Learning optimal integration of spatial and temporal information in noisy chemotaxis

Alonso, Albert; Kirkegaard, Julius B

Published in:
PNAS Nexus

DOI:
10.1093/pnasnexus/pgae235

Publication date:
2024

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Learning optimal integration of spatial and temporal information in noisy chemotaxis

Albert Alonso and Julius B. Kirkegaard

*Department of Computer Science, University of Copenhagen, Copenhagen 2100, Denmark
bNiels Bohr Institute, University of Copenhagen, Copenhagen 2100, Denmark

Abstract

We investigate the boundary between chemotaxis driven by spatial estimation of gradients and chemotaxis driven by temporal estimation. While it is well known that spatial chemotaxis becomes disadvantageous for small organisms at high noise levels, it is unclear whether there is a discontinuous switch of optimal strategies or a continuous transition exists. Here, we employ deep reinforcement learning to study the possible integration of spatial and temporal information in an a priori unconstrained manner. We parameterize such a combined chemotactic policy by a recurrent neural network and evaluate it using a minimal theoretical model of a chemotactic cell. By comparing with constrained variants of the policy, we show that it converges to purely temporal and spatial strategies at small and large cell sizes, respectively. We find that the transition between the regimes is continuous, with the combined strategy outperforming in the transition region both the constrained variants as well as models that explicitly integrate spatial and temporal information. Finally, by utilizing the attribution method of integrated gradients, we show that the policy relies on a nontrivial combination of spatially and temporally derived gradient information in a ratio that varies dynamically during the chemotactic trajectories.

Keywords: chemotaxis, navigation, sensing, reinforcement learning, machine learning

Significance Statement

Motile, microscopic organisms have evolved various strategies to perform chemotaxis: motion biased in a desirable direction. Large, slow-moving organisms typically possess chemotactic strategies that rely on spatial estimation of the chemoattractant gradient, while small, fast-moving organisms have evolved strategies that rely on temporally derived information. We employ deep reinforcement learning and demonstrate that in the intermediate range of parameters, strategies exist that can use a nontrivial combination of spatially and temporally derived information and that these strategies are superior to ones that only employ a single sensing type. Our findings suggest that microorganism adaptation to environmental conditions can include adapting the chemotaxis strategy itself.

Introduction

Chemotaxis, the directed motion of organisms towards or away from chemical cues, is a fundamental biological mechanism that spans biological kingdoms. For instance, prokaryotes rely on chemotaxis to find nutrients, avoid toxins, or even optimize oxygen and pH levels by sensing molecular cues (1–3). Single-celled eukaryotes show similar chemotactic traits (4), and countless biological processes in multicellular eukaryotes are supported by chemotaxis such as the fighting of bacterial infections by white blood cells, the positioning of stem cells during early embryonic development, and formation of multicellular structures in slime mold development (4–6). Likewise, a hallmark of cancer metastasis is the chemotaxis of tumor cells towards blood vessels (7).

However, the ubiquity of chemotaxis in biology does not imply uniformity in the mechanisms that underlie the navigation. At the scale of microorganisms, the fluctuations of the molecules that bind to the cells’ receptors are non-negligible and impose physical limits on the accuracy of the measurements and, thus, navigation. Chemotaxis is typically dichotomized into spatial and temporal strategies (Fig. 1C)(8, 10, 11). Larger cells, usually eukaryotes, primarily exploit spatial sensing, harnessing their size to directly perceive chemical concentration gradients (12), whereas smaller cells like bacteria are known to adopt temporal sensing, detecting alterations in chemical concentrations temporally to deduce information on the gradient’s direction, as the fluctuations across their body render spatial sensing useless. Cells are able to internally calculate temporal information due to the prolonged presence of the chemical signals within their body, which, in the absence of spatial cues, can provide information about the concentration changes experienced by the cell over time, especially as the cell moves through the environment (13–15). These
differences in sensing mechanisms have direct consequences for the possible types of navigation decision processes.

This binary classification enables detailed analysis of the distinct forms of chemotaxis within each category. However, as the optimal strategy is dependent on continuously varying parameters such as the size and velocity of the organism as well as the chemoattractant concentration, it leaves the question of whether organisms can utilize an integration of both spatial and temporal sensing mechanisms in their chemotactic strategies (16), and whether such a combination would be preferential in intermediate ranges of these parameters. Interestingly, it has been shown that cells thought only to use spatial sensing also rely on temporal information during chemotaxis when given periodic waves of chemoattractant (17, 18). Static temporal averaging of previous measurements has also been shown to reduce sensing noise on cells placed in shallow concentration gradients (12), however, this does not take into account the effect of the motile cell itself reacting to the measurements. Previous work has proposed a more complex inclusion of both types of sensing to develop new strategies without being able to outperform single sensing strategies (19), showcasing that efficient integration of both strategies is probably nontrivial.

Here, we employ deep reinforcement learning (DRL) to discover optimal chemotactic strategies that can combine spatial and temporal sensing. Previous work has successfully made use of DRL to find optimal strategies for self-propelled agents exploiting the flow in fluid environments (20) and for studying the tracking policies of flying insects relying on memory from noisy measurements to locate food or other insects (21). Similarly, machine learning has been utilized for demonstrating the optimality of known chemotactic strategies (22, 23).

We propose a minimal chemotactic single-cell model and use modern policy optimization techniques (24) to identify the strategy that minimizes the time it takes the cell to reach a source of chemoattractant. The model cell is endowed with distinct sensors that enable spatial gradient estimation and is given an internal memory state that allows temporal information to be derived. Based on a combination of these inputs, the cell must modify its orientation, a mapping that we leave largely unconstrained by employing deep neural networks.

We demonstrate the existence of a better performant chemotactic strategy that nontrivially combines spatial and temporal sensing. Specifically, we pinpoint a range of cell sizes where a combined sensing strategy outperforms optimal single sensing strategies. We then concentrate our analysis on this interface, comparing it to analytical ones and offering both qualitative and quantitative insights on the internal dynamics of the optimal navigational policy.

Methods

The simulation model

We study a exponentially decaying, 2D distribution $C(x)$ of chemoattractant particles with a concentration peak at $x = 0$.

$$C(x) = C_0 \exp(-\lambda |x|).$$ (1)
In Supplementary Material, we further give examples of algebraic and Bessel function concentration profiles which can e.g. arise from decaying and diffusing particles emanating from a central static source,

$$D\nabla^2 C(x) - \kappa C(x) + \rho \delta(x) = 0.$$  

(2)

Here, $D$ is the particle diffusion coefficient, and $\kappa$ is a particle decay rate, which sets a length scale $\lambda = \sqrt{D\kappa}$. We take $\lambda = 0.032 \mu$m$^{-1}$, given typical values of $D = 100 \mu$m$^2$/s and $\kappa = 0.1$ s$^{-1}$ (19), and study $C_0$ varying from $C_0 = 16 \mu$m$^{-2}$ to $10 \cdot C_0$, which sets the signal-to-noise ratio of the system and places the cell in the fundamental limit of sensing regime. $\rho$ is the rate of particle release at $x = 0$. Even though we assume a steady state profile for a static source, the cell will experience a change in concentration as it navigates the environment.

Our cell model consists of a circular disk of radius $R$, equipped with $K$ sensors uniformly spaced around its surface (Fig. 1A), whose objective is to reach the source of the chemoattractant by controlling its direction of motion depending on the environmental measurements. The cell senses the environment through molecules binding to cell-surface receptors. Still, in the interest of keeping our model as simple as possible, we neglect the complex receptor dynamics of receptor binding and unbinding (25). Thus, we assume each of the $K$ sensors to possess a detection area of radius $r_s = R \sin(\pi/K)$ such that the entire surface of the cell is covered. We fix $K = 5$ for all our experiments.

Our model cell never stops and moves forward at a constant speed, which we arbitrarily set to $v = 5 \mu$m/s, with its trajectory orientation $\theta(t)$ being modified both by its own actions as well as due to rotational noise,

$$d\theta = a \cdot dt + \sqrt{2D_\theta} \cdot dW.$$  

(3)

Here, $a_i$ is the output of the cell’s navigational policy $\pi$, and the second term is Wiener noise with rotational diffusion coefficient $D_\theta$. Rotational diffusion forces the cells’ navigation policies to react to the sensor signals at least on a time scale $1/D_\theta$ (26). In our experiments, we use $D_\theta = 0.025$/s as an average value on microorganisms of our sizes (8) and use a time-stepping of $\Delta t = 0.1$ s to solve the stochastic equations.

We model the cell receptors as perfect instruments (25), meaning that at each time step of our simulation, the sensors measure the exact number of molecules inside their sensor range instantaneously. This induces fluctuations in measurements with a signal-to-noise ratio that increases with concentration. Thus, nutrient-deprived environments with a low number of detected particles are noisy, and nutrient-rich environments are more deterministic.

We approximate the particle count within each sensor’s area as a stochastic process sampling from a Poisson distribution. Exploiting the nearly constant particle density over the detection area, we use,

$$E(M_i) = \int_A C(x) \cdot dA \approx C(d_i) \cdot \pi d_i^2.$$  

(4)

$$d_i \sim \text{Poisson}(E(M_i)).$$  

(5)

$d_i$ is the radial distance of the receptor center to the source of the chemoattractant.

Simulations are initialized at random distances $d_0$ from the source with random orientations $\theta_0$ and crucially with a rate of particle release $\rho$, which we sample in the range $\rho_0$ and $10 \cdot \rho_0$.

These random initializations ensure that the cell agents cannot overtrain to specific molecule counts and specific trajectories but rather need to generalize across noise levels and become adaptable to varying concentration profiles.

Finally, we do biologically inspired preprocessing of the receptor input by transforming according to the Weber–Fechner law (27),

$$m_i = \log(M_i + 1).$$  

(6)

While this could have been learned directly from the data, it conveniently brings the neural network input to a tightly constrained domain that is more suitable for DRL, and also means that noise in $m_i$ decreases not just relative to the signal but also in absolute numbers as $\rho$ increases.

The policy

The internal mechanisms of a chemotactic cell involve a complex set of biochemical spatio-temporal reactions. Here, we do not model these reactions explicitly, but instead model directly an input-output approximator, the cell policy $\pi$. This policy maps an internal state $s_\pi$, which in the simplest case could just be the vector of instantaneous measurements, to an action $a_i$. We parameterize the function using artificial neural networks (ANN) to minimize expressive restrictions on the learned policy.

To estimate the cell policy, we assume that it is an optimizer of efficient chemotaxis, which we define as minimizing the time it takes to reach a certain distance from the source, as the faster a cell reaches a source, the less competition with other cells it will encounter. More precisely, at the end of a simulation, we calculate a reward by

$$R = \frac{t_{\text{max}} - t}{t_{\text{max}}} + \max(-1, \frac{\delta - d}{d_0 - d}).$$  

(7)

where $d_0$ and $d$ are the initial and final distance to the source (which will be $d = \delta$ if the source has been reached), respectively, and $t$ is the simulation duration. In the case of not reaching the source, $r = t_{\text{max}}$. We include distance information as part of the reward reshaping technique to still gather information when the cell is not able to reach the source before $t_{\text{max}}$. The first term is a normalized reward for getting to the source fast, and the second is a bootstrapping reward that punishes cells that do not reach within the required distance $\delta$ of the source. The reward is normalized between $[-1, 1]$, as is convention in reinforcement learning. We perform episodic rewards instead of rewarding every action as we found the combined episodic rewards to converge to better solutions. Simulations terminate when the cell has reached a distance $\delta$ to the source or the simulation time has exceeded $t_{\text{max}}$, thus only one term of the reward expression is nonzero at the end of the episode; with the distance reward dominating early in training and the time reward at the end of training.

To find the optimal ANN policy, we employ Proximal Policy Optimization (PPO) (24), which adapts the policy $\pi$ in order to maximize the average reward. We study three variants of the agents (Fig. 1D): one policy we restrict to act purely on instantaneous spatial information. This is enforced by simply designing the neural network to be a pure feedforward network—from measurements $[m_i^0, m_i^1, \ldots, m_i^n]$ to output $a_i$. Likewise, we design a purely temporal network, which does not receive spatial information but rather the average of all receptors ($m_i$). Instead, this agent must rely on memory to provide temporal information on the particle gradients. This is achieved by introducing a recurrent layer into the policy neural network, which emulates the biochemical memory of real cells. Finally, we study a combined agent, which
has access to both spatially resolved measurements and has memory that can be used to derive temporal information. This agent can execute pure spatial and pure temporal strategies but can furthermore act on any combination of this information. Network details are given in Supplementary Material.

Our networks also output an estimate of the final reward $V_t$ (Fig. 1D), which the PPO algorithms use to speed up convergence but which does not influence the policy once trained. Further, as the nature of PPO’s exploration strategy adds noise to the policy output, we also recurrently feed the cell’s action back into the temporal policies, which aids the training in reaching a deterministic strategy without hindering stochastic exploration.

**Results**

**Optimizing for noise-robust strategies**

Our deep reinforcement learning approach is designed principally to work at all noise levels. In nutrient-rich environments, where the input to the agents is not corrupted by noise, our DRL framework converges quickly to effective temporal and spatial strategies. Resulting trajectories in these environments are close to deterministic as the noise from measurements gets reduced, and fewer mistakes in orientation corrections tend to occur. In those scenarios, spatial-based gradient estimation is effective in directly locating the source of chemoattractant, and noise due to rotational diffusion does not pose a challenge for the cell, which only needs to follow the strength of the sensors (Fig. 2C). Likewise, the optimal temporal sensing strategy at high concentrations is easily understood as it continuously measures the change in concentration and increases the turn when the concentration starts diminishing. As the temporal strategy contains no information about the sensors’ positions, it has to spontaneously break its rotational symmetry, which is exemplified in the resulting left-turning shown in Fig. 2A, resembling, e.g. the chirality of sperm chemotaxis trajectories (28).

In contrast, in the low concentration limit, the input to the cell receptors is extremely noisy (Fig. 2D), and the identification of optimal strategies becomes less clear. Yet, our DRL approach is able to identify working strategies both using purely spatial and purely temporal sensing mechanisms (Fig. 2B). Qualitatively, we note that the identified low-concentration temporal strategy behaves very robustly against noise, as its trajectory remains smooth despite its stochastic input. This can be interpreted as low reactivity, which also showcases itself as the temporal strategy only slowly adapts its trajectory as it nears the source. In comparison, the spatial strategy is very reactive, and while this makes it susceptible to the stochastic input, it enables it to quickly adapt its orientation once it nears the source and the concentration is relatively high. Finally, we observe the first hint that the combined strategy can outperform the two: it shows low reactivity when far from the source and high reactivity once in its proximity. Low reactivity in shallow and noisy concentration profiles allows the cell to rely on persistence to avoid getting trapped, increasing the likelihood of reaching a region where measurements convey more information. This strategy has also been observed in situations where cells are less likely to instantly turn around, enhancing their chemotactic response by reducing their reactivity (29).

While deterministic policies are fast to identify, the information that reinforces policies in the low concentration limit is much more stochastic, making the optimization process harder. To enable learning in this very noisy regime, our reinforcement...
learning steps rely on averaging the result of thousands of runs and require millions of simulations to converge to a solution (see Supplementary Table). To make this feasible, we developed a custom end-to-end RL implementation that runs exclusively on GPUs (see Data Availability).

We note that DRL is not guaranteed to find the globally optimal policy. However, we find that independent runs of the DRL training procedure result in the same policies, which hints that the obtained local optima could be global.

**Smooth transition between a temporal and a spatial strategy**

For evaluation, we define a strategy’s chemotactic efficiency $\eta$ by how fast the cell reaches the source compared to the minimal time a cell of speed $v$ would take to reach it from the same initial position (note that this is independent of $t_{\text{max}}$ which was used for training). Thus, the efficiency of a strategy is given by

$$\eta = \left( \frac{d_0 - \delta}{v \cdot \tau} \right), \quad (8)$$

where $\tau$ is the time it takes the cell to reach the source threshold distance $\delta$ and $d_0$ is the initial distance to the source, and the average is taken over all realizations.

We train our three variants, spatial ($S$), temporal ($\tau$), and combined ($C$), on the same simulation parameters at different cell sizes and proceed to calculate their efficiencies (Fig. 3A). At small sizes, where the positional sensor information becomes indistinguishable due to the noise, both $\tau$ and $C$ policies show the same performance. This is in accordance with previous studies showing that small cells are incapable of sensing gradients along their own body due to the fluctuations in measurements ($16$). Nevertheless, as the cell size increases, $C$ starts to outperform $\tau$, indicating that the tiny amount of available gradient information, as observed by the poor performance of $S$, can somehow be integrated into a temporally dominated strategy to improve its performance. At large cell sizes, $S$ dominates $\tau$, and while a gap still remains between $S$ and $C$ at large $R$, it shows convergence towards the same strategy. Thus, the sensors need not rely heavily on old measurements to estimate the gradient accurately at the largest scales. At intermediate cell sizes, we find that the optimal strategy is not purely spatial or temporal. In detail, we observe a smooth transition between strategies, indicating that there is a continuous integration of information stemming from spatial input and memory. Despite being dominated by noise, as illustrated in Fig. 2D, $C$ is capable of taking advantage of the measurement differences between the different receptors on the cell surface to improve its efficiency.

To explore this integration, we now focus on this intermediate region where both $S$ and $\tau$ perform similarly yet are outperformed by $C$, at $R \approx 2 \mu m$.

Inspecting the distribution of arrival times as shown in Fig. 3B for $R = 2 \mu m$, we observe a clear difference in skewness between $\tau$ and $S$. The distribution of arrival times in $S$ has long tails since cells that start far away from the source are experiencing very low concentrations of molecules, which disproportionately affect the spatial strategy. In contrast, $\tau$ shows very few cells that reach the source quickly, as this strategy relies on building memory. Interestingly, the cells that use $C$ are both fast and do not get trapped, having both benefits of the other variants.

To evaluate the optimality of the found strategy $C$, we compare it against commonly proposed strategies that use memory kernels to integrate temporal information into spatial strategies. Likewise, we explore a switching strategy in which cells start by using the noise-robust $\tau$ strategy and later switch to using the reactive $S$ strategy at a set threshold. This incorporates the advantages of each variant as shown in Fig. 3B. In all cases, we find that the RL learned strategy outcompetes these simpler explicit strategies (see Supplementary Information).

### Integrating temporal and spatial information

Having established that $C$ can integrate spatial and temporal information to outcompete both $\tau$ and $S$, we move on to studying the internals of $C$ directly. The policy $\pi$ is a highly nonlinear, recursive function which we have parameterized by deep learning neural networks—this at the cost of lack of interpretability. Nonetheless, numerous techniques have been developed to gain insight into the internals of a trained neural network, for instance, by estimating the importance of the input variables. One of the most elegant techniques to study this attribution problem is the method of integrated gradients (IG) ($30$), which calculates the importance of feature $x_i$ as

$$I_i = |x_i - x'| \int_0^{x_0} \frac{\partial \pi(x_0 + a(x_0 - x'))}{\partial x_i} \; da, \quad (9)$$

where $x'$ is a baseline, which we here simply take to be no input $x' = 0$. IG is sensitive meaning $I_i$ is nonzero if and only if $x_i$ contributes to the output, and satisfies completeness such that the attributions sum to the output, i.e. $\sum_i I_i = \pi(0) + \sum_i I_i$.

We use IG to understand how the cell relies on previous measurements transmitted to it by the hidden state $h_{-1}$, compared to current measurements $m_i$ from the receptors. We define $U_h$ as the relative importance of memory,

$$U_h = \frac{\sum_i |I_i|}{\sum_i |I_i|}, \quad (10)$$
where the contribution of the hidden state inputs is normalized by the sum of contributions of hidden state components (memory) and instant molecule measurements. Memory usage is well defined between 1, i.e. the cell only relies on memory for its decision without considering the current measures, and 0, where the cell disregards the memory information and instead only depends on instant measurements. Thus, a sensor is shown, as all have the same profile as per the designed symmetry. The red arrow indicates the swimming direction. Curves are obtained by averaging over \( \sim 10^3 \) trajectories with initial conditions sampled similarly to previous plots. B) Sensor contributes to the combined policy for different cell sizes. C) Trajectory visualization of both small (top) and large (bottom) cells. See Supplementary Material for a plot with all three variants.

Figure 5A shows the IG attributions of measurements for the purely spatial, the purely temporal, and the combined strategy at \( R = 2 \mu m \). As the cell diagram indicates, a positive IG value translates into a contribution for a positive reorientation and a negative value vice versa. For our model, this translates positive and negative contributions to pushing the cell to turn left or right, respectively. On a pure spatial strategy, the sensors work in opposition, and previous measurements obviously do not contribute. In contrast, all sensors contribute the same on a pure temporal strategy, but previous measurements oppose current measurements. Curiously, the shape of the contributions highly resembles the bi-lobed shape of the chemotactic memory kernel measured experimentally on the impulse responses of E. coli bacteria (26).

Similar to the spatial strategy, the combined strategy shows sensors working in opposition, but the left–right symmetry is broken and compensated by temporal variance, with one side dominating early and the other side contributing late. This sensor signature of the combined strategy makes explicit the nontrivial combination of information it is utilizing, and while these curves are merely IG components, they are indicative of a nonlinear combination of information that asymmetrically merges spatial and temporal processing (Fig. 5A). We observe a transition from temporal towards spatial information processing by looking at how measurements are integrated into the combined policy for different sizes (Fig. 5B). This similarity is clearly observed when comparing trajectories of the purely temporal and pure spatial policies with the combined one at the respective extreme cell sizes (Fig. 5C).
Discussion

In this study, we have explored the theoretical possibilities in chemotaxis that arise when traditional limitations are relaxed, i.e. when spatial and temporal strategies are not studied in isolation. Our findings show that the borders of binary classifications of chemotaxis strategies can be blurred by suitable integration of spatial and temporal information. In particular, we have shown that for cells with the ability to sense across their bodies as well as having memory access, there is a navigation strategy that outperforms those with only one sensing ability. Therefore, our results show that relying solely on unimodal chemotactic strategies to evaluate microorganisms’ efficiencies may indicate only a lower bound of their actual chemotaxis performance, and integrating both sensing information instead may yield more accurate estimates. Without imposing any constraints on the policy, we have seen the optimal solution to converge to known policies (purely spatial, purely temporal) in the limits where it is known that one sensing mechanism clearly provides faster information on the chemical gradient. Here, we explored this as a function of cell size and found that for large cells, the emerged combined strategy converges to relying only on spatial information, whereas for small microorganisms, the gradient information is strictly obtained on temporal differences. In the intermediate range, we found no sudden switch in strategy, but instead, the transition between them is continuous and smooth, where information is slowly being integrated by the cell into its decision process. We expect our mapping of optimal strategies during this transition may provide some guidance to understand how larger, more complex cells exploit temporal information to improve their sensing capabilities.

Our general perspective on chemotaxis is achieved by employing artificial neural networks and optimizing these by reinforcement learning. The drawback to this is that the obtained strategies are difficult to interpret. Yet, by comparing analytical strategies and employing integrated gradients to study feature attribution, we find that the optimal strategy that employs both spatial and temporal information is not a simple combination of known strategies, nor is its integration of information types trivial. Our analysis reveals that memory usage varies with cell size and concentration and changes dynamically throughout trajectories. This is akin to the well-known phenomenon that cells adapt their measurement sensitivity to local concentration (31), but here, we find that in an optimal setting, the navigation strategy itself must also dynamically adapt.

Using DRL to study chemotaxis in the noise-dominated regime is computationally challenging, as it requires a large number of simulations that must dynamically be run during training. Our custom approach runs simulations and training on GPU, avoiding slow system-to-device transfers. Here, we have employed this approach to study a simple chemotactic agent in two dimensions. An interesting avenue for future research is the move to three dimensions, where the space of possible strategies is qualitatively different. Likewise, it could be interesting to consider the consequences of a nonstatic source of chemoattractant or heterogeneous environments and discover their effect on a combined chemotactic policy. Similarly, it is of interest to extend our minimal cell model to specificities of particular organisms, such as a thorough modeling receptor dynamics (32), the inclusion of stochastic tumbles of peritrichously flagellated bacteria (15), or more complex behaviors as the ones seen in C. elegans (33). Concurrent to our work, agents capable of both spatial and temporal sensing mechanisms have been investigated using an information theoretical approach (34), deriving complementary insights.

Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions.

Supplementary Material

Supplementary material is available at PNAS Nexus online.

Funding

This project has received funding from the Novo Nordisk Foundation Grant Agreement NNF20OC0062047.

Author Contributions

A.A. and J.B.K. conducted the formal analysis, investigation, methodology, and writing and editing of the manuscript. A.A. was responsible for software development and visualization, and J.B.K. for conceptualization, supervision, funding acquisition, and project administration.

Preprints

This manuscript was posted on a preprint: 10.48550/arXiv.2310.10531

Data Availability

The original code and trained networks are available at Zenodo at https://zenodo.org/records/12544224 with DOI/accession number 10.5281/zenodo.8383156. The code for performing training and running the simulations and the scripts to evaluate the strategies can also be found at https://github.com/kirkegaardlab/chemoxrl.

References


