Integrative metamodeling of early T-cell antigen receptor (TCR) signaling accounts for nanoscale TCR activation patterns

Keywords: T cell, immunological synapse, Bayesian metamodeling, T cell activation, kinetic segregation, LCK, T cell receptor, CD45, peptide-MHC

# Abstract (<250 words)

T cells trigger highly specific and sensitive effector responses against foreign pathogens by recognizing cognate antigens presented by Antigen-presenting cells. Various microscopic processes that give rise to rapid and robust TCR activation have been proposed, but they each individually fail to capture all critical data related to TCR activation. Here, we apply Bayesian metamodeling for integrating multiple aspects of TCR triggering from fragmented input models of varying representations and scales. Specifically, we integrate models of the organization, activity, interactions, and dynamics of the signaling molecules TCR, CD45 and Lck, and of membrane shape and elasticity. Inputs include physical models, Monte-Carlo simulations, single molecule localization microscopy and published results. Our metamodel accounts for nanoscale dynamic patterns of TCR activation that could not be accounted for by the individual partial models. The scalability and flexibility of our metamodeling approach can serve to iteratively expand its predictive power for T cell activation and other biological systems.

# Significance Statement (<120 words)

T cell signaling is determined by a complex interplay of multiple microscopic processes. Metamodelling can integrate fragmented models of such processes and account for yet unexplained nanoscale activation patterns.

Introduction

T cells mount an immune response by recognizing foreign cognate antigens, presented on the surface of antigen-presenting cells (APCs). For that, the T cells form a dynamic interface with the APCs, known as the immune synapse (IS). Within this synapse signaling molecules dynamically organize to facilitate signal activation and regulation [PMID: 21179118].

Recently, multiple nanoscale patterns of these molecules have been observed, including clustering of molecules, and the segregation of receptors from glycoproteins (such as CD45). The abundant CD45 glycoproteins are phosphatases and can quench the TCR signal. Their physical separation from TCRs has been proposed and then shown to promote TCR activation (a model known as ‘kinetic segregation’ [DOI: 10.1038/ni1389]). These patterns have been shown to affect TCR signaling.

Multiple additional models have been proposed to account for the various aspects of T cell activation. 5. Such mechanisms include receptor clustering 6,7, conformational changes of receptor chains 8,9, dynamic formation of signaling complexes 10,11, cooperativity in triggering within nanoclusters 12,13, physical segregation of glycoprotein-phosphatases occurring in early tight contacts with APCs 14,15 where signalling molecules are enriched 16, effects of cell topography 17, etc.

Still, each of these models cannot capture the entirety of data related to TCR-dependent T cell activation [Ref]. The integration of such models have been proposed as a necessary step for achieving comprehensive and predictive understanding TCR activation [PMID: 21127503].

Arguably,integrating all available data sources on T-cell activation and its underlying mechanisms is impractical using existing data integration approaches. However, it is possible to integrate partial data subsets, and use these data subsets to inform partial models of narrower aspects of T-cell activation (components, subsystems, functions) [[3–16]](https://paperpile.com/c/G64qnd/wWBFZ+FVhg5+gEpPz+gFDLy+cnEhQ+XtLSQ+NsxBm+3HfWJ+zLDv2+y0mPD+NoxYt+OdYrN+7tFz5+SJJl7). Thus, we propose that Big Data integration can be divided-and-conquered if we shift the focus from direct data integration to model integration. For that, we turn to Bayesian metamodeling [[17]](https://paperpile.com/c/G64qnd/Y5bI5) (Fig. 2). Through this approach data can be collected from diverse data sources. Following data collection, the data is broken into **partial data subsets;** the partial data subsets are used as input for constructing **partial models**; each of these partial models is converted to a **probabilistic surrogate model** over some variables of the corresponding partial models (e.g. using a Bayesian network or a generative deep-learning model); the surrogate models are then coupled through imposition of **statistical coupling restraints** resulting in a a **Bayesian metamodel** - a single **joint probability distribution function (PDF)** over variables from all surrogate models. Finally, hyperparameters of partial input models are updated using their posterior estimates in the Bayesian metamodel. Critically, the output metamodel integrates all data used to inform any of the partial models.

To make informed decisions, T cells integrate multiple cues, which are sensed by specific receptors and induce signaling pathways. The detailed mechanisms by which T cells process and integrate multiple signals into accurate and reliable cellular responses remain incompletely understood. For example, in the context of pathogen recognition, T cells were shown to physically probe the surface of antigen-presenting cells (APCs) for cognate foreign antigens through their T-cell antigen receptors (TCRs). **However, the simple affinity of TCR-antigen interactions cannot explain the robustness of T-cell decision-making.** Detailed studies of TCR-dependent signaling using diverse experimental and theoretical methods have resulted in evidence for multiple possible mechanisms, which likely occur simultaneously to modulate the cell response. Such mechanisms pertain to diverse temporal and spatial scales, and depend on specific context and environment, such as **micro- to nano-scale clustering of receptors and signaling molecules,** **cooperativity,** mechano-sensing, and more. The integration of these partial mechanisms into a unified view of T-cell recognition is critically missing, leading to a lack of comprehensive and predictive understanding of such critical T-cell decision-making.

Here, we created a metamodel of T-cell recognition, based on the following models, using the following assumptions and free parameters. Currently, the first input model describes the major aspects of the kinetic segregation model, and combines them with information on the activity of LCK (TBD). If integrated correctly, they are expected to produce the observed pattern of phosphorylation and bull’s-eye spatial densities of TCRs and CD45s.

Our metamodel accounts for nanoscale dynamic patterns of TCR activation that could not be accounted for by the partial models. The scalability and flexibility of this model can serve to iteratively expand its predictive power for T cell activation and other biological systems.

Construction of the metamodel

Models are the central units of scientific theorizing and are necessary tools for an integrative understanding of heterogeneous data types. Our goal is to construct a unified model of T-cell decision-making, informed by pertinent data and prior knowledge. We create this model using Bayesian metamodeling, a novel divide-and-conquer approach for integrative whole-cell modeling. Using this formulation, we break the overall modeling problem into smaller and thus more-tractable modeling tasks. This includes: (A) collecting experimental and theoretical data using state-of-the-art imaging and complementary methods; (B) constructing multiple models describing aspects of T-cell decision-making; (C) integrating these partial models into a unified, quantitative, predictive, testable, and scalable metamodel. We then iterate through steps A-to-C to expand and improve both our experimental results, the partial models, and the unified metamodel (see Planned Activities for A-C and proof-of-concept below).

(A) Systematic and model-driven data collection across scales (Fig. 2; outer ring). We collect data on T cells and follow their decisions related to antigen recognition and differentiation. For that we will employ a broad set of imaging and other modalities, utilizing the diverse expertise of the center and committee members, as well as collaborators, as outlined in detail in Planned Activities.

(B) Constructing individual models of separate aspects of T-cell decision-making (Fig. 2; middle ring). Models, based on our data and literature, will be described using different datasets, representations, and pertaining to different parts and scales of the cell. Due to the scalable nature of our modeling approach, the list of integrated models is expected to expand iteratively over time as our scientific network expands and more experimental data becomes available.

(C) Assembling a metamodel of cellular decision-making (Fig. 2; inner ring). We will convert each partial model into a unified statistical representation, namely a probability density function (PDF). For this, we will use appropriate statistical modeling and machine learning approaches, including probabilistic graphical models and generative deep neural networks. We will then assemble the partial models in their unified statistical representations into a comprehensive metamodel of the cell, and harmonize them with each other using principles of Bayesian statistics. Importantly, the partial models can be constructed and computed independently; and can be integrated regardless of their scales by relying on statistical relations. This facilitates the sharing of data, resources, expertise and models by network members, and maximizes the modeling accuracy, precision, completeness, and scalability

The specific models

Our meta-model integrates multiple partial models. Each partial model captures specific mechanisms that contribute to signaling downstream the TCR. Specifically, we model the surface of T cells as they undergo an interaction with an antigen presenting cell (APC), forming an immune synapse (IS). Note that our focus is on the dynamic organization of molecules within the T cells. The partial models include the TCRs, glycoproteins (esp. CD45), key kinases (Lck), and the T cell membrane. We now describe each of the models and related results that serve as inputs to the metamodeling briefly, elaborate descriptions are provided in SI.

**Model 1** computes the spatiotemporal patterning of a population of TCR and CD45 molecules embedded in the plasma membrane of T cells during T-cell activation through the IS. The model aims to recapitulate imaging measurements of TCR and CD45 patterning when TCRs interact with peptide-MHC molecules on APCs or with molecular mimic, leading to TCR triggering [(Razvag et al. 2018; Neve-Oz et al. 2018; Chang et al. 2016)](https://paperpile.com/c/CRnYld/fjMT+H6vM+9Ity). The formation of these interfaces leads to reorganization of surface molecules, changes in the topography of the plasma membrane, and TCR activation. The tight adhesion of the two cells at the immune synapse (or the adhesion of the T cell to the mimic coverslip) causes physical segregation between the TCRs and CD45. This segregation occurs due to the smaller size of the molecular complex of the TCR-pMHC relative to the bigger size of the bulky CD45 (and other glycoproteins). Since CD45 is also a phosphatase, it has been suggested to dynamically quench the TCR signal through its dephosphorylation. Thus, the physical segregation of these molecules may promote TCR activation - a model known as ‘kinetic segregation’ [Ref Van-der Merwe+Davis].

Here, we specifically used Reaction-diffusion Markov-Chain Monte-Carlo (MCMC) simulations to capture the positions of these molecules at the tight interface. For this model, we used InterCells - our previously described computational model and simulation [(Neve-Oz et al. 2018)](https://paperpile.com/c/CRnYld/H6vM).

Briefly, our model captures the assumptions of the Kinetic Segregation model, but it does not account for TCR phosphorylation by Lck and downstream TCR spatiotemporal signaling.

**Model 2** describes the spatial patterning of a population of active Lck molecules, indicated as Lck\* [Refs Acuto, Gaus]. It computes Lck\* distribution given the density of CD45 molecules by using Monte-Carlo simulations. The free parameters for the model include the diffusion coefficient of Lck and the probability of spontaneous decay of Lck\* back to its inactive state. The input variables include mean (μ) and standard deviations (σ) of CD45 distribution. The output variables include probabilistic parameters that describe the radial distribution of steady state Lck\*.

Thus - this model captures Lck activation state at the immune synapse, but not the spatiotemporal organization of CD45 and the TCR.

**Model 3** computes the two-dimensional spatial patterning of phosphorylated TCR molecules (indicated as TCR\*) across the plasma membrane, as a function of the spatial distribution of TCR, CD45 and Lck\* molecules.

## Results

1. None of the three models captures pTCR patterning at initial contact
2. Prior information from KS and LCK models is sufficient to restore the observed pTCR pattern (roughy)
3. Metamodeling using surrogate for KS, LCK and pTCR\* models formalizes this observation, providing a quantitative probabilistic framework (including confidence intervals etc.) to couple the three models, contextualize them
4. Confronting metamodel with experimental data to infer the most probable model parameters for KS, LCK and pTCR\* models

## Discussion

T cell activation is a complex process that likely incorporates multiple microscopic processes. While the processes are inherently stochastic, there seems to be a high level of spatiotemporal organization and orchestration of the events that give rise to TCR-dependent signaling and cell activation. Here, we focused on the earliest events of TCR signaling. Specifically we attempted to integrate multiple partial models to account for previously unexplained patterns of TCR activation in early contacts. We employed a new approach to modeling, called metamodelling, that can seamlessly integrate models of various origin and spatiotemporal scales. We first showed that only the integration of multiple partial models can capture the previously observed phosphorylation pattern of TCR in the surroundings of initial tight contacts of the T cell at the IS. We then confronted the model with microscopy data and used the metamodel to infer the most probable parameters of the individual models. Such parameters include: ….

Fernandes et al have proposed that the KS model in early contacts should produce a uniform activation pattern of TCR The In model3 a CD45. Still, imaging by Razvag et al [PMID: 31825832] has shown the TCR activation is enriched at the periphery of early and tight T cells contacts with activating coverslips. Our metamodeling shows that the integration of multiple partial models was essential for capturing this patterning of TCR activation at the early contacts.

Spatiotemporal organization of molecules determines the rates of local interactions. Thus, such interactions are a common regulating mechanism for a wide range of cellular processes, including cell sensing, signaling pathways, metabolic networks, transcription, translation, and more. Spatiotemporal dependencies and stochasticisity (e.g. of molecular diffusion, interactions, etc) directly translate to system complexity. This complexity becomes daunting, esp. when one tries to gain fundamental understanding of the system working. Metamodeling has a unique and natural capability to describe complex systems in a tractable and scalable fashion. Thus, while our metamodelling approach was applied here to a concrete example of TCR activation, it could naturally grow to account for additional cellular systems.

Materials and Methods

The software, input files, and example output files for the present work are available at <https://github.cs.huji.ac.il/ravehb-lab/immune-synapse-metamodeling>. The metamodel was implemented using the PyMC3 package in Python{REF; ,https://docs.pymc.io/} (tested on version XX.XX; Python vXX.XX). For an outline of the approach, see Results. The specifications of the three input models to metamodeling and the technical details of their conversion to surrogate models and their coupling through metamodeling are described in detail in the SI Appendix: Supplementary Text 1.

## References

Fernandes, <https://doi.org/10.1073/pnas.1817255116>

## 6. Schamel, W. W. A. A. *et al.* Coexistence of multivalent and monovalent TCRs explains high sensitivity and wide range of response. *J. Exp. Med.* **202,** 493–503 (2005).

## 7. Lillemeier, B. F. *et al.* TCR and Lat are expressed on separate protein islands on T cell membranes and concatenate during activation. *Nat. Immunol.* **11,** 90–96 (2010).

## 8. Blanco, R., Alarcón, B. & Alarcon, B. TCR nanoclusters as the framework for transmission of conformational changes and cooperativity. *Front. Immunol.* **3,** 1–7 (2012).

## 9. Xu, C. *et al.* Regulation of T cell receptor activation by dynamic membrane binding of the CD3epsilon cytoplasmic tyrosine-based motif. *Cell* **135,** 702–713 (2008).

## 10. Sherman, E. *et al.* Hierarchical nanostructure and synergy of multimolecular signalling complexes. *Nat. Commun.* **7,** 12161 (2016).

## 11. Barr, V. A. *et al.* Development of nanoscale structure in LAT-based signaling complexes. *J. Cell Sci.* **Accepted f,** (2016).

## 12. Neve-Oz, Y., Razvag, Y., Sajman, J. & Sherman, E. Mechanisms of localized activation of the T cell antigen receptor inside clusters. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853,** 810–821 (2015).

## 13. Pageon, S. V *et al.* Functional role of T-cell receptor nanoclusters in signal initiation and antigen discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **113,** (2016).

## 14. Chang, V. T. *et al.* Initiation of T cell signaling by CD45 segregation at ‘close contacts’. *Nat. Immunol.* **17,** (2016).

## 15. Razvag, Y., Neve-Oz, Y., Sajman, J., Reches, M. & Sherman, E. Nanoscale kinetic segregation of TCR and CD45 in engaged microvilli facilitates early T cell activation. *Nat. Commun.* **9,** 732 (2018).

## 16. Jung, Y. *et al.* Three-dimensional localization of T-cell receptors in relation to microvilli using a combination of superresolution microscopies. *Proc. Natl. Acad. Sci. U. S. A.* **113,** (2016).

## 17. Razvag, Y., Neve-Oz, Y., Sherman, E. & Reches, M. Nanoscale Topography-Rigidity Correlation at the Surface of T Cells. *ACS Nano* **13,** 346–356 (2019).

## 

## Figures

**Fig. 1? - Intro**: biological/biophysical question - set up the challenge. Describe models (briefly) and data sources

**Fig. 2 - the approach** - combining three models, based on known literature and reasonable assumptions; learning of surrogate models

**Fig. 3 Results** (construction of model - everything before validation)

1. Building of a surrogate model (training)

**Fig. 4 - Validation**,

(A) model predictions, etc. - comparison to validation data

(B) a new integrated model of T cell molecular patterning

(C) Testable prospective prediction - what/if (try to connect to open questions in the field)

## Supporting Figures

Fig. S1A,B - Model 1 - A) Segregation of TCR and CD45 B) Conversion to surrogate

Fig. S2A,B - Model 2 - A) (CD45-dependent) Lck activity B) Conversion to surrogate

Fig. S3A,B - Model 3 - A) (Spatial patterns of) TCR activation B) Conversion to surrogate

# Supplementary Information

## Models

In this section, we describe each of the input models used for metamodeling, their conversion to probabilistic surrogate models, and their coupling. For each of the input models, we distinguish the free parameters, the input variables, and the output variables. All models were implemented in either Matlab or Python (tested on Matlab version R2020b, Python version 3.7.6 installd using Conda version 4.9.2). All surrogate models were implemented in Python (tested using the same version). The code for all models, surrogate models, and coupled surrogate models is provided in a designated Github repository (<https://github.cs.huji.ac.il/ravehb-lab/immune-synapse-metamodeling>), in subfolders InputModels, SurrogateModels, and CoupledModels, respectively.

### Model 1

**Input and output.** Model 1 computes the spatiotemporal patterning of a population of TCR and CD45 molecules embedded in the plasma membrane of T-cell during T-cell activation. The input to the model includes the model parameters and the initial configuration of the T-cell plasma membrane, the TCR and CD45 molecules on the T cell membrane, and the peptide-MHC (pMHC) complexes on the APC membrane (Table 1). The output of the model is the spatiotemporal trajectory of the model’s configuration, as follows.

**Model configuration.** A model configuration is defined as follows (Fig. S1). A section of the T-cell membrane at the IS is represented by a grid of nxn squares of axa nm. Each TCR or CD45 molecule occupies a single grid square, without overlap. In addition, each TCR and CD45 molecule has a certain vertical length hTCR and hCD45. A section of the APC membrane facing the T-cell membrane at the IS is represented by a two-dimensional grid of identical dimensions. Each pMHC complex occupies a single grid point of the APC membrane. Finally, the T-cell and APC membranes can be both curved; the distance of the T-cell membrane from the APC at each grid square is specified by . In all simulations, the TCR and CD45 molecules were initially located in circular clusters (fig. S1. C), based on {DOI: 10.1038/s41467-018-03127-w}, although in principle, this is not a requirement of the model.

**Model interactions:** We assume local equilibrium, and thus, the interactions in the model are described using a Hamiltonian over a model configuration :

is the total bending energy of the T-cell and APC membranes due to changes in , and are the spring energies associated with compressing the vertical length of the CD45 and TCR molecules, respectively, and is the binding energy of a molecule to a molecule on the opposite membrane.

.

is the area of one square, is the membrane bending rigidity.

and

is the spring constant of the molecule.

.

.

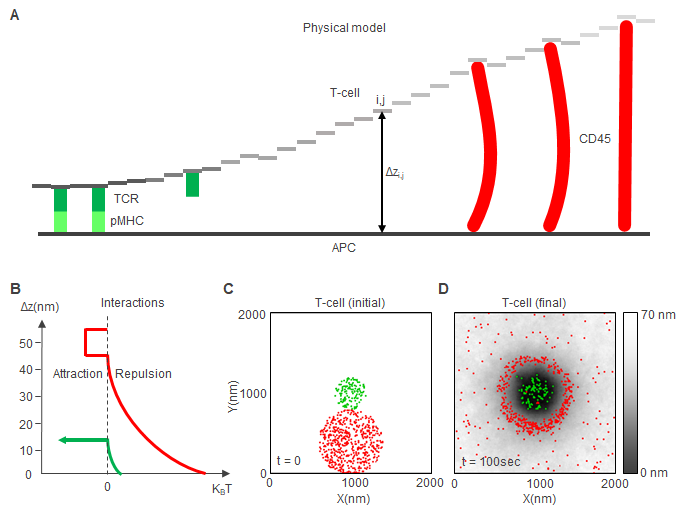
Planar molecules interactions:The planar interactions of the molecules is implemented in the property that two molecules on the same membrane can not be at the same square at the same time (interacting like hard spheres). TCR molecules can have a self-clustering factor, Pon, that ...

Vertical molecules interactions: Receptor-ligand interactions occur when a TCR and a peptide-MHC have the same planar location. In this case they are forced to be bound and at that location the inter-membranes distance, Δz, will be 13 nm. Molecule-membrane interactions: CD45 molecules interact with the APC membrane as repulsive springs when Δz < CD45 resting length.

**Model dynamics.** We evaluate the model dynamics using Reaction-diffusion Markov-Chain Monte-Carlo {DOI: [10.1039/B902017A](https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1039%2FB902017A), DOI:[10.1093/BIOMET/57.1.97](https://doi.org/10.1093/BIOMET%2F57.1.97)}.A typical simulation runs for 10,000 iterations. To simulate Brownian diffusion of each molecule along the membrane, we sample a random molecular step in a random direction (uniformly sampled between and radians), and with a magnitude that is the absolute value of a normally distributed scalar with mean 0.0 and standard-deviation ; *D* is the diffusion coefficient in units of . Therefore we can treat the propagation of the simulation as propagation in time with time step = . We used periodic boundary conditions (molecules that exit at one side enter on the other side). At every iteration, all the TCR and CD45 molecules are moved to a new location with random step ; the move is accepted/rejected using the Metropolis criterion: it is accepted with . if the energy at the new state is lower than the energy at the old state . If the energy at the new state is higher, the attempt is accepted with . In addition, the inter-membranes distance at every grid square is changed by a normally-distributed random step with mean = 0 nm and standard deviation = 1 nm.

Dynamics rules for one iteration: Molecules dynamics: (we are using periodic boundary conditions - if a molecule exists the simulation array it enters on the opposite side): Attempting to move all TCR and CD45 molecules simultaneously to new locations. If a molecule attempts to move into a new square that was occupied in the previous iteration it is rejected. If more than one molecule attempts to move into the same new square all these attempts are rejected. The molecules that were accepted by criterions 1.2 and 1.3 change their height, h, to the new at their new locations and are accepted or rejected according to the Metropolis criterion for molecules.

Membrane dynamics: All square attempt to move up or down to a new . The of square i,j that is occupied by a TCR and a pMHC is forced to have. The rest of the new are accepted or rejected according to the Metropolis criterion for membranes.



**Fig. S1. A. side view of the simulation setup.** B. Interaction potentials: The inter-membranes distance, where a TCR and a pMHC are in the same square, will be fixed at Δz = 13nm. CD45 acts as spring with a spring constant of 0.1 kT/nm2 and resting length of 50nm. C. initial conditions of the simulations. NTCR ~ 125, NCD45 ~ 500, The inter-membranes distance, Δz = 70nm (white color) except at the locations of the molecules, where for the CD45 locations Δz = 50nm (resting length of CD45 molecule) and for the TCR locations Δz = 13nm (length of bound αCD3-TCR). pMHC molecules (not shown here) are uniformly scattered over the APC membrane with surface density of 300/μm2. D. molecules distribution and topography after 10,000 iterations (100 sec). The colorbar represents inter-membranes distance.

Table 1: Free parameters of Model 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Description** | **Default value** | **Units** | **References** |
| Array size | size width x size length. | 2000x2000 |  | - |
| a | Area unit that divides the array to square units that has approximately the lateral size of a molecule. | 10 |  | WL,  InterCells |
| Δt | Iteration length. | 0.01 |  | InterCells |
| Niter | Number of iterations | 10,000 | - | InterCells |
| h0,TCR | Vertical resting length of the TCR. | 13 |  | WL |
| h0,CD45 | Vertical resting length of the CD45. | 50 |  | WL |
| NTCR | Total number of TCR molecules. | ~125 |  | - |
| NCD45 | Total number of CD45 molecules. | ~500 |  | - |
| DTCR | Diffusion coefficient of TCR molecules. | 10,000 |  | WL |
| DCD45 | Diffusion coefficient of CD45 molecules. | 11,000 |  | WL |
| κ (kappa) | Membrane rigidity. | 25 |  | WL |
| k | Spring constant | 10 κ/a2 |  | WL |
| uCD45 | Binding energy of CD45 to APC membrane | -10 |  |  |
| uTCR | Binding energy of TCR to APC membrane | -10 |  |  |
| uTCR-pMHC | Binding energy of TCR to pMHC | Effectively -inf |  |  |
| Pon,TCR | Probability of self clustering of TCR when in contact with another TCR | 0.995 |  | InterCells |
| Input parameters (initial configuration): | | | | |
| TCR initial distribution | Circle with uniformly distributed molecules. | Center = (1000,1000)  R = 200 |  | - |
| CD45 initial distribution | Circle with uniformly distributed molecules. | Center = (1000,400)  R = 400 |  | - |
| Δz0 | Initial membrane height. It is constant everywhere except at the locations of the molecules. | 70 |  | InterCells |
| Δz0,TCR | Initial membrane height at the locations of TCR molecules. | h0,TCR |  | InterCells |
| Δz0,CD45 | Initial membrane height at the locations of CD45 molecules. | h0,CD45 |  | InterCells |

#### 

#### Surrogate model:

To learn a surrogate model for Model 1, we first mapped its outputs for different input parameter values (Fig. S2A-C); we created a probabilistic graphical model {REF} that describes statistical relations among its variables in parameterized form (Fig. S3; Table S1A,B); next, we fitted the probabilistic graphical model to recapitulate the statistical relations between the model parameters and its values. We now explain each of these steps in detail.

**Parameter phasespace mapping:** We evaluated the model using 20 different values of membrane rigidity, κ, from κ=5 kT/nm2 to κ=100 kT/nm2 at 5 kT/nm2 intervals, using an identical initial configuration each time (Fig. S1F). As reported earlier [(Neve-Oz et al. 2018)](https://paperpile.com/c/CRnYld/H6vM), and as expected from the Kinetic Segregation model {DOI: 10.1038/s41467-018-03127-w}, on which Model 1 is based, the CD45 and TCR molecules formed concentric outer and inner rings, respectively, with a low-density depletion zone separating between the two rings (Fig. S1G); this result holds regardless of the precise choice of model parameters or the initial configuration of the TCR and CD45 molecules. Therefore, we analyzed the following three properties at 10 different simulation time points for different input parameters: the width of the ring of TCR molecules (Fig. S2A), the width of the ring of CD45 molecules (S2B), and the width of a low-density depletion zone that typically emerges between the TCR and the CD45 rings (Fig. S2C). The width was computed using pair-correlation analysis, as explained below {REF Intercells 2018, Razvag; else?}.

**Surrogate probabilistic graphical model.** We created a probabilistic graphical model, specifically a Bayesian network {Friedman, Koller} describing statistical relations among the membrane rigidity parameter κ, the time *t*, and the widths of the TCR and CD45 rings and the depletion zone between them.

Based on the output heatmaps we create a surrogate model for each. In the surrogate model we made an approximation of an output heatmap by using a two dimensional surface z(t,κ) that is a function of random variables (RV’s) (Table S2A). To run the Bayesian network we used the PyMC3 [(Salvatier J., Wiecki T.V., Fonnesbeck...)](https://paperpile.com/c/CRnYld/1vzI) which is a probabilistic programming package for Python that allows users to fit Bayesian models using a variety of numerical methods, most notably Markov chain Monte Carlo (MCMC) and variational inference”. We followed the following process:

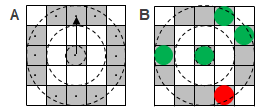
1. Form an equation that can describe the characteristics of the output heatmap.
2. Choose random variables that can describe the heatmap correctly with wide enough margins. We use the observed t and κ as input.
3. Run the PyMC3 package to make a ‘sanity check’ of the results and see if the resulting random variable can fit the data.
4. If the result passes the ‘sanity check’ we can refer to it as a surrogate model. We train the surrogate model by running the model with the learned random variable over a batch of t and κ values that we choose. The resulting trained model is independent of the data.

We used PyMC3 v3.9.3 with a ‘NUTS’ sampler and 4 Markov chains of 2000 steps.

To create the trained model we ran the model over a batch of t and κ. t was from 10 to 100 sec with 10 sec intervals. κ was from 5 kT/nm2 to 100 kT/nm2 with 5 kT/nm2 intervals.

For details see table S2A.

**Pair correlation analysis:** We used univariate pair-correlation analysis to compute the width of the TCR and CD45 clusters, and a bivariate pair-correlation analysis to compute the width of the depletion zone. The correlation function was described earlier [(Contreras and Valenzuela 1986)](https://paperpile.com/c/CRnYld/SXqn); briefly, where is the overall density. ( is total number of points and is total area). . (is the number of points on a ring with radius centered at point . is the area of a ring with radius and width . The algorithm we used uses the squares of the grid; a ‘ring’ is consisted of squares that their centers are within and where (square size). In this case is the number of squares that makes the ring (Fig. S##).



We calculated the univariate pair correlation functions, g11(r), g22(r) and the bivariate pair correlation function g12(r).We characterized each g(r) curve by a single value. For g11(r) and g22(r) we used the width at Δg/2 where Δg/2 is the difference between minimum and maximum of g(r) (fig. S1 H). For g12(r) we used the same method but with the distance to the nearest curve intersection as the depletion distance (fig S1 I). With these three single value data sets we created three maps (fig. S2 A-C).

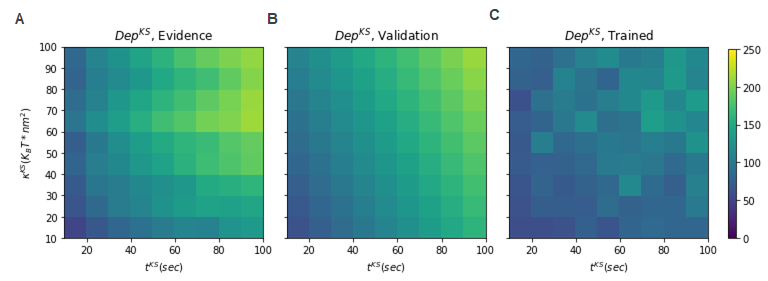
Table S2A: random variables for surrogate model 1, before training.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Name | Description | Distribution | Distribution parameters | Unit |
| Random  variables | t | time |  |  | *sec* |
| κ | Membrane rigidity |  |  | *KT\*nm2* |
|  |  | Depletion range between TCR and CD45 |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |

able 2B: random variable for surrogate model 2, after training.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Name | Description | Distribution | Distribution parameters | Unit |
| Random  variables | t | time |  |  | *sec* |
| κ | Membrane rigidity |  |  | *KT\*nm2* |
|  |  | Depletion range between TCR and CD45 |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |

Fig. S2 - surrogate model (heatmap before after, maybe PGM topology)



**Fig. S2. Getting a single value that characterizes the depletion range between TCR and CD45.** S2A) Heatmap of the evidence for the depletion distance between TCR and CD45. S2B) Heatmap used for validation of the and equations and random variables that we chose. S2C) Heatmap of the fitted model after running the PyMC3 package with the validated equations and random variables.

### 

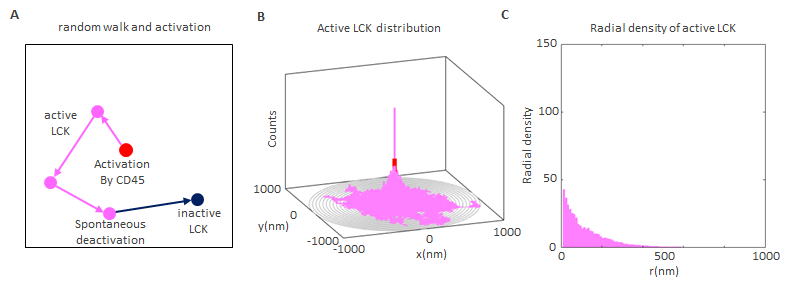
### Model 2

Input model

**Model configuration.** The membrane is represented as a two-dimensional grid of 200 x 200 squares of 10 nm x 10 nm each. The free parameters and input variables of the model are described in Table S3. At the center of the array there is a CD45 molecule (activating location). All Lck molecules start their path at the center of the array as an active Lck. As the Lck\* molecules diffuse away from the center with a diffusion coefficient DLck it is spontaneously deactivated with probability Poff at every iteration.

**Model simulation.** A typical simulation runs for 1000 iterations. To simulate Brownian diffusion of each molecule along the membrane, we sample a molecular step in a random direction (uniformly sampled between and radians), and with a magnitude that is the absolute value of a normally distributed scalar with mean 0.0 and standard-deviation ; *D* is the diffusion coefficient in units of . Therefore we can treat the propagation of the simulation as propagation in time with time step = . In all simulations, we used. We used periodic boundary conditions (molecules that exit at one side enter on the other side). At every iteration all the molecules ‘jump’ to a new location with no limiting conditions.

### **Fig. S3** –Model 2 – Lck activation (LA) – Physical model + MC simulations



**Fig. S3. Lck activation as a result of an interaction with CD45**. S3A. The red point marks the location of a CD45 molecule. All Lck’s paths begin at an active state at the CD45 location. As they propagate in a two dimensional random walk they are being spontaneously deactivated with a probability Poff at every step. S3B. The magenta paths mark the locations of active Lck and the heights of the bars mark how many times active Lck molecules passed through this location (bin) during the simulation. S3C. Radial distribution of the locations of active Lck. The radial distribution is calculated as the sum over angles of the active Lck around the location of CD45 (gray rings in B). The radial distribution behaves as an exponential decay function and depends on the DLck and PoffLck\*.

Table S3: Free parameters of Model 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Description** | **Default value** | **Units** | **References** |
| Array size | size width x size length. | 2000x2000 |  | - |
| a | Area unit that divides the array to square units that has approximately the lateral size of a molecule. | 10 |  | WL,  InterCells |
| Δt | Iteration length. | 0.01 |  | InterCells |
| Niter | Number of iterations | 1000 |  | InterCells |
| NLck | Total number of TCR molecules. | 1000 |  | - |
| DLck | Diffusion coefficient of TCR molecules. | 10,000 |  | WL |
| PoffLck | Probability of Lck deativation per one iteration | 0.01 |  | - |

#### 

#### Surrogate model:

To learn a surrogate model for Model 2, we first mapped its outputs for different input parameter values (Fig. S3). We created a probabilistic graphical model {REF} that describes statistical relations among its variables in parameterized form (Fig. XXX; Table S4); and then, we fitted the probabilistic graphical model to recapitulate the statistical relations between the model parameters and its values. We now explain each of these steps in detail.

**Parameter phase space mapping:** Specifically, we evaluated the model using 13 different values of diffusion coefficient, D and 11 values of Poff. , where is from -3 to 0 with intervals of 0.25. , where is from -5 to 0 with intervals of 0.5. For details see tables S4A,B.

Table S4A: random variables for surrogate model 2, before training. TBD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model 2 – Lck activation (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  | log10 of diffusion coefficient of Lck\* |  |  |  |
|  | log10 of deactivation probability of LCK\* |  |  |  |
|  | Sigmoid minimum |  |  |  |
|  | Sigmoid maximum |  |  |  |
|  | Poff sigmoid center |  |  | - |
|  | Poff sigmoid devisor |  |  | - |
|  | D sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Noise\Uncertainty |  |  |  |
|  | Distribution width of Lck\* |  | μ= , |  |

Table S4B: random variables for surrogate model 2, after training. TBD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model 2 – Lck activation (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  | log10 of diffusion coefficient of Lck\* |  |  |  |
|  | log10 of deactivation probability of LCK\* |  |  |  |
|  | Sigmoid minimum |  |  |  |
|  | Sigmoid maximum |  |  |  |
|  | Poff sigmoid center |  |  | - |
|  | Poff sigmoid devisor |  |  | - |
|  | D sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Noise\Uncertainty |  |  |  |
|  | Distribution width of Lck\* |  | μ= , |  |

### 

### Fig. S4.

#### Surrogate model:

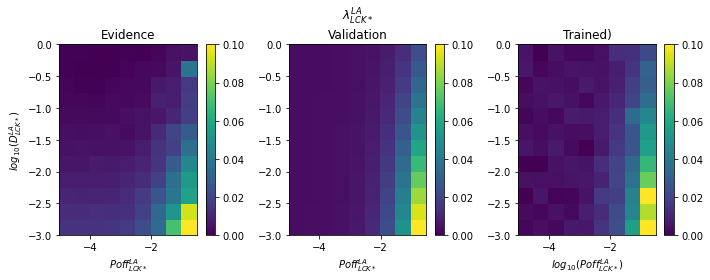


Fig. S4. A) untrained model: a heatmap that shows the decay coefficient of Lck\* as a function of the diffusion coefficient, , and the deactivation probability per iteration, . B) Validation of the chosen equations and random variables for the trained model. C) Trained model calculated with the trained parameters.

### Model 3

**Input and output.** The model computes the spatial distribution of active (phosphorylated) TCR molecules (TCR\*) on the T cell membrane from the spatial distributions of all TCR, CD45 molecules , and LCK\* probability distribution that collectively regulates TCR activation.

**Model configuration.** The model configuration is defined by the spatial positions of the TCR and CD45 molecules on the membrane, represented using a 400 x 400 grid of 10 nm x 10 nm squares (Fig. S4A). Every grid square contains at most one molecule of each type. In this model the molecules are fixed. The probability distribution of Lck\* around a single CD45 molecule is an exponential decay distribution with radial symmetry (Fig. S4B). The overall Lck\* probability distribution around all the CD45 molecules is the sum of all the individual Lck\* probability distributions (Fig. S4C).

**Model interactions.** The interactions in this model include the deactivation of TCR molecules by CD45 molecules, and the activation of TCR molecules by LCK\* molecules. Both interactions depend on the intermolecular distances. The interaction parameters are described in Table XX.

**Model evaluation.** The probability of a TCR being phosphorylated is proportional the value of the Lck\* probability distribution at the TCR location (Fig. S4D-E).

The TCR\* probability distribution is computed

from the other three distributions: a TCR is phosphorylated if it is in the phosphorylation range (10nm) from a Lck\*. a TCR is dephosphorylated if it is in the dephosphorylation range (10nm) from a CD45. In this model, TCR\* value can be 1 or 0.

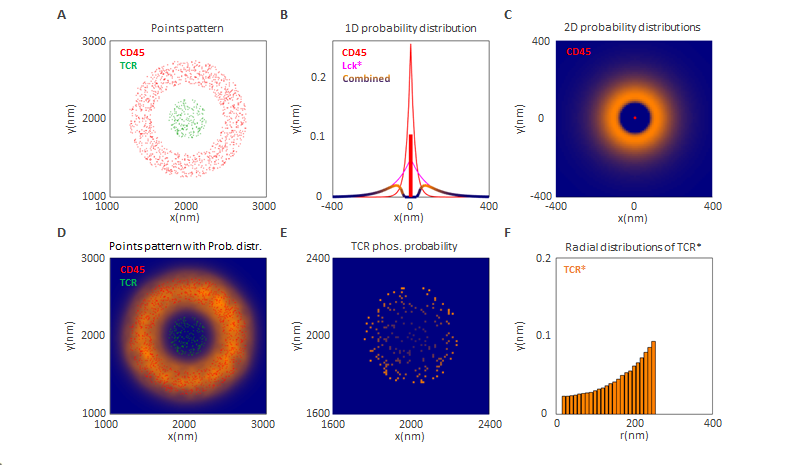


Figure S4A. Point patterns of TCR (green), and CD45 (red). We assume that the TCR molecules are arranged in a circle with radius r at the center of the array, the CD45 molecules are arranged in a concentric ring around the center. S4B. Probability distribution of active Lck (Lck\*) (magenta shade) after being activated by CD45 (red point). The distribution has radial symmetry and a radial exponential decay. S4C. Sum of the Lck\* probability distributions over all the CD45 molecules. S4D. Enlarged view of probability distribution in C relative to the TCR locations. S4E. multiplication of probability distribution in D by the locations of TCR (value at TCR locations = 1, otherwise = 0). S4F. Radial distribution histograms centered at (2000,2000) nm. Green - radial distribution of TCR locations. Parula gradient - TCR phosphorylation probability (not in scale). Orange - ratio of phosphorylation probability to TCR distribution (proportional to Lck\* probability distribution) (not in scale..

Table xx: Free parameters of Model 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Description | Default value | Units | References |
| Array size | Two dimensional grid of squares | 4000x4000 | nm | - |
| a | Square size | 10 | nm |  |
| TCR distribution | A ring with inner radius, r1 and outer radius r2 with uniform distribution. | r1 = 0  r2 = 250 | nm | - |
| CD45 distribution | A ring with inner radius, r1 and outer radius r2 with uniform distribution. | r1 = 450  r2 = 750 | nm | - |
| Single Lck\* distribution | Two dimensional exponential-decay probability distribution with radial symmetry around the location of a CD45. | λ = 0.01 | nm-1 | - |
| Total Lck\* distribution |  |  |  |  |
| Interaction |  |  |  |  |
| Interaction |  |  |  |  |

### 

#### 

#### 

#### Surrogate model:

Description TBD - Random text: The configuration space is defined by four vectors, one for each molecular species. Each vector describes the mean molecular density per μm2 as a function of the distance from a shared center, using 200 x 0.01 μm bins, running from 0 μm to 2 μm (fig. S4 B). batch parameters

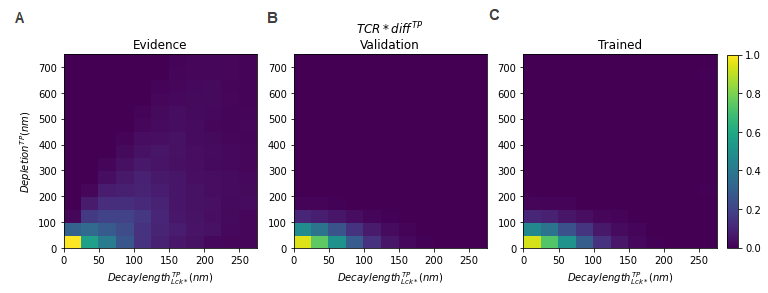
Table S6A: random variables for surrogate model 1, before training.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model 3 – pTCR (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  | log10 of diffusion coefficient of Lck\* |  |  |  |
|  | log10 of deactivation probability of LCK\* |  |  |  |
|  | Sigmoid minimum |  |  | nm |
|  | Sigmoid maximum |  |  | nm |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Uncertainty in |  |  | nm |
|  | Distribution width of Lck\* |  | μ= , | nm |

Table S6B: random variable for surrogate model 3, after training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model32 – Lck activation (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  | log10 of diffusion coefficient of Lck\* |  |  |  |
|  | log10 of deactivation probability of LCK\* |  |  |  |
|  | Sigmoid minimum |  |  | nm |
|  | Sigmoid maximum |  |  | nm |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Uncertainty in |  |  | nm |
|  | Distribution width of Lck\* |  | μ= , | nm |

Figure S6: surrogate model 3



TBD

## Coupling

Description TBD

Table S7A (coupling variables depending on model variables)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model32 – Lck activation (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Table S7B (model variables now depending on coupling variables)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model32 – Lck activation (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Figure S7: topology of coupled PGM? (maybe only in main text)

Backpropagation

# Miscellaneous

* Immunological synapse and T cell activation are important, resulting from complex interactions among molecular populations at multiple spatial and temporal scales.
* Molecular patterning occurs during T cell activation, but how and why it contributes to specific and robust recognition of foreign antigens is not entirely understood.
* Various models tried to explain these patterns, many of which correctly predict certain aspects of molecular patterning at the immune synapse, but none can explain all observations - concrete example: recently observed phospho-TCR patterning
* We wanted to see if a combination of several existing models could provide an explanation.
* We did this using our recent Bayesian metamodeling framework, which enables modular integration of input models of varying representations and scales.
* Specifically, we created a metamodel that harmonizes three models, each based on information available from the literature (on its own):
  + (1) The kinetic segregation model, which partially explains spatial organization of T cell receptors and CD45 molecules;
  + (2) A model of spatial distribution of activated LCK due to dephosphorylation by CD 45 (TODO: be more precise here, the 394 and 505 tyrosines, various theories about activation and deactivation)
  + (3) A model of simultaneous activation and inhibition of TCRs due to phosphorylation and dephosphorylation by LCK and CD45 molecules, respectively, given certain spatial distributions of TCR and active LCK molecules
* The metamodel shows that the combination of the three model, each making simple and easily justifiable assumptions, explains the observed phosphorylation pattern of TCRs.
* OPTIMISTIC ENDING - WHY ITS IMPORTANT AND WHAT DO WE LEARN FROM THIS, AND THIS WILL LEAD IN FUTURE TO…

Dynamics rules for one iteration: Molecules dynamics:

1. Updating molecules location (we are using periodic boundary conditions - if a molecule exists the simulation array it enters on the opposite side):
   1. Attempting to move all TCR and CD45 molecules simultaneously to new locations.
   2. If a molecule attempts to move into a new square that was occupied in the previous iteration it is rejected.
   3. If more than one molecule attempts to move into the same new square all these attempts are rejected.
   4. The molecules that were accepted by criterions 1.2 and 1.3 change their height, h, to the new at their new locations and are accepted or rejected according to the Metropolis criterion for molecules.
2. Update inter-membranes distance:
   1. All square attempt to move up or down to a new .
   2. The of square i,j that is occupied by a TCR and a pMHC is forced to have.
   3. The rest of the new are accepted or rejected according to the Metropolis criterion for membranes.

## 

## Random text

The input models

**Model 1:** Guests at a wedding. the guests are: TCR, CD45, (pMHC), APC

The constraints are: Membrane curvature, pMHC distribution, steric clashes between CD45 and membrane. LCK in this model is probably distributed uniformly (regardless of phosphorylated/dephosphorylated state)

**Free parameters:** Membrane rigidity

**Input:** initial TCR distribution, initial CD45 distribution, initial inter-membrane distance

**Output:** final TCR distribution, final CD45 distribution, final inter-membrane distance

Fig. S1.2.

**Model 2:** Phosphorylation patterns of “guests” as a function of their steady-state spatial distributions. Free parameters include the range and magnitude of the phosphorylation, given spatial proximity to other species (“guest types”).

**Free parameters:** short-range dephosphorylation range of TCRs by CD45, short-range phosphorylation range of TCRs by LCK\*.

**Input:** spatial distribution of TCRs (from model1), spatial distribution of CD45 (from model1), spatial distribution of LCK\* (from model3).

**Output:** phosphorylation patterns of TCRs

**Model 3:**

Reaction diffusion of LCK and its dephosphorylation by CD45. LCK is represented explicitly, in either phosphorylated (inactive) or dephosphorylated (active, LCK\*) state. It diffuses (diffusion reaction or Brownian dynamics simulations), forming a gradient of LCK\*. Phosphorylated spontaneously with a certain rate constant (free parameter) - in future, phosphorylation might be modeled explicitly. Dephosphorylated by short-range CD45 interaction

**Free parameters:** Diffusion coefficient of LCK, dephosphorylation parameters for CD45 → LCK, Spontaneous phosphorylation LCK\* → LCK.

**Input:** spatial distribution of CD45, concentration of LCK.

**Output:** spatial distribution of LCK\* (dephosphorylated).

**Participants, times, dimensions and boundaries:**

Simulation dimensions: x range = 2000nm, z range = 100nm, Time range is between 1sec and 100sec, Time step, Δt = 0.01sec.

**Membranes**: Thickness = 4nm (does not play a role here). Length = x range, rigidity, κ = 25KBT. The membrane is divided to patches with size, a = 10nm.

**TCR**: A plasma membrane molecule, have parts outside the membrane and inside the membrane. Size x = 10nm, Size z = 13nm (includes MHC that is not in the model right now). Diffusion coefficient = 0.011μm2/sec. [ref].

**pTCR**: A property of **TCR**, the level of phosphorylation of the **TCR**. Can have values from 0 to 10. Right now it can be 0 or 1. The **pTCR** is phosphorylated by **LCK\*** (active LCK) and dephsophorylated by **CD45**.

**CD45**: A plasma membrane molecule, have parts outside the membrane and inside the membrane. Size x = 10nm, Size z = 50nm. Diffusion coefficient = 0.01μm2/sec. [ref].

**CD45** dephophorylates (activates) **LCK** to **LCK\*** and dephsphorylates **pTCR**.

**LCK**: Moves in 2D on the inner side of the membrane. Size x = 5nm (). Size z = 5nm (). Diffusion coefficient = 0.3μm2/sec. [ref]

**LCK\***: A property of **LCK**, it canan have values of 0 or 1. The **LCK** is dephosphorylated (activated) by **CD45** when they are in contact range. **LCK\*** phosphorylates the **TCR** when they are in contact range.

Fig. 1.1a

Fig. 1.1b

Fig. 1.1. Summary of the physical setup. In-scale x, z and sizes of molecules, z range ~ 100nm, x range ~ 700nm. Fig. 1.1a. T-cell and APC before contact, Fig 1.1b. T-cell and APC it contact.

### 

### **Model 1**

Model1 contains the membranes of the T-cell and the APC, TCR molecules and CD45 molecules. The 'physical' simulation range is -1000nm < x < 1000nm and 0nm < z < 100nm. Each membrane has a rigidity K (resistance to bending). The T-cell membrane is on top, the APC membrane is on the bottom. The two membranes do not interact directly but via their surface molecules, in this case TCR and CD45. The TCRs are short (~13nm) and bind to MHC molecules on the APC (at this stage no MHC are present so we assume the TCRs bind to the APC membrane). The CD45 molecules are long (~50nm) and act as repulsive springs pushing the opposite APC membrane. As a result of the length difference, the binding and pushing and the membranes rigidity, the short and long molecules are separated and organize in a state of lower energy. To simplify the model, the inter-membrane distance ,Δz, is forced to be equal to the length of the molecule at that location. The only free parameter is the membrane rigidity.

**Model 1:**

**Free parameters:** Membrane rigidity

**Input:** initial TCR distribution, initial CD45 distribution, initial inter-membrane distance

**Output:** final TCR distribution, final CD45 distribution, final inter-membrane distance

**Initial conditions**:

Bottom membrane is assumed to flat at all times. Top membrane is basically flat at z=70nm except at the locations of molecules where the membrane is forced to be at z=molecule\_z. Both types of molecules, TCR and CD45, belong to the T-cell and therefore always attached to its membrane. On the hand, to simplify the dynamics, we assume that the end of the molecules are always in contact with the APC (z=0). The top membrane edges (+-1000nm) can be held fixed at 70nm.

To demonstrate the dynamics we distribute the TCR molecules around the center and the CD45 molecules are distributed on both sides of the TCRs.

**Propagation**: the simulation propagates at iterations of 0.01sec. At every iteration two things occur: the membrane height is updated and the molecules locations are updated. Every patch of the membrane attempts to change its height according to a normal distribution with μ = 0 and σ = dz = 1nm. The attempt is accepted by the membrane bending energy and the Metropolis criterion. Membrane bending energy:

Metropolis criterion:

The same idea is applied to the movement of molecules: at every step all the molecules attempt to jump to a new location, μ = 0 and σ = σmol.

σTCR = sqrt(2\*DiffTCR\*Δt) = sqrt(2\*0.011μm2/sec\*0.01sec) = 0.0148μm = 14.8nm.

σCD45 = sqrt(2\*DiffCD45\*Δt) = sqrt(2\*0.01μm2/sec\*0.01sec) = 0.0141μm = 14.1nm.

Two molecules can not occupy the same patch at the same time.

Boundary conditions: if a molecule attempts to jump out of bounds it comes back in on the other side.

Fig. 1.2a. Initial conditions of simulation.

Fig 1.2b. final conditions of the simulation.

Table S2A: random variables for surrogate model 1, before training.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Name | Description | Distribution | Distribution parameters | Unit |
| Random  variables | t | time |  |  | *sec* |
| κ | Membrane rigidity |  |  | *KT\*nm2* |
|  |  | Distribution Width of TCR |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |
|  |  | Distribution Width of CD45 |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |
|  |  | Depletion range between TCR and CD45 |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |

Table 2B: random variable for surrogate model 2, after training.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Name | Description | Distribution | Distribution parameters | Unit |
| Random  variables | t | time |  |  | *sec* |
| κ | Membrane rigidity |  |  | *KT\*nm2* |
|  |  | Distribution Width of TCR |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |
|  |  | Distribution Width of CD45 |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |
|  |  | Depletion range between TCR and CD45 |  | , | *nm* |
|  | Surface intercept |  |  | *nm* |
|  | Slope for t |  |  | *nm/sec* |
|  | Slope for k |  |  | *1/(nm\*KT)* |
|  | Uncertainty in |  |  | *nm* |

### 

### **Model 2**

In model2 the inputs are TCR distribution (M1), CD45 distribution (M1) and LCK\* distribution (M3), the output is pTCR distribution. The free parameters are the TCR-LCK\* phosphorylation range and the TCR-CD45 dephosphorylation range.

**Free parameters:** TCR-LCK\* phosphorylation range, TCR-CD45 dephosphorylation range

**Input:** TCR distribution (M1), CD45 distribution (M1), LCK\* distribution (M3)

**Output:** pTCR distribution.

A TCR molecule is located at the center (x=0) of the simulation range. One parameter is σTCR-LCK\*phos range, the other parameter is σTCR-CD45 dephosphorylation range. Both parameters have 6 values from 0 to 50nm.

### 

### **Model 3**

In model3 the inputs are CD45 distribution (M1) and LCK distribution, the output is LCK\* distribition. The free parameters are LCK diffusion and LCK Poff (the probability of deactivation).

**Free parameters:** LCK diffusion, LCK Poff.

**Input:** CD45 distribution (M1), LCK distribution.

**Output:** LCK\* distribution.

In model3 a CD45 molecule is located at the center (x=0) of the simulation range . An LCK molecule starts a 1D random walk of 101 steps from the location of the CD45. The random walk steps are normally distributed with μ = 0 and σ = σLCK. LCK can have two states, active (LCK∗), or inactive (LCK). We assume an activation probability of 1 when an LCK is in activation range from the CD45. The activation range is 5nm from the location of the CD45. The LCK\* is randomly deactivated with probability Poff and is reactivated when it reenters the activation range of the CD45. In order to get a smoother distribution we run 100 independent LCKs simultaneously. Every time an LCK is active it gets a value of 1 and we save the LCKs locations and activeness state. The locations and binary values are binned into 500 bins to a histogram of LCK\*. LCK Poff values are:

10linspace(−5,0,19), LCK Diff values are: 10linspace(0,2.3,20). σLCK = sqrt(2\*DiffLCK\*Δt). A typical σLCK = sqrt(2\*0.3μm2/sec\*0.01sec) = 0.077μm = 77nm.

Fig. 3.2.

Pactive = (1-Poff)n, where n = number of iterations from last activation

Fig. 3.2.

Fig. 3.3.

Table S6A: random variables for surrogate model 1, before training.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model 3 – Lck activation (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  | log10 of diffusion coefficient of Lck\* |  |  |  |
|  | log10 of deactivation probability of LCK\* |  |  |  |
|  | Sigmoid minimum |  |  | nm |
|  | Sigmoid maximum |  |  | nm |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Uncertainty in |  |  | nm |
|  | Distribution width of Lck\* |  | μ= , | nm |

Table S6B: random variable for surrogate model 2, after training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model32 – Lck activation (LA) | | | | | |
| Type | Name | Description | Distribution | Distribution parameters | Units |
| Random variables |  | log10 of diffusion coefficient of Lck\* |  |  |  |
|  | log10 of deactivation probability of LCK\* |  |  |  |
|  | Sigmoid minimum |  |  | nm |
|  | Sigmoid maximum |  |  | nm |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Poff sigmoid center |  |  | - |
|  | D sigmoid devisor |  |  | - |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Used for sigmoid equation |  |  |  |
|  | Uncertainty in |  |  | nm |
|  | Distribution width of Lck\* |  | μ= , | nm |

