**Supplementary Note**

**Model-based analysis of single-molecule tracking data using Spot-On.** To analyze the spaSPT data, we used our previously described kinetic modeling approach (Spot-On)[34, 36]. Briefly, we analyze each replicate separately and the bound fractions and free diffusion coefficients are reported as the mean +/- standard deviation. We merge the data from all cells (~9-10) for each replicate, compile histograms of displacements and then fit the displacement cumulative distribution functions for 7 time points using a two-state model that assumes that Pol II can either exist in an immobile (e.g. chromatin-associated) or freely diffusive state:

$$P(r,\Delta\tau) = F_{\mathrm{BOUND}}\frac{r}{2(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}} + Z_{\mathrm{CORR}}(\Delta\tau)(1-F_{\mathrm{BOUND}})\frac{r}{2(D_{\mathrm{FREE}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{FREE}}\Delta\tau+\sigma^2)}}$$

where:

$$Z_{\mathrm{CORR}}(\Delta\tau) = \frac{1}{\Delta z}\int_{-\Delta z/2}^{\Delta z/2}\left\{1 - \sum_{n=0}^{\infty}(-1)^n\left[\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z}{2}-z}{\sqrt{4D_{\mathrm{FREE}}\Delta\tau}}\right)+\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z}{2}+z}{\sqrt{4D_{\mathrm{FREE}}\Delta\tau}}\right)\right]\right\}\mathrm{d}z$$

and:

$$\Delta z = 0.700\ \mu\mathrm{m} + 0.20805\mathrm{s}^{-1/2}\sqrt{D} + 0.20336\ \mu\mathrm{m}$$

Here, $F_{\mathrm{BOUND}}$ is the fraction of molecules that are bound to chromatin, $D_{\mathrm{BOUND}}$ is the diffusion coefficient of chromatin bound molecules, $D_{\mathrm{FREE}}$ is the diffusion coefficient of freely diffusing molecules, $r$ is the displacement length, $\Delta\tau$ is the lag time between frames, $\Delta z$ is the axial detection range, $\sigma$ is the localization error (35 nm) and $Z_{\mathrm{CORR}}$ corrects for defocalization bias (i.e. the fact that freely diffusing molecules gradually move out-of-focus, but chromatin bound molecules do not). Model fitting and parameter optimization was performed using a non-linear least squares algorithm (Levenberg-Marquardt) implemented in the Matlab version of Spot-On (v1.0; GitLab tag 92cdf210) and the following parameters: dZ=0.7 μm; GapsAllowed=1; TimePoints: 8; JumpsToConsider=4; ModelFit=2;

NumberOfStates=2; FitLocError=0; D_Free_2State=[0.4;25]; D_Bound_2State=[0.00001;0.05].

**Diffusion coefficient calculations.** The observed free diffusion coefficients obtained from fitting the spaSPT data with the Spot-On model (Brownian motion) were 3.74 +/- 0.178 μm$^2$/s, 2.97 +/- 0.0912 μm$^2$/s and 2.34 +/- 0.049 μm$^2$/s for the 25R, 52R and 70R versions of Halo-Rpb1, respectively (mean +/- standard error). Given that the molecular weight of e.g. 25R is lower, one would expect the diffusion coefficient to be higher. To estimate whether this large difference could be explained by size alone or whether it might be due to reduced multivalent interactions, we consider the Stokes-Einstein relation according to which the diffusion coefficient is given by:

$$D = \frac{k_B T}{6\pi\eta r}$$

where $k_B$ is Boltzmann's constant, $T$ is the absolute temperature, $\eta$ is the viscosity of the liquid (the nucleoplasm here; assumed to be the same for 25R, 52R and 70R) and $r$ is the radius. The Stokes-Einstein equation assumes the particle to be a sphere and accordingly the radius is given by the volume, V:

$$r = \sqrt[3]{\frac{3V}{4\pi}}$$

In turn, the volume is related to the mass, $m$, and density, $\rho$:

$$V = \frac{m}{\rho} = \frac{MW}{\rho N_A}$$

where $N_A$ is Avogadro's constant and $MW$ is the molecular weight in atomic mass units (Daltons). Thus, the diffusion coefficient is related to the molecular weight by:

$$D = \frac{k_B T}{6\pi\eta \sqrt[3]{\frac{3MW}{4\pi\rho N_A}}}$$

Thus using 25R and 52R as the example, the ratio between the diffusion coefficients of 25R and 52R Halo-Rpb1 (assuming that the density is the same) is:

$$\frac{D_{\text{52R-Rpb1}}}{D_{\text{25R-Rpb1}}} = \frac{\dfrac{k_B T}{6\pi\eta\sqrt[3]{\dfrac{3MW_{\text{52R-Rpb1}}}{4\pi\rho N_A}}}}{\dfrac{k_B T}{6\pi\eta\sqrt[3]{\dfrac{3MW_{\text{25R-Rpb1}}}{4\pi\rho N_A}}}} = \sqrt[3]{\frac{MW_{\text{25R-Rpb1}}}{MW_{\text{52R-Rpb1}}}}$$

According to UniProt (P24928) the molecular weight of wild-type Rpb1 is 217.2 kDa (52R). The molecular weight of the HaloTag is 33.6 kDa. Thus, the molecular weight of Halo-Rpb1-52R is ~250.8 kDa, the molecular weight of Halo-Rpb1-25R is ~230.9 kDa and the molecular weight of Halo-Rpb1-70R is ~258.1 kDa. Thus, the expected difference in diffusion coefficients is:

$$\frac{D_{\text{52R-Rpb1}}}{D_{\text{25R-Rpb1}_{\text{EXPECTED}}}} = \sqrt[3]{\frac{MW_{\text{25R-Rpb1}}}{MW_{\text{52R-Rpb1}}}} = \sqrt[3]{\frac{230.9\ \text{kDa}}{250.8\ \text{kDa}}} = 0.973$$

We can compare this to the experimentally observed ratio:

$$\frac{D_{\text{52R-Rpb1}}}{D_{\text{25R-Rpb1}_{\text{OBSERVED}}}} = \frac{2.97\,\frac{\mu\text{m}^2}{\text{s}}}{3.74\,\frac{\mu\text{m}^2}{\text{s}}} = 0.794$$

It becomes clear that size/mass difference alone cannot explain the large difference between the diffusion coefficients that we observe in cells. To be comprehensive, below we list the Stokes-Einstein expected and observed diffusion coefficient ratios for all the combinations:

| Comparison | Stokes-Einstein expectation | Observed ratio |
|---|---|---|
| 25R vs. 52R | 0.973 | 0.794 |
| 25R vs. 70R | 0.964 | 0.626 |
| 52R vs. 70R | 0.991 | 0.789 |

For all three combinations, the observed ratio cannot be explained by the change in size/mass. Instead this indicates a higher propensity of the full-length CTD to engage in intermolecular interactions. Moreover, in the above calculations we have just considered the change in the mass of Rpb1. In reality, Rpb1 is likely diffusing as part of the Pol II holocomplex, thus the relative difference due to the smaller CTD (e.g. ~20 kDa between 25R and 52R) is actually much smaller than the calculations

using only Rpb1 would suggest and thus the expected difference in diffusion coefficients due to mass/size would be even much closer to 1. We conclude that the mass/size difference between the 25R, 52R and 70R Pol II enzymes cannot explain their observed differences in diffusion coefficients.