Online Supplemental Material to

"Activated STAT1 transcription factors conduct distinct saltatory movements in the cell nucleus" by Jasmin Speil, Eugen Baumgart, Jan-Peter Siebrasse, Roman Veith, Uwe Vinkemeier, and Ulrich Kubitscheck

Estimation of the spatial variance of bound molecules

Fig. S1 demonstrates how the variance σ_b^2 was experimentally determined from obviously fixed molecules. The trajectory showed a clear transition from mobile to immobile phases (Fig. S1A). This was reflected in a distinct decrease of the positional variance, Fig S1B. We screened all (smoothed) experimental trajectories for long, distinct binding events, and from these we calculated the mean positional variance. We obtained σ_b^2 =0.0005 μm^2 corresponding to a mean localization precision of σ_b =22 nm from 15 such events.

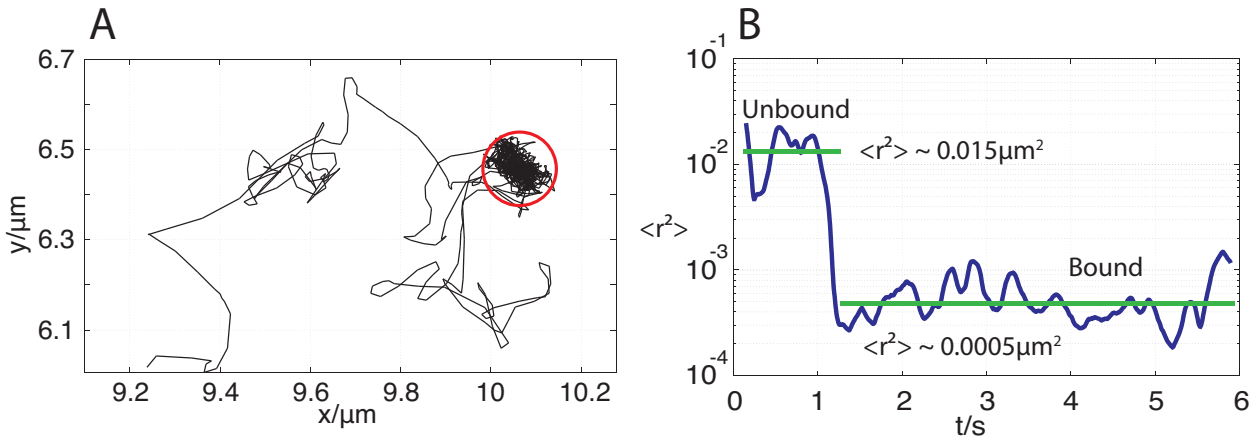

Fig. S1 Intranuclear binding and positional variance

(A) Example of a long trajectory with a distinct binding event. The complete trajectory ran over 1200 frames corresponding to six seconds. In the bound state, the STAT1 molecule could be tracked for five seconds (encircled in red). (B) The variance along the trajectory measured in a 30 frames wide window dropped off as the molecule became immobile at a putative binding site. The horizontal green lines indicate the mean variances in the unbound and bound trajectory sections. Although the STAT1 molecule was obviously fixed, the trajectory appeared slightly elongated. This could be due to slow DNA movement or a drift of the whole sample.

Analysis of binding times from single molecule trajectories

The approach introduced in this paper was based on previous treatments of this problem (Meilhac et al, 2006; Simson et al, 1995; Saxton, 1993). We performed extensive Monte Carlo simulations with particles switching between a free and bound state in order to examine the applicability of the method. Furthermore, we examined the parameter range, for which the determination of bound and free phases was reliable, and the "recovered" binding times approximated a known "input" binding time.

(i) Description of simulation and algorithms

We performed random walk simulations with mobile particles undergoing changes between a "bound" and a "free" state with diffusion coefficients D_b and D_f, respectively. The bound state was modelled in two different ways. Firstly, the particle position was fluctuating around a mean value and secondly, the particles performed just a very slow diffusion. The latter accounted for the fact that in our STAT1 experiments we did not only observe perfect immobilization of bound particles, but in cases also a slight drift possibly caused by sliding motions on DNA, by an overall DNA movement or by stage drift. An alternative way to characterize the respective mobility was by means of the spatial variance that is related to D by $\sigma^2 = 4D\Delta t$. It has the advantage that an estimate can directly be calculated from a set of spatial coordinates of a trajectory. The implementation of the complete simulation and analysis was realized in MATLAB.

The step sizes in x- and y-direction during Brownian diffusion were obtained from a normal distribution random number generator (randn in Matlab) with a standard deviation of $\sqrt{2D_f\Delta t}$. As time step Δt we used the inverse imaging frame rate, 5 ms. For the bound state with zero movement of the mean position, the standard deviation of the random numbers around the mean was $\sqrt{2D_b\Delta t}$. The same was used for the bound state with slow Brownian motion, but in this case new positions were calculated by adding the steps in x and y to the previous position. The probability for switching between free and bound states was $k_{on}\Delta t$ and $k_{off}\Delta t$, where k_{on} and k_{off} were association and dissociation constants. These were determined as follows. Before each simulated time step, a random number z from an equal distribution in the interval [0, 1] was drawn. If $z < k_{on}\Delta t$, the particle switched to the bound state. Similarly, the particle switched from the bound to the free state was performed, if $z < k_{off}\Delta t$. The binding time was determined

from $t_b = 1/k_{off}$. k_{on} can be calculated from k_{off}, since we assumed for simplicity an equal probability for the bound and free states.

(ii) Parameter range, for which the determination of binding times was reliable

The above simulated single molecule trajectories contained phases of free diffusion and bound states. The duration of binding 1/ k_{off} was a simulation parameter. Bound and free states were characterized by the diffusion constants D_b and D_f or the values of the spatial variance, σ_b^2 and σ_f^2, respectively. In the simulation the ratio σ_b^2/σ_f^2 could be varied at will between maximally unity (i.e. free and bound particles were practically indistinguishable in their mobility) or approach zero (i.e. motion of the bound particles was negligible compared to the free ones), and also the duration of the bound state. For meaningful simulations of two *separate* states obviously $D_b \leq 2D_f$. Simulation result was a trajectory with bound and free segments. Next we tested how well an analysis of the trajectory would recover the assumed bound and free phases of the dynamics. To this end we used the spatial variance of the positions along the trajectory in a window of length n (n≥3), and calculated the gliding variance σ_n^2 along the trajectory.

Obviously, during bound phases $\sigma_n^2 \approx \sigma_b^2$ and the parameter L_E defined by $L_E = \frac{\sigma_b^2}{\sigma_n^2}$ was near unity, whereas during free trajectory segments $\sigma_n^2 \approx \sigma_f^2$ and $L_E = \frac{\sigma_b^2}{\sigma_n^2} \approx \frac{\sigma_b^2}{\sigma_f^2}$, which was the minimal value of L_E. Upon switching from the mobile to the bound state and back L_E passed the "critical" value in the middle between unity and minimum, $L_c = \frac{1}{2} + \frac{\sigma_b^2}{2\sigma_f^2}$

We defined this L_E-value as threshold in the trajectory analysis: whenever $L_E > L_c$ we defined the respective position as a "bound". Below we checked, how the results depended on the exact choice of this threshold L_c.

We simulated trajectories of 10^5 jumps and kept track of the bound phases. Then we analyzed the trajectory using the criterion $L_E > L_c$. From this we determined the percentage of correctly assigned positions of the trajectory. This analysis was performed for the minimal possible window size, n=3, as a function of σ_b^2/σ_f^2 and for various binding durations. The simulations showed that the tracking analysis procedure could identify bound molecules. We checked, what

percentage of binding events was correctly identified as a function of σ_b^2/σ_f^2 . In Fig. S2 we considered cases, in which the ratio of the variance of bound vs. free molecules differed by a factor of 1/100 to 0.5. This analysis was performed for binding durations ranging from 10 ms to 2 s. As expected, the method was more reliable for small ratios σ_b^2/σ_f^2 , and for long binding events. For $\sigma_b^2/\sigma_f^2 \le 0.15$ and binding durations longer than 0.1 s more than 80% of the bound positions were correctly assigned.

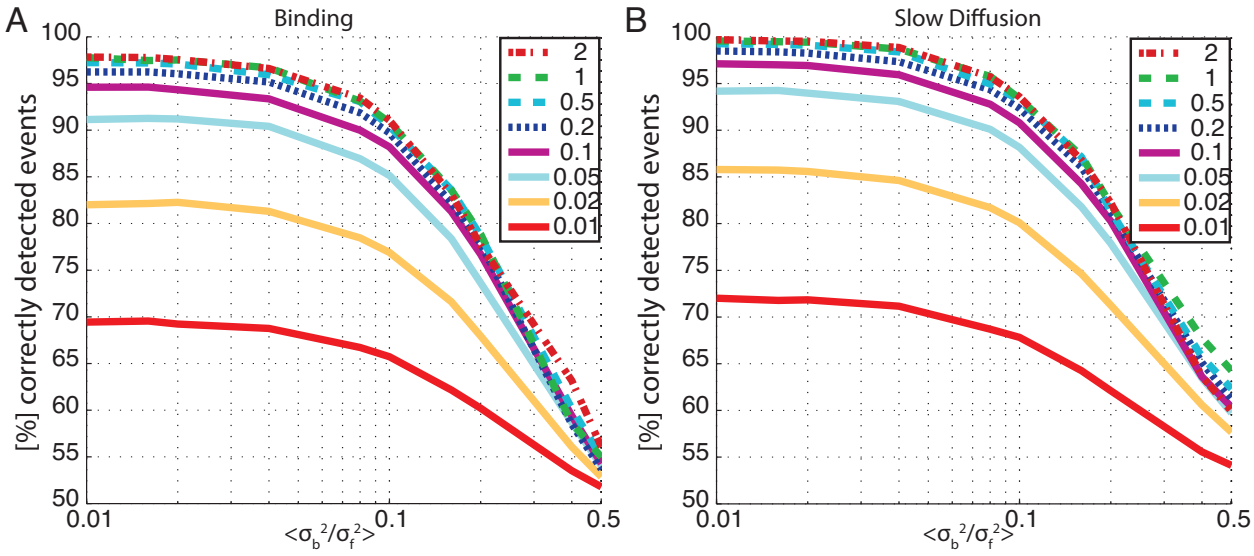

Fig. S2 Percentage of correctly identified particle states

Percentage of binding events that was correctly identified as a function of σ_b^2/σ_f^2 . (A) In the bound state the particle position fluctuated around a mean value and (B) the particles performed a slow diffusion in the bound state. The lines display the results for various binding durations t_b as indicated in the insert. The time unit was seconds. A simulation step was performed every 0.005 s. k_{on} was calculated from $k_{off}=1/t_b$, since we assumed an equal probability for the bound and free states.

(III) Recovered binding times versus true binding times

It was not sufficient to demonstrate that the majority of positions could correctly be assigned to bound or free trajectory segments. It also had to be checked that the true binding time could correctly be recovered, because to achieve this a number of *subsequent* positions must be assigned correctly as "bound" or "free".

Simulations were carried out for a wide range of binding times and the case that the diffusion coefficients of bound and unbound fractions differed by one order of magnitude. We checked that the absolute D values had no impact on the results.

It should be noted that the analysis of STAT1 binding times was based on trajectories, which were determined in image data smoothed by a Gaussian kernel with a standard deviation of 1 pixel in x, y and time. Clearly, this kernel "slowed down" the molecular dynamics and therefore, the absolute values of the diffusion coefficients employed in the binding analysis did not correspond to those of the unsmoothed data as given in Table 1. The aim of this analysis, however, was not to quantify molecular dynamics, but to identify trajectory segments indicating molecular binding.

The simulation parameters were set to D_b =0.05 µm²/s, D_f=0.5 µm²/s. These values were comparable to the motion of the bound and slow mobility fraction from the *smoothed* STAT1 data, a further mobility component was about 10-fold faster. The time between subsequent steps of a particle was set to Δt=0.005 s corresponding to the image integration time of the experiment. Trajectories with 10^5 jumps were simulated, which spent identical times in the bound and in the free state, respectively.

These simulations were also used to determine the validity of the threshold for the detection of binding, L_c, to yield the best accuracy in the detection of binding events. For this purpose, we ran simulations with the parameters mentioned above and binding times ranging from 5 ms to 1 s. We searched for the L_c that minimized the sum of differences between simulated and recovered binding times in this temporal range. The differences were normalized to the simulated binding time. L_c was varied between 0.3 and 0.7. We obtained best results for $L_c = 0.5$. Obviously, $L_c = 0.5 \approx \frac{1}{2} + \frac{\sigma_b^2}{2\sigma_f^2} = 0.55$ as expected for this case.

The binding times that were recovered by the analysis for different input values in the simulation were plotted in log-log plots (see Fig. S3). A quite accurate determination of the simulated binding time was achieved for times in the range of 50 ms up to 1 second. The deviating

analysis results for shorter binding times were caused by the averaging process along the time window for quantifying the variance. Binding times smaller than 20 ms corresponded to less than four frames, and were below the time scale that was meaningfully accessible by our experiments. The investigation of short binding events below this limit would require higher frame rates. Very long binding events tended to be split by noise in the L_E-plot leading to an underestimation of the true binding time. The length of the sliding average window determined the temporal range, for which the binding time could be recovered most accurately. Increasing the window for calculating the gliding variance from n=3 to 5 or 7 yielded more accurate results for long binding times.

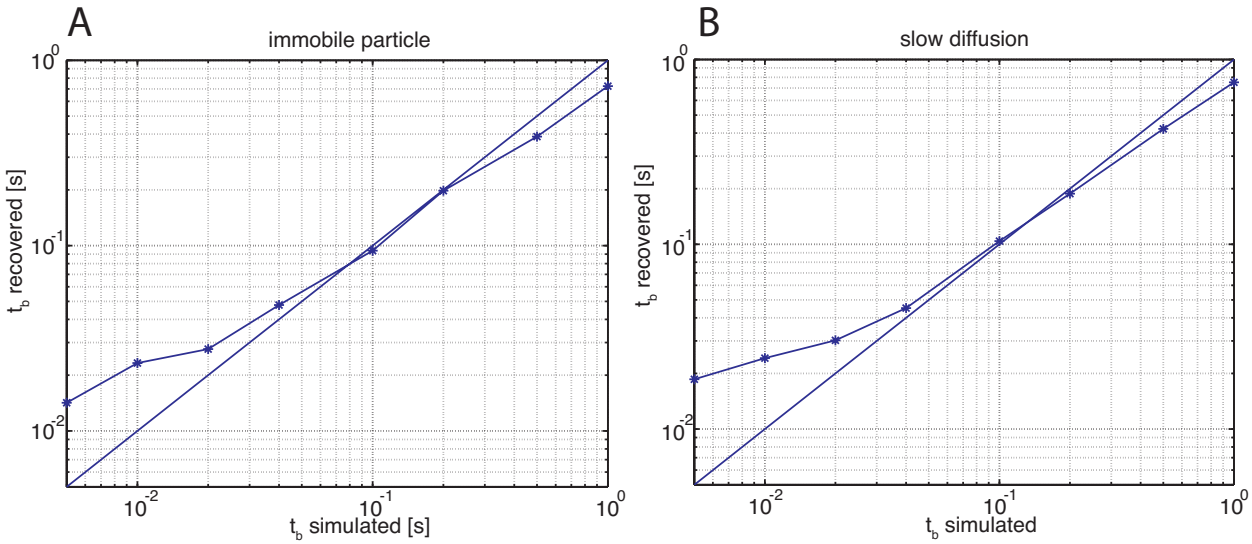

Fig. S3 Recovered binding times versus the simulated binding time.

(A) In the bound state the particle position fluctuated around a mean value and (B) the particles performed a slow diffusion in the bound state. The coincidence with the diagonal is a measure for the accuracy of the detection algorithm. Binding events were detected with L_c=0.5 and n=3, what yielded the smallest sum of weighted differences between simulated and recovered binding times. Δt=0.005 s, D_b =0.05 µm²/s, D_f=0.5 µm²/s.

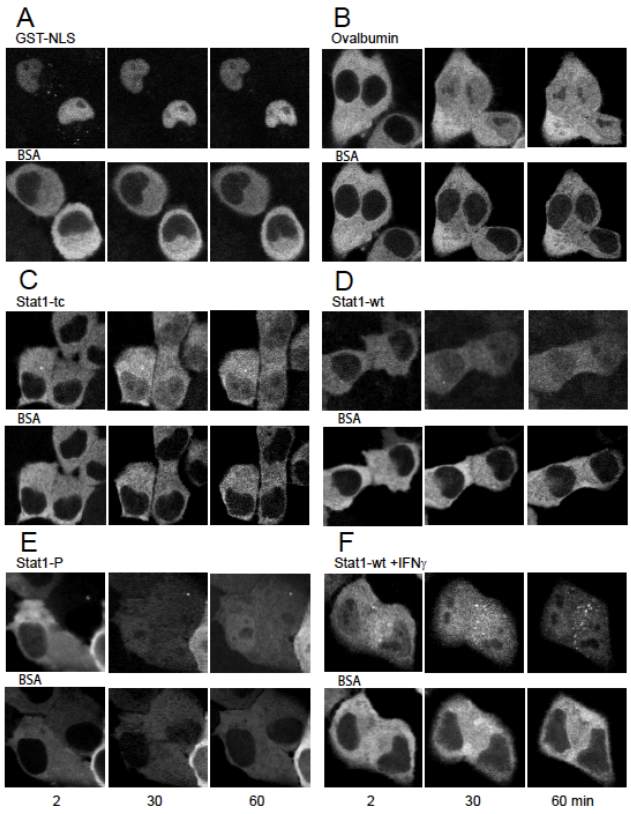

Fig. S4: Intracellular transport and distribution of fluorescent STAT1

The images show the time course of the co-injection of ATTON647 conjugated transport substrates (A-F, upper images) and AF488 conjugated bovine serum albumin (A-F, lower images, BSA) into the cytoplasm of HeLa S3 cells at 37°C. The time after microinjection is given in minutes below the graphics. The distributions of the injected substrates were followed for the period of one hour within the cells. (A) translocation process of an import control protein, GST-NLS; (B) control protein ovalbumin (for details, see Speil and Kubitscheck, 2010); (C) STAT1-tc; (D) wt-STAT1; (E) STAT1-P and (F) wt-STAT1 during IFNγ activation. The co-injected BSA-AF488 marked the site of injection and served as control of nuclear envelope integrity. Ovalbumin, wt-STAT1 and STAT1-tc translocated into the nucleus. Nuclear accumulation, in contrast, was observed only with GST-NLS, STAT1-P, and wt-STAT1 after incubation with IFNγ (5ng/ml). Object field size, 50x50 μm².

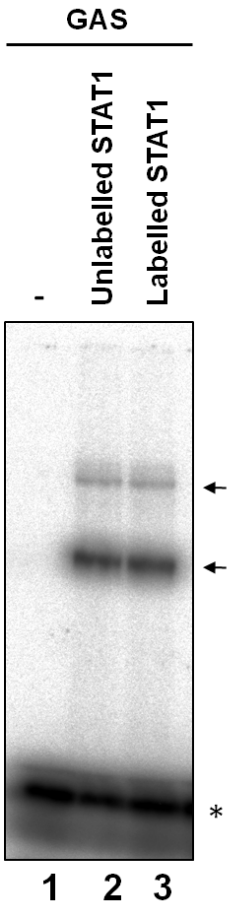

Fig. S5 Labelling of STAT1 with an amine-reactive fluorescent dye did not alter DNA binding

The figure shows electrophoretic mobility shift analyses (EMSA) using highly purified tyrosine701-phosphorylated STAT1 (2.5 nM) before (lane 2) and after (lane 3) labelling with succinimidyl ester-activated Alexa488. STAT1 was not included in lane 1. The positions of dimeric (lower arrow) and tetrameric (upper arrow) STAT1/DNA binding complexes are indicated, as well as the unbound probe (*). Radioactively labelled DNA contained a single high-affinity M67 STAT1 binding site and was used at a concentration of 1 nM. Representative of two independent experiments.

Summary of trajectory data and fitting results

A summary of the fitting results and detailed information concerning the analysed data were given in Table S2. Fits with exponential decay function were applied to all histograms (Fig. 5). For activated STAT1 inside the nucleus a mono-exponential obviously failed to describe the data (Fig. 5E and F), and a double-exponential decay function was fitted in these cases. The number of molecular jumps, the number of binding events and bound molecule positions for these data sets strongly increased inside the nucleus in comparison to the cytoplasm.

Data Set	No. Trajs	Total Jumps	Bound Jumps	Fraction of Bound Jumps	Binding Events	$\tau_{b,1}$/s	$\sigma_{b,1}$/s	$\tau_{b,2}$/s	$\sigma_{b,2}$/s
STAT1-tc Cyt.	152	15630	6207	0.4	72	0.02	0.02	1.8	3
STAT1-tc Nuc.	7	275	0	-	0	*insufficient data*			
wt-STAT1 Cyt.	161	12958	4422	0.34	70	0.04	0.005	4	10
wt-STAT1 Nuc.	8	312	153	-	7	*insufficient data*			
STAT1-p (Cyt)	93	10150	3791	0.37	77	0.03	0.003	10	90
STAT1-p (Nuc)	193	22101	13873	0.62	246	0.02	0.002	0.5	0.3
wt-STAT1 IFNγ (Cyt)	66	6331	2106	0.33	40	0.03	0.004	0.7	2.7
wt-STAT1 IFNγ (Nuc)	259	26461	14661	0.55	381	0.03	0.01	0.5	0.4

Table S1: Trajectory statistics and fitting results

The table summarizes trajectory statistics and the fitting results for the case that a double-exponential decay function was used to describe the data. The first four columns contain the number of analyzed trajectories within the particular data sets, the total number of molecular jumps evaluated, the number of bound positions according to the analysis of the L_E-plots, the fraction of bound jumps related to the total number, and the number of binding events detected. The next four columns report the fitting results for the binding times $\tau_{b,i}$ of the two exponential decay functions together with the respective standard deviations, $\sigma_{b,i}$.

Dependence of binding duration τ_2 on the value of L_c

The lengths of the immobile trajectory segments depended on the exact choice for L_c. We varied L_c for the trajectories of IFN-activated STAT1 and STAT1-P within the nuclei, and determined new binding time histograms. Again, the histograms were fitted by double-exponential decay functions. The table S1 reported the retrieved binding times τ_2.

L_c	τ_2 STAT1& IFNγ [s]	τ_2 STAT1-P [s]
0,05	0,08±0,03	0,12±0,09
0,09	0,4±0,033	0,12±0,09
0,20	0,47±0,25	0,45±0,34
0,30	0,44±0,25	0,49±0,27
0,40	0,50±0,28	0,50±0,27
0,50	0,63±0,37	0,48±0,19
0,60	0,37±0,16	0,37±0,12
0,70	0,30±0,10	0,45±0,16
0,80	0,36±0,12	0,32±0,10
1,00	0,23±0,05	0,25±0,06

Table S2: Dependence of the retrieved binding time τ_2 on L_c.

L_c was varied from 0.05 to 1. For small and for large values of L_c the determined value of τ_2 became smaller. For L_c close to zero long sections of the trajectories were defined as „bound“. However, since we counted only the durations of those trajectory sections, for that we could detect both the beginning and the end of the bound phase (meaning where the ratio σ_b^2/σ_f^2 first exceeded L_c and then fell below it), only a few short „binding events“ remained detectable. For large values of L_c the sections that exceed the threshold become shorter and shorter due to stochastic reasons. Therefore, the detected binding times become increasingly shorter. A choice of L_c close to 0.5 yielded a stable result for τ_2.