OpenPathSampling: A flexible Python framework for rare events

David W.H. Swenson, van 't Hoff Institute for Molecular Sciences, University of Amsterdam

Rare Events

When metastable states are separated by a large free energy barrier, the transition between them is a **rare event**. Properties such as the mechanism and rate of the transition require significant sampling of the transition region. However, the Boltzmann distribution means that the vast majority of time is spent in the metastable states, not the transition region.

Path Sampling

Direct molecular dynamics is, at best, an inefficient way to study a rare event; at worst, intractable. **Path sampling** methods perform a Monte Carlo simulation in the space of paths (trajectories), enabling efficient simulation of rare events by focusing the simulation effort on the transition region and reducing the simulation inside the metastable states.

OpenPathSampling

OpenPathSampling is a recently-developed software package for studying rare events. It includes many tools for performing and analyzing path sampling simulations, as well as tools from other trajectory-based approaches to rare events.

http://openpathsampling.org
http://github.com/openpathsampling/openpathsampling
Twitter: @pathsampling

Transition Path Sampling

TPS is a path sampling approach that uses trajectories that connect the metastable states. While some TPS-based approaches can extract kinetic information such as rates, TPS is best-suited for studying the mechanisms of complex systems.

Many tools for analyzing mechanisms are based on initial sampling using a variant of TPS.

OpenPathSampling supports several variants of TPS:

- fixed length
- flexible length
- multiple state

OPS supports several variants of the shooting move (for both TPS and TIS):

- one-way shooting
- two-way shooting
- uniform or biased shooting point selection

Dellago et al. J. Chem. Phys. **108** 1964 (1998).

Transition Interface Sampling

TIS is a path sampling method that is specifically designed to calculate rates. In TIS, the space is foliated by several "interfaces" along an approximate reaction coordinate. The rate is determined based on the successive crossing probabilities of ensembles defined by the interfaces:

$$k_{AB} = \phi_{A_0} \prod_{i=0}^{m-1} P_A(\lambda_{i+1}|\lambda_i) \; P_A(B|\lambda_m)$$

flux   crossing prob.   conditional transition prob.

TIS is forgiving with respect to the proposed order parameter. In addition, path from TIS can be reweighted to obtain free energy landscapes in any collective variables of interest, without re-running the sampling.

OpenPathSampling supports several variants of TIS:

- replica exchange
- multiple state
- multiple interface set
- single replica

van Erp, Moroni, Bolhuis. J. Chem. Phys. **118** 7762 (2003).

Committor Analysis

The committor for a given configuration is the probability that a trajectory launched from there (with Boltzmann-distributed velocities) will land in the "product" state. It is an excellent approximation for the reaction coordinate.

OpenPathSampling includes tools to efficiently calculate the committor. Using OPS's monitoring function, the OPS committor simulation ensures that the trajectory stops when a trajectory enters a state, thus avoiding wasted computing effort.

External contributors are building modules based on the committor simulation to add calculations such as reactive flux (Bennet-Chandler) calculation of the rate.

Direct Simulation

OpenPathSampling can also manage direct molecular dynamics simulations. This is particularly useful for using OPS calculate the flux (or, in some cases, the rate) for a process, since the OPS state definitions can be re-used across simulations and are uniquely identified (tracking provenance). The direct simulation module has the option of not saving snapshots, which can be useful for very long simulations.

Analysis tools

In addition to sampling rare events, OPS includes many tools for analyzing rare events, including path density plots, analysis of fluxes, crossing probabilities, and rates from TIS, analysis of replica exchange behavior (trip times, replica flow, mixing matrix) for replica exchange TIS.

It also includes tools to facilitate creation of new state definitions based on user-annotated trajectories, and tools to identify trajectories that follow proposed mechanisms or classify trajectories according to a specified reaction channel. It is easy to add more analysis tools, and external contributors have already created some, such as reaction coordinate analysis using maximum likelihood analysis.

OPS interfaces with other codes

OpenPathSampling is designed in interface with existing software packages and libraries, in order to provide familiar functionality to users. This includes general scientific software as well as tools for specifically for molecular dynamics.
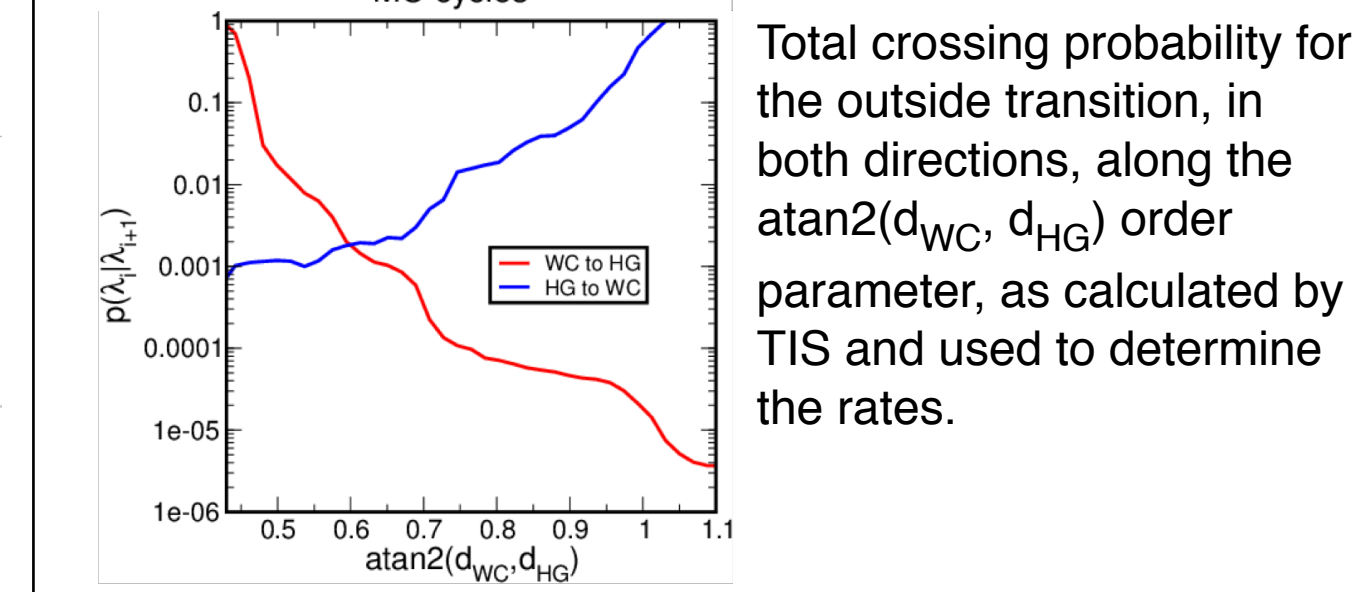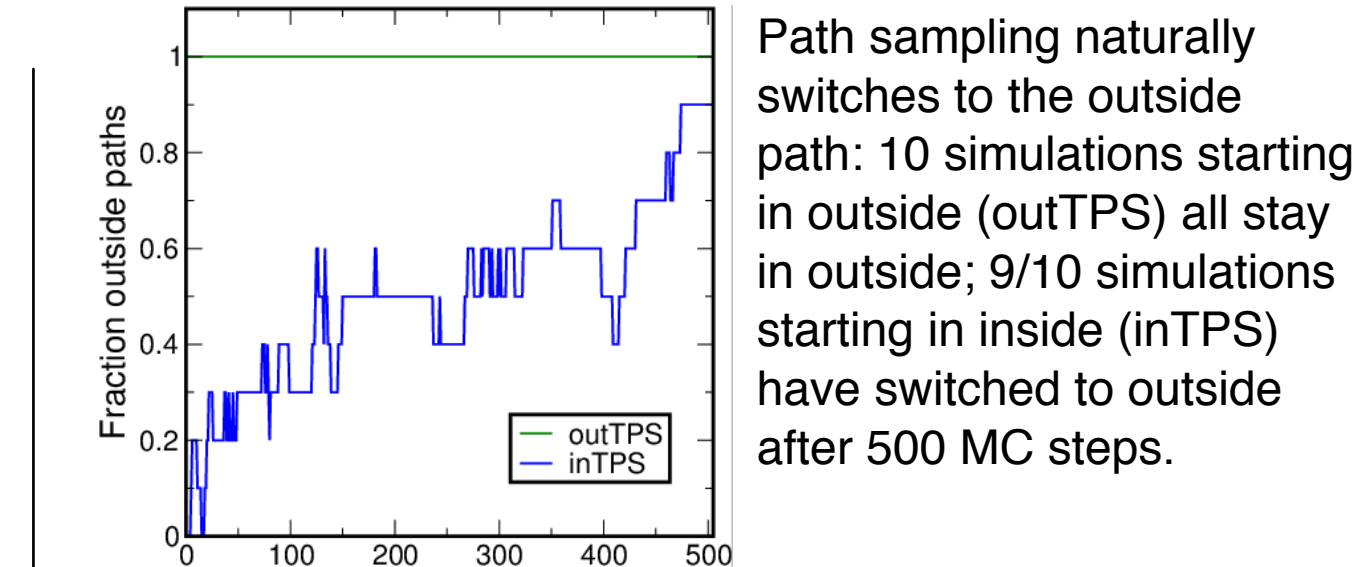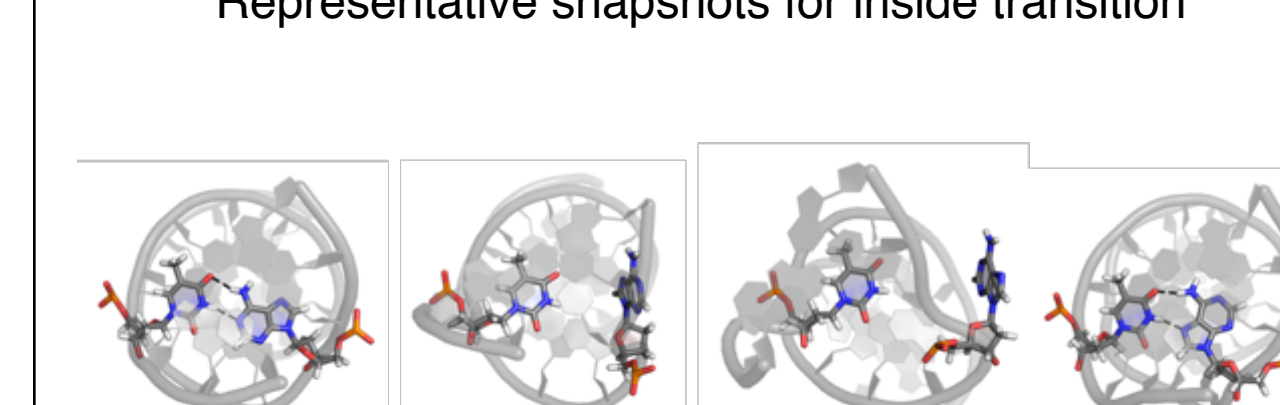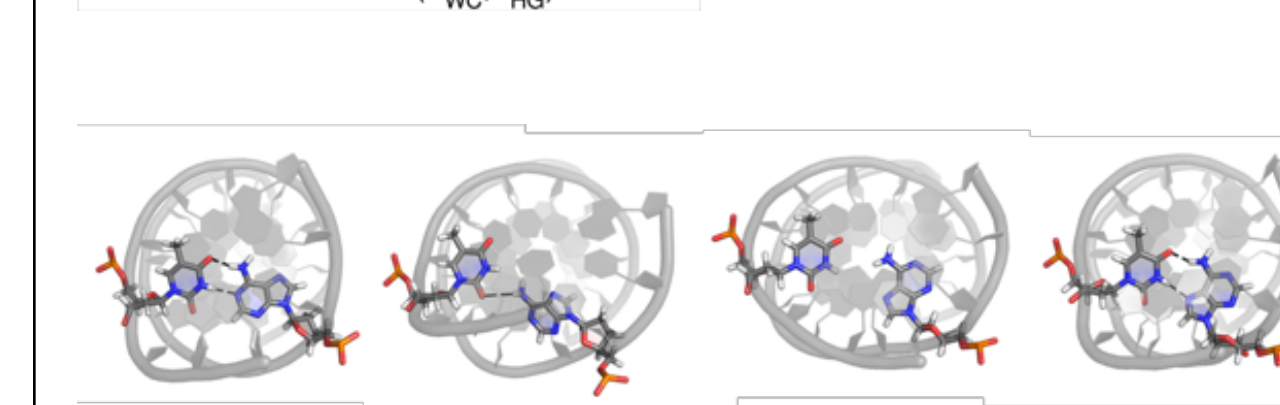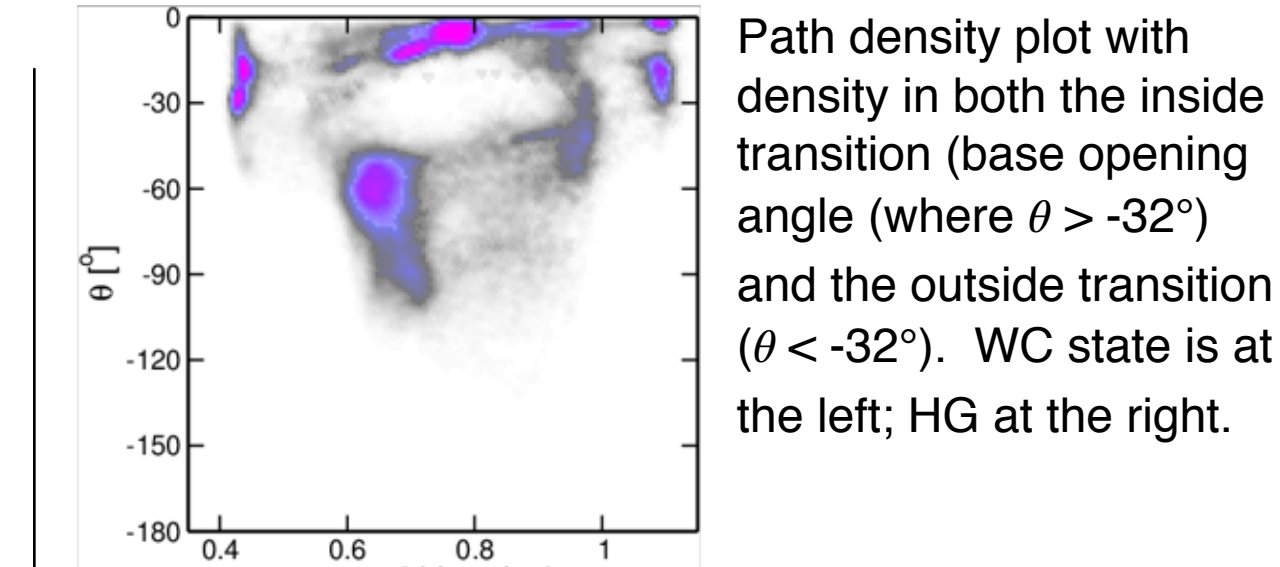

Watson-Crick to Hoogsteen Conversion in DNA

In addition to the well-known Watson-Crick (WC) motif, DNA at physiological conditions can be found in the Hoogsteen (HG) motif, where the purine is flipped 180° relative to Watson-Crick. This motif plays a role in some DNA replication processes, and may be important in many other biological processes.

Watson-Crick   Hoogsteen
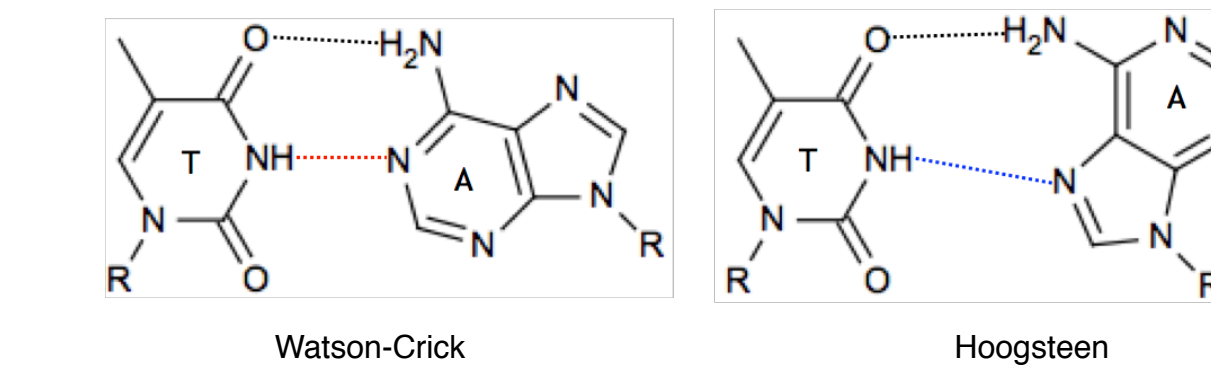
In TPS simulations, we observe multiple mechanisms (channels) for this transition. In particular, we distinguish between an "inside" mechanism, where the purine rotates while inside the double helix structure, and an "outside" mechanism, where the purine exits the helical structure (base flipping) before rotating and re-entering the helix. Previous work had emphasized the importance of the inside transition, but we see evidence that the outside transition is preferred.

The characteristic hydrogen bond distances d_{WC} and d_{HG} are illustrated in the figure. We use atan2(d_{WC}, d_{HG}) as an approximate reaction coordinate (order parameter) for TIS.


Path density plot with density in both the inside transition (base opening angle (where θ > -32°) and the outside transition (θ < -32°). WC state is at the left; HG at the right.

Representative snapshots for inside transition

Representative snapshots for outside transition

Path sampling naturally switches to the outside path: 10 simulations starting in outside transition (outTPS) all stay in outside; 9/10 simulations starting in inside (inTPS) have switched to outside after 500 MC steps.

Total crossing probability for the outside transition, in both directions, along the atan2(d_{WC}, d_{HG}) order parameter, as calculated by TIS and used to determine the rates.

	expt	TIS
$k_{WC \to HG}$ (s^{-1})	14.2 ± 1.03	742
$k_{HG \to WC}$ (s^{-1})	3670 ± 200	1.6·10^5
ΔG (k$_B$T)	5.5	5.4

Comparing rates/free energy differences from TIS and from experiment. Rates are overestimated, but free energies are very good.

Vreede, Bolhuis, DWHS. In prep.

Differences in the dynamics of the oncogenic K-Ras mutant Q61L

Mutations in the Ras protein family play a role in 15% of all cancers. The GTPase K-Ras is a member of this family, and 85% of oncogenic Ras mutations are found in K-Ras. Experiments suggest that GTP-bound K-Ras exists in two substates: an inactive state 1, in which two loop regions are flexible, and an active (and less flexible) state 2. One of the oncogenic mutations, Q61L, involves a mutation in the switch 2 region. This work asks the question: **How does a mutation (such as Q61L in K-Ras) change the fundamental dynamics of a protein in a way that can lead to cancer?**


state 2   state 1
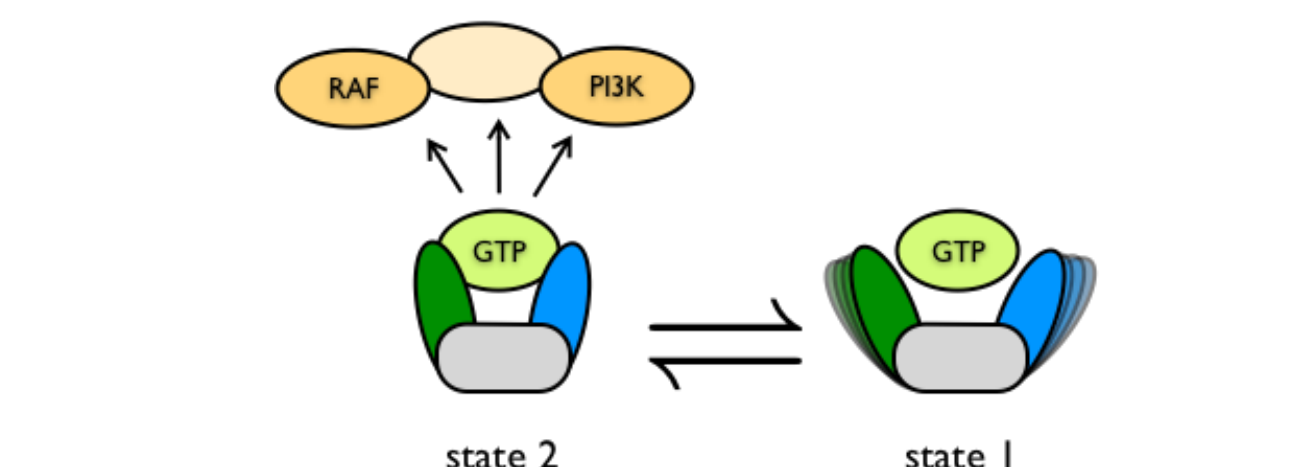
The K-Ras mutant Q61L, with the switch 1, switch 2, and alpha helix 3 regions highlighted. The residue 61L is drawn in licorice, as is α3 and the GTP.

Path sampling identified 3 metastable states for switch 2 in GTP-bound K-Ras: with the switch bound to the GTP, bound to the α3, or unbound.
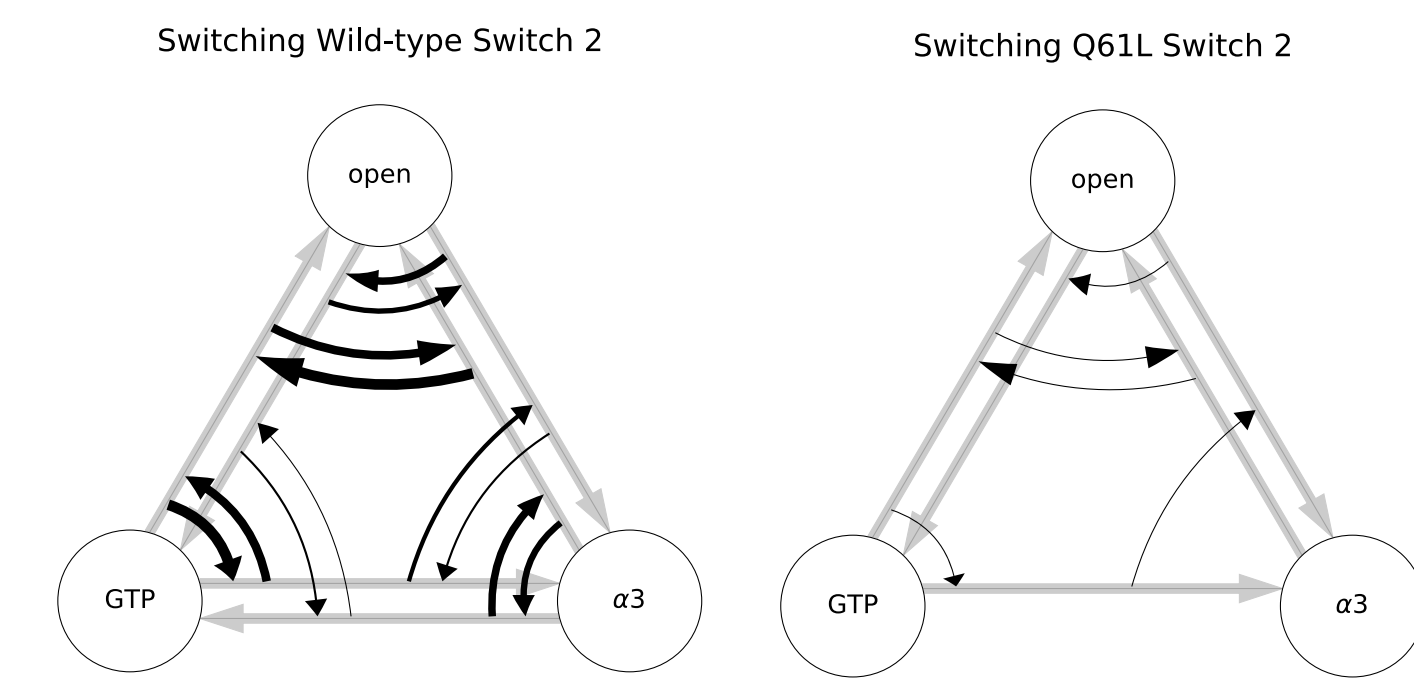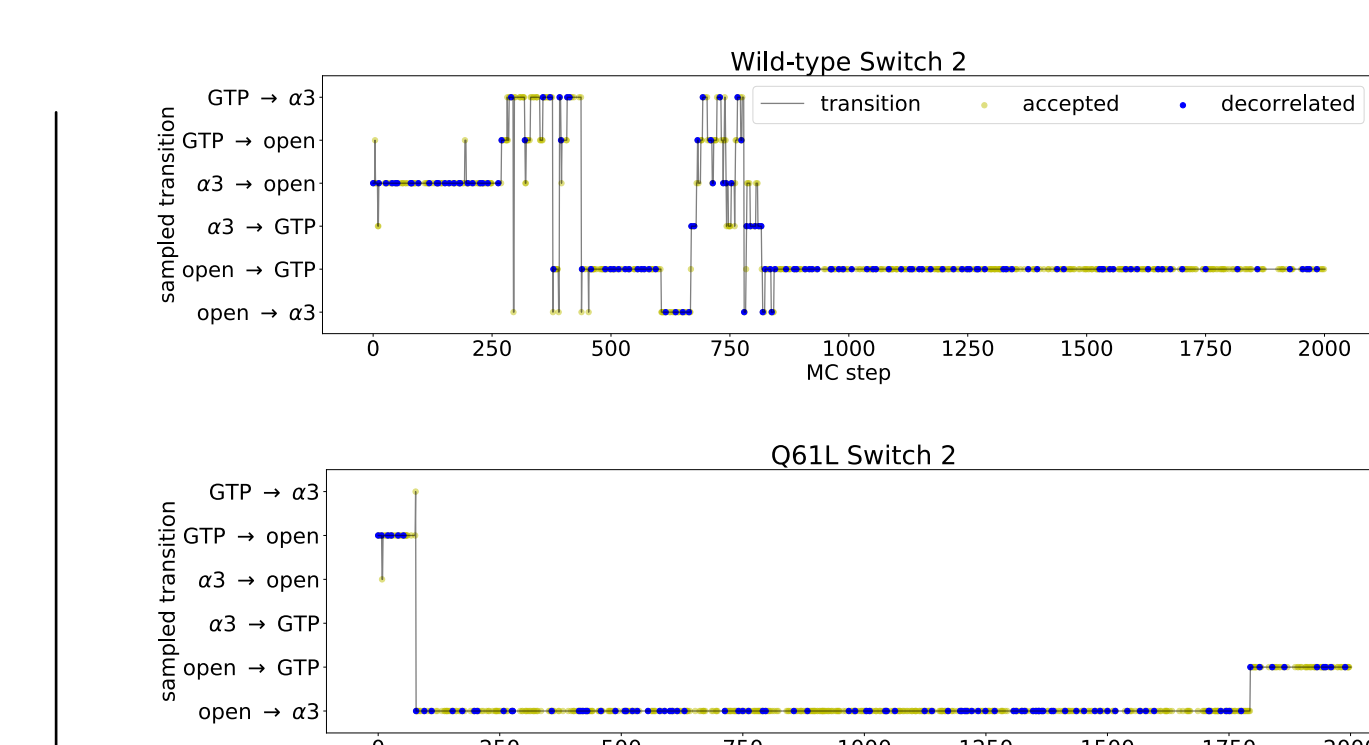
Transition switching in the TPS simulations of wild type and Q61L K-Ras. The three substates lead to six possible transitions. In a multiple state TPS simulation, the transition being sampled can change as a function of the Monte Carlo step. Yellow dots mark accepted trajectories; blue dots mark successive decorrelated trajectories (no frames in common.)

Another view of transition switching in the TPS simulations. Circles indicate the metastable states, grey arrows indicate the sampled transitions, and curved black arrows indicate the number of switches between transitions. The width of the curved black arrows is proportional to the number of switches. There is much more switching in the wild type than in the mutant.
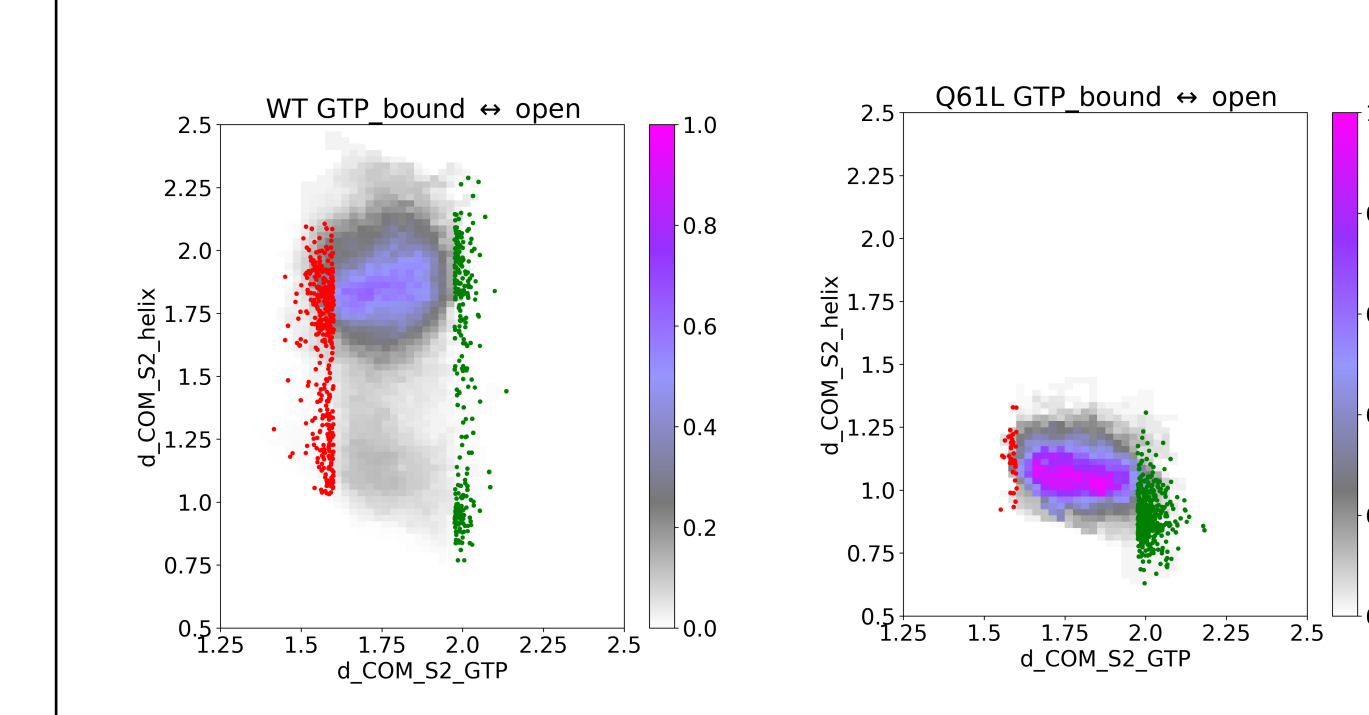
Path density plots for the wild type and Q61L mutant, for paths making the transition from GTP-bound to unbound (open) or vice versa. The horizontal axis measures the distance between switch 2 and GTP, which separates the states (bound shown in red dots to the left; unbound in green dots to the right). The vertical axis measures the distance from switch 2 to the α3 helix. In the wild type, two channels are sampled: one with a low value of switch 2-α3 distance, and one with a high value. Only the low value is sampled in the mutant. The low distance to α3 corresponds to sliding along the helix to unbind; the high distance corresponds to direct solvation of the switch region.

The Q61L mutation makes direct solvation unlikely, leading to more time in the closed, active "state 2," causing abnormal cell growth and proliferation.

Roet, Hooft, Bolhuis, DWHS, Vreede. In prep.