Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins

Helgi Ingolfsson
Lawrence Livermore National Laboratory  https://orcid.org/0000-0002-7613-9143

Chris Neale
Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Timothy Carpenter
Lawrence Livermore National Laboratory  https://orcid.org/0000-0001-7848-9983

Rebika Shrestha
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

Cesar Lopez
Los Alamos National Laboratory

Timothy Tran
Frederick National Laboratory for Cancer Research

Tomas Oppelstrup
Lawrence Livermore National Laboratory

Harsh Bhatia
Computing Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Liam Stanton
Department of Mathematics and Statistics, San José State University, San José, CA.

Xiaohua Zhang
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.  https://orcid.org/0000-0003-0102-6352

Shiv Sundram
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Francesco Di Natale
Lawrence Livermore National Laboratory, Livermore, CA.  https://orcid.org/0000-0002-1954-170X

Animesh Agarwal
Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Gautham Dharuman
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Sara Kokkila Schumacher
Data Centric Systems, IBM T.J. Watson Research Center, Yorktown Heights, NY.
Thomas Turbyville
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

Gulcin Gulten
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

Que Van
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory
https://orcid.org/0000-0002-2919-7375

Debanjan Goswami
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

Frantz Jean-Francios
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

Constance Agamasu
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

De Chen
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

Jeevapani Hettige
Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Timothy Travers
Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Sumantra Sarkar
Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM.

Michael Surh
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Yue Yang
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Adam Moody
Computing Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Shusen Liu
Lawrence Livermore National Laboratory

Brian Van Essen
Computing Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Arthur Voter
Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM.

Arvind Ramanathan
Computing, Environment & Life Sciences (CELS) Directorate, Argonne National Laboratory, Lemont, IL.

Nicolas Hengartner
Los Alamos National Laboratory

Dhirendra Simanshu
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.
Andrew Stephen  
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.

Peer-Timo Bremer  
Computing Directorate, Lawrence Livermore National Laboratory, Livermore, CA.  
https://orcid.org/0000-0003-4107-3831

S Gnanakaran  
Los Alamos National Laboratory  
https://orcid.org/0000-0002-9368-3044

James Glosli  
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Felice Lightstone  
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

Frank McCormick  
University of California, San Francisco

Dwight Nissley  
RAS Initiative, The Cancer Research Technology Program, Frederick National Laboratory, Frederick, MD.  
https://orcid.org/0000-0001-7523-116X

Frederick Streitz (✉ streitz1@llnl.gov)  
Physical and Life Sciences (PLS) Directorate, Lawrence Livermore National Laboratory, Livermore, CA.  
https://orcid.org/0000-0003-0924-037X

Article

Keywords: RAS, machine learning, lipid-dependent dynamics

DOI: https://doi.org/10.21203/rs.3.rs-50842/v1

License: ☺️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

RAS is a signaling protein associated with the cell membrane that is mutated in 30% of human cancers. RAS signaling has been proposed to be regulated by dynamic heterogeneity of the cell membrane. Investigating such a mechanism requires near-atomic detail at macroscopic temporal and spatial scales, which is not possible with conventional computational or experimental techniques. We demonstrate here a multiscale simulation infrastructure that uses machine learning to create a scale-bridging ensemble of over 100,000 simulations of active wild-type KRAS on a complex, asymmetric membrane. Initialized and validated with experimental data (including a new structure of active wild-type KRAS), these simulations represent a substantial advance in the ability to characterize RAS-membrane biology. We report distinctive patterns of local lipid composition that correlate with interfacially promiscuous RAS multimerization. These lipid fingerprints are coupled to RAS dynamics, predicted to influence effector binding, and therefore may be a mechanism for regulating cell signaling cascades.

Introduction

RAS driven cancers are common(1), difficult to treat(2), and a major cause of death worldwide(3). KRAS, the isoform most frequently associated with disease, is mutated in nearly all pancreatic cancers, and often in lung and colorectal cancers(4, 5). Only recently, with the development of covalent inhibitors of the G12C mutant(6), has direct targeting of RAS been successful, and more broadly applicable inhibitors are needed.

The RAS subfamily comprises peripheral membrane proteins with a conserved globular GTP-binding domain (G-domain)(7) that is tethered to the cell membrane by a prenylated ~ 20 residue C-terminal domain called the hypervariable region (HVR)(8, 9). RAS proteins function as molecular switches whose active conformations, stabilized by GTP binding, interact with several protein effectors to control cell growth, proliferation, differentiation, and migration(10). Constitutive activation of oncogenic RAS perturbs several cellular signaling cascades, including the MAPK pathway, which RAS accesses via activation of RAF kinase at the plasma membrane (PM).

There is substantial interest in assessing the ability of RAS molecules to dimerize(11) or co-localize(12) at the membrane, because RAS-dependent RAF activation requires dimerization of RAF(13–15). Although wild-type KRAS4b, a common splice variant of KRAS (hereafter referred to as RAS), does not dimerize on two-component supported lipid bilayers(16), it preferentially co-localizes with anionic lipids in the liquid-disordered domains of giant unilamellar vesicles(17) and the oncogenic G12V mutant clusters on the scale of tens of nanometers in extracted PM sheets(12, 18).

Preferential interaction of RAS with anionic lipids is mediated by 11 positively charged lysines in its HVR(19). However, charge complementarity is insufficient to fully describe RAS nanoclustering, which is exquisitely sensitive to lipid composition(20). Even less is known about the influence of RAS-lipid
coupling on RAS self-assembly and effector activation. While several feasible dimer interfaces have been reported(11, 21–24), how RAS forms dimers, if at all, remains a major area of interest.

The fundamental challenge of investigating RAS activation events is that many of the proposed mechanisms involve time- and lengths scales currently not accessible. For example, functional events in RAS dynamics may preferentially depend on local depletion or enrichment of specific lipids are extremely difficult to observe directly either in computational or biological experiments. Experimentally, we use single particle tracking (SPT) to follow HaloTag-conjugated RAS via total internal reflection fluorescence (TIRF) microscopy in live HeLa cells. While the broad heterogeneity in lateral diffusion observed by SPT (Fig. 1A) is indicative of multiple RAS sub-populations that have distinct patterns of interaction with lipids and other cellular components, it provides little insights on local lipid environments and their effect on RAS behavior. Similarly, we can use detailed molecular dynamics (MD) simulations to probe specific lipid environments, but systems large enough to support substantial fluctuations in lipid and protein concentrations cannot be practically simulated long enough to observe relevant fluctuations. Instead, we present a new multiscale infrastructure that directly couples a macro scale continuum simulation capable of observing RAS clustering and long-range lipid rearrangement, with a massive ensemble of micro scale MD simulations that provides detailed insights into local dynamics. Both scales are connected through a machine learning (ML) informed dynamic sampling process(25) which enables mapping of findings from the MD simulations onto the continuum simulation, resulting in micro scale insights that are observable over macro spatial and temporal scales.

More specifically, to enable simulations of active wild-type RAS, we solve its crystal structure bound to the GTP analog GppNHp at a resolution of 2.5 Å (Fig. 1B). We chose to focus the effort on wild-type GTP-loaded KRAS4b because GTP hydrolysis will not occur on the time-scale of the simulations, and because available structures of wild-type or oncogenic KRAS4b bound to the RAS binding domain (RBD) of RAF1 are similar. This current work will serve as the foundation for understanding activation of RAF by GTP-loaded RAS. We mimic the composition of a biologically-relevant PM(26) by employing an asymmetric 8-lipid mixture(27) of cholesterol, phosphatidylcholines (PC), phosphatidylethanolamines (PE), phosphatidylinerine (PS), phosphatidylinositol bisphosphate (PIP2), and sphingomyelin (SM), with varying acyl chain length and saturation (Fig. 1C, S1.2.1).  

**Multiscale Simulation Framework**

This paper introduces the Multiscale Machine-learned Modeling Infrastructure (MuMMI; S1.1, S1.2). MuMMI employs a macro model based on dynamic density functional theory (DDFT) running on CPUs to explore 300 RAS proteins on a $1 \times 1 \mu m^2$ lipid bilayer for 150 µs (Fig. 1D). Orchestrated by a massively parallel workflow(28), a machine learning algorithm(25) selects $30 \times 30 \text{nm}^2$ macro model sub-regions (hereafter patches) with one or more RAS that are evaluated in more detail by coarse-grained (CG) Martini(29) MD simulation (Fig. 1C-1F). These CG simulations are performed on GPUs(30), analyzed in situ, and used to continuously refine macro model parameters. To favor the instantiation of CG simulations with novel compositions, MuMMI encodes local lipid concentrations in a 15D latent space.
using an autoencoder-style neural network and spawns new simulations of candidate patches that are most dissimilar to existing simulations (Fig. 1C-1F). This approach enables broad sampling of local lipid composition while maintaining enough information about non-selected patches to accurately reconstruct the true probabilities of compositional fluctuations (S1.1.5, S2.2.3, S2.3.2).

MuMMI represents an emerging new capability that delivers a macro scale model at experimentally relevant time- and length-scales with effective micro-scale resolution. MuMMI simulations presented in this study contain an ensemble of 119,686 unique CG simulations, with one or more RAS molecules, totaling a simulation time of 206 ms (Fig. 1H, S2.1). These CG simulations capture and quantify RAS-lipid dynamics across local compositions observed in the much larger macro simulation.

**Ras-lipid Co-localization**

Lipid fingerprints are spatially resolved lipid concentration signatures around RAS proteins in the PM. The macro simulation reveals lipid redistribution around RAS (Fig. 2A) indicative of discrete lipid fingerprints, and a propensity for RAS to self-associate (Fig. 2B). The latter is consistent with experimental RAS clustering as seen in atomic force microscopy (AFM) experiments with simple lipid mixtures(17), though clusters observed in that study were much larger than seen in our simulations.

To further explore the connection between the local lipid environment and RAS clustering we use function preserving projections (FPP)(31) to find a linear embedding of all patches that maximally correlates the number of RAS in a patch with local lipid concentrations (Fig. 2C, S2.8). A subselection within the high/low RAS regions of the embedding is used to find unimodal lipid compositions that are either RAS depleted (low RAS co-localization or LRC) or enriched (high RAS co-localization or HRC) as classified in Fig. 2D, Table S7. In particular, the HRC is enriched in PIP2 and depleted in cholesterol and polyunsaturated lipids, whereas the LRC exhibits opposite trends. These findings on our in silico artificial membrane corroborate RAS’ preferential association with PIP2 on experimental artificial membranes(32, 33). However, discrepancy still exists between reconstituted systems and the observed enrichment of PS and phosphatidic acids in RAS nanoclusters on intact-cell PM sheets(17). While some lipid concentrations, such as 15.5% PIP2 in the HRC, are extreme for the global PM, our macro scale model indicates their presence in spatially localized regions due to concentration fluctuations and protein-mediated enrichment. Surface plasmon resonance (SPR) confirms that, compared to the initial average lipid composition (ARC), RAS binds more weakly to the LRC and more strongly to the HRC (Fig. 2E), consistent with specific lipid concentrations influencing RAS recruitment and co-localization.

**Ras Diffusion And Preferential Interaction With Pip2**

Diffusion coefficients of inner leaflet lipids are broadly distributed among the 88,392 unique CG simulations with one RAS molecule (Fig. 3A, S2.7). This variation reflects the dependence of lipid diffusion on membrane composition. For instance, a 10% change in cholesterol content slows the diffusion of all lipids except PIP2 by a factor of 2 (Figs. S42, S43). For a more complete discussion of the
impacts of membrane composition, see S2.3.2. The ARC diffusion rates calculated from fluorescence lifetime correlation spectroscopy (FLCS) are shown in Fig. 3B. All lipids except PIP2 diffuse at similar average rates in FLCS and in the ensemble of simulations, whereas RAS and PIP2 diffuse 2–3 times slower (Figs. 3A, 3B) and direct interaction with RAS slows the diffusion of PIP2 (Fig. 3C). In CG simulations with multiple RAS molecules, increased aggregation of RAS is associated with local enrichment of PIP2 and depletion of PAPS (Fig. 3D), emphasizing the potential importance of charge density. RAS aggregation also slows the diffusion of both RAS and PIP2 (Fig. 3E).

To evaluate whether lipid composition drives RAS multimerization, we use SPT to study supported lipid bilayers with LRC, ARC and HRC compositions. Mean square displacement (MSD) plots from 50,000-100,000 tracks obtained on each bilayer show that RAS undergoes anomalous diffusion that is slower and more confined on HRC compared to LRC and ARC (Fig. 3F). Hidden Markov modeling (HMM) of the SPT data identifies three interconverting states in which RAS diffusion is slow ~ 0.1 µm²/s, medium ~ 0.5 µm²/s, and fast ~ 4 µm²/s (S2.8.5). This HMM analysis reveals that the lipid-dependence of RAS diffusion (Fig. 3F) results from a decrease in slow and medium diffusion rates and an increased population of slowly diffusing RAS in the HRC (S2.8.5). One interpretation of these data is that RAS and PIP2 strongly associate and slow down as they cluster, although other factors can also slow diffusion.

Taken together, simulation and experiments identify a lipid composition (HRC) that recruits more RAS (Figs. 2C, 2E) and favors RAS multimerization (Fig. 2C), slowing its diffusion (Fig. 3F). Lipid-driven RAS clustering provides an attractive hypothesis as an initial organizational step in nucleating a molecular signaling platform for efficient recruitment of RAS effectors.

Ras-ras Association

RAS multimers are hypothesized to function as signaling hotspots, and defining RAS-RAS interaction is critical to understanding the role of co-localization in signaling competence(12). Indeed, specific RAS-RAS interfaces have been identified in previously published works(11, 21, 22, 24, 34). However, analysis of 10,939 MuMMI simulations with RAS-RAS contacts reveals a broad distribution of interfacial arrangements (Fig. 4A), including published interfaces (Fig. S33). Although initial events leading to RAS complexation tend to involve the HVR (Fig. 4B) and associated PIP2 (Figs. 4D, 4E), these orientational propensities become less pronounced upon G-domain contact (Fig. 4B-4D). The resulting modest (~ 2-fold) interfacial preferences are inconsistent with the existence of a distinct dimer interface (Fig. 4C). As such, distinct lipid fingerprints may facilitate initial RAS association through pre-orientation of the protein. Due to these multiple potential interfaces and heterogeneity seen in RAS-RAS association (Fig. 4C), it is problematic to deconvolute a non-rotationally averaged lipid fingerprint for RAS multimers. Nevertheless, the local lipid environments of RAS aggregates are enriched/depleted in different lipid species, including enriched in PIP2 and depleted of ordered lipids in comparison to RAS monomers (Figs. 3D, 4D, 4E, S58). This lipid-RAS codependence is consistent with a signaling mechanism in which lipid-RAS interactions concentrate RAS prior to RAF binding, and in which subsequent RAF dimerization is facilitated by this RAF-independent RAS colocalization.
Lipid Dependence Of Ras Orientation And Competence For Effector Binding

RAS is a closely tethered peripheral membrane protein, and the orientation of the G-domain controls access to its effector binding site through reversible membrane-based occlusion\(^{(35, 36)}\). We define the orientation of the RAS G-domain with respect to the membrane surface\(^{(37)}\). Briefly, a tilt angle defines the deflection of the long axis of helix 5 from the bilayer normal, and a rotation angle defines the direction in which that tilt occurs (Fig. 5A). CG simulations with one RAS reveal statistical confinement of RAS to specific orientations; moreover, Markov state modeling (MSM) of RAS orientation reveals three kinetic states of RAS orientation in which tilting brings G-domain helices 3–5 (\(\alpha\) state) or \(\beta\)-sheets 1–3 and switch I (\(\beta\) and \(\beta'\) states) toward the membrane (Figs. 5B-5F, S2.4-S2.5).

The most populated RAS orientations are incompatible with effector binding (Fig. 5G). We use the RBD of RAF (PDB ID: 4G0N) as the relevant effector and observe that \(\alpha\) states project the effector binding interface of RAS away from the membrane (Figs. 5B, 5C, 5D), possibly positioning RAS to bind its effectors\(^{(38, 39)}\). Conversely, effector binding is occluded in most of \(\beta'\) and half of the \(\beta\) state orientations (Figs. 5B, 5E-5G). Previously described exposed and occluded orientations\(^{(38–40)}\) are included in the \(\alpha\) and \(\beta/\beta'\) states, respectively, and an experimentally-driven model of RAS-RAF association straddles the dividing line between the \(\alpha\) and \(\beta\) states\(^{(39)}\). HMM analysis indicates that direct transitions between \(\beta\) and \(\alpha\) are relatively rare (Fig. 5H) and interconversion between these states typically occurs via the \(\beta\) state intermediate.

Consistent with the macro simulation, CG simulations reveal distinct RAS lipid fingerprints, which exhibit spatially complex patterns of lipid enrichment and depletion around the G-domain, HVR, and farnesyl anchor (Fig. S36, S2.6). A variety of protein-lipid interactions contribute to the modulation of G-domain orientation by PM lipids, and distinct lipid fingerprints observed for the different RAS orientational states (Fig. 5I) are specific enough that a convolutional neural network can predict RAS states with an accuracy of 82\% based on inner leaflet spatial lipid distributions (Fig. 5J, S2.6.2). Among the complex lipid signatures, regression analysis shows that enrichment/depletion of different lipids species affects the likelihood of transitions between states (Fig. 5K, S2.6.3). For example, changes in the number of PIP2 lipids interacting with the G-domain emerge as a good predictor of transitions between \(\beta\) and \(\beta'\) states, an orientational change that can modulate effector occlusion (Fig. 5B).

Significantly, occlusion of the RAF-binding interface by the membrane is relieved by PIP2 at molar concentrations above 2\% (Fig. 5M). The kinetic barrier between \(\beta\) and \(\beta'\) states is related to a specific PIP2 binding site at the junction between the G-domain and the HVR, near H166 (Fig. 5I). The exchange between metastable \(\beta\) and \(\beta'\) states (Figs. 5B, 5H), appears to be controlled by reversible membrane adhesion of this region of the HVR (Fig. 5L), which must disengage to permit extensive G-domain tilting and transition from \(\beta\) to \(\beta'\) (Fig. 5N). In this work, we identify molecular details by which lipid-RAS interactions regulate occlusion, revealing a plausible strategy for inhibition of RAS function.
Discussion

The unique capabilities of MuMMI to operate and provide insights at two vastly different temporal and spatial scales in a coordinated manner represent a fundamentally new technology in computational biology. Furthermore, this study highlights the necessity of such a capability by demonstrating the importance of considering membrane proteins and lipids collectively, as a single unit with two highly interdependent components. Only through the combination of the macro scale sampling of diverse local lipid environments with the micro scale dynamics at the molecular level can we observe these distinct RAS-lipid interaction patterns, which were subsequently used to design the supporting experiments.

The combination of simulation and experiment reveal that RAS not only remodels its local lipid environment, but that it is itself regulated by these distinct lipid patterns. For instance, an increased concentration of GTP-loaded KRAS4b is associated with PIP2 enrichment and cholesterol depletion, while PIP2 drives functionally relevant orientational changes in RAS. The differential influence of lipid fingerprints on interactions of RAS with regulators such as SOS(41) and effectors such as RAF, PI3K, and RALGDS are unknown, but there is clear potential for lipid patterning to define recruitment sites for discrete effectors and thereby segregate RAS-mediated signal transduction pathways.

The current work finds RAS-RAS interactions to be interfacially nonspecific and validates all previously proposed interfaces(11, 21, 22, 24, 34) as part of a broad ensemble of possible interactions. This interfacial promiscuity suggests that RAS multimer formation is mediated by lipids rather than specific interfacial contacts and leads us to conclude that non-enzymatically-relevant mutations, some of which impact tumor growth(11), may impact the dynamics of RAS multimer formation through changes in protein-lipid interactions.

MuMMI uses cohesive multiscale coupling, which in this application reveals both the broad scope and fine details of membrane remodeling that underlies functionally relevant RAS-lipid dynamics. More importantly, MuMMI has been designed to be generalizable, and as such opens new avenues for potential research and expands the range of questions that can be addressed using biomolecular simulation.

Declarations

Acknowledgments.

This work was performed under the auspices of the U.S. Department of Energy (DOE) by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, Los Alamos National Laboratory (LANL) under Contract DE-AC5206NA25396, Oak Ridge National Laboratory under Contract DE-AC05-00OR22725, Argonne National Laboratory under Contract DE-AC02-06-CH11357, and FNLCR under Contract HHSN261200800001E. Release: LA-UR-19-32061, LLNL-JRNL-799963. We thank Siewert-Jan Marrink and Carsten F.E. Schroer for advice on Martini and GTP parameters, and Matt Drew, Shelley Perkins and Vanessa Wall from the National Cancer Institute (NCI) RAS Initiative, Jeff Clogston from Nanotechnology Characterization Laboratory, and Will Heinz from Optical Microscopy Analysis.
Laboratory, Frederick National Laboratory for Cancer Research (FNLCR), for cloning, expression and protein purification.

**Funding:**

This work has been supported in part by the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program established by the U.S. DOE and the NCI of the National Institutes of Health. We thank the Livermore Institutional Grand Challenge and LANL Institutional computing for computing time. The LANL Institutional Computing Program is supported by the U.S. DOE National Nuclear Security Administration under Contract No. DE-AC52-06NA25396.

**Author contributions:**


**Competing interests:**

F.M. is a consultant for AduroBiotech, Amgen, Daiichi Ltd., PellePharm, Pfizer, Inc., PMV Pharma, and Portola Pharmaceuticals. F.M. is a consultant for and cofounder of Avidity, BridgeBio, KGen, and Quartz.

**Data and materials availability:**

All macro and micro simulation input and parameter files with an example micro simulation are available at https://bbs.llnl.gov/xxx.html. Simulation data will be made available upon reasonable requests, size of all raw data is hundreds of TB.

**Code availability:**

MuMMI is composed of numerus sub-components both freely available 3rd party applications and custom codes; of those ddcMD (github.com/LLNL/ddcMD), ddcMDconverter (github.com/LLNL/ddcMDconverter), Maestro (github.com/LLNL/maestrowf), and Flux (github.com/flux-framework) are already available open source.
References


27. H. I. Ingólfsson et al., Capturing biologically complex tissue-specific membranes at different levels of compositional complexity. *Journal of Physical Chemistry B (Under review)*, (2020).


**Figures**

**Figure 1**

Experimental input and computational approach. (A) Diffusion mapping of single molecules of KRAS4b tethered to or within 100 nm of the PM in a 16×16 μm² region of a live HeLa cell accumulated over 10 s. (B) Crystal structures of wild-type KRAS in active (green and blue; GppNHp-bound) and inactive (grey; GppCH2p-bound configurations). (C) Average macro model lipid composition. (D-F) The Multiscale Machine-learned Modeling Infrastructure (MuMMI). (D) Representative snapshots of each of the different lipid distributions in the inner leaflet of a 0.3×0.3 μm² region of the full 1×1 μm² macro simulation; color saturation indicates local lipid density. (E) Schematic illustrating latent space encoding of lipid composition in 30×30 nm² membrane patches. From the candidate patches (blue and green), those that are most dissimilar (green) to existing (white) CG simulations are selected and used to spawn new CG
simulations. (F) Representative CG simulation systems (water not shown). (G) Improvement of macro model parameter inputs from feedback iteration. (H) Distribution of CG simulation duration and (inset) number of RAS per patch.

Figure 2

Lipid-dependence of RAS co-localization. (A) Representative macro model inner leaflet lipid densities around RAS, shown separately (small boxes) and together (large central box). Color saturation indicates local lipid density. (B) Population ratio of RAS multimer sizes observed in the macro simulation vs. a random uniform distribution. (C) Average number of RAS in macro model regions (radius 5 nm) along the primary embedding dimension (PED) from FPP analysis (S2.8.2). Vertical lines denote thresholds used to define high-RAS (HRC), initial average (ARC), and low-RAS (LRC) lipid compositions. (D) Distributions of inner leaflet lipid concentrations for all patches with RAS (black), HRC (red), and LRC (blue). The ARC is represented by a dashed vertical line. (E) SPR sensorgrams of RAS adhesion to liposomes with the LRC, ARC and HRC lipid compositions. Each subplot contains multiple traces representing distinct RAS concentrations (two-fold dilutions, 60-0.1 μM).
Figure 3

Diffusion of RAS and lipids. (A, B) Distributions of lateral diffusion rates, $D$, for lipids and RAS in (A) CG simulations and (B) FLCS on the ARC. (C) Values of $D$ for PIP2 conditioned on interaction with monomeric RAS. (D) Proportion of each lipid type in RAS’s first solvation shell (within 0.55 nm), normalized by its molar fraction in the ARC, as a function of RAS aggregate size. (E) Values of $D$ as a function of RAS aggregate size. (F) Mean square displacement (MSD) curves of JF646-labeled RAS from SPT on supported lipid bilayers with the LRC, ARC, and HRC lipid compositions.
Figure 4

RAS-RAS interactions. (A) Distributions of radii of gyration based on RAS aggregation number. (B) Angular arrangement of the (left) HVR and (right) G-domain with respect to the G-domain of another RAS molecule for various minimum intermolecular distances, \(d_{\text{min}}\). (C) Spatial density of one RAS protein around another RAS protein from all CG simulations of two RAS with initial \(d_{\text{min}} > 4.5\) nm (greyscale heatmap), and the average COM→T35 orientation vector of the non-central RAS (colored arrows). For
clarity, arrows are only shown for bins with >10 data points, and where the vector length is >50% of the maximum observed length. (D) PIP2 remodeling based on RAS-RAS G-domain separation, dG-G, shown with G-domain centers of mass on the x-axis. Data show PIP2 density (red heatmaps), PIP2 density integrated over -4 nm < y < 4 nm (black lines), and a model reflecting translation of static PIP2 density distributions (blue dashed lines). Differences between the integrated density and the static translation model indicates regions of enrichment and depletion of PIP2 lipids during RAS-RAS association. (E) Preferential binding coefficients showing the enrichment of PIP2 among lipids in contact with RAS, δPIP2, as a function of the number of PIP2 per CG patch, shown separately for RAS dimers and monomers. The larger δPIP2 values for dimers indicates that dimers formation is favored by higher PIP2 concentration.

Figure 5

RAS orientation, competence for effector binding, and lipid interactions. (A) Definition of tilt and rotation angles based on a reference orientation with the long axis of G-domain helix 5 perpendicular to the membrane surface. (B) Kinetic states of G-domain orientation. Dots represent MSM microstates and color associates sampled orientations with the nearest microstate. Blue isocontours define the likelihood of membrane-based occlusion of RAF binding. (C-F) Representative configurations of α, β, and β' states illustrating orientation-dependent competence for effector binding. (G) G-domain disposition in simulations with one RAS. (H) MSM state populations and transition rates. (I) Relative lipid densities around monomeric RAS. The center of mass of the G-domain is at the origin and the backbone bead of C-
terminal HVR residue C185 is on the positive x-axis. (J) Confusion matrix showing the accuracy at which a neural network predicts RAS state based on local lipid composition. (K) Local lipid composition is predictive for transitions between states. Transition probabilities increase when noted lipid concentrations (+) increase and (-) decrease (see S2.6.3). (L) State-specific distributions of the displacement of the H166 backbone bead from the bilayer center, dzH166. (M) RAF-occlusion vs. PIP2 content. (N) G-domain tilt vs. dzH166.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.pdf](#)