

# Embarrassingly Parallel Analysis Of T Immune Cell Surface Receptors

Tommi Muller

## Introduction

T cells are the agents in the immune system which track down pathogens that have infiltrated the body. They are known to release tags from their cell surface called *Lymphocyte Function-Associated Antigen 1* (or *LFA-1*) which attaches to invaders and marks them for the immune system to identify and destroy. LFA-1 receptors are also involved in the complex, subtle, and currently poorly understood communication between T cells and other cells in the immune system. One way to understand their role is to study their motion on the T cell surface.

The trajectories of many LFA-1 receptors on a T cell surface can be observed using a microscopic technique called *Total Internal Reflection Fluorescence (TIRF)* microscopy. This involves tagging fluorescent markers onto the LFA-1 receptors so that when the cell surface is illuminated, the fluorescent markers light-up. Using a high-speed video camera, the motion of the glowing markers on the surface can be recorded and analyzed. The positions of the markers suggest the locations of the LFA-1 receptors.

Previous studies have shown that the movements of LFA-1 fluctuate continually. As it traverses the membrane, an LFA-1 receptor can slow down, then accelerate, and change speed later. It is understood that through complex cell communication, a T cell can plant anchoring proteins on its surface, causing the LFA-1 receptors to sometimes slow down. The research team, consisting of Daniel Coombs, Michael Gold, Rebeca Cardim Falcão, Libin Abraham, Josh Scurll, and Rhys Chappell, produced a mathematical model, called the *Dodo 2-State Model*, which assumes that the LFA-1 receptors can each be in one of two states: a fast-moving state and a slow-moving state. The receptors can also transition between these two states. The model calculates the diffusion coefficients of both the fast and slow states of the receptors using the TIRF data.

Obtaining results from the Dodo 2-State Model, however, took days on a supercomputer due to the vast size of the data.

## Aim

The goal of this USRA project was to find a solution to the impractically long computation time of the Dodo 2-State Model so that it runs in under an hour.

## Methods

The original approach to calculate the diffusion coefficients in the Dodo 2-State Model essentially comes down to solving an optimization problem. There is a likelihood function which calculates the likelihood of observing the two diffusion coefficients - one for the fast state and one for the slow state - given all of the sequences of steps from all of the fluorescent markers in the TIRF data. The objective is to find the values of these two diffusion coefficients that will maximize this likelihood function; i.e. the two diffusion coefficients which most likely fit the data.

The optimization technique used in the model was the *Metropolis-Hastings algorithm*. Imagine starting at the bottom of a rugged mountain while blind-folded. Using the Metropolis-Hastings algorithm to maximize the likelihood function is analogous to reaching the top of the mountain by taking steps whose length and direction randomly depend on how high one climbed on the previous step. A result called the *Markov Chain Convergence Theorem* asserts that with enough steps, the algorithm converges to the global maximum of the likelihood function, producing a distribution of diffusion coefficients close to the maximum. With very large data, however, convergence can take a long time.

Some other optimization techniques which were used instead of the Metropolis-Hastings algorithm in hopes of running the model faster were, for example, *gradient descent* and the *Riemannian Manifold algorithm*. The gradient descent can be thought of as climbing the mountain, still blind-folded, by following the hill of steepest ascent. This algorithm, while fast, does not necessarily locate the global maximum and was found to be unreliable. The Riemannian Manifold algorithm uses techniques from differential geometry to predict the best direction to find the maximum based on the curvature of the likelihood function. This was an accurate algorithm but was not sufficient to reduce the computation time.

## Results

The solution found was to parallelize the computation of the Metropolis-Hastings algorithm. This means that the computation was done not just on one processor, or computer, but among several, in a coordinated manner. A heuristic called *Embarrassingly Parallel Monte Carlo Markov Chain*, from a paper by Neiswanger, Wang, and Xing (2013), was used to reduce the analysis by the Dodo 2-State Model of the TIRF data down to 10 minutes. The Embarrassingly Parallel heuristic is as follows:

Firstly, divide the TIRF data into equal chunks for each processor. Secondly, run the Metropolis-Hastings algorithm on each processor independently to obtain a distribution of the close-to-maximum diffusion coefficients on each processor. Lastly, randomly sample diffusion coefficients from each distribution obtained by each processor to produce a full distribution of close-to-maximum diffusion coefficients. This full distribution “recovers” the original distribution, as if the full data was used rather than the divided chunks.

## Further Research

The assumption of the Dodo 2-State Model that the LFA-1 receptors can only be in either a fast or a slow state was to simplify calculations. In reality, there may be more states, such as a “moderate” speed state. Recently, a new model called the *Infinite Hidden Markov Model* could potentially be used to find how many states fit the TIRF data well and the diffusion coefficients for each of these states.

## Acknowledgements

I would like to thank Professor Daniel Coombs for providing me with this opportunity to work on an exciting project and be involved in research. I want to thank Rebeca Cardim Falcão and Dr. Libin Abraham for their continual mentorship and support, and Professor Michael Gold, Josh Scurll, and Rhys Chappell for their support in the project.