Extracting landscape features from single particle trajectories

Ádám Halász[1], Brandon Clark[1], Ouri Maler[1,3], Jeremy Edwards[2]

[1]Dept. of Mathematics, West Virginia University, Morgantown, WV, USA
[2]Dept. of Chemistry & Chemical Biology, University of New Mexico, Albuquerque, NM, USA
[3]*current address:* VERIMAG / Université Grenoble Alpes

HSB 2019 – 6th International Workshop on Hybrid Systems Biology,
Charles University, Prague, Czechia, April 6-7, 2019

1. Background, motivation, or why do this at all?

Signal initiation: biological significance

Modeling of signal initiation

Experimental modalities

Phenomenology, importance of landscape

Challenge of model validation using high resolution data

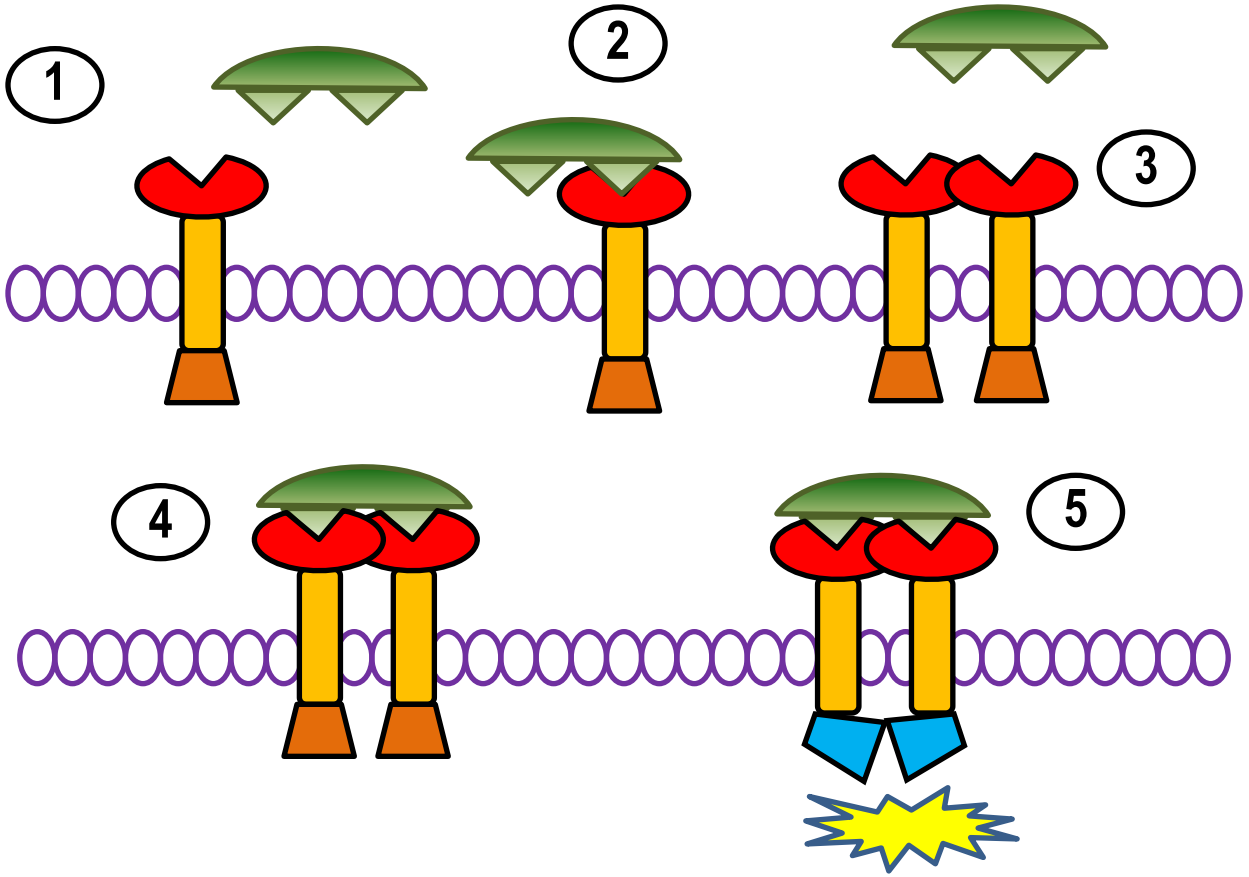
Signal initiation by membrane bound receptors

- Relevant to cancers, immune conditions
- EGF (ErbB2, ErbB3), VEGF, pre-B, FcεRI
- Receptors located on the cell membrane
- Ligand (“signal”) binds to receptors
- A sequence of transformations results in downstream signal propagation

Signal initiation by membrane bound receptors

VEGF signal initiation relies on **ligand induced dimerization**

$$\boldsymbol{R + V \leftrightarrow VR} \;\; ; \;\; \boldsymbol{VR + R \leftrightarrow \Delta} \; ; \;\; \boldsymbol{\Delta \leftrightarrow \Delta^{*}}$$



Complex reaction patterns

Ligand induced oligomerization of receptors

EGF / ErbB: $LR + LR \rightarrow (LR)_2 \rightarrow (LR)(LR^*)$

preB: $R + R \rightarrow RR;\ RR + R \rightarrow RRR \rightarrow \cdots$

Cross-phosphorylation of bound receptors

$$\left(LR^{(i)}\right)\left(LR^{(j)}\right) \rightarrow \left(LR^{(i)}\right)\left(LR^{(j*)}\right)$$

Successive phosphorylation events (**kinetic proofreading**)

$$(RP)(R) \rightarrow (RP)(R^*) \rightarrow (RP^*)(R^*)$$

Complexity:

$$\begin{bmatrix} k \\ \text{receptor} \\ \text{types} \end{bmatrix} \times \begin{bmatrix} m \\ \text{possible} \\ \text{proteins} \end{bmatrix} \times \begin{bmatrix} n \\ \text{phos} \\ \text{states} \end{bmatrix} \times \begin{bmatrix} \text{oligomer size} \\ \text{phos of proteins} \\ \text{labeling} \end{bmatrix}$$

Stochastic, **rule-** and **agent-based** representation (“on the fly” species)

Modeling (stochastic, rule based, "network free")

Complexity: large set of reactions $(A + B \leftrightarrow C)$ or $(A \leftrightarrow A')$,

Many are the same transformation of 1-2 basic species

$$LRR^* \rightarrow LRR \;\; ; LR^* \rightarrow LR \; ; \;\; R^* \rightarrow R$$

$$R + R \rightarrow RR; \;\; LR + R \rightarrow LRR \; ; LR^* \rightarrow LR^*R$$

Good idea: Identify basic species and "rules"*

$$\{L, R\} \; ; \{L + R \leftrightarrow LR; R + R \leftrightarrow RR; R \leftrightarrow R^*\}$$

(a) Create **list of species** and reactions, track amounts of each

(b) Agent based approach: track the **state of each copy** of the basic species (makes sense when **evolution** is **stochastic**)

(* as done in Kappa, also BioNetGen / NFSim)

Experimental collaboration

Wilson & Lidke labs at U. of New Mexico

- Super-resolution optical microscopy
- Receptors labelled with fluorescent quantum dots
- Labelled particles are detected based on the light (photons) they emit
- Location (centroid) is determined by fitting the distribution of detected light
- Resolution: $\mathcal{O}(10\mathrm{nm})$ spatial / 20+ frames per second
- Trajectories of particles are reconstructed using dedicated software (HMM etc.)

Context

We use spatially resolved simulations of the reaction networks to compare with microscopy data

- Extract parameters (e.g. dimerization / dissociation rates)
- Infer underlying landscape
- Use calibrated simulations to make predictions
- The data is one of the major sources of parameters for the simulation

2. Analysis of trajectories & domain reconstruction

Brownian motion – distributions & tests

Anomalous diffusion

Confinement – experimental evidence, possible impact

Results from observed & simulated trajectories

Domain reconstruction algorithm

Results with reconstructed domains

Analysis of Jump Size Distributions

Brownian motion [equiv. to diffusion: $\rho_t = D(\rho_{xx} + \rho_{yy})$]:

- displacements $\Delta x, \Delta y$ follow a normal with $\boxed{\sigma^2 = 2D\Delta t}$:

$$f_{xy}(x, y) = \frac{1}{4\pi Dt} e^{\left(-(\Delta x^2 + \Delta y^2)/4Dt\right)}$$

- the **square displacement*** $s \equiv \left|\overrightarrow{\Delta r}\right|^2 = \Delta x^2 + \Delta y^2$ follows

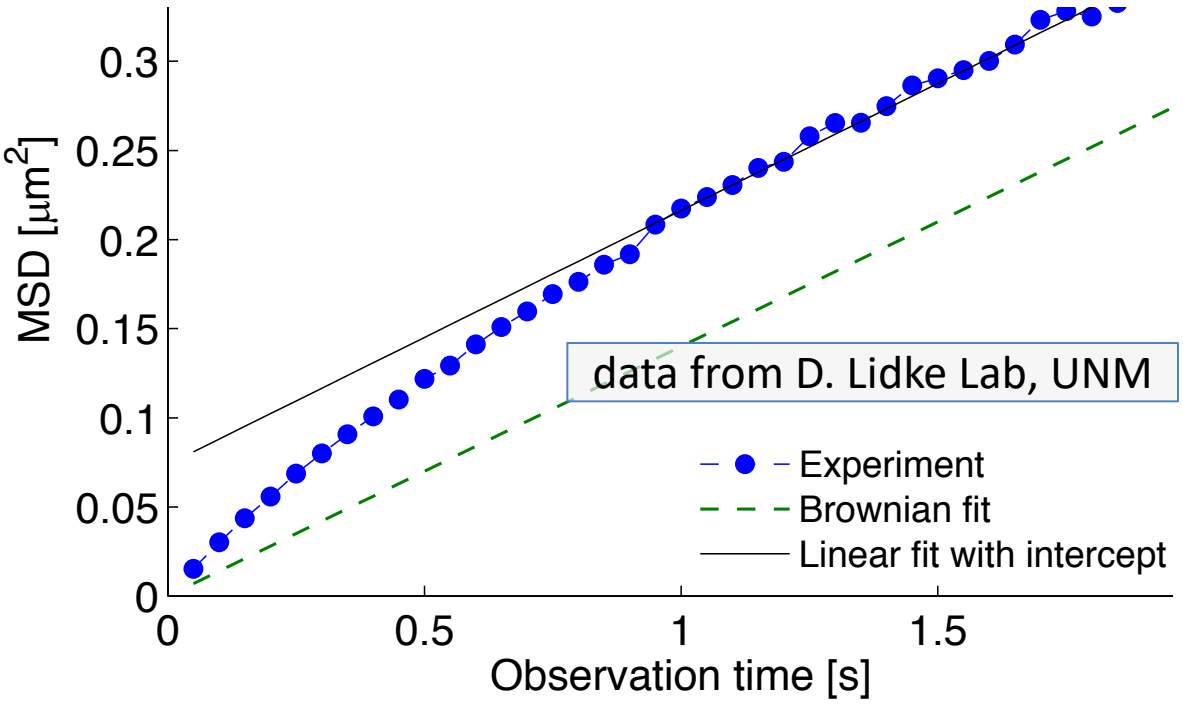
$$f_s(\Delta r^2) = \frac{\exp(-\Delta r^2/4Dt)}{4Dt} \Leftrightarrow \boxed{f_s(s) = \frac{\exp(-s/2\sigma^2)}{2\sigma^2}}$$

- the **mean square displacement (MSD)** : $\boxed{\langle \Delta r^2 \rangle = 4Dt}$

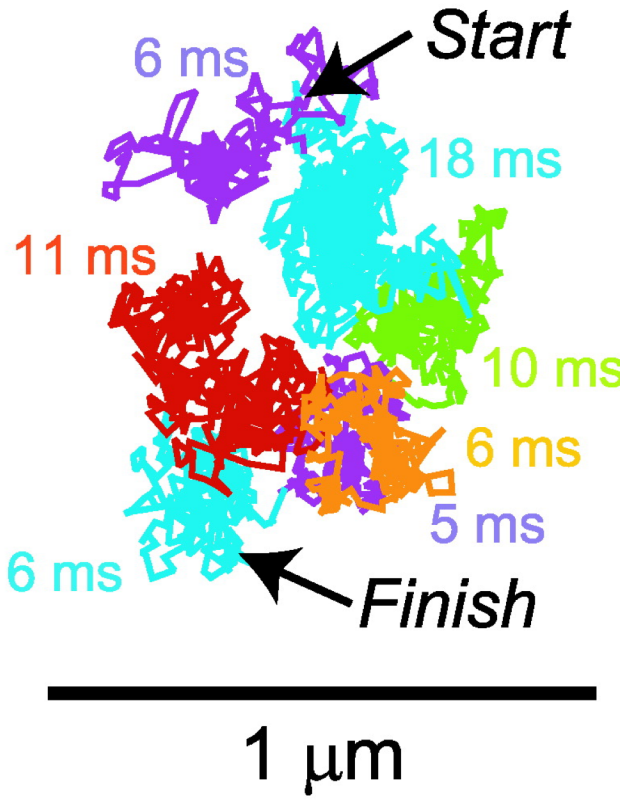
*[$\iint \cdots dxdy = \int \cdots rdr \int d\phi = 2\pi \int \cdots rdr = \pi \int \cdots ds$]

Reconstructed trajectories: Anomalous Diffusion

- Mean Square Displacements (expected proportional to t) have a decreasing slope



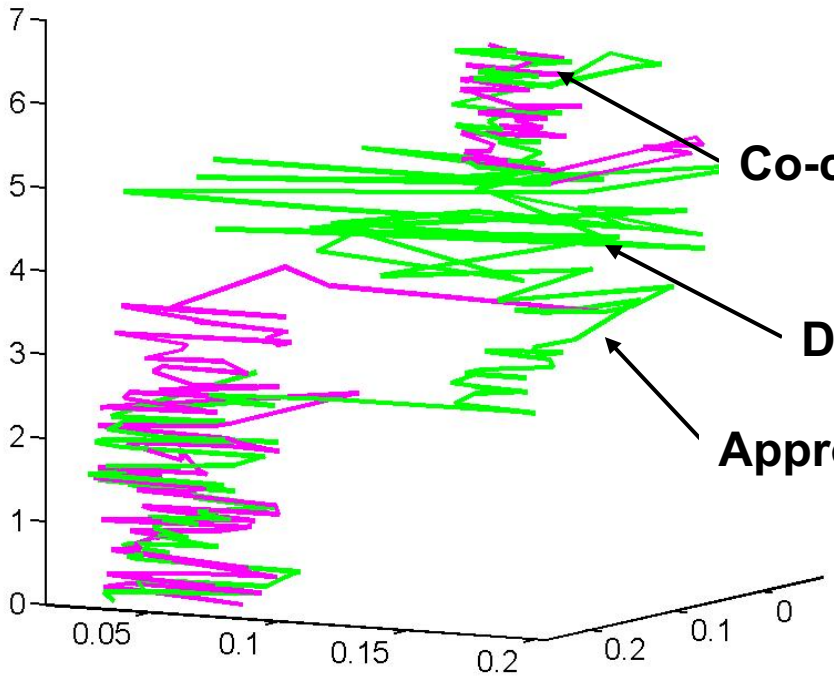
Kusumi, Annu. Rev. Biophys. & Biomol.Struct. (2005)



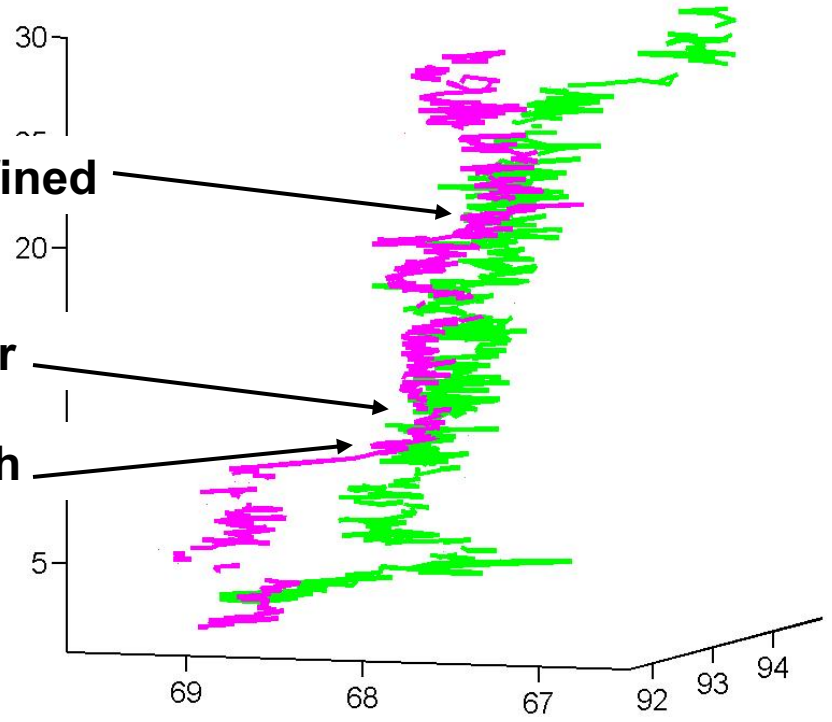
- There is ample evidence of a non-uniform movement, typically described as transient confinement

Co-confinement vs. dimerization

Simulation



Experimental Data



Co-confined

Dimer

Approach

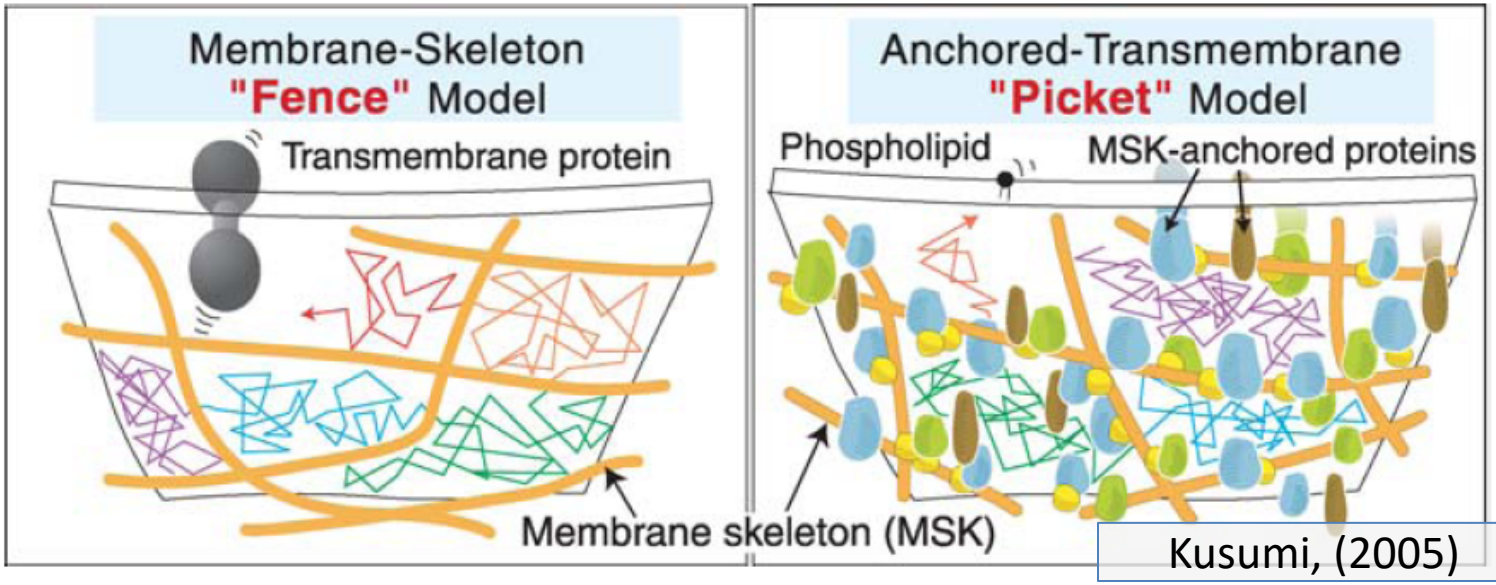
McCabe et al, Biophys. J., (2011)

Low-Nam et al, Nature Struct. & Mol. Biol., (2011)

Confinement, microdomains, cytoskeleton

Single particle tracking shows confinement over short timescales

- A likely explanation is interaction with the cytoskeleton, which impedes movement of membrane proteins
- The transient localization is due to microdomains, induced by elements of the cytoskeleton
- Kusumi's work (early 2000's) is being revisited but the phenomenon of transient confinement is widely established



Kusumi, (2005)

Confinement, microdomains, cytoskeleton

Single particle tracking shows confinement over short timescales

- A likely explanation is interaction with the cytoskeleton, which impedes movement of membrane proteins
- The transient localization is due to microdomains, induced by elements of the cytoskeleton
- Kusumi's work (early 2000's) is being revisited but the phenomenon of transient confinement is widely established

Aspects of practical interest for modeling:

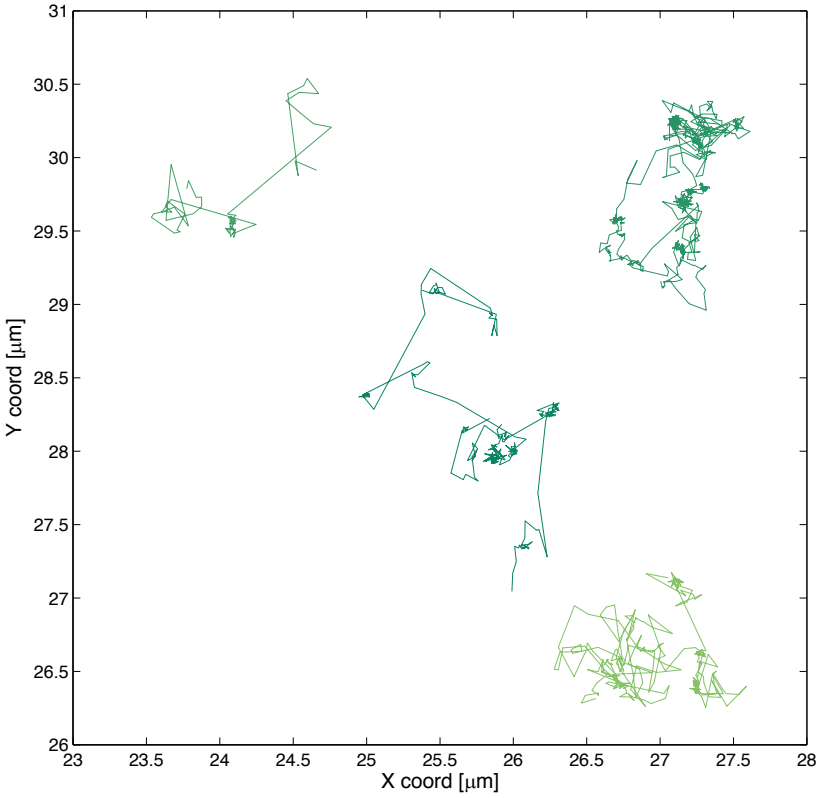
- Impact of microdomains on signal initiation kinetics
  - Receptor oligomerization
  - Effectiveness of (scarce) kinases involved in activation
- The physical mechanism that gives rise to microdomains

Both require a way to reliably **identify confining domains**

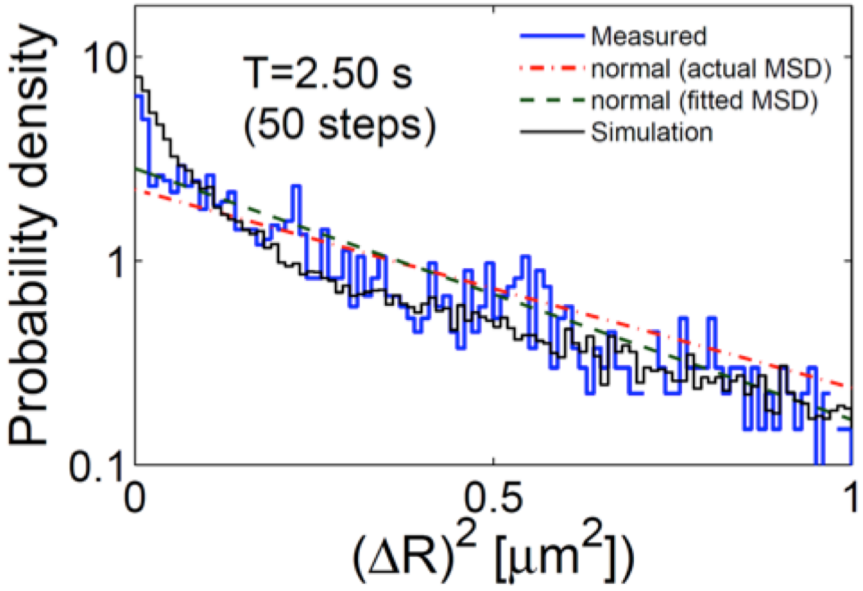
Modeling 1: Understanding the distributions

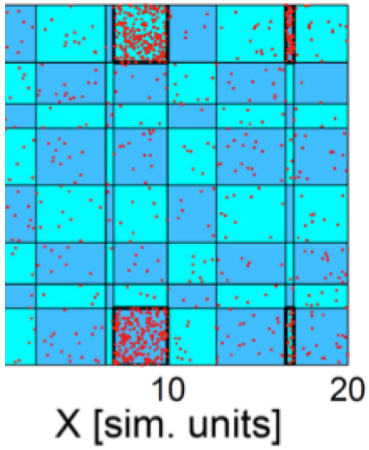
Distribution of jump sizes ➔

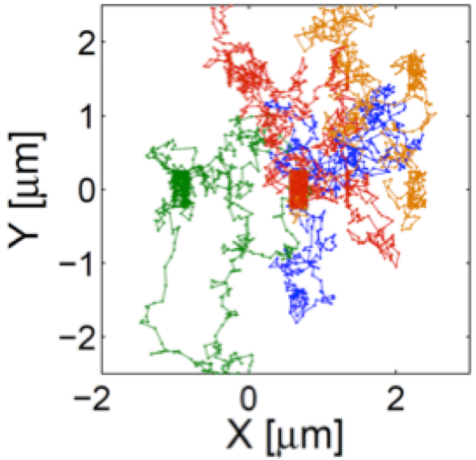
Actual ➘ and simulated ➘ trajectories



data from D. Lidke Lab, UNM







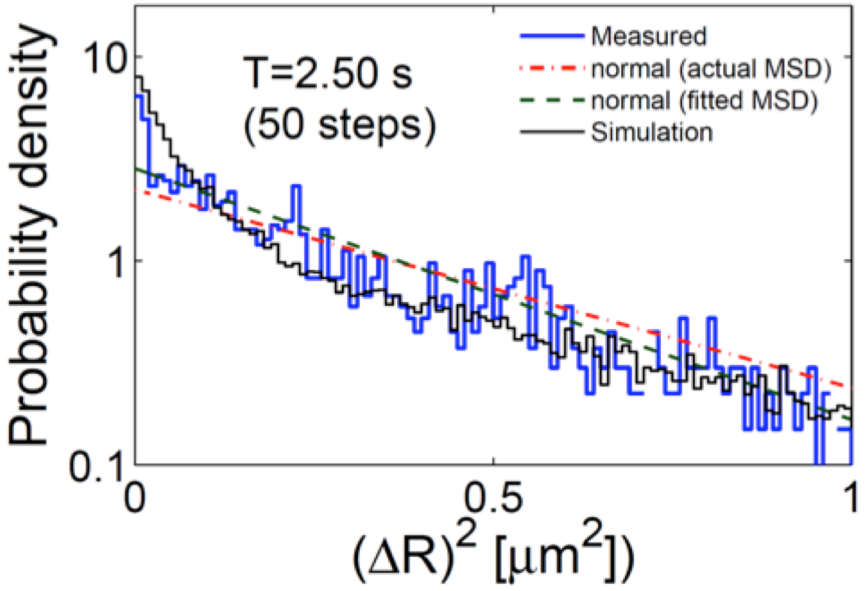
Modeling 1: Understanding the distributions

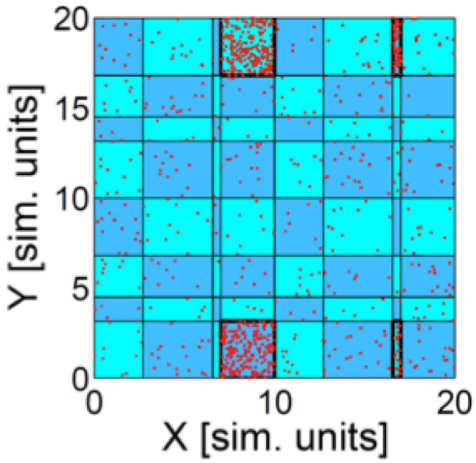
Distribution of jump sizes ➔

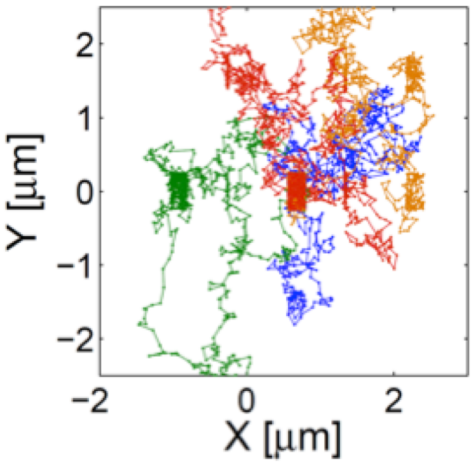
Actual 🡻 and simulated 🡾 trajectories

Simulated **confining domains** 🡾

data from D. Lidke Lab, UNM



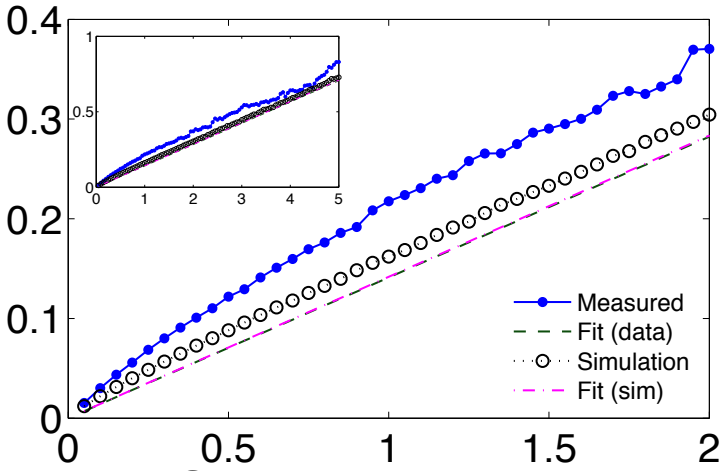


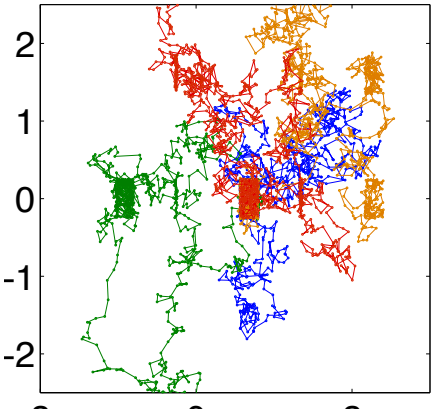


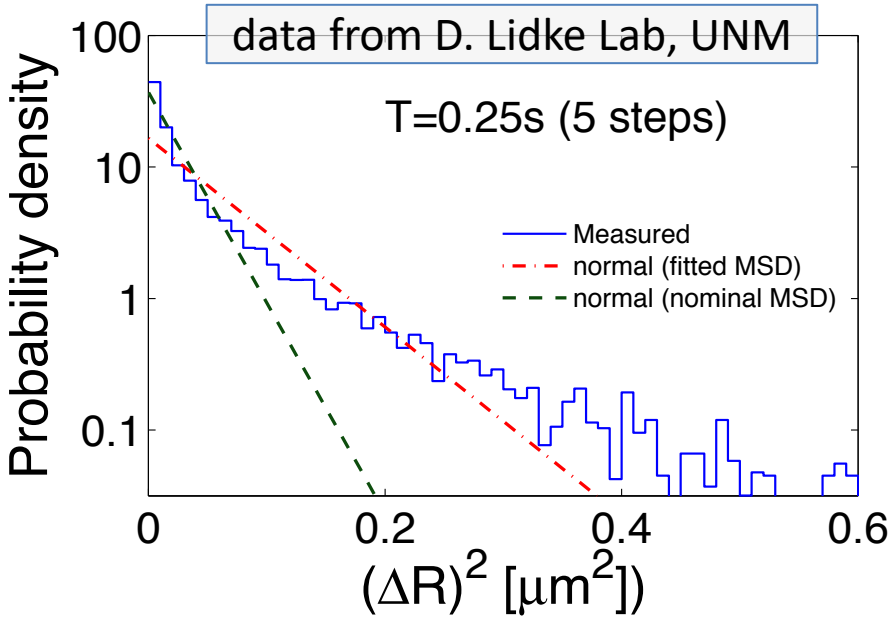
Modeling 1: Diffusion with confining domains

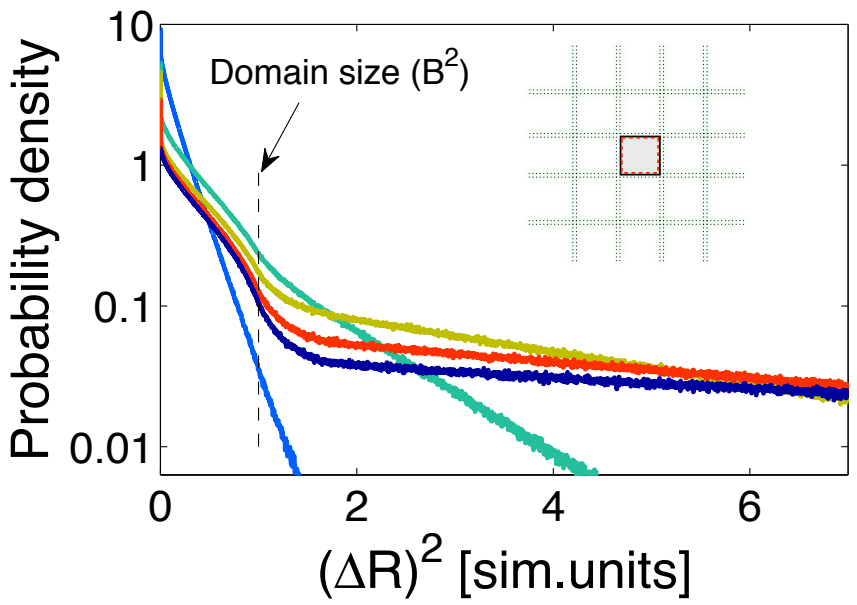
$$P(r^2) = \frac{1}{2\sigma_{xy}^2} \exp\left(-\frac{r^2}{2\sigma_{xy}^2}\right)$$

$$\langle r^2 \rangle = 2\sigma_{xy}^2(t) = 4Dt$$





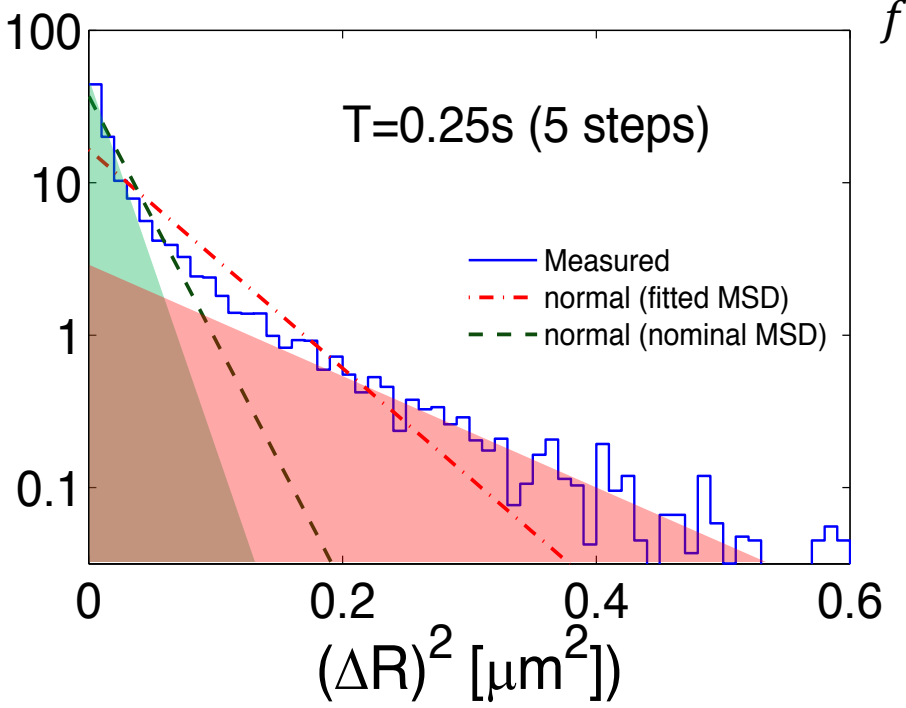




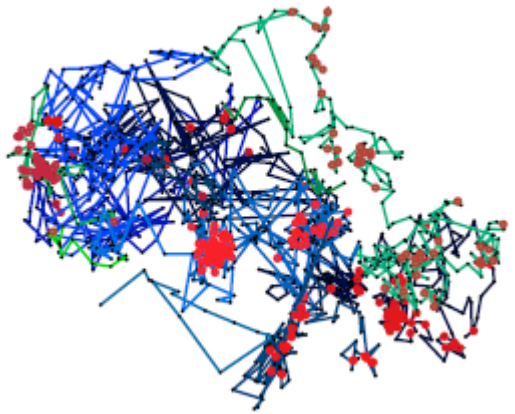
Modeling 1: Understanding the distributions

The 'hockey stick' distribution of jump sizes reflects the existence of at least two populations of receptors

- Faster moving → molecules outside domains, diffusing freely
- Slower moving → molecules confined in domains



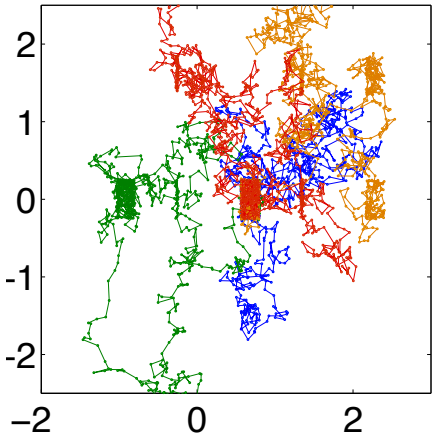
$$f(s) \approx Ce^{-s^2/2\sigma^2} \rightarrow \quad f(s) \approx \sum_j c_j e^{-s^2/2\sigma_j^2}$$

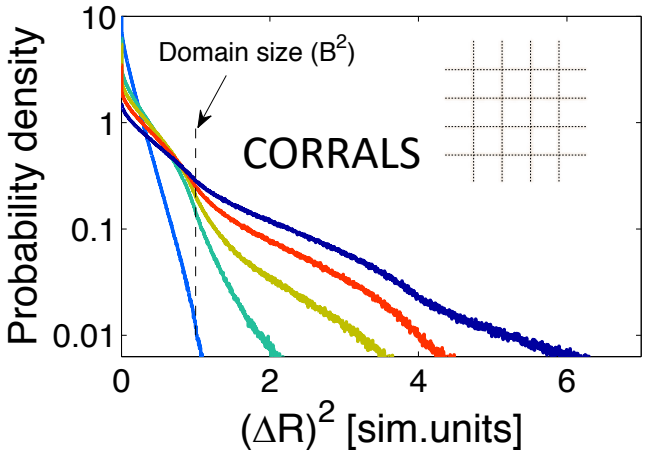


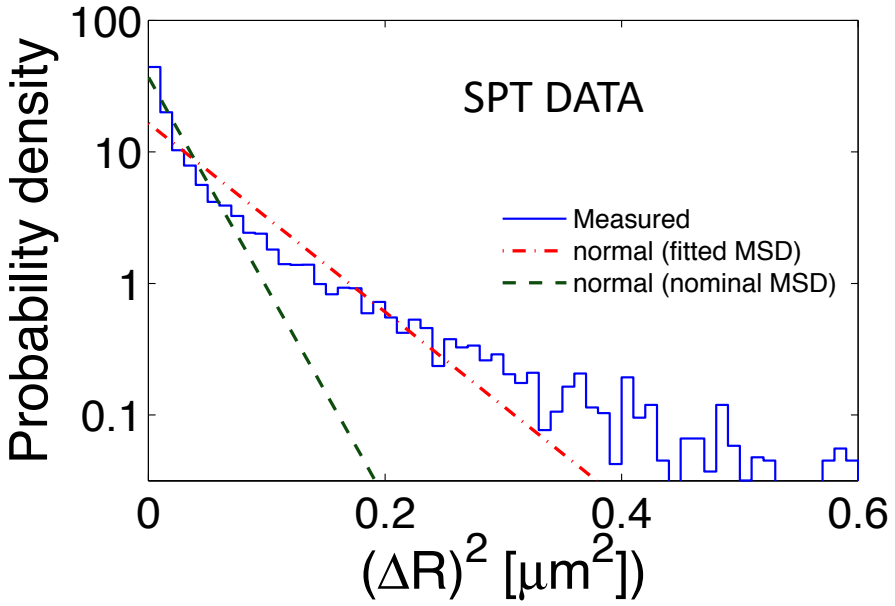
Modeling 1: Confining Domains vs. Corrals

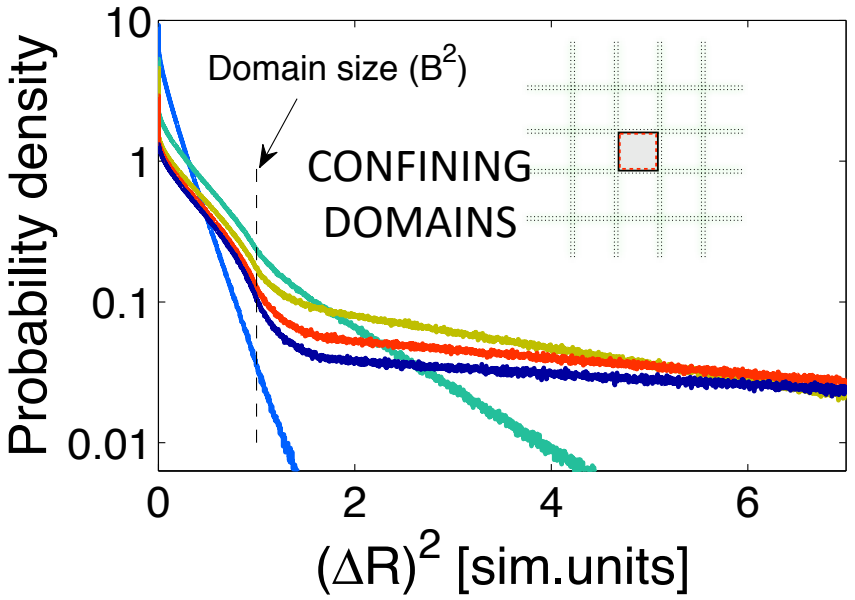
Brownian motion simulations in a landscape:

- Confining domains versus corrals
- The upward curved shape is reproduced only by confining domains
- Both reproduce the MSD time dependence (qualitatively)









Concerns

Only qualitative match

How do we know that the populations are not distinct molecule types?

- the same molecule can switch from one regime to another (confined / free)
- what if there are several types of molecules, some always fast, some always slow

Distributions were sensitive to the shape of the simulated domains

Why not **identify the domains**?

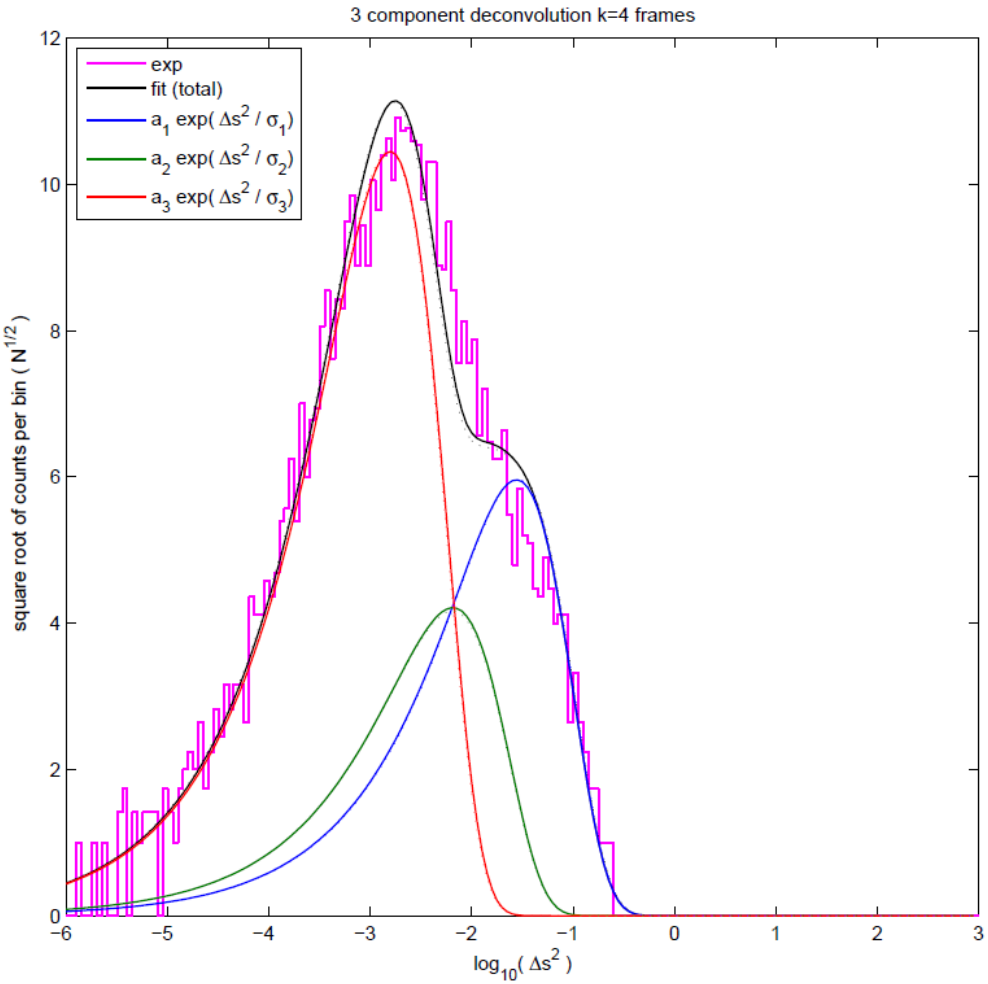
Analyzing the jump size distributions

More careful decomposition into sum of exponentials

Log binning

Error estimation based on number of counts per bin

Simulated annealing fit of two or three exponentials, for each numbe of steps



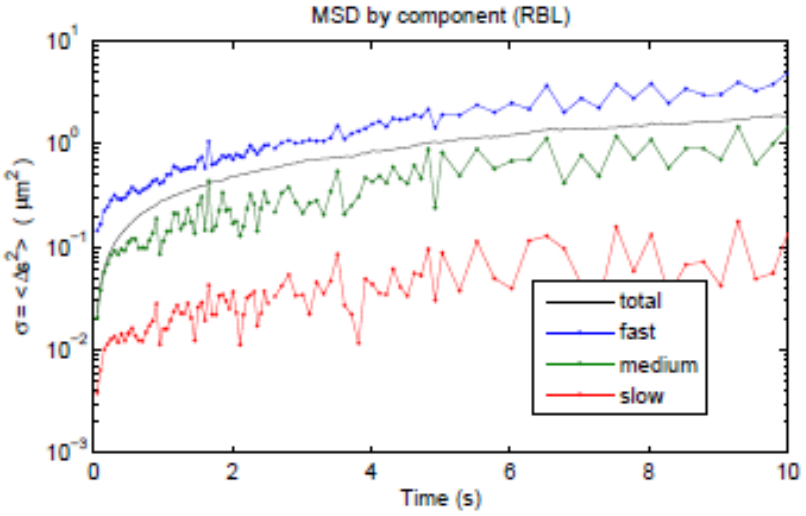
Analyzing the jump size distributions

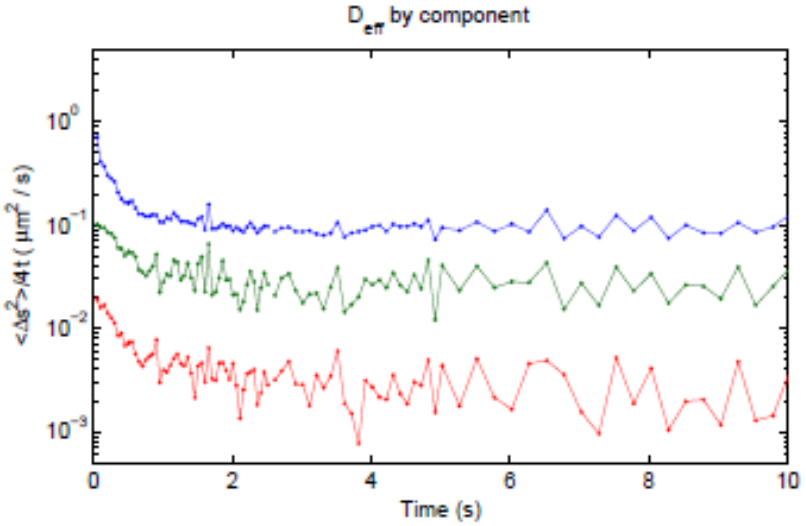
More careful decomposition into sum of exponentials

Log binning

Error estimation based on number of counts per bin

Simulated annealing fit of two or three exponentials, for each number of steps





Modeling 2: Domain Reconstruction

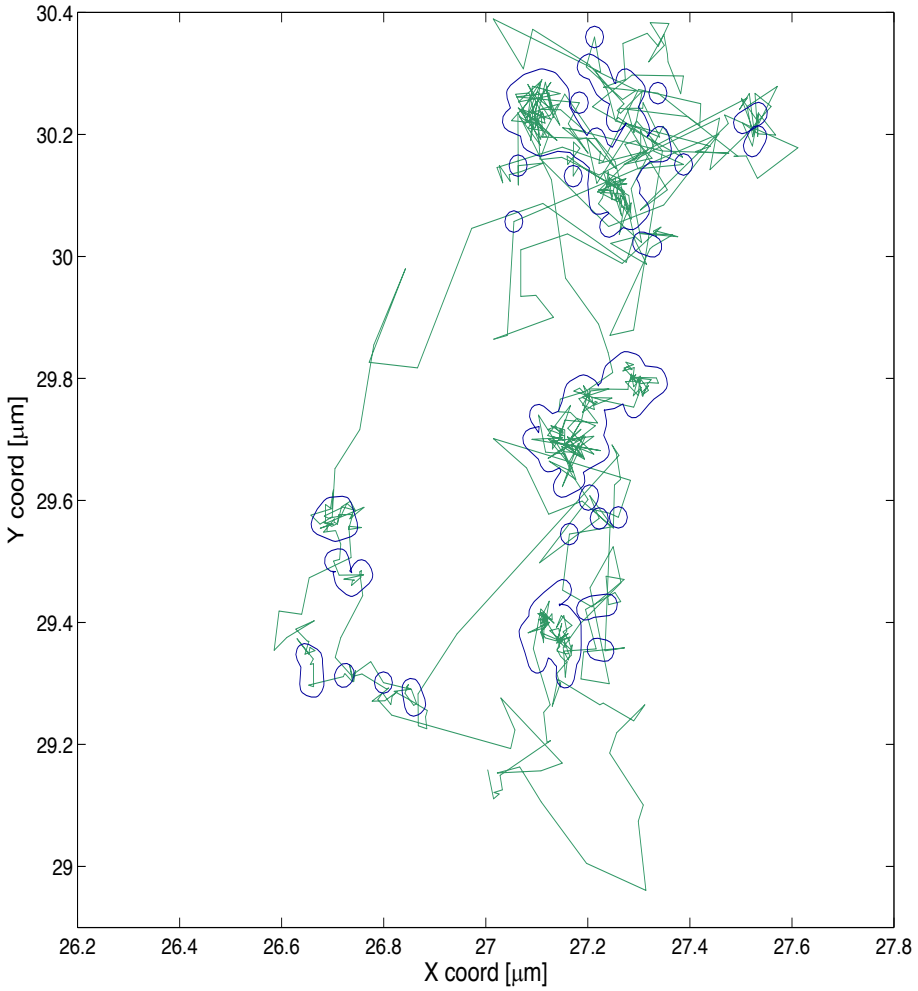
Estimate the likelihood that a given point in an SPT trajectory is part of the confined population or not

Attempt **to reconstruct** the confining domains that modulate the movement of the particles.

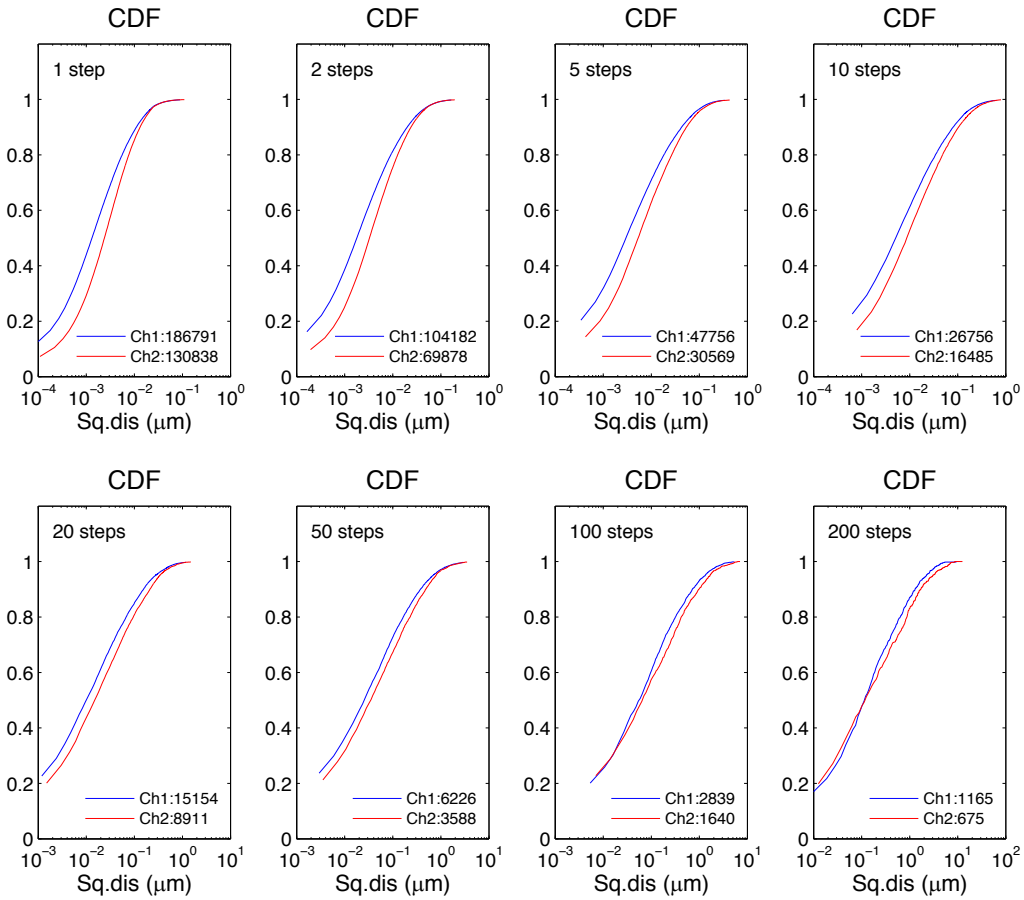
Modeling 2: Domain Reconstruction

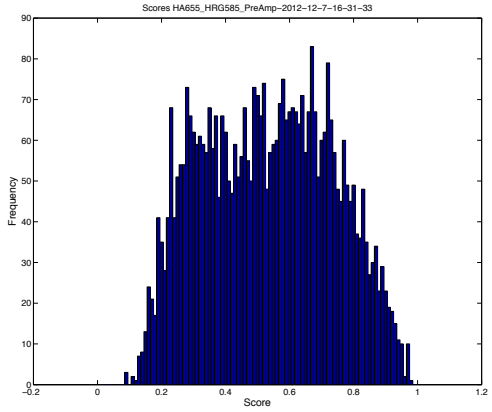
1. Construct a distribution of jump sizes for a selection of step (frame interval) numbers for the entire sample
2. Define a joint score as weighted average of the percentage rank for each point
3. Identify the sub-population of slower points
4. Cluster the identified points
5. Construct a contour around each cluster

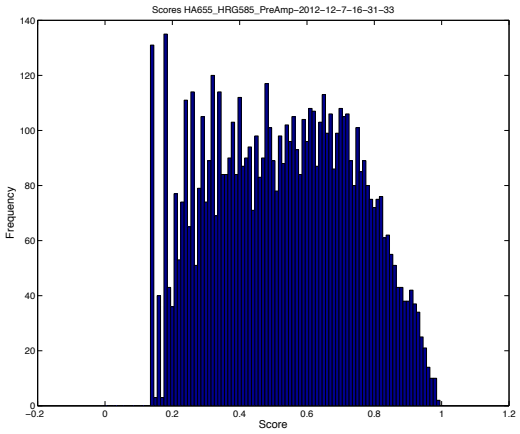
Domains from contours



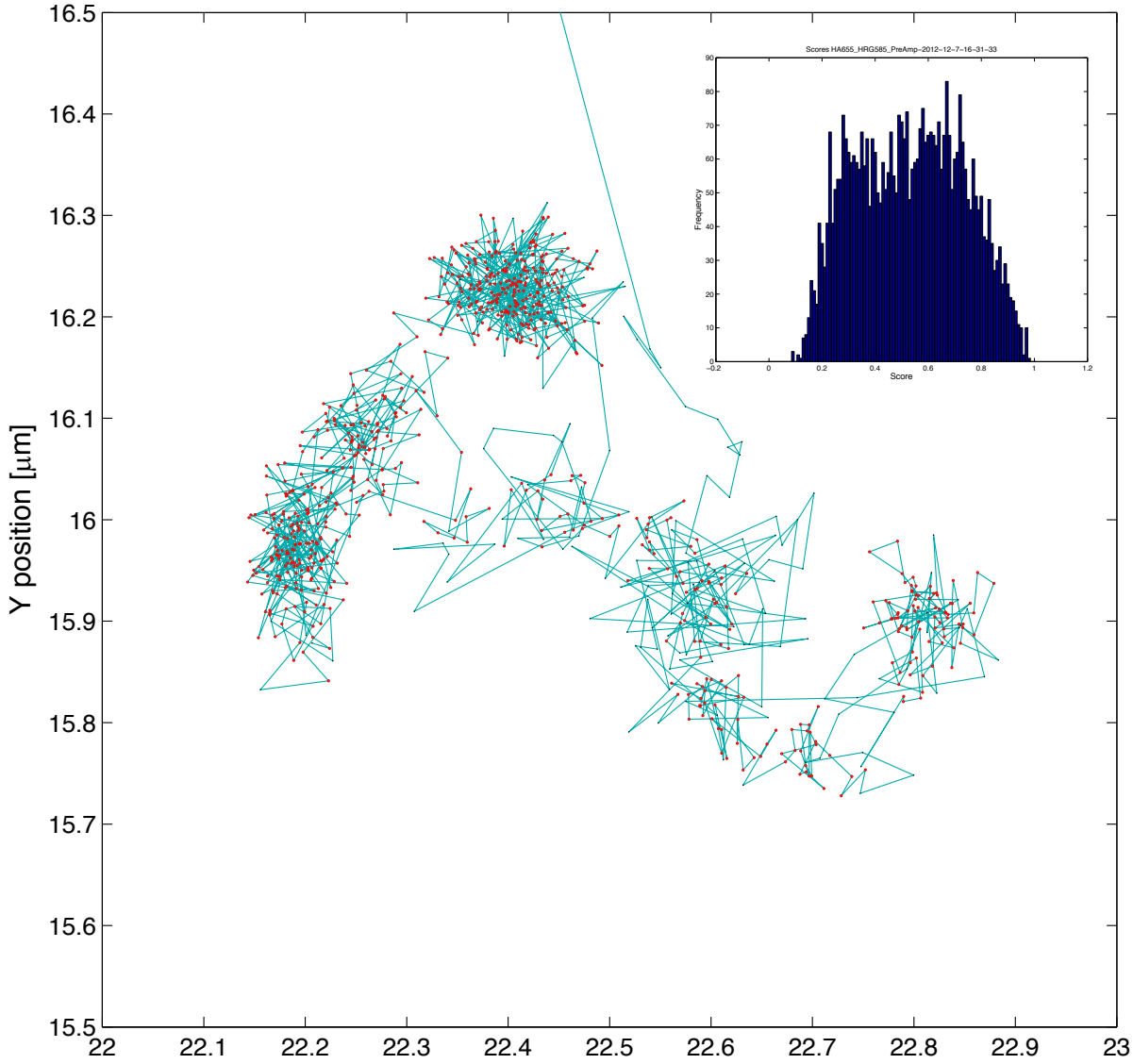
Step size distributions: build a cumulative score





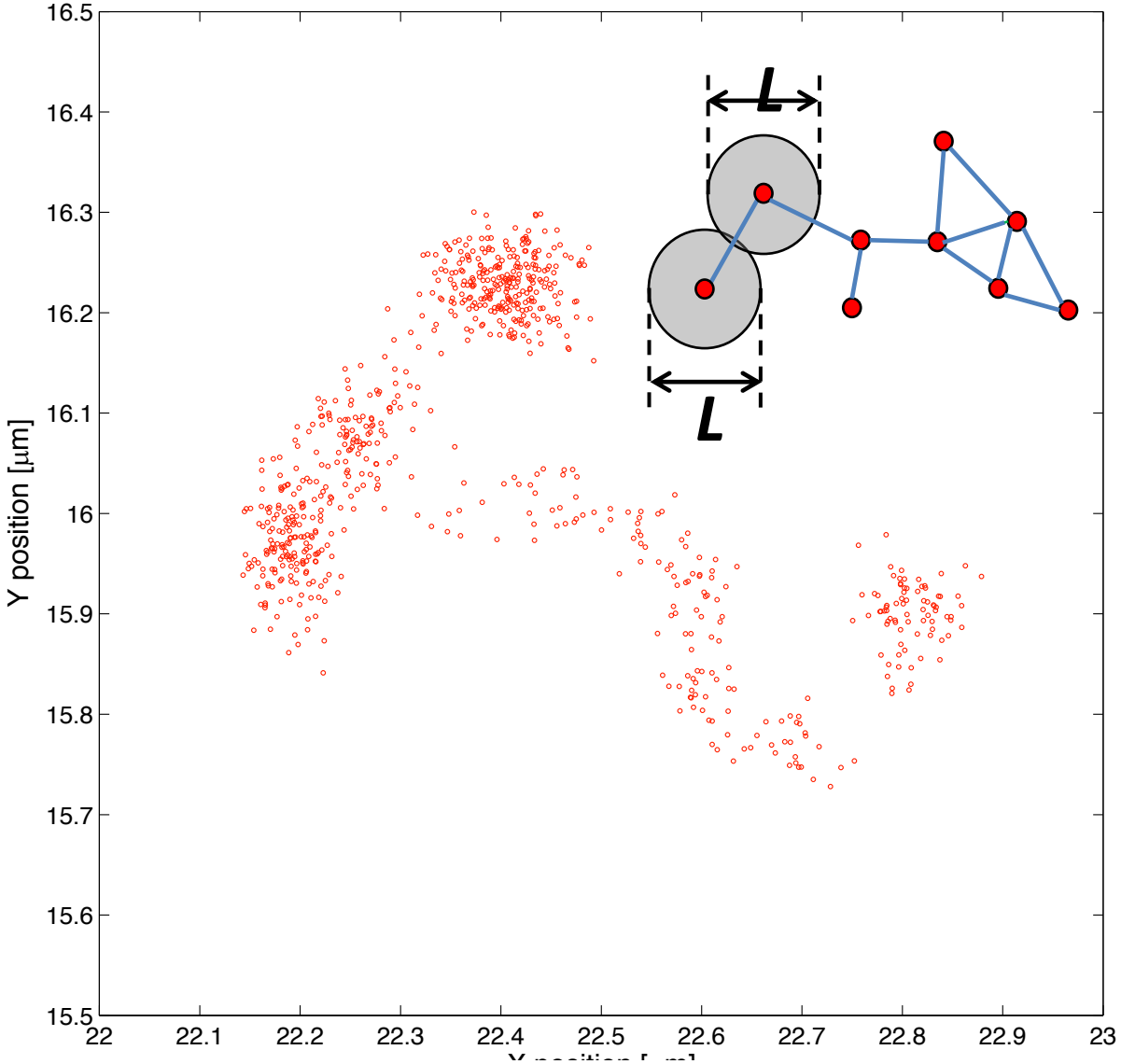


Domain Reconstruction: Find slow points



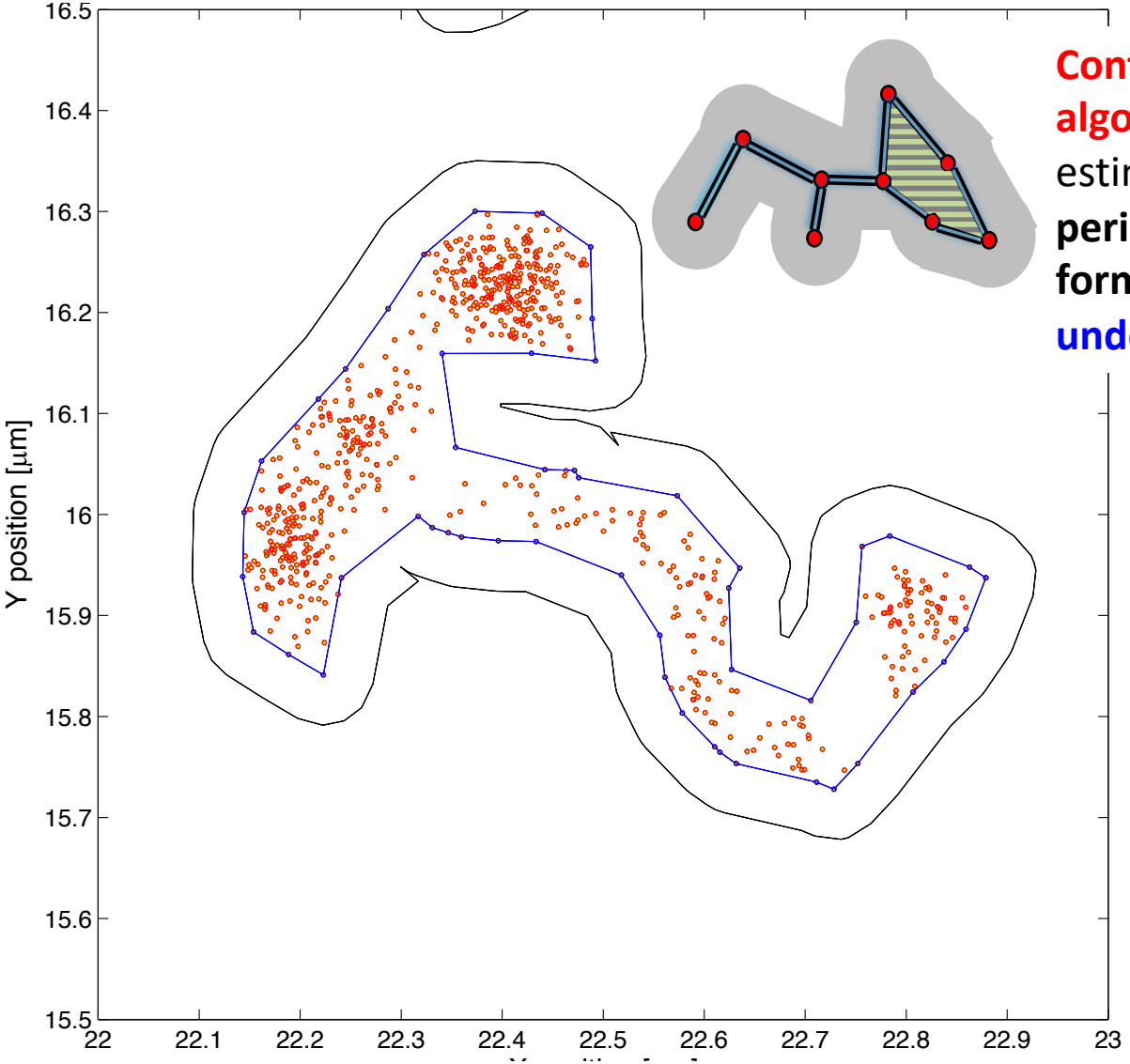
Slow points identified from a combined score based on jump sizes over several frame intervals

Domain Reconstruction: Cluster slow points



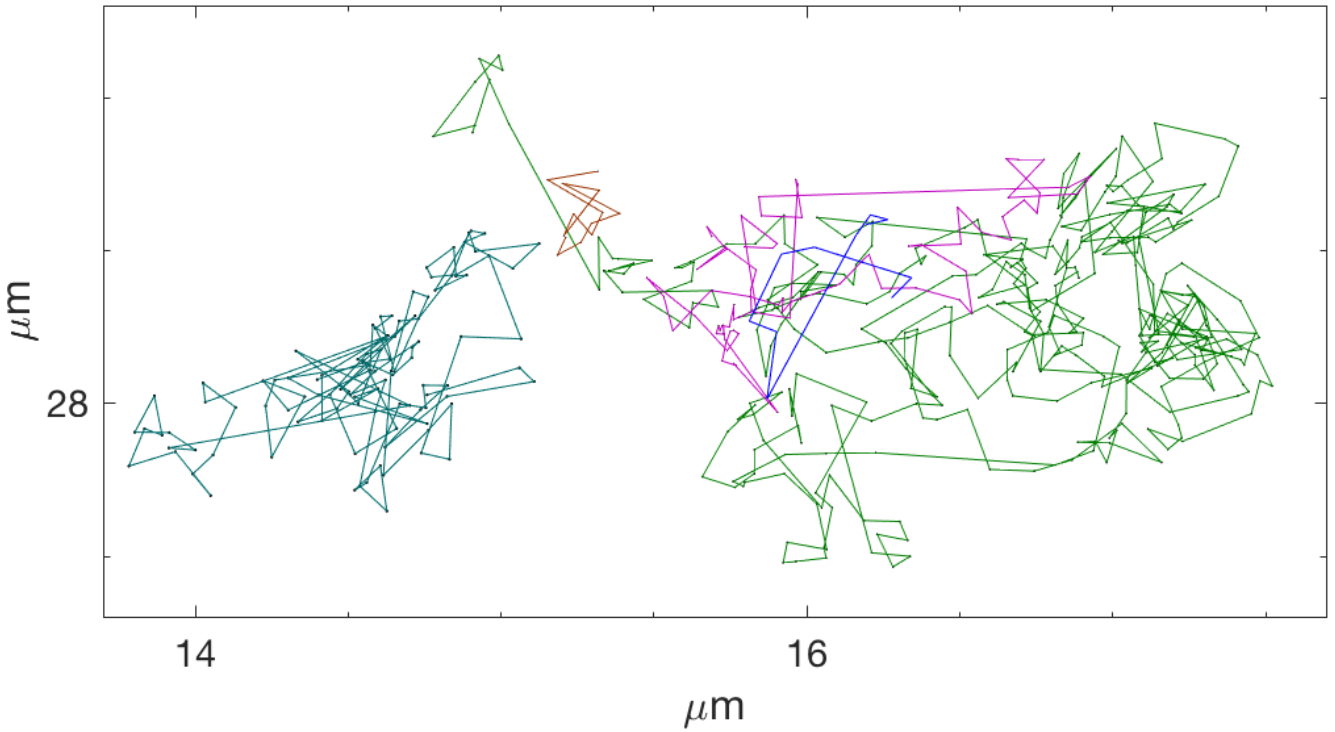
Clusters of points identified via distance based clustering

Domain Reconstruction: Footprint

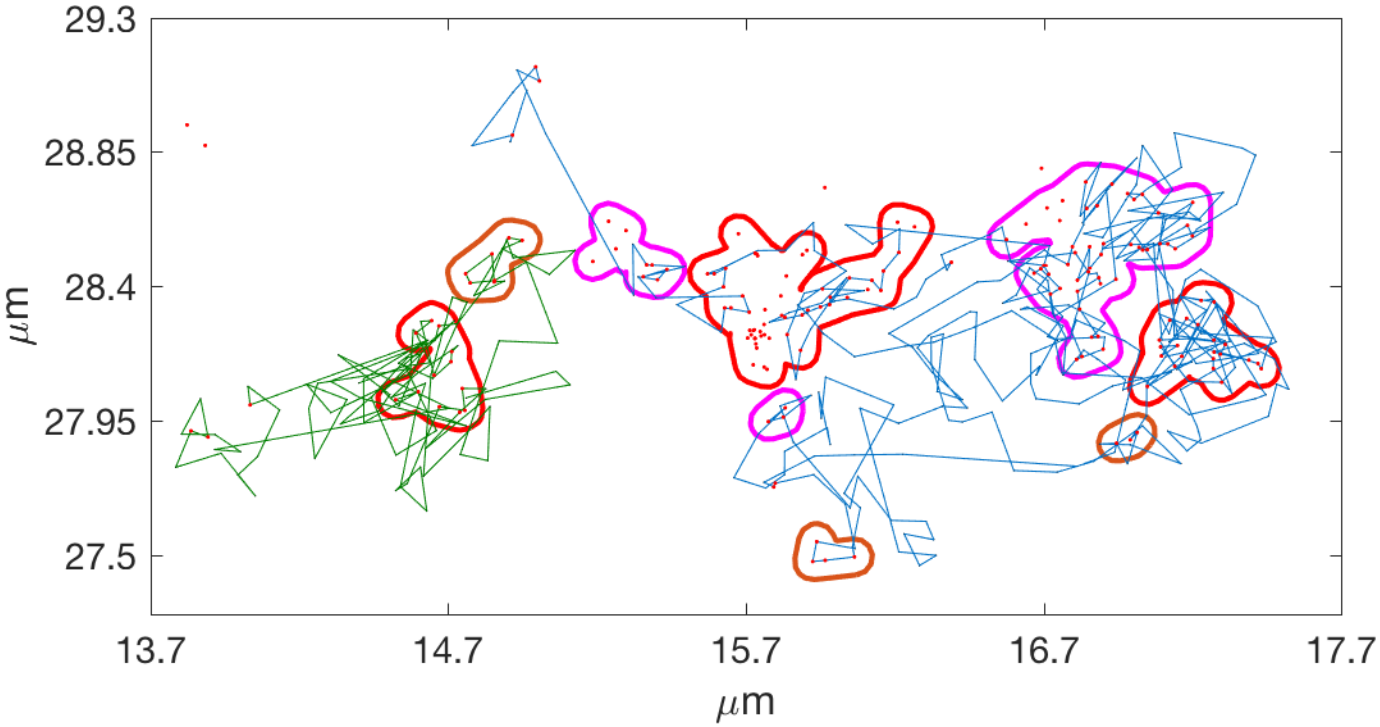


Contour building algorithm - estimates **area**, **perimeter**, and **form factor of underlying domains**

Domain reconstruction (pre-B)



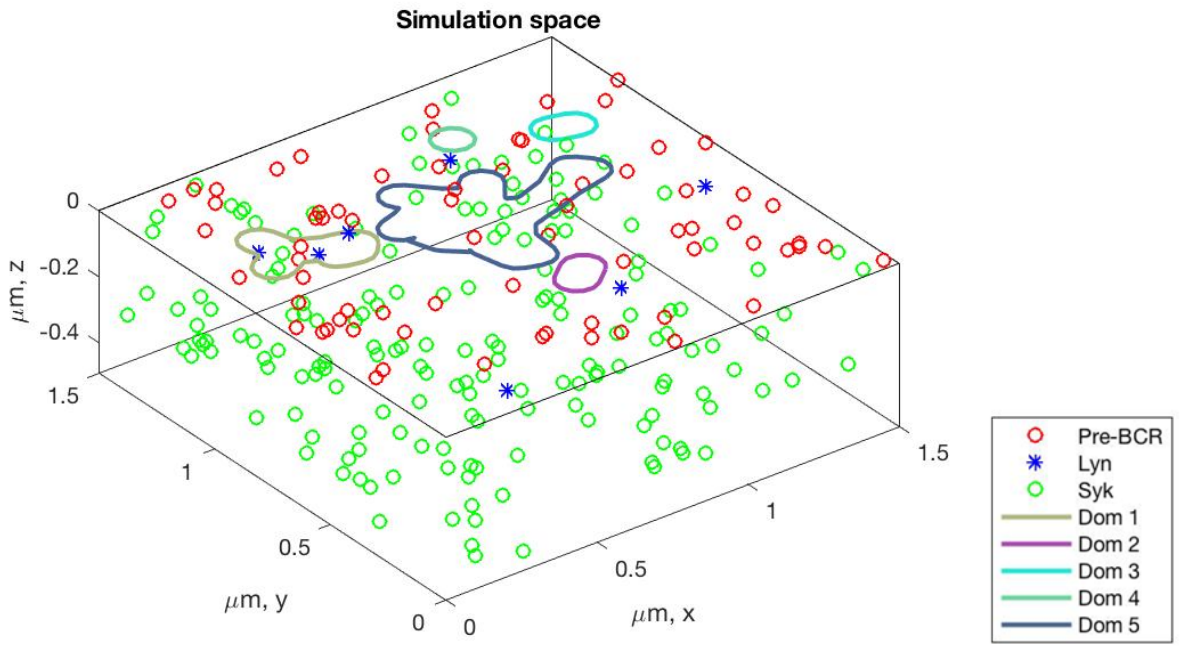
Domain reconstruction (pre-B)



Simulations with domains

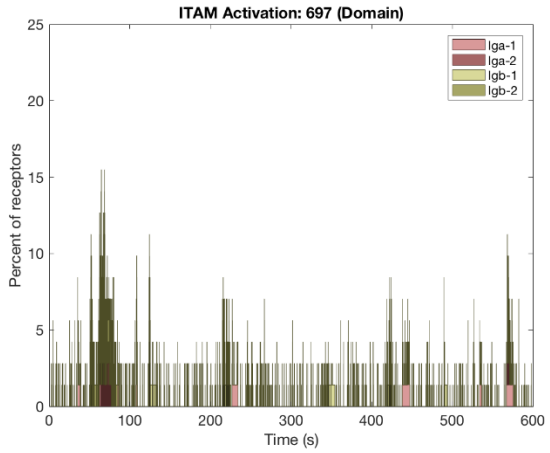
- Pre-B cell receptors
  - Receptors have two receptor binding domains
  - May form higher oligomers
  - Two additional proteins (kinases), Lyn and Syk
  - Phosphorylation through cross-activation mechanism involving three entities
- The system also has domains
- Trying to understand two patient samples
  - different levels of signaling in the absence of ligand
- Kerketta et al., in preparation / submitted

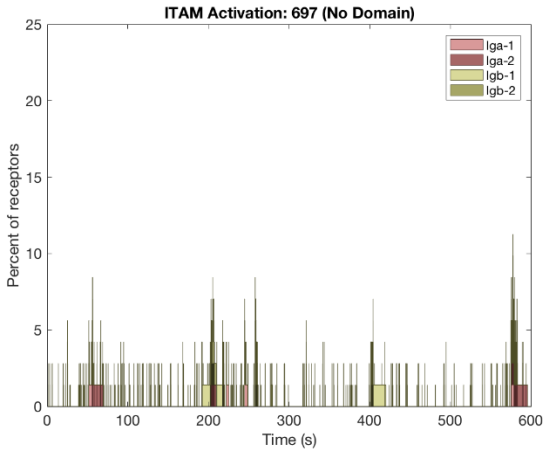
Simulations with domains



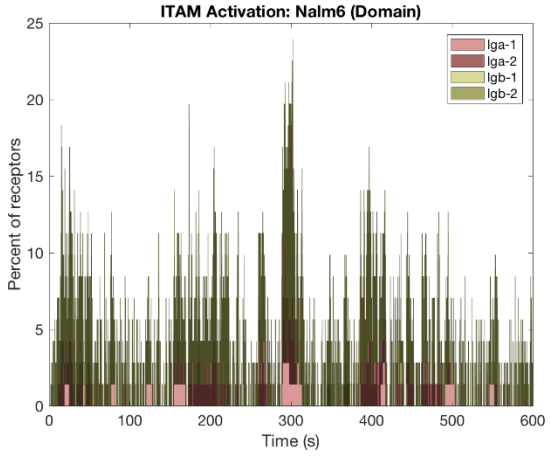
Kerketta et al. (in preparation)

Simulations with domains

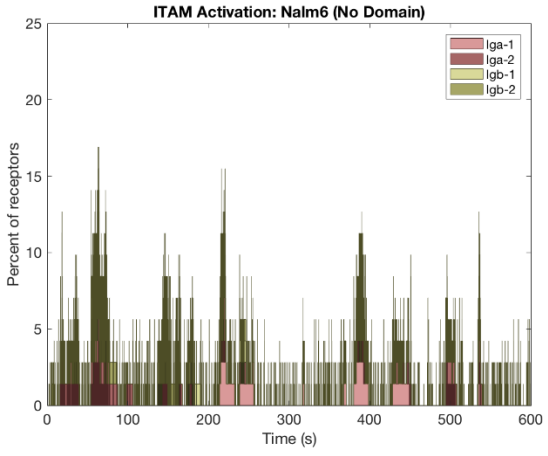




C

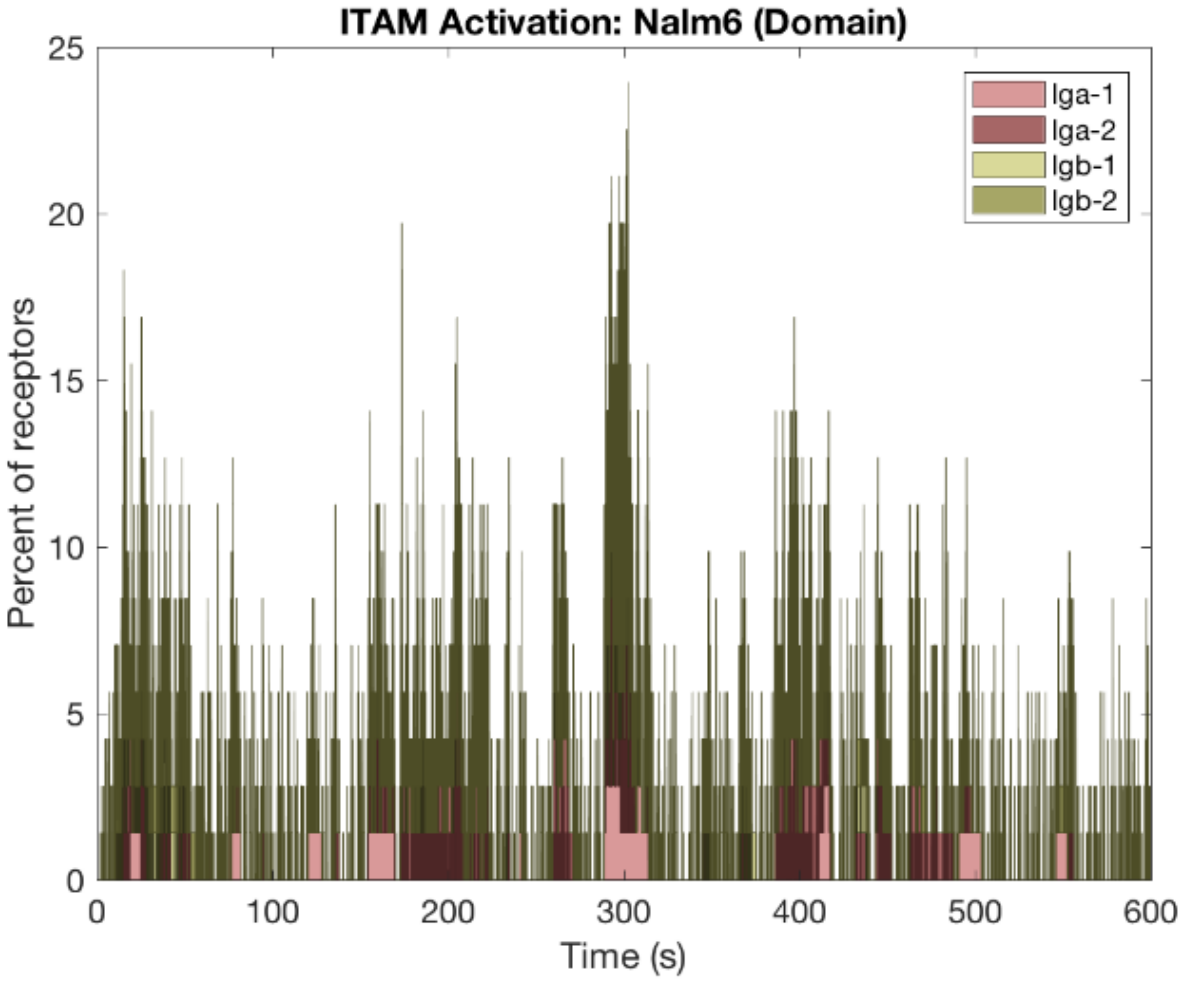


D



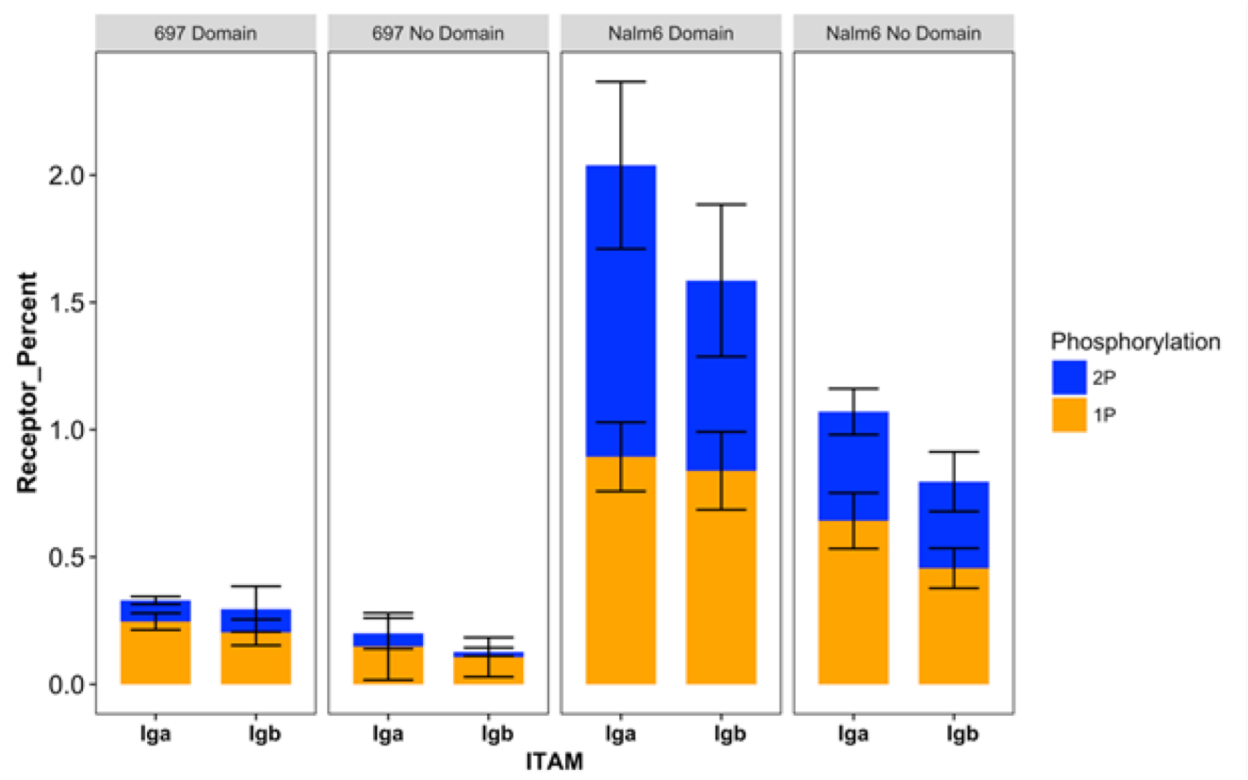
Kerketta et al. (in preparation)

Simulations with domains



Kerketta et al. (in preparation)

Simulations with domains



Kerketta et al. (in preparation)

3. Improvements and extensions

Close the validation loop

Identification of domains **and** intrinsic mobility changes

Better characterization of the landscape

Combine with identification of binding

Closing the validation loop

We analyze trajectories and infer confining domains

- Algorithm* depends on a lot of parameters
- In particular, the weights used in constructing the cumulative score

The reconstructed domains are used in a spatially resolved simulation

Additional details, such as dimer on-and off-rates, are estimated from experimental data

- Mapping from observed rates to ”true” rates requires simulations

Closing the validation loop

We analyze trajectories and infer confining domains which are used in a spatially resolved simulations

Next:

- Extract synthetic experimental data from simulations
- Run synthetic data through domain reconstruction / parameter estimation procedures
- Optimize the procedures by comparing the input and the output.

Closing the validation loop…

Simulations with random barriers, localization density