Statistical Learning on LAT Phase Transition Complexes

Eric Gai

May 9, 2021

1 Introduction

T cells are an important class of cells involved in adaptive immunity. They feature a specialized signaling pathway that begins with the T cell receptor (TCR) recognizing antigen peptide-bound major histocompatibility complexes (pMHC) and results in cytokine secretion, cell proliferation, and cytotoxic activity (in the case of killer T cells). T cells are impressively sensitive and specifc, with the potential to be activated by single pMHC molecules and the ability to distinguish between peptides with single amino acid substitutions [6]. Following TCR-pMHC interaction, the kinase ZAP-70 is recruited to the CD3ζ chain of the intracellular domain of the TCR and activated. ZAP-70 phosphorylates tyrosine residues on the linker for activation of T cells (LAT). Three of these phosphotyrosines can interact with the SH2 domain of Grb2. Each Grb2 molecule consists of one SH2 domain and two SH3 domains that can interact with the proline rich region of SOS (SOS-PR). SOS-PR can interact with several Grb2 molecules at the same time, effectively crosslinking LAT molecules into a condensed phase transition complex [5] (Fig. 1). Recent modeling of LAT complexes suggests that phosphorylation and interaction at all three phosphotyrosine sites are required for condensation [12]. LAT serves as a signaling hub for many pathways involved in T cell signaling.

Previous studies have shown that exclusion of the phosphatase CD45 from T cell and antigen presenting cell contact sites is necessary [11] and sufficient [4] for TCR activation. In addition, CD45 has been shown to be excluded from LAT clusters [13]. With phase transitions a classical example of nonlinear phenomena, condensed LAT complexes have been shown to drive downstream nonlinear resposnes, most recently demonstrated through an increase in the dwell time of SOS, implicating LAT phase transition as a kinetic proofreading step that increases T cell signaling fidelity [9]. It is then natural to ask whether the formation of LAT clusters itself is a process that is driven nonlinearly in the cellular context. Given that LAT phase transition is dependent upon its phosphorylation and the spatial exclusion of CD45, the balance of ZAP-70 and CD45 concentrations might be an important driver of the nonlinear phase transition behaviour.

Biological phase separation behaviour has garnered considerable attention in recent years. In solution, phase separated domains organize cellular functions such as transcriptional regulation and ribosomal RNA production [1]. More recently, phase separation has been recognized as an important mechanism for membrane signal transduction [3].

Machine learning approaches large datasets with the goal of extracting important information and making predictions based on them. Developed based on statistical theory, it has found wide ranging applications and gained widespread attention in recent years [7]. There are mainly two classes of machine learning problems: supervised learning, where labels are given for a set of training data and the goal is to predict labels for unlabeled data, and unsupervised learning, where no labels are involved and the algorithm must find some pattern in the data. Recently, machine learning methods, both supervised and unsupervised, have been applied to studies of phase transtions with fruitful results, from identification of order parameters, discovery of condensation phenomenology, to analysis of critical temperatures [8, 14, 2]. It is then worth considering whether application of these methods to biological condensate systems will yield similar insights.

Here, we build a model for LAT condensates using Monte Carlo methods, and perform unsupervised learning to examine what useful information can be gleaned from this system.

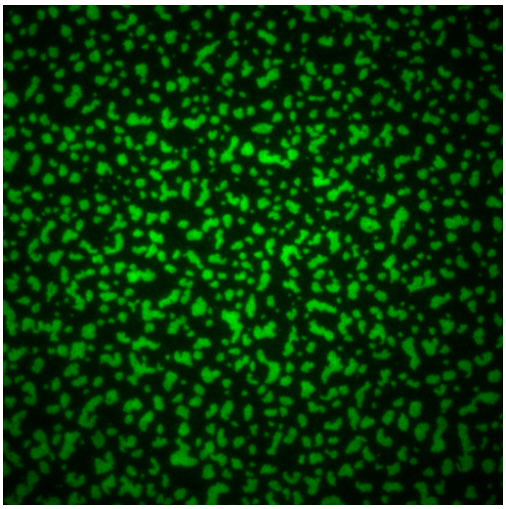


Figure 1: Phase-separated LAT condensates observed by total internal reflection microscopy.

2 Theory

Here we describe some classical theories of phase transitions, as well as describe the theoretical formulation for dimensionality reduction as applied in this work.

2.1 Landau-Ginzburg Phase Transition Theory

Here we derive a mean-field treatment of the LAT condensation following the framework of the Landau-Ginzburg phase transition theory. This development closely follows the material from Lecture 25 of this course.

Given some phosphatase-kinase competition pressure, let the average phosphorylation state of all LAT molecules be $\langle m \rangle$ with

$$0 \leq \langle m \rangle \leq 3$$

For some position $\mathbf{r}$, we can define a difference from the average

$$\phi(\mathbf{r}) = \delta m(\mathbf{r}) = m(\mathbf{r}) - \langle m \rangle$$

and

$$\langle \phi(\mathbf{r}) \rangle = \langle m \rangle - \langle m \rangle = 0$$

We can define a probability distribution $p[\phi(\mathbf{r})]$ and thus a free energy given by

$$\mathcal{H}[\phi(\mathbf{r})] = -k_B T \ln(p[\phi(\mathbf{r})])$$

Performing a cumulant expansion to fourth order,

$$\beta\mathcal{H} = \beta h_0 + \int_{\mathbf{r}} \Gamma_1(\mathbf{r})\phi(\mathbf{r}) + \frac{1}{2}\int_{\mathbf{r}}\int_{\mathbf{r}'} \Gamma_2(\mathbf{r},\mathbf{r}')\phi(\mathbf{r})\phi(\mathbf{r}') + \frac{1}{3!}\int_{\mathbf{r}}\int_{\mathbf{r}'}\int_{\mathbf{r}''} \Gamma_3(\mathbf{r},\mathbf{r}',\mathbf{r}'')\phi(\mathbf{r})\phi(\mathbf{r}')\phi(\mathbf{r}'') + \frac{1}{4!}\int_{\mathbf{r}}\int_{\mathbf{r}'}\int_{\mathbf{r}''}\int_{\mathbf{r}'''} \Gamma_4(\mathbf{r},\mathbf{r}',\mathbf{r}'',\mathbf{r}''')\phi(\mathbf{r})\phi(\mathbf{r}')\phi(\mathbf{r}'')\phi(\mathbf{r}''')$$

Since the free energy is at a minimum for $\phi = 0$, the free energy should be approximately even about that point. Therefore we can drop the first and third order terms, as well as the constant factor as we can pick an arbitrary reference point for the free energy. This yields

$$\beta\mathcal{H} = \frac{1}{2}\int_{\mathbf{r}}\int_{\mathbf{r}'} \Gamma_2(\mathbf{r},\mathbf{r}')\phi(\mathbf{r})\phi(\mathbf{r}') + \frac{1}{4!}\int_{\mathbf{r}}\int_{\mathbf{r}'}\int_{\mathbf{r}''}\int_{\mathbf{r}'''} \Gamma_4(\mathbf{r},\mathbf{r}',\mathbf{r}'',\mathbf{r}''')\phi(\mathbf{r})\phi(\mathbf{r}')\phi(\mathbf{r}'')\phi(\mathbf{r}''')$$

A Fourier transform yields

$$\beta\mathcal{H} = \frac{1}{2}\int_{\mathbf{k}} \hat{\phi}(\mathbf{k})\hat{\phi}(-\mathbf{k})\hat{\Gamma_2}(|\mathbf{k}|) + \frac{1}{4!}\int_{\mathbf{k}}\int_{\mathbf{k}'}\int_{\mathbf{k}''} \hat{m}(\mathbf{k})\hat{\phi}(\mathbf{k}')\hat{\phi}(\mathbf{k}'')\hat{\phi}(-\mathbf{k}-\mathbf{k}'-\mathbf{k}'')\hat{\Gamma_4}(\mathbf{k},\mathbf{k}',\mathbf{k}'')$$

Expanding about the low frequency modes where $\mathbf{k} = 0$,

$$\hat{\Gamma_2}(|\mathbf{k}|) = a + \gamma k^2$$

$$\hat{\Gamma_4}(\mathbf{k},\mathbf{k}',\mathbf{k}'') = 6u$$

which yields

$$\beta\mathcal{H} = \int_{\mathbf{q}} \frac{1}{2}(a + \gamma k^2)\hat{\phi}^2(\mathbf{k}) + \frac{1}{4}u\hat{\phi}^4(\mathbf{k})$$

An inverse Fourier transform then gives the Landau free energy of the system

$$\beta\mathcal{H} = \int_{\mathbf{r}} \frac{1}{2}a\phi^2(\mathbf{r}) + \frac{1}{4}u\phi^4(\mathbf{r}) + \frac{1}{2}\gamma|\nabla\phi(\mathbf{r})|^2$$

2.2 Principal Component Analysis

Here we discuss the mathematical formulation of principal component analysis, which is a linear dimensionality reduction technique. The main ideas of the following derivation come from Johnathan Shewchuk's lecture notes for CS189.

Suppose we have n sample points $\mathbf{x_1}...\mathbf{x_n}$ that are each d-dimensional. We can define the mean of these points as

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i}$$

We can then center the data points according to

$$\mathbf{z_i} = \mathbf{x_i} - \boldsymbol{\mu}$$

The centered points $\mathbf{z_1}...\mathbf{z_n}$ then have zero mean.

These centered data points can be used to define a design matrix

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z_1} \\ . \\ . \\ . \\ \mathbf{z_n} \end{bmatrix}$$

Since the mean of these points is zero, the $d \times d$ covariance matrix is given as

$$\boldsymbol{\Sigma} = \mathbf{Z^T Z}$$

This is matrix is positive semidefinite, hence admitting a set of orthonormal eigenvectors $\mathbf{v_1}...\mathbf{v_d}$ that span the $\mathbb{R}^d$ feature space with corresponding nonnegative eigenvalues $\lambda_1...\lambda_d$.

Without loss of generality, we can let the eigenvectors be ordered such that

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{d-1} \geq \lambda_d$$

The variance of the data points in a particular direction given by an unit vector $\mathbf{w}$ is

$$\sigma_{\mathbf{w}} = \sum_{i=1}^{n}(\mathbf{z_i}\cdot\mathbf{w})^2 = \begin{bmatrix}\mathbf{z_1}\cdot\mathbf{w} & . & . & . & \mathbf{z_n}\cdot\mathbf{w}\end{bmatrix}\begin{bmatrix}\mathbf{z_1}\cdot\mathbf{w} \\ . \\ . \\ . \\ \mathbf{z_n}\cdot\mathbf{w}\end{bmatrix} =$$

$$(\mathbf{Z}\cdot\mathbf{w})^\mathbf{T}\mathbf{Z}\cdot\mathbf{w} = \mathbf{w^T Z^T Z w} = \mathbf{w^T}\boldsymbol{\Sigma}\mathbf{w}$$

Then the maximum value of $\sigma_{\mathbf{w}}$ is achieved for

$$\mathbf{w} = \mathbf{v_1}$$

and

$$\max \sigma_{\mathbf{w}} = \lambda_1$$

This implies that $\mathbf{v_1}$ gives the direction where the data points have the greatest variance, and $\mathbf{v_2}$ gives an orthogonal direction with the next greatest variance, and so forth. These are referred to as the "principal components" of the dataset. In addition, we can project each point $\mathbf{z_i}$ to a lower d' dimensional space (hence the name "dimensionality reduction") by applying the transform

$$\mathbf{z_i'} = \begin{bmatrix}\mathbf{v_1} \\ . \\ . \\ . \\ \mathbf{v_{d'}}\end{bmatrix}\mathbf{z_i} = \mathbf{V'z_i}$$

which can be easily visualized for $d' = 2$ or $d' = 3$.

This method transforms high-dimensional data into fewer dimensions where the variation within the dataset is most significant. This allows us to extract information from highly redundant datasets, like a large collection of lattice sites. We can define an explained variance ratio

$$\tilde{\lambda_i} = \frac{\lambda_i}{\sum_{j=0}^{n} \lambda_j}$$

which assigns a significance to each principal component.

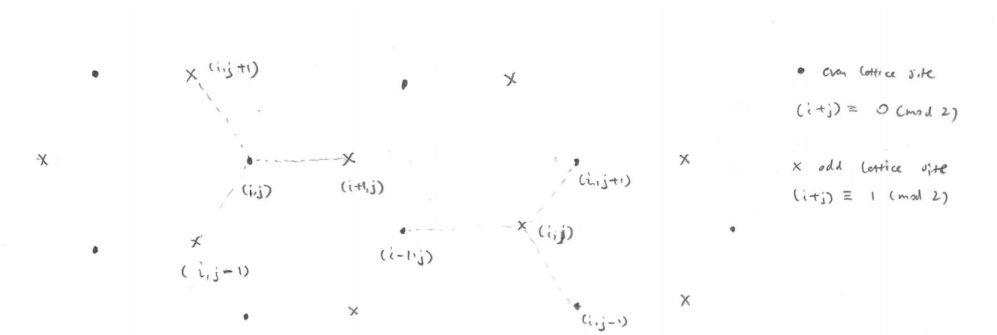


Figure 2: Setup of interacting hexagonal lattice for Monte Carlo simulations. Each lattice site is assigned a parity according to the parity of the sum of its coordinates. Lattice sites that interact depend on the parity of the specific lattice site.

3 Simulation

Short of only having fully phosphorlyated LAT molecules, this system is rather complex and difficult to solve analytically. Given that bonding of two adjacent LAT molecules depend on how each of them are already bonded to other LAT molecules, it is difficult to even write an expression for the energy of the system given a configuration of LAT molecules of specific phosphorylation states. Therefore, we built a Monte Carlo model for this molecular system as described below.

3.1 Monte Carlo Implementation

A grand canonical Monte Carlo system was used to simulate LAT molecules in a specified membrane region. The LAT molecules are fixed in position on a 20×20 hexagonal lattice with each lattice site interacting with three adjacent sites given according to the parity of the lattice site in question, which is in turn given by the parity of the sum of coordinates of the lattice site (Fig. 2). Periodic boundary conditions are applied to simulate a larger system. The supplied system parameters include a temperature given as the thermodynamic inverse temperature $\beta = \dfrac{1}{k_B T}$, an energy associated with forming a bond between two adjacent lattice sites E, and chemical potentials of unphosphorylated, monophosphorylated, diphosporylated, and fully phosphorylated LAT molecules, μ_0, μ_1, μ_2, μ_3, respectively.

Each proposed Monte Carlo move can add or remove a phosphate on a LAT molecule, or add or remove a bond between two adjacent lattice sites. Phosphate groups can only be added when a LAT molecule is not fully phosphorylated (in other words, there is a free unphosphorylated tyrosine residue). Phosphate groups can only be removed when LAT is not fully unphosphorylated and the phosphorylated residues are free (not bound to an adjacent site). A bond between adjacent lattice sites can be formed only if there are free phosphate groups on both LAT molecules.

Acceptance of proposed moves is determined by a Metropolis criteria. For phosphorylation and dephosphorylation moves, the acceptance probability is given by

$$p_{\text{Acc}} = \min\{1, e^{\beta(\mu' - \mu)}\}$$

where μ' is the chemical potential of the proposed state following the move and μ is the chemical potential of the current state. For bonding and unbonding moves, the acceptance probability is given by

$$p_{\text{Acc}} = \min\{1, e^{\pm \beta E}\}$$

with positive exponential for bond breaking and negative exponential for bond forming.

3.2 Derivation of Steady State Chemical Potentials

Here we derive steady state chemical potentials for LAT phosphorylation states to analyze the effect of kinase-phosphatase balance in LAT condensation.

Assuming both the kinase and phosphatase have equal rates on all the phosphotyrosine sites of LAT, the chemical equations for individual LAT molecules can be written as

$$\begin{aligned}
\mathrm{LAT{-}0\,pY} &\underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{LAT{-}1\,pY} \\
\mathrm{LAT{-}1\,pY} &\underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{LAT{-}2\,pY} \\
\mathrm{LAT{-}2\,pY} &\underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{LAT{-}3\,pY}
\end{aligned}$$

where

$$\begin{aligned}
k_+ &= k_{\mathrm{kinase}}[\mathrm{kinase}] \\
k_- &= k_{\mathrm{phosphatase}}[\mathrm{phosphatase}]
\end{aligned}$$

The kinetic rates under the steady state approximation can be written as

$$\begin{aligned}
\frac{\mathrm{d}[\mathrm{LAT}-0\mathrm{pY}]}{\mathrm{d}t} &= k_-[\mathrm{LAT}-1\mathrm{pY}] - k_+[\mathrm{LAT}-0\mathrm{pY}] = 0 \\
\frac{\mathrm{d}[\mathrm{LAT}-1\mathrm{pY}]}{\mathrm{d}t} &= k_-[\mathrm{LAT}-2\mathrm{pY}] - k_+[\mathrm{LAT}-1\mathrm{pY}] - k_-[\mathrm{LAT}-1\mathrm{pY}] + k_+[\mathrm{LAT}-0\mathrm{pY}] = 0 \\
\frac{\mathrm{d}[\mathrm{LAT}-2\mathrm{pY}]}{\mathrm{d}t} &= k_-[\mathrm{LAT}-3\mathrm{pY}] - k_+[\mathrm{LAT}-2\mathrm{pY}] - k_-[\mathrm{LAT}-2\mathrm{pY}] + k_+[\mathrm{LAT}-1\mathrm{pY}] = 0 \\
\frac{\mathrm{d}[\mathrm{LAT}-3\mathrm{pY}]}{\mathrm{d}t} &= k_+[\mathrm{LAT}-2\mathrm{pY}] - k_-[\mathrm{LAT}-3\mathrm{pY}] = 0
\end{aligned}$$

Solving the system yields

$$\begin{aligned}
\frac{[\mathrm{LAT}-1\mathrm{pY}]}{[\mathrm{LAT}-0\mathrm{pY}]} &= \frac{k_+}{k_-} \\
\frac{[\mathrm{LAT}-2\mathrm{pY}]}{[\mathrm{LAT}-0\mathrm{pY}]} &= \frac{k_+^2}{k_-^2} \\
\frac{[\mathrm{LAT}-3\mathrm{pY}]}{[\mathrm{LAT}-0\mathrm{pY}]} &= \frac{k_+^3}{k_-^3}
\end{aligned}$$

Since there are three relevant tyrosine sites, there is an additional entropy factor for these relations. There are three ways to arrange phosphotyrosine sites for monophosphorylated LAT, and likewise for diphosphorylated LAT. For unphosphorylated and fully phosphorylated LAT, there is only one configuration. Letting $\Phi = \frac{k_+}{k_-}$,

$$\begin{aligned}
\frac{[\mathrm{LAT}-1\mathrm{pY}]}{[\mathrm{LAT}-0\mathrm{pY}]} &= 3\Phi \\
\frac{[\mathrm{LAT}-2\mathrm{pY}]}{[\mathrm{LAT}-0\mathrm{pY}]} &= 3\Phi^2 \\
\frac{[\mathrm{LAT}-3\mathrm{pY}]}{[\mathrm{LAT}-0\mathrm{pY}]} &= \Phi^3
\end{aligned}$$

We can assume that the different phosphorylated species are similar enough as molecules that the reference potential $\mu^{(0)}$ is the same. In addition, working in the grand canonical ensemble, we can assume that all species exist as an ideal solution in the reservoir so the chemical potential can be written as

$$\mu = \mu^{(0)} + k_B T \ln x$$

where x is the molar composition. Then we can write

$$\begin{aligned}
\mu_1 - \mu_0 &= k_B T \ln x_1 - k_B T \ln x_2 = k_B T \ln\left(\frac{x_1}{x_2}\right) = k_B T \ln\left(\frac{[\mathrm{LAT}-1\mathrm{pY}]}{[\mathrm{LAT}-0\mathrm{pY}]}\right) = \frac{\ln(3\Phi)}{\beta} \\
\mu_2 - \mu_0 &= \frac{\ln\left(3\Phi^2\right)}{\beta} \\
\mu_3 - \mu_0 &= \frac{3\ln\Phi}{\beta}
\end{aligned}$$

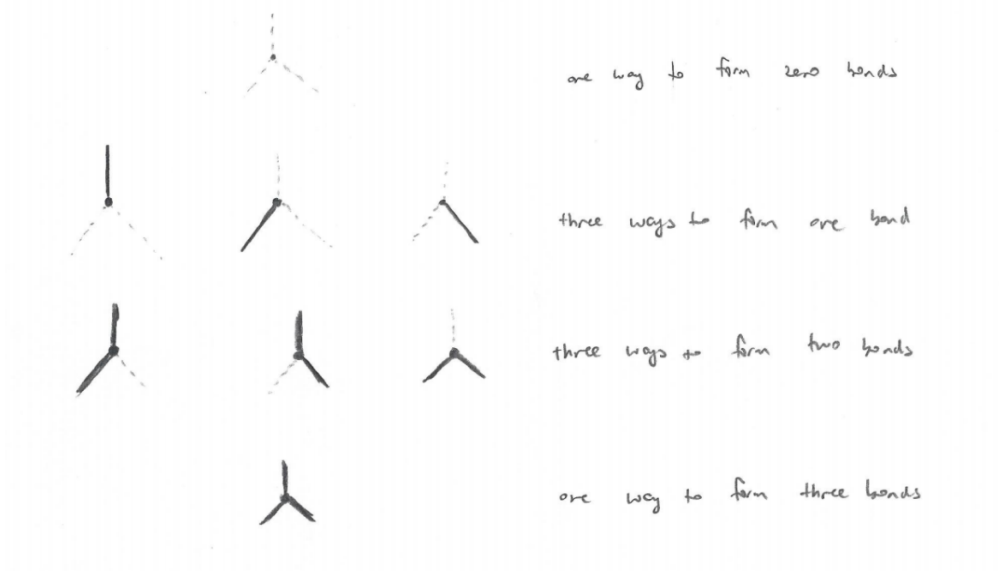


Figure 3: Ways to arrange bonds for each lattice site of LAT.

where μ_i is the chemical potential for LAT with i phosphotyrosine residues. For purposes of the simulation, since only the difference in chemical potential is relevant for acceptance, we can take the chemical potential of the unphosphorylated LAT molecule μ_0 to be zero without loss of generality. We arrive at

$$\mu_1 = \frac{\ln(3\Phi)}{\beta}$$
$$\mu_2 = \frac{\ln\left(3\Phi^2\right)}{\beta}$$
$$\mu_3 = \frac{3\ln\Phi}{\beta}$$

3.3 Insights by the Naked Eye

3.3.1 Distribution of Phosphorylation States

Even with zero-energy bonds and equal chemical potential for all phosphorylation states, the complexity of this system is already apparent. In this case, it manifests as an entropy effect due to the combination of bonds. In particular, more phosphorylated species have more ways to form bonds with adjacent sites (Fig. 3). Based on simple counting, one would expect a 1:4:7:8 ratio for the four different phosphorylation states. However, the simulation shows that the proportion of low phosphorylation states is underestimated and the proportion of high phosphorylation states is overestimated (Fig. 4). This can be attributed to the inability of higher phosphorylation states to access all possible bonding configurations due to limited adjacent lattice sites with free phosphates.

3.3.2 Temperature Transition

Changing the temperature of the system while holding the chemical potentials constant leads to an apparent increase in more phosphorylated LAT molecules. They seem to form condensed domains with irregular shapes remniscent of those observed in experiments. At higher temperature, the distribution of phosphorylated states approaches those of the system with zero-bond energy (Fig. 5).

However, despite the apparent condensation effect, there seems to not be any abrupt transition in the order parameters considered. Both the composition of the lattice and the total bonding energy changes smoothly through the condensation (Fig. 6). This is rather peculiar in the context of a phase transition,

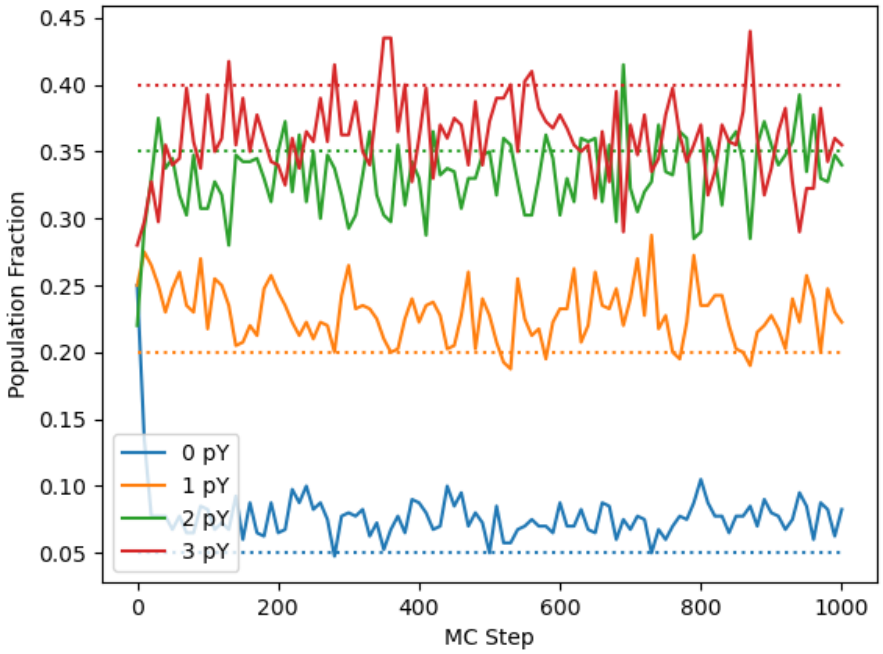


Figure 4: Comparison of simulated trajectory of distribution of phosphorylation states with naive expectation based on bond combinations. Solid lines represent simulated trajectories and dotted lines represent the naive expectations.

which is usually characterized by a discontinuity in some order parameter. Further work is needed to understand the nature of this transition.

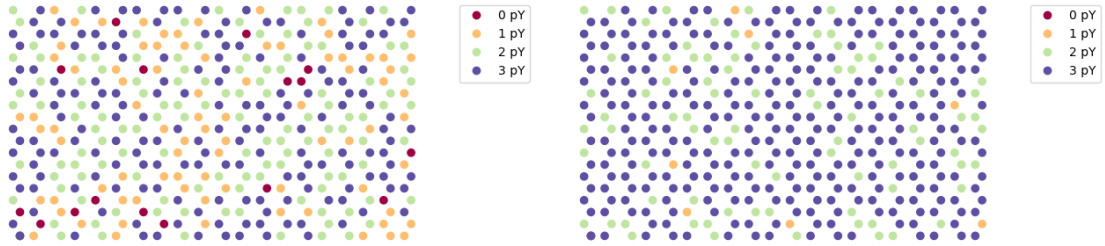
3.3.3 Kinase-Phosphatase Balance

Varying the kinase-phosphatase balance ratio Φ leads to even more dramatic shifts in the lattice configurations. When phosphatase pressure is high (Φ is small), virtually all LAT molecules are unphosphorylated. When there is close to no phosphatase pressure (Φ is large), almost all LAT molecules are fully phosphorylated and fully condense in the lattice (Fig. 7). The balance of phosphorylation levels, on the other hand, is distinct from that observed when varying temperature. The intermediate phosphorylated species exhibit a rise and fall pattern, demonstrating the nonlinear effect on the phosphorylation levels due to multisite phosphorylation (Fig. 8a). The level of fully phosphorylated LAT, which is the only species capable of full condensation, exhibits a particularly nonlinear response to the kinase-phosphatase balance, with a Hill coefficient of $n = 1.48 \pm 0.01$ when fitted to a Hill equation, suggesting high sensitivity to kinase-phosphatase balance (Fig. 9). The total bond energy of the system also exhibits a nonlinear response to the kinase-phosphatase balance, albeit not nearly as strong as the level of fully phosphorylated LAT. Taken together, these results suggest that the formation of fully condensated LAT complexes makes the system much more sensitive to kinase-phosphatase balance, potentially contributing to the sensitivity of the overall T cell activation pathway.

3.4 Insights by Statistical Learning

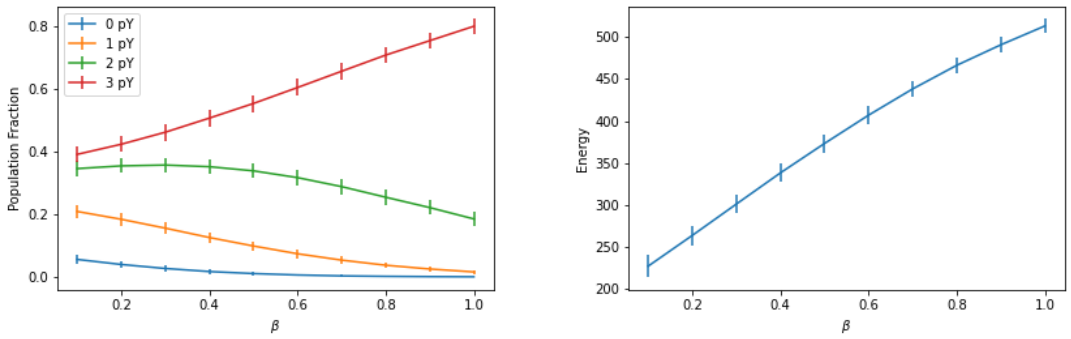
3.4.1 Temperature Transition

Applying PCA to configurations of the lattice at different temperatures results in a highly varying first principal component (10% explained variance), with all other components rather insignificant and barely



(a) (b)

Figure 5: Typical configurations of the lattice at different temperatures. (a) $\beta = 0.1$ (b) $\beta = 1.0$.



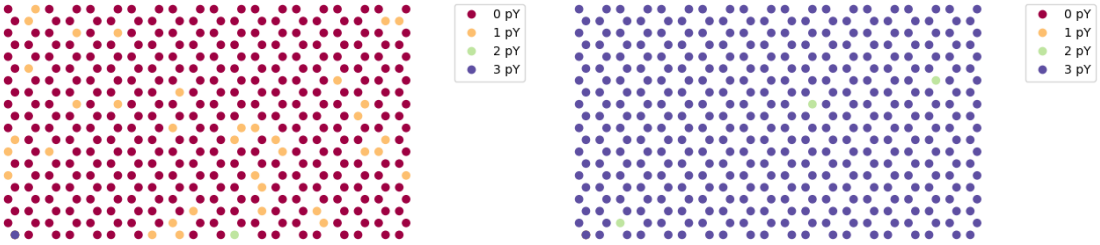
(a) (b)

Figure 6: Average composition of the lattice and average energy of the lattice varying the inverse temperature β. Data is represented as mean±S.D.

more than what one would expect from a totally uncorrelated configuration (∼0.03 compared to 0.025) (Fig. 10). This principal component has approximately equal weight for all lattice sites, suggesting a uniform increase or decrease in the phosphorylation state of the lattice. Higher principal components seems to indicate some degree of local correlation among certain lattice sites, where higher phosphorylation states of one lattice site is associated with higher phoshporylation states of adjacent lattice sites, although one would expect all of these correlations to be averaged out for larger sample sizes (Fig. 11). Transforming the datapoints into principal component space reveals the strong correlation of the first principal component with the temperature (Fig. 12). In particular, this suggests that the most significantly varying parameter with temperature is indeed the average phosphorylation state of the lattice. One also notices the larger variance of higher principal components at higher temperature, which is expected for the higher temperatures to have greater fluctuation.

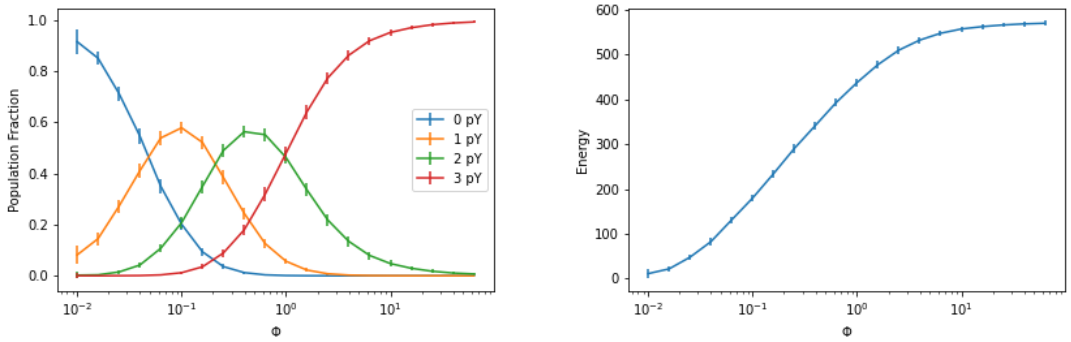
3.4.2 Kinase-Phosphatase Balance

Similar to the temperature transition, the principal components featured a highly varying $\mathbf{v_1}$ and rather insignificant higher components (Fig. 13). $\tilde{\lambda_1}$ is now up to over 80%, which might be reasonably expected since the activity of phosphorylating and dephosphorylating enzymes are directly being tuned. In particular, the first principal component describing the overall phosphorylation state is almost identical to that in the temperature transition, with higher components similarly capturing some local correlations (Fig. 14). In



(a) (b)

Figure 7: Typical configurations of the lattice at different kinase-phosphatase balance levels. (a) $\Phi = 0.01$ (b) $\Phi = 63$.



(a) (b)

Figure 8: Average composition of the lattice and average energy of the lattice varying the kinase-phosphatase activity ratio Φ. Data is represented as mean±S.D.

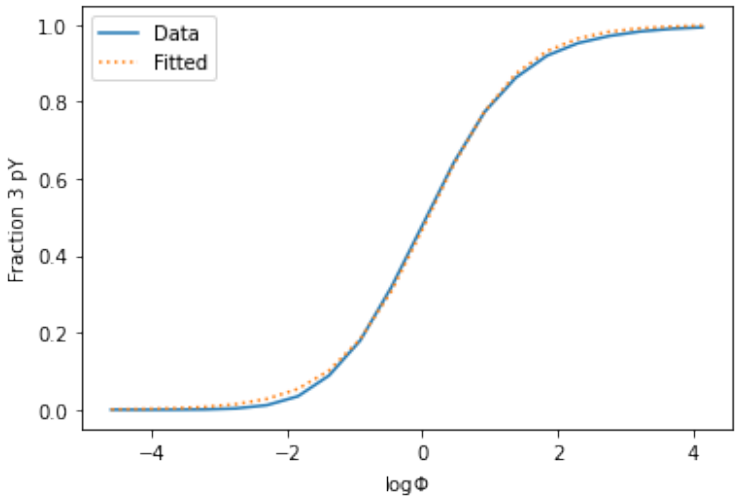


Figure 9: Hill equation fitting of the response response curve of the proportion of fully phosphorylated LAT to kinase-phosphatase balance.

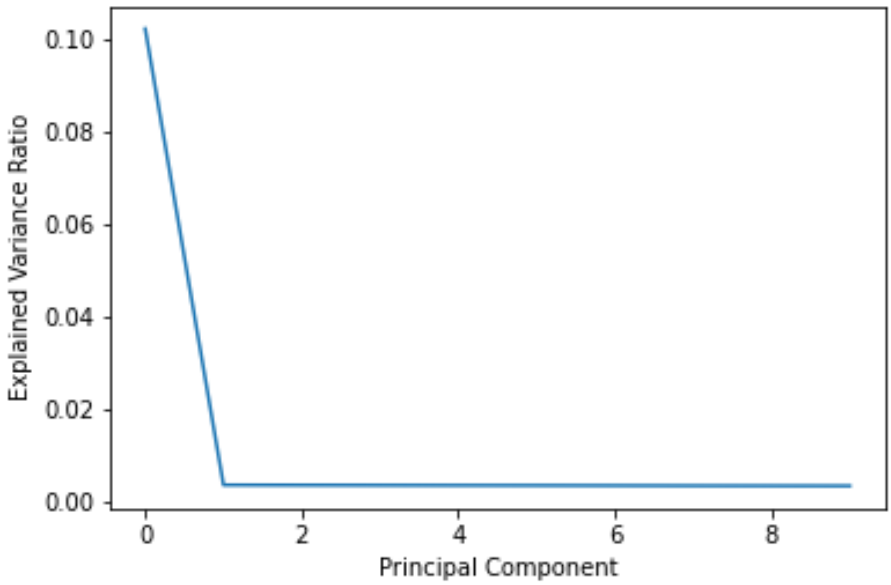
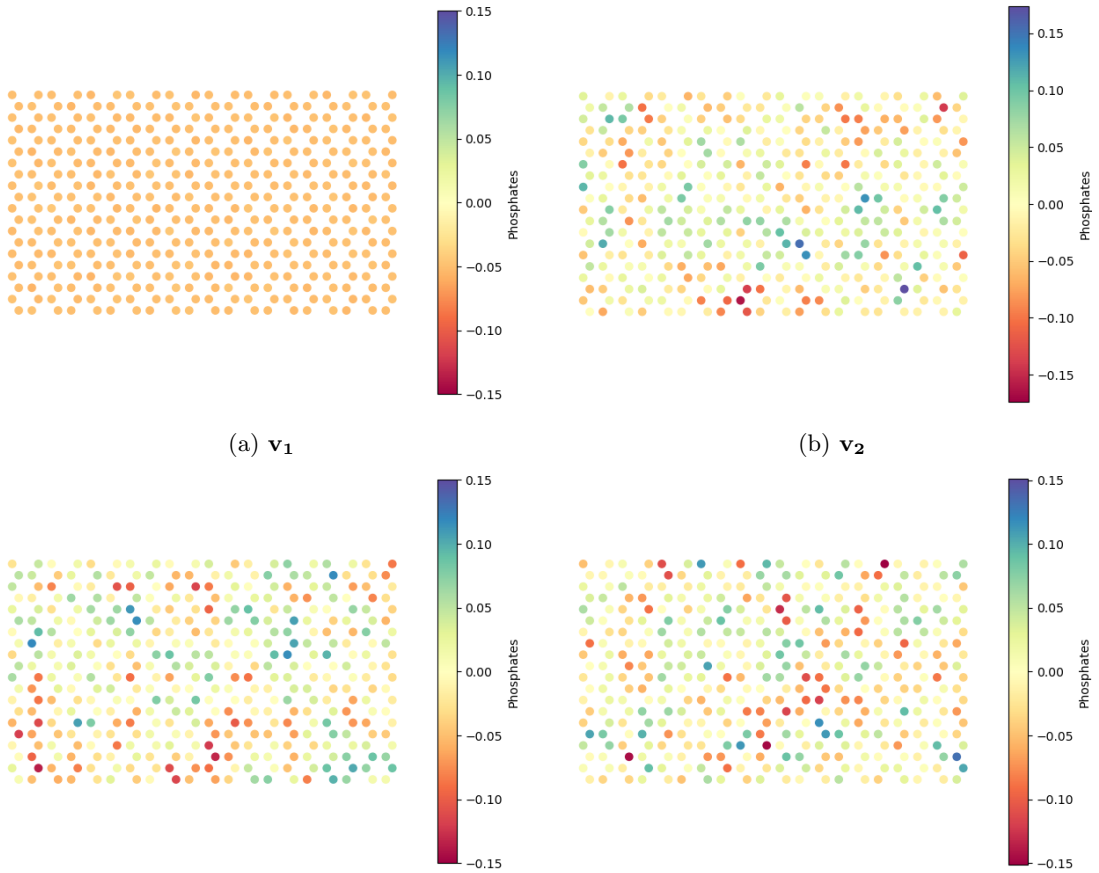


Figure 10: Explained variance ratio for temperature-varying configurations of the lattice.

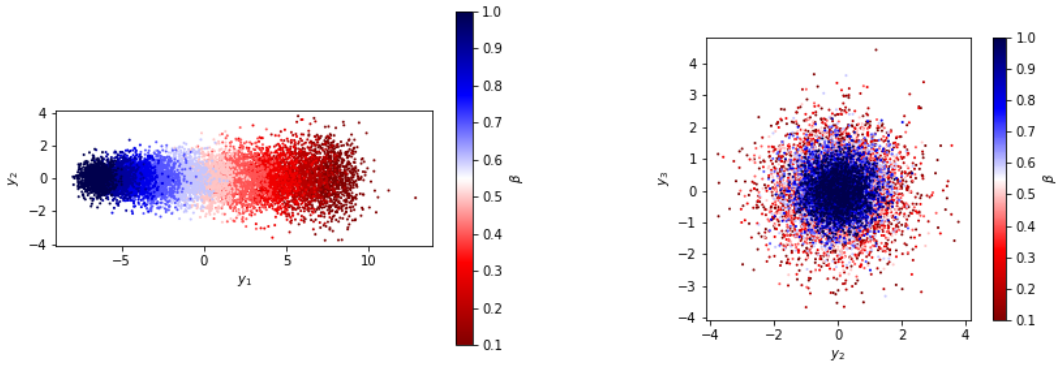
the principal component space, the correlation with Φ is, as expected due to the large $\tilde{\lambda_1}$, very strong. The configurations at different values of Φ form distinct clusters in the space of $\mathbf{v_1}$. Variation in the directions of the other principal components are largest for intermediate values of Φ (Fig. 15), indicating rather uniform configurations of the lattice when the balance of the competition reaction strongly favours the kinase or the phosphatase.



(a) $\mathbf{v_1}$ (b) $\mathbf{v_2}$

(c) $\mathbf{v_3}$ (d) $\mathbf{v_4}$

Figure 11: Principal components of lattice configurations at varying temperatures.



(a) (b)

Figure 12: Configurations of the lattice at varying temperatures plotted in the principal component space.

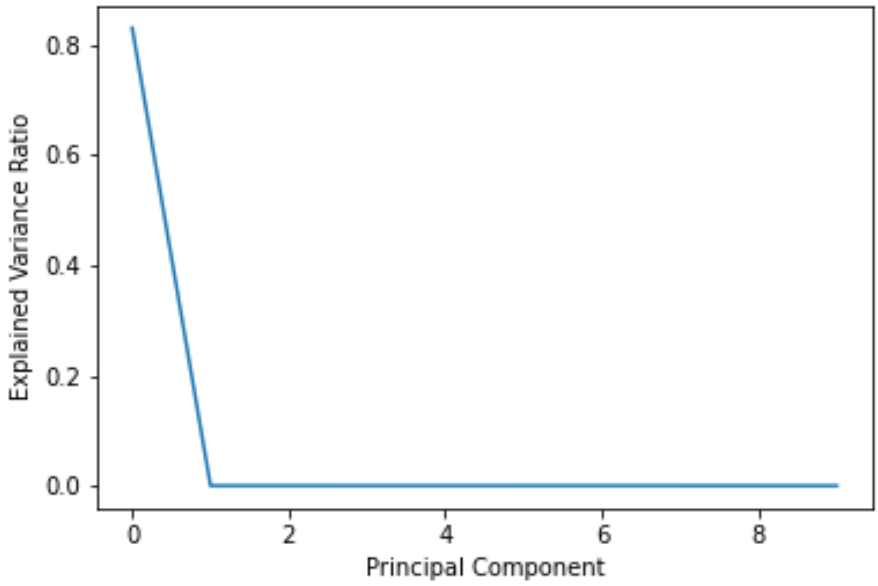
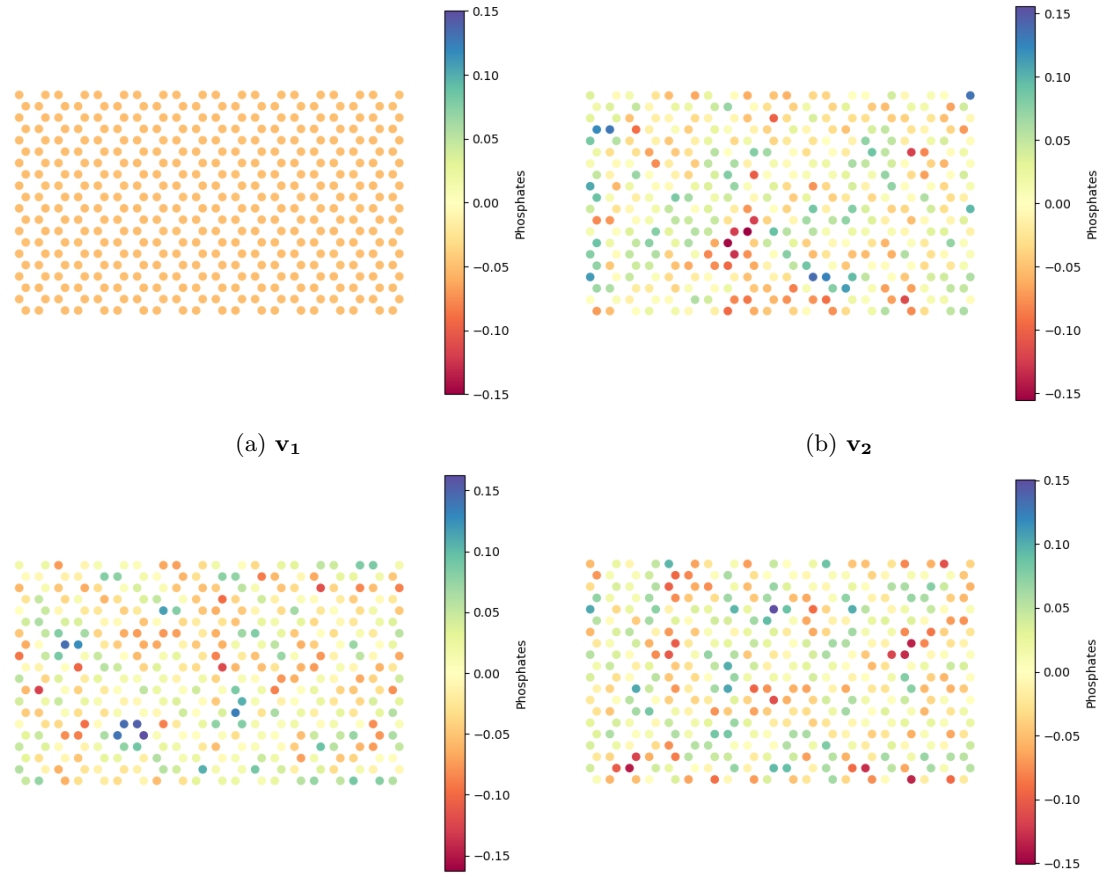


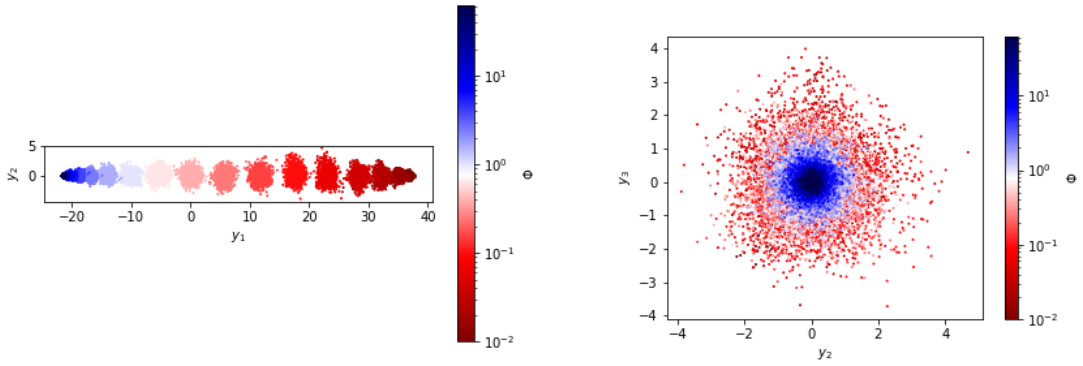
Figure 13: Explained variance ratio for kinase-phosphatase balance-varying configurations of the lattice.



(a) $\mathbf{v_1}$ (b) $\mathbf{v_2}$

(c) $\mathbf{v_3}$ (d) $\mathbf{v_4}$

Figure 14: Principal components of lattice configurations at varying kinase-phosphatase balance.



(a) (b)

Figure 15: Configurations of the lattice at varying kinase-phosphatase balance plotted in the principal component space.

4 Conclusion

In this work, we have constructed a computational model for LAT condensation and gleaned some insight into this system based on both the simulation itself as well as through methods of statistical learning. The sharp transitions typical of phase transition behaviour was notably absent under the conditions considered despite a recapitulation of the condensation behaviour. The average phosphorylation state of the lattice system was identified as a key order parameter for the condensation. Full condensation was driven nonlinearly by the kinase-phosphatase balance and mediated by multisite phoshporylation on LAT.

There are, however, several aspects of the LAT signaling system that were not considered in this study. Experiments suggest that tyrosine sites become more accessible when other tyrosine sites are phosphorylated, which can be regarded as an allosteric cooperativity that can increase the sensitivity of the signaling pathway [10]. In addition, the simulations here were performed with bond energy $E = -3k_BT$, while the experimentally determined value of $-15k_BT$ would presumably allow condensation at higher temperature and lower levels of kinase activity. The spatial exclusion of phosphatase from the LAT clusters was also not considered in the analysis above, which can potentially further contribute to the nonlinearity of the LAT condensation. Incorporation of these elements may yield more accurate results with greater relevance in the biological context.

The findings from PCA, while key to the behaviour of the system, are somewhat trivial. Characteristics of phase transitions such as long range correlation were noticeably absent from the principal component lattice vectors, which can be attributed to the fact that PCA treats each lattice site as an independent location spatially. Hence it might be useful to implement alternative machine learning methods such as convolutional neural networks that capture spatial correlation by its nature.

References

[1] Joel Berry, Clifford P Brangwynne, and Mikko Haataja. Physical principles of intracellular organization via active and passive phase transitions. *Reports on Progress in Physics*, 81(4):046601, February 2018.

[2] Juan Carrasquilla and Roger G. Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, February 2017.

[3] Lindsay B. Case, Jonathon A. Ditlev, and Michael K. Rosen. Regulation of transmembrane signaling by phase separation. *Annual Review of Biophysics*, 48(1):465–494, May 2019.

[4] Veronica T Chang, Ricardo A Fernandes, Kristina A Ganzinger, Steven F Lee, Christian Siebold, James McColl, Peter Jönsson, Matthieu Palayret, Karl Harlos, Charlotte H Coles, E Yvonne Jones, Yuan Lui, Elizabeth Huang, Robert J C Gilbert, David Klenerman, A Radu Aricescu, and Simon J Davis. Initiation of t cell signaling by CD45 segregation at 'close contacts'. *Nature Immunology*, 17(5):574–582, March 2016.

[5] Adam H. Courtney, Wan-Lin Lo, and Arthur Weiss. TCR signaling: Mechanisms of initiation and propagation. *Trends in Biochemical Sciences*, 43(2):108–123, February 2018.

[6] Andrew J.T. George, Jaroslav Stark, and Cliburn Chan. Understanding specificity and sensitivity of t-cell recognition. *Trends in Immunology*, 26(12):653–659, December 2005.

[7] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[8] Wenjian Hu, Rajiv R. P. Singh, and Richard T. Scalettar. Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination. *Physical Review E*, 95(6), June 2017.

[9] William Y. C. Huang, Steven Alvarez, Yasushi Kondo, Young Kwang Lee, Jean K. Chung, Hiu Yue Monatrice Lam, Kabir H. Biswas, John Kuriyan, and Jay T. Groves. A molecular assembly phase transition and kinetic proofreading modulate ras activation by SOS. *Science*, 363(6431):1098–1103, March 2019.

[10] William Y. C. Huang, Jonathon A. Ditlev, Han-Kuei Chiang, Michael K. Rosen, and Jay T. Groves. Allosteric modulation of grb2 recruitment to the intrinsically disordered scaffold protein, LAT, by remote site phosphorylation. *Journal of the American Chemical Society*, 139(49):18009–18015, November 2017.

[11] John R. James and Ronald D. Vale. Biophysical mechanism of t-cell receptor triggering in a reconstituted system. *Nature*, 487(7405):64–69, June 2012.

[12] Ambarish Nag, Michael I. Monine, James R. Faeder, and Byron Goldstein. Aggregation of membrane proteins by cytosolic cross-linkers: Theory and simulation of the LAT-grb2-SOS1 system. *Biophysical Journal*, 96(7):2604–2623, April 2009.

[13] Xiaolei Su, Jonathon A. Ditlev, Enfu Hui, Wenmin Xing, Sudeep Banjade, Julia Okrut, David S. King, Jack Taunton, Michael K. Rosen, and Ronald D. Vale. Phase separation of signaling molecules promotes t cell receptor signal transduction. *Science*, 352(6285):595–599, April 2016.

[14] Lei Wang. Discovering phase transitions with unsupervised learning. *Physical Review B*, 94(19), November 2016.