M. G. Izban, D. S. Luse, SIL-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. *J. Biol. Chem.* **268**, 12864–12873 (1993).
46. D. Wang *et al.*, Structural basis of transcription: Backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324**, 1203–1206 (2009).
47. C. Dose, M. E. Farkas, D. M. Chenoweth, P. B. Dervan, Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit. *J. Am. Chem. Soc.* **130**, 6859–6866 (2008).
48. M. Minoshima, T. Bando, S. Sasaki, J. Fujimoto, H. Sugiyama, Pyrrole-imidazole hairpin polyamides with high affinity at 5'-CGCG-3' DNA sequence; influence of cytosine methylation on binding. *Nucleic Acids Res.* **36**, 2889–2894 (2008).
49. J. E. Bradner, D. Hnisz, R. A. Young, Transcriptional addiction in cancer. *Cell* **168**, 629–643 (2017).
50. M. Lambert, S. Jambon, S. Depauw, M. H. David-Cordonnier, Targeting transcription factors for cancer treatment. *Molecules* **23**, 1479 (2018).
51. J. H. Bushweller, Targeting transcription factors in cancer – From undruggable to reality. *Nat. Rev. Cancer* **19**, 611–624 (2019).
52. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).