Supporting Information for

RNAPII Subunit Dynamics

Michael A. Tycon[†], Matthew K. Daddysman[†], and Christopher J. Fecko[†]*

[†]Department of Chemistry

University of North Carolina at Chapel Hill

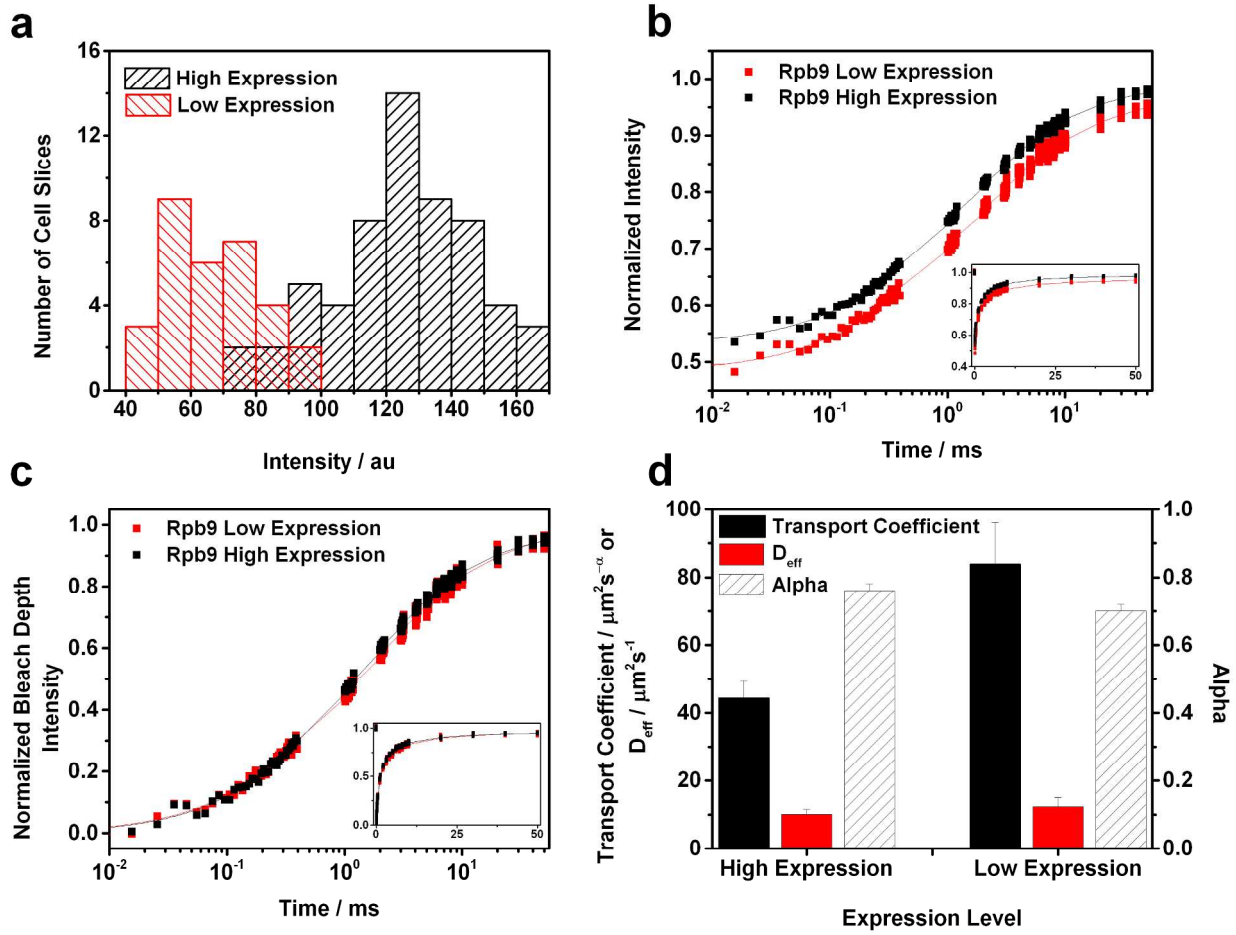
Chapel Hill, North Carolina 27599-3290

Contents:

A. High expression levels of fusion proteins are not responsible for the observed anomalous diffusion:

The Rpb3-GFP and Rpb9-GFP fusion proteins are exogenous insertions expressed under the control of the GAL4 driver system and believed to be functional due to recruitment to HSP promoter sites [1]. As a result they are highly over-expressed compared to the native, untagged RNAPII subunits. To test if the over-expression was creating a population of unincorporated subunit that was being manifest as apparent anomalous diffusion, we crossed our Rpb9-GFP with a GAL4 driver under the control of a heat shock induced promoter (Bloomington Stock Center #1799).(**d**) The expression level of this cross, Rpb9-GFPx1799, can be lowered by raising the fly larvae at 18°C (red bars) and was determined to reduce expression levels by up to 50% compared to the Rpb9-GFPxH2B-mRFP line raised at 22°C (black bars). The mean expression levels of these two populations were found to be statistically different ($p<0.001$). While this construct did not have the chromatin labeled by the H2B-mRFP histone protein, the Rpb9-GFP showed strong exclusion from chromatin regions (determined previously) still enabling us to restrict the FRAP analysis to the interchromatin space. (**a**) The FRAP recoveries and (**b**) normalized recoveries for the high (black) and low (red) Rpb9-GFP expression levels flies are shown. (**c**) Within experimental error, the effective diffusion coefficient and anomlity value of the reduced expression line matched the results found using the Rpb9-GFPxH2B-mRFP line. Thus we are confident that the over expression is not responsible for the anomalous diffusion. This could not be repeated for the Rpb3-GFP construct since it is expressed by a GAL4 driver sequence previously bred into the fly line.

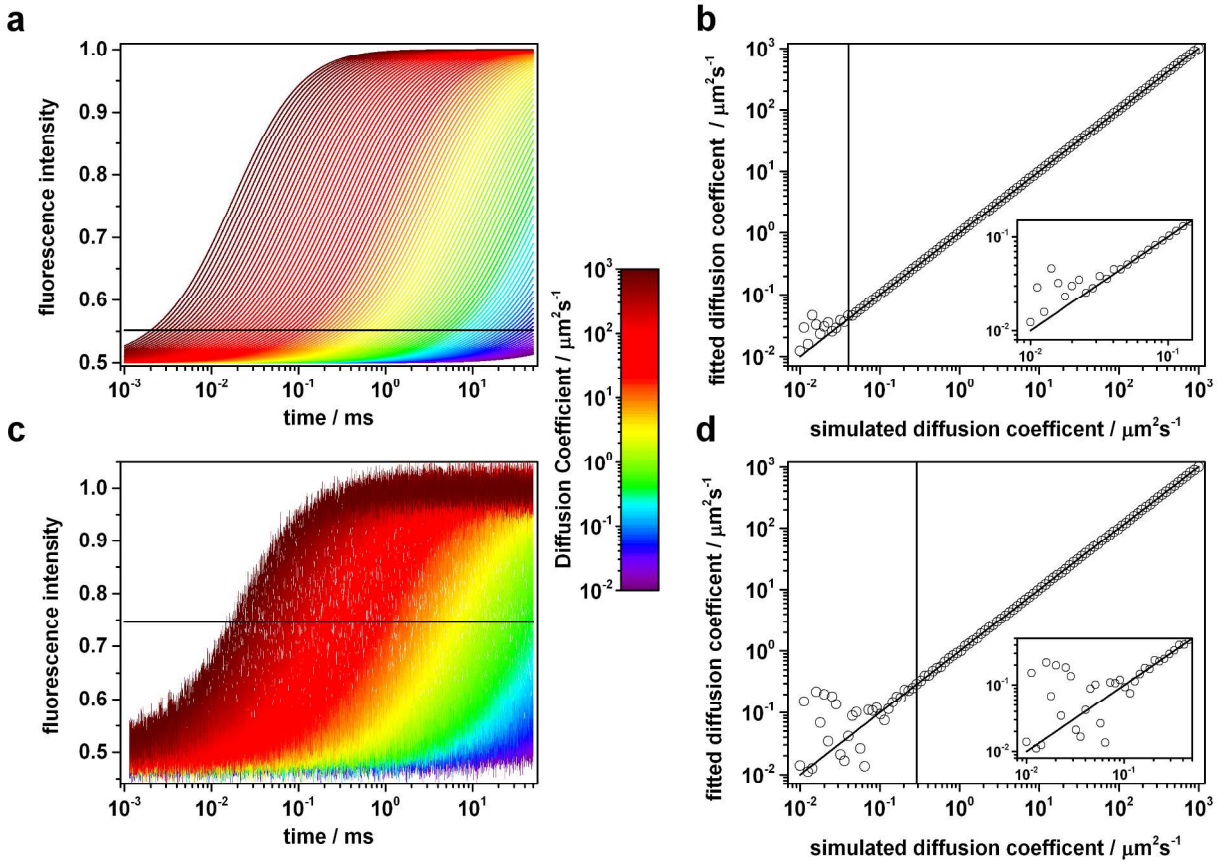
Supplementary Figure 1:



B. Determining the Resolution of the Point FRAP Method:

For slow moving species, determining the diffusion coefficient is difficult if the FRAP curve does not fully recovery to the pre-bleach level on the time course of the measurement. Despite the rapid time resolution of our data collection method, we are limited in how slow a diffusion component we can accurately measure by the 50 ms time duration of our recovery collection. If Brownian diffusion is assumed, our fitting algorithm estimates the final recovery extent based on the slope of the FRAP curve once it begins to level off. Further, the estimation of the recovery extent will strongly affect the estimated diffusion coefficient. For very slow moving species, the recovery will be very shallow and the algorithm is unable to accurately estimate the diffusion coefficient. This became a significant concern when applying the distribution model [2] as a threshold for reliable determination of diffusion coefficients needed to be established. We chose to empirically evaluate which diffusion coefficients were reliable by applying our fitting algorithm to simulated data and determining where the estimated diffusion coefficients began to deviate from the input value. (**a**) FRAP recovery curves were simulated that correspond to diffusion coefficients from 0.01 to 1000 $\mu m^2/s$. As can be seen, the majority of the curves exhibit a significant recovery, but the slow moving components are nearly flat on the 50 ms timescale of the simulation. (**b**) The fitting algorithm was applied to each curve and the estimated diffusion coefficient was plotted against the initial input value. We determined the diffusion coefficient estimation was accurate with as little as 10.3% recovery (**a**-horizontal black line), corresponding to a diffusion coefficient of 0.04 um^2/s (**b**-vertical black line). (**c**) Next, white noise was added to the FRAP curves resulting in simulated data with a signal to noise ratio (SNR) of 35 dB. This SNR corresponds well to our experimental FRAP data. Again, we applied the fitting algorithm to the noisy data and compared the estimated diffusion coefficients to the input values. At this SNR, the estimations begin to deviate once the recovery is less than 47.6% complete (**c**-horizontal black line), corresponding to a diffusion coefficient of 0.29 um^2/s (**d**-vertical black line). Thus we can see the accuracy of the fitting depends on the SNR of the data. Erring on the side of caution, we rejected any diffusion components that showed less than a 50% recovery. This method outlines a framework for evaluating the robustness of a FRAP fitting method as long as the SNR of the data can accurately be estimated.

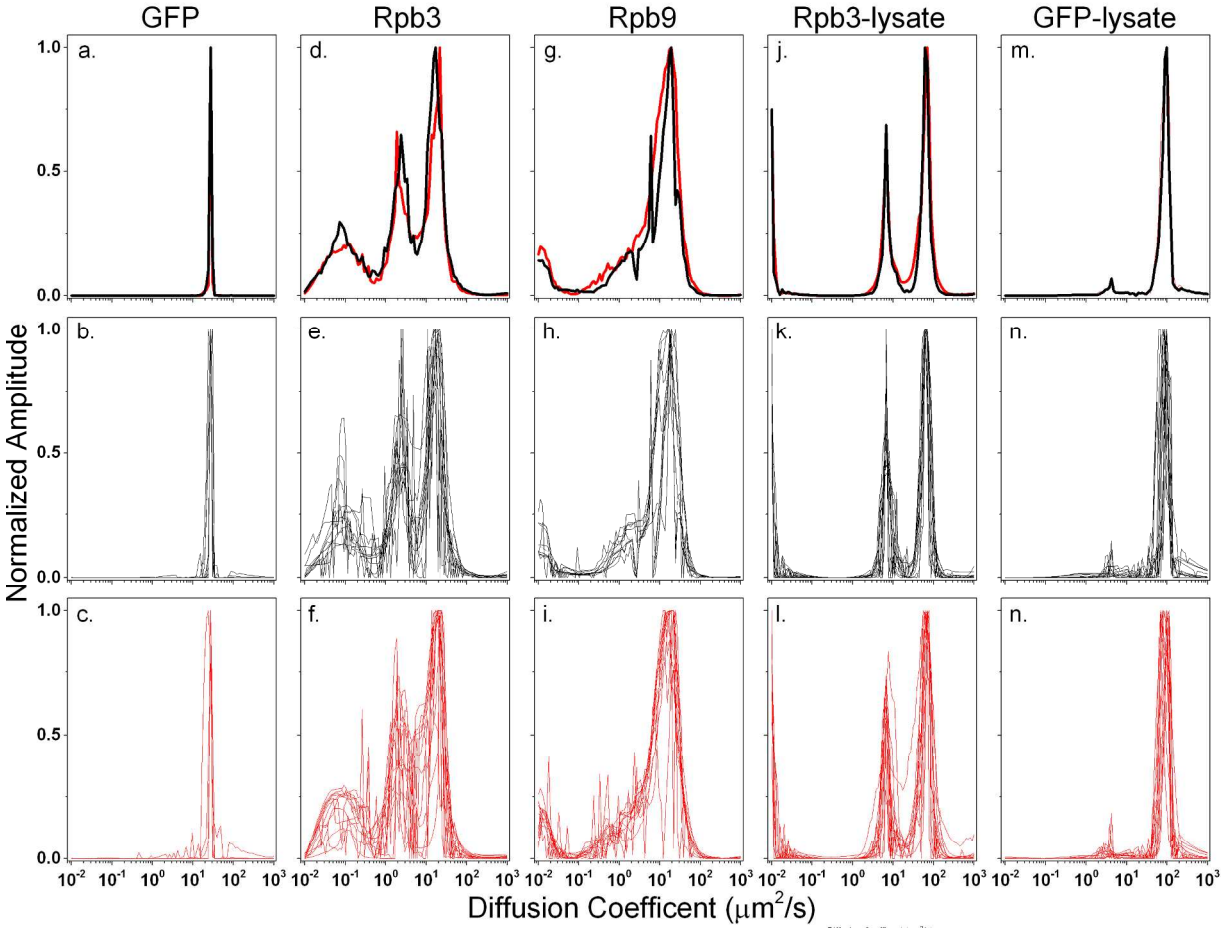
Supplementary Figure 2:



C. Establishing the Robustness of the Distribution Model on Experimental Data

As presented in the Results and Discussion, the Rpb3 datasets indicate a bimodal distribution. We wanted to ensure the robustness of the Distribution model to predict bimodal distributions without a bias predicated on the initial component amplitudes. To achieve this, we tested the output of the Distribution model in response to different initial amplitude profiles, as well as different fitting protocols. Four sets of initial conditions were tested: (1,2-Gaussian) shaped the initial amplitudes in a Gaussian envelope with 35 or 15 dB noise added, (3,4-Flat) provided 35 or 15 dB Gaussian white noise as the input. To test for reproducibility, each input condition was tested three times. In the first, unbiased implementation (panels **b,e,h,k,n**), the input profile amplitudes were floated to achieve a best-fit to the FRAP data. The output distribution was then smoothed with a median filter. This process was repeated five times until the fit residuals no longer improved. The last step omitted smoothing to prevent distorting the output. All the outputs are overlaid indicating the similarity regardless of input profile. Next, the effect of biasing the distribution to a single component by implementing a Gaussian smoothing step was tested. A five-step procedure was used, but in contrast to the previous method, between the third and fourth smoothing steps the output was fit to a Gaussian envelope. The final fit output was not forced to a Gaussian to reveal the most stable output. The fitting outputs from all twelve input distributions are shown (panels **c,f,l,n**); again the outputs are (1) very similar and (2) show the same structure as the un-biased fitting method. The results of the twelve outputs for both fitting methods were averaged and compared (panels **a,d,g,j,m**), indicating nearly identical distributions. This indicates that random noise on the input does not affect the output and the distribution fit find the most stable output. This test was significant for the Rpb3 distribution results. If biasing the output to one component altered the final output away from a bimodal fit, then the distribution model algorithm could not be considered robust. However, since even when the fit was forced to conform to a single peak it still "stepped away" to a bimodal fit on the next iteration, the fitting method was considered stable.

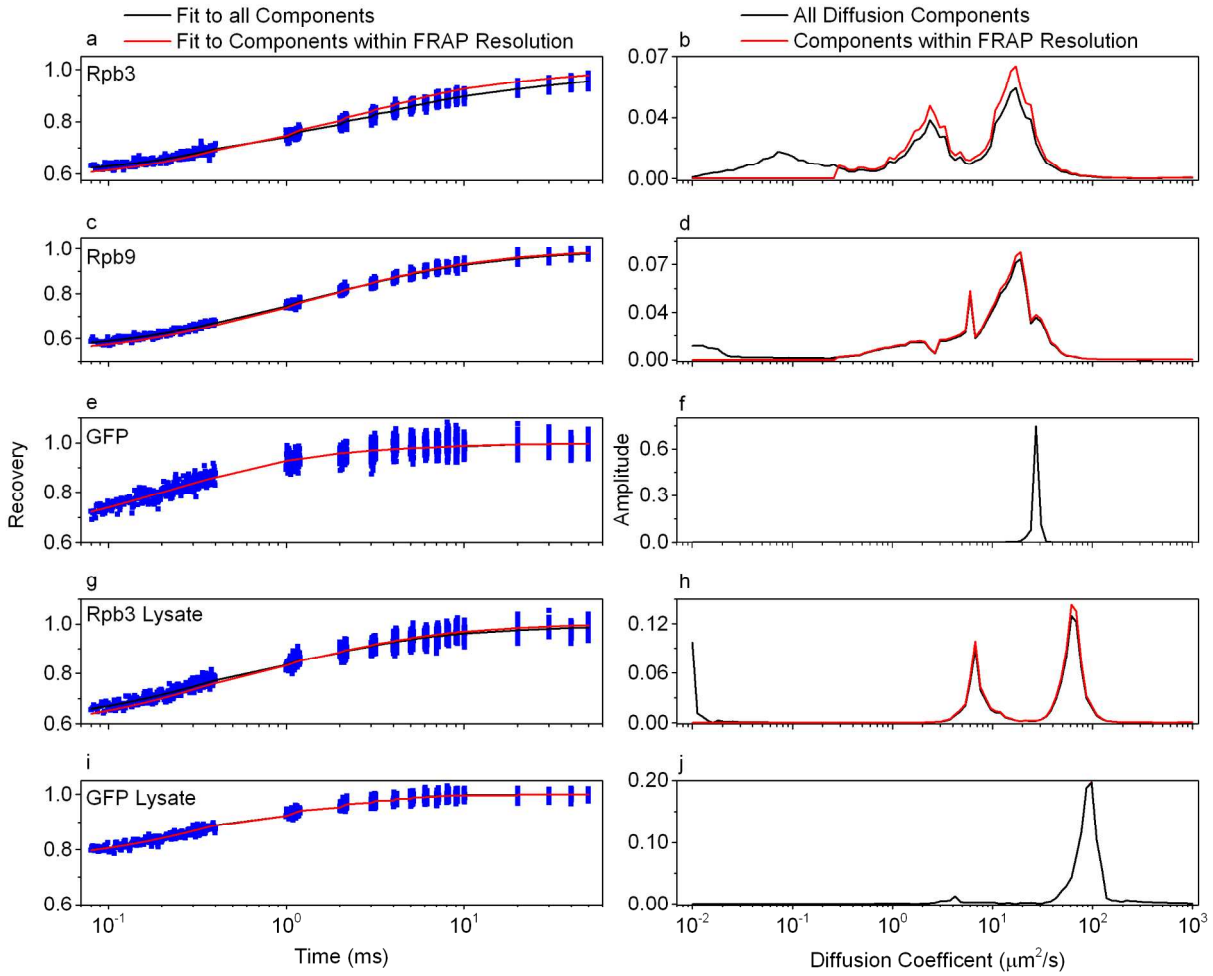
Supplementary Figure 3:



D. Fit quality excluding diffusion components under FRAP resolution

After confirming that the Distribution modeling can robustly determine the number of components that comprise a FRAP curve and having established the FRAP resolution limit, we chose to investigate how accurately the retained components recapitulated the original data. The output distributions (panels **b,d,f,h,j**, black lines) were truncated at 0.30 $\mu m^2/s$ (red lines), and renormalized so the total distribution summed to unity. This slightly increased the amplitudes of the retained components. These truncated distributions were used to establish a fit to the data (panels **a,c,e,g,I**, fit to all components black line, fit to truncated distribution red line). For the Rpb3 *in vivo* data, the retained components do alter the recovery dynamics, shifting the curve to a faster recovery. For all other samples, the fits are unchanged.

Supplementary Figure 4:



E. FRAP fitting results for each dataset

For each experiment, several datasets were collected and the resulting raw data averaged together to yield finalized data with a high SNR. The finalized data was fit with the apparent anomalous diffusion and distribution models. To ensure that the averaging of several datasets did not distort the final results, each individual dataset was fit with the apparent anomalous diffusion model. The results are compiled below. Typically, the subset of the finalized data shows nearly the same anomlity and effective diffusion coefficient, but the 95% confidence error intervals are larger than if the datasets are compiled. As shown, averaging the fit outputs of the subsets is not identical to fitting the averaged data. This method is in line with how the data was analyzed in Daddysman et al. 2011.

Supplementary Table 1:

Conditions	Sample	Set	Gamma (um2/s[a])	D (um^2/s)	Alpha
***In vivo* (live polytenes)**	GFP	I	-	32.7±16.1	0.99±0.08
		II	-	36.2±20.1	1.00±0.09
		III	-	27.5±20.1	1.00±0.12
		Ensemble*	-	32.0±6.0	1.00
	Rpb3	I	70.8±11.7	21.0±4.5	0.78±0.06
		II	37.3±15.9	6.2±3.9	0.73±0.07
		III	54.1±33.3	4.4±5.0	0.64±0.10
		IV	105.6±37.3	7.4±5.1	0.58±0.06
		V	271.9±130.0	9.2±13.6	0.43±0.08
		VI	90.3±23.5	5.0±2.5	0.57±0.05
		Ensemble*	69.1±10.5	5.5±1.4	0.62±0.03
	Rpb9	I	45.7±7.2	7.9±1.7	0.73±0.03
		II	38.9±14.3	9.70±4.8	0.78±0.06
		III	30.7±8.9	7.6±2.9	0.78±0.05
		IV	46.8±7.3	12.8±2.6	0.78±0.02
		Ensemble*	44.4±5.0	10.0±1.5	0.76±0.02
***In vitro* (cell lysate)**	GFP	I	98.0±50.0	79.8±43.0	0.96±0.07
		II	75.1±33.8	71.1±32.5	0.99±0.07
		Ensemble*	112.2±37.5	79.1±29.0	0.92±0.05
	Rpb3	I	69.4±11.3	43.8±7.85	0.91±0.05
		II	246±136.7	41.2±40.1	0.65±0.08
		III	85.4±37.4	30.6±17.2	0.81±0.07
		IV	115.4±45.9	23.2±13.7	0.72±0.06
		Ensemble*	150±36.4	33.0±11.7	0.72±0.04
In vivo Low Expression Level	Rpb9	I	83.9±12.2	12.3±2.65	0.70±0.02
		II	118.3±21.9	10.7±3.3	0.65±0.03
		Ensemble*	97.3±12.1	11.7±2.2	0.67±0.02

*Parameters resulting from fitting the average of all the listed datasets. This procedure improves the fitting results by increasing the SNR of the data.

244 References

245 (1) Yao, J.; Ardehali, M. B.; Fecko, C. J.; Webb, W. W.; Lis, J. T. *Molecular Cell* **2007***, 28*, 978-990.

246 (2) Periasamy,N., Verkman, A.S *Biophys. J.* **1998***, 75*, 557.

247