

Activation Pathways of Neotensin Receptor 1 Elucidated Using Statistical Machine Learning

Prakarsh Yadav and Amir Barati Farimani*

Cite This: *ACS Chem. Neurosci.* 2022, 13, 1333–1341

Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Neotensin receptor 1 (NTSR1) is a G-protein coupled receptor (GPCR) that mediates many biological processes through its interaction with the neotensin (NTS) peptide. The NTSR1 protein is a clinically significant target as it is involved in the proliferation of cancer cells. Understanding the activation mechanism of NTSR1 is an important prerequisite for exploring the therapeutic potential of targeting NTSR1 and the development of drug molecules specific to NTSR1. Previous studies have been aimed at elucidating the structure of NTSR1 in the active and inactive conformations; however, the intermediate molecular pathway for NTSR1 activation dynamics is largely unknown. In this study, we performed extensive molecular dynamics (MD) simulations of the NTSR1 protein and analyzed its kinetic conformational changes to determine the microswitches that drive NTSR1 activation. To biophysically interpret the high-dimensional simulation trajectories, we used Markov state models and machine learning to elucidate the important and detailed conformational changes in NTSR1. Through the analysis of identified microswitches, we propose a mechanistic pathway for NTSR1 activation.

KEYWORDS: *neotensin receptor, machine learning, conformational changes, activation mechanism*

INTRODUCTION

Neotensin (NTS) is an oligopeptide, which plays an important role in the nervous system and the regulation of digestive processes in the gut.^{1,2} NTS has high clinical and physiological significance as it is involved in pathogenesis of diseases such as schizophrenia and Parkinson's disease and proliferation of cancer cells.^{3–5} In addition, NTS is involved in the regulation of hypothermia, antinociception, and dopamine neurotransmission.^{6–9} The NTS peptide binds to the neotensin receptor (NTSR), which is a G-protein coupled receptor (GPCR). Among the three types of NTSRs, NTSR1 belongs to the class A GPCR family, and it is the main target for NTS. NTSR1 is activated by the binding of NTS, which acts as its agonist (Figure 1 and Supporting Information Movie 1). The activated NTSR1 becomes phosphorylated to facilitate signal transduction via the G-protein, and later, NTSR1 binds to β -arrestin to become desensitized.^{12,13} The understanding of the molecular mechanism of NTSR1 activation is crucial for rational drug design to target NTSR1 for therapeutic purposes. Toward this goal, many research groups have focused their efforts to determine the structure of NTSR1 at an atomic level resolution.^{14–19} The protein structures of NTSR1 have been determined using crystallography, which provide the snapshots of the receptor in the active and inactive conformations.^{15,16} These static states are not sufficient to identify the key conformational changes that determine the transition between

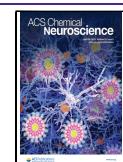
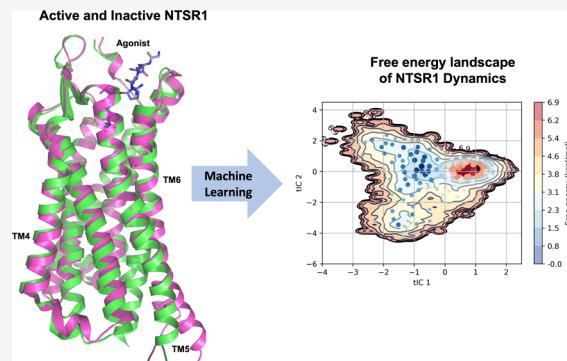
active and inactive states.²⁰ Molecular dynamics (MD) simulations can elucidate the temporal motion of atoms at a high resolution, which is essential for understanding the GPCR dynamics.²¹ MD simulations start with a static structure and enable the study of the motion of every atom in the protein and how they interact with the agonist.²²

It is necessary to identify the intermediate states—important residue conformations—along the activation pathway. Characterization of these residues and the associated conformational changes is key in small-molecule design and in achieving the desired activation pathway. However, identification of these microswitches is very time-consuming and expensive using experimental methods, such as mutagenesis. MD simulations generate trajectories based on the physical laws for all atoms, resulting in detailed dynamics of proteins and the protein conformations. Since the simulations are conducted at a femtosecond resolution, a large volume of high-dimensional data is generated, and this large volume of data needs to be analyzed to characterize the kinetics and conformational

Received: March 8, 2022

Accepted: March 24, 2022

Published: April 5, 2022



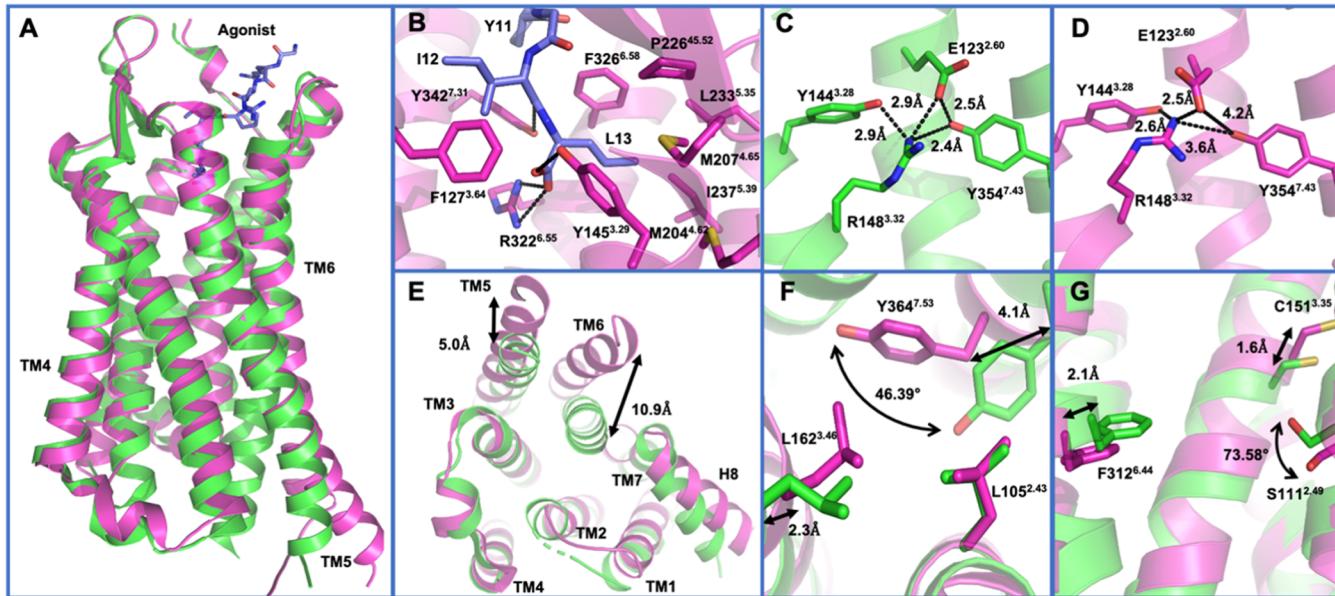


Figure 1. (A) Side view of PDB structures of NTSR1 in active (pink; PDB: 6UP7) and inactive (green; PDB: 3ZEV) conformations with the agonist (blue; NTS_{8–13}). (B) Network of electrostatic and hydrophobic interactions that stabilize I12 and L13 of the agonist. (C) Four-residue network of E123^{2.60}–Y144^{3.28}–R148^{3.32}–Y354^{7.43} in the inactive state. (D) Four-residue network of E123^{2.60}–Y144^{3.28}–R148^{3.32}–Y354^{7.43} in the active state. (E) Cytoplasmic view of NTSR1 in active (pink; PDB: 6UP7) and inactive (green; PDB: 3ZEV) conformations with the agonist (blue; NTS_{8–13}). TM5 and TM6 have a large outward displacement and adopt a conformation similar to the canonical activation of NTSR1. The displacement of TM5 and TM6 is shown as black arrows. (F) Differences in the conformation of Y364^{7.53} of the NPxxY region and neighboring residues (L105^{2.43} and L162^{3.46}). (G) Differences in the conformation of key residues in GPCR dynamics (S111^{2.49}, C151^{3.35}, and F312^{6.44}).

changes of the protein. The analysis of such a volume of data is not feasible by visualization of simulation trajectories; therefore, unsupervised ML techniques such as dimensionality reduction will be required. The commonly used dimensionality reduction method, principal component analysis (PCA), attempts to capture the maximum variance in the data, but the maximum variance axes may not necessarily contain the slowest processes. The identification of such processes and microswitches is required to detect the kinetically stable intermediate conformations of the protein. To overcome the limitation of PCA to identify the slowest processes and the microswitches from simulation data, machine learning (ML) methods can be used to reduce the dimensions of the data to identify key conformational changes in the intermediate states along the activation pathway. Time-structure independent components analysis (tICA) is a method for finding the slowest relaxing degrees of freedom in time series data, such as simulation trajectories, which can be formed from linear combinations from a set of input degrees of freedom.^{23,24}

To determine the activation pathway and identify the conformational changes, we combined MD simulations with unsupervised ML (Figure 1). In this work, we have simulated the active and inactive conformations of the NTSR1 protein in the presence and absence of the NTS peptide (Figure 1A). By using dimensionality reduction techniques, we analyzed the simulation trajectories and identified the important contributing features (important residues in the activation pathway). Finally, after analyzing the motion of the most important residues (determined through tICA), we propose a sequence of important conformational changes (determined through transfer entropy) that govern one of the parallel paths of the NTSR1 activation mechanism.

■ CURRENT KNOWLEDGE ABOUT THE NTSR1 STRUCTURE

The binding of NTS causes conformational changes in NTSR1, leading to the characteristic outward movement of TM5 (5.0 Å) and TM6 (10.9 Å, Figure 1E). The outward movement of TM5 and TM6 is the activation signature of class A GPCRs.²⁵ The active and inactive conformations of NTSR1 also result in differences in the conformation of individual residues.^{15,16} The hydrophobic amino acids of NTS (I12 and L13) are stabilized by interactions with the hydrophobic and aromatic residues of NTSR1 (Figure 1B).^{15,18} A four-residue electrostatic interaction network (E123^{2.60}–Y144^{3.28}–R148^{3.32}–Y354^{7.43}) is present in both the active (pink) and inactive (green) conformations (Figure 1C,D).^{16,18} In addition, Y364^{7.53} of the NPxxY region, in TM7, shows a translation of 4.1 Å (Ca distance) and a rotation of 46.39° (angle between Ca and hydroxyl) toward TM3 in the active conformation (Figure 1F).²⁵ These conformational changes are the notable features of the NTSR1 activation pathway.

■ METHODS

The MD simulations for the NTSR1 protein were performed using NAMD.²⁶ The simulations setup consisted of the receptor protein (inactive state: PDB 4GRV¹⁶ and active state: PDB 6UP7¹⁸), lipid bilayer (POPC), water (TIP3P), and ions (NaCl). To obtain a charge neutral system, the counter ions were added to simulation box, and the system molarity was set to 0.15 M (Figures S1 and S2). The POPC membrane (80 Å × 80 Å) was aligned such that it was parallel to the XY plane and was padded with a solvent box of size 25 Å on the Z axis. The receptor protein was aligned to the center of the POPC membrane. The structure of the active and inactive states of the protein contained the NTS peptide. To simulate the apo protein, two additional systems were created (active and inactive states) without the NTS peptide, with the same POPC membrane and solvent box. The simulation system was equilibrated in three steps. First, the POPC molecules were equilibrated

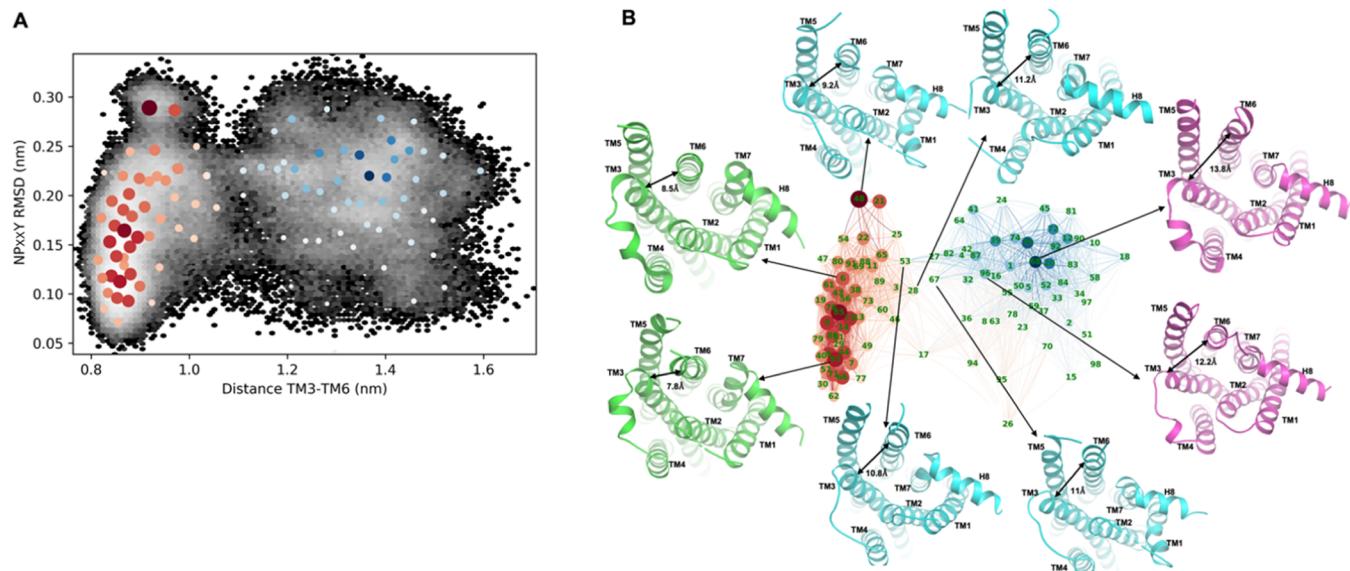


Figure 2. (A) Density projection with the overlaid MSM of the TM3–TM6 distance vs the NPxxY region rmsd. The circles indicate microstate centers, the size of the circle shows the population size, and the slowest processes are shown as transitions from red states to blue states. The TM3–TM6 distance axis shows the presence of two clusters, with active states having a higher TM3–TM6 distance (~ 1.4 nm) and inactive states having a smaller TM3–TM6 distance (~ 0.9 nm). (B) Network representing the Markov states with the NPxxY vs TM3–TM6 distance. Each node represents a Markov state, and each connecting edge is the transition probability between the corresponding nodes. The structures are the representative states of the corresponding population of Markov states (inactive conformation: green, intermediate states: cyan, and active conformation: pink). The graph nodes and edges are colored such that the slowest processes are shown as transitions from red states to blue states.

for 100 ps, while constraining the solvent and protein. Second, the solvent molecules were equilibrated, while constraining the POPC molecules and protein, for 100 ps. Finally, the protein molecule was equilibrated, while constraining the POPC molecules and the solvent, for 100 ps. These steps of equilibration were performed for 100,000 steps with a timestep of 1 fs. The system was then equilibrated under the NPT ensemble for 1 ns at 1 atm pressure and 310 K temperature. The Nosé–Hoover Langevin piston pressure control was used with a target pressure of 1 atm, an oscillation period of 200 fs, and a damping time scale of 50 fs.²⁷ The Langevin thermostat was used with the damping coefficient of 1 ps⁻¹. The system was then simulated under the NPT ensemble for 200 ns with a time step of 1 fs, while collecting and saving the data every 50,000 steps (50 ps).

In total, 68 simulations were run with a total of 16 μ s of trajectories (Table S5). The simulations consisted of four different conformations, active, active with the agonist, inactive, and inactive with the agonist. Each of the conformations served as an initial structure for the simulation of NTSR1 and were initialized with different velocities. The simulations were carried out in five generations to ensure sufficient sampling of the conformational landscape (Figure S6). The aggregated trajectories from generation 5 simulations were used for subsequent data analysis. The analysis was performed using a custom Python code, which is based on the MSMBuilder²⁸ and MDTraj²⁹ python libraries. Necessary modifications to the libraries were carried out to extract features from the trajectories [root mean square deviation (rmsd), dihedral angles, and contact distances]. The features were extracted corresponding only to the residues composed of the transmembrane region of the protein (Figure S3). To compute the contact distances between residues, a neighbor list was defined such that the closest heavy atom distance was computed with all residues within a 6 Å distance of either crystal structure to account for all significant nonbonded interactions. Dihedral angles were calculated for the receptor at every timestep by calculating the Φ and Ψ angles between the adjacent residues of the transmembrane region. To determine the reaction coordinates of the protein, the sparse-tICA algorithm (Figures S4 and S5) was applied to the extracted features (dihedral angles, lag = 25 ns, and contact distances, lag = 75 ns) of the TM region (Figure S3). We selected the sparse-tICA algorithm over normal tICA because it generates a sparse, interpretable, and regularized representation of the

data (Figures S9 and S10). We used a random forest (RF) regressor³⁰ to calculate the contribution of individual residual features for the calculation of the low-dimension projection of simulation data. We then ranked the learnt feature importance to determine the highest contributing residues. Then, a k -means model with 100 clusters ($k = 100$)³¹ was trained on the first two tICs of the tICA projection to identify the clusters in the projection. A Markov state model (MSM) was constructed with a lag time of 25 ns to determine the slowest processes.³² Transfer entropy between the MSM population centers for contact distances was calculated using the MDEntropy python library.³³ Transfer entropy between features X and Y ($T_{X \rightarrow Y}$) indicates the reduction in uncertainty of Y when conditioned on previous values of X . Higher values of transfer entropy from X to Y ($T_{X \rightarrow Y}$) than from Y to X ($T_{Y \rightarrow X}$) indicates that event X is more likely to occur before event Y .³⁴ Transfer entropy can be calculated from Shannon entropy $H(X)$ using the relation

$$T_{X \rightarrow Y} = H(Y_t | Y_{t-1:t-L}) - H(Y_t | Y_{t-1:t-L}, X_{t-1:t-L})$$

RESULTS AND DISCUSSION

The simulation data for large transmembrane proteins, such as NTSR1, is high-dimensional, which poses a great challenge for identification of the pharmacologically and physiologically significant conformational changes in the protein dynamics. It is necessary to decorrelate the slowest processes in protein dynamics from the long-timescale MD simulations to identify the microswitches and their corresponding conformational changes in NTSR1 dynamics. Toward this objective, we used the time-lagged independent components analysis (tICA) to find linear combinations of the features extracted from MD simulations and identify the dominant components (tIC1 and tIC2). These tICs can be used to generate a two-dimensional projection of the simulation features (for example: TM3–TM6 distance vs NPxxY region rmsd, dihedral angles, and contact distances) and construct the density projection for them.³⁵ By using the density projection of features from our tICA and the

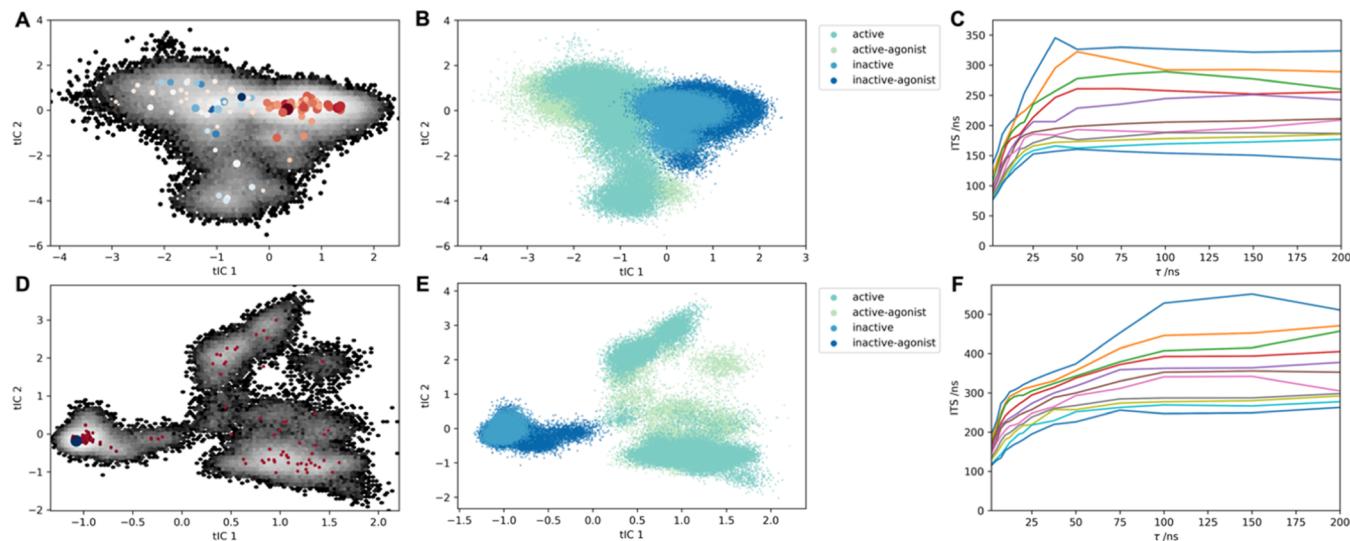


Figure 3. MSM overlaid on the density projection of the first two dominant tICs of (A) dihedral angles and contact distances (D) of the transmembrane regions of NTSR1. The circles indicate microstate centers, the size of the circle shows the population size, and the slowest processes are shown as transitions from red states to blue states. Visualization of different types of simulations in the sparse-tICA projection of (B) dihedral angles and (E) contact distances, where cyan circles represent the active conformation without the agonist, green circles represent the active states with the agonist, light blue circles represent the inactive states without the agonist, and blue circles represent the inactive states with the agonist. ITS as a function of various MSMs. (E) MSM for the sparse-tICA projection of dihedral angles showing the convergence of the slowest processes near 25–50 ns. (F) MSM for the tICA projection of contact distances showing the convergence of the slowest processes near 75–100 ns.

interpretation of the MSM of NTSR1 trajectories, we have identified conformational changes along the activation pathway, which are depicted in Figure 1.

We simulated the NTSR1 protein in the active and inactive conformations for $16\ \mu s$ (four different conformations: active, active with the agonist, inactive, and inactive with the agonist, Table S5). Using the $16\ \mu s$ simulation data, we first investigated the NPxxY region rmsd and the TM3–TM6 distance. The NPxxY rmsd versus TM3–TM6 distance has been used frequently in previous studies to distinguish between the alternative conformations of the GPCR proteins.²⁰ We created a density projection of NPxxY rmsd versus TM3–TM6 distance and discretized the various conformations on this projection using the k -means clustering algorithm ($k = 100$, Figure 2A). We visualized the cluster centers of the k -means model as a graph network representation (generated using the NetworkX python library,³⁶ Figure 2B). The nodes of the graph are the various discrete conformations of NTSR1 at the cluster centers, and the edges between them are the transition probabilities associated with each state (determined by the MSM of cluster centers). Furthermore, we tagged the nodes of the graph by giving them an identifying index and displaying the representative NTSR1 conformation corresponding to randomly selected nodes (Figure 2B, green: inactive, cyan: intermediate, and pink: active). The density projection and the network diagram showed that there exist many different parallel paths of NTSR1 activation that the protein may sample. Additionally, two distinct groups of NTSR1 conformations can be identified along the TM3–TM6 distance axis (active and inactive). The cluster center with index 28 (TM3–TM6 distance = $11.2\ \text{\AA}$; distance between $C\alpha$ atoms of terminal residues of each TM) represents an intermediate conformation that lies at the border of the active and inactive conformation density clusters (Figure 2B). The density projection of NPxxY region rmsd versus TM3–TM6 distance allowed us to identify and validate the features of

NTSR1 dynamics that have been identified for other class A GPCRs.^{20,22,25}

The NPxxY region rmsd versus TM3–TM6 distance plot allows us to cluster and distinguish between the active and inactive conformations of NTSR1. However, such coarse feature extraction does not give us insights into the slowest residue-level conformational changes (microswitches) that govern the transition between the active and inactive conformations. We used the dihedral angles (Φ and Ψ) and contact distance of all the residues in the transmembrane helices of NTSR1 from the simulation trajectories as features to identify the microswitches. The changes in dihedral angles and contact distances are an important feature of protein structure dynamics as they capture both the large and small conformational changes in the protein structure. However, they are high-dimensional features (dihedral angles have 420 dimensions and contact distances have 2314 dimensions), therefore, we used the sparse-tICA model to project them in two dimensions (density projection) by using the dominant tICs (tIC1 and tIC2, Figure 3A and 3D). We selected sparse-tICA over the normal tICA algorithm because sparse-tICA is able to generate sparse, interpretable, and regularized low-dimensional representation of the dihedral features. To increase the biological interpretability of the density projection, we colored the data points based on the initial state of that trajectory. This clustered the projection into four classes: active state, active state with the agonist, inactive state, and the inactive state with the agonist (Figure 3B,E). We performed k -means clustering ($k = 100$) of the tIC1 and tIC2 projections of dihedral angles and the contact distances to discretize and cluster NTSR1. The cluster centers from k -means were resized based on the population of protein conformations they represent, and then, the MSM of these clusters was generated to identify the slowest processes on the density projection. The MSM was overlaid onto the dihedral angle and the contact distance density projection (Figure 3A,D). To determine the ideal lag time for MSM generation, which

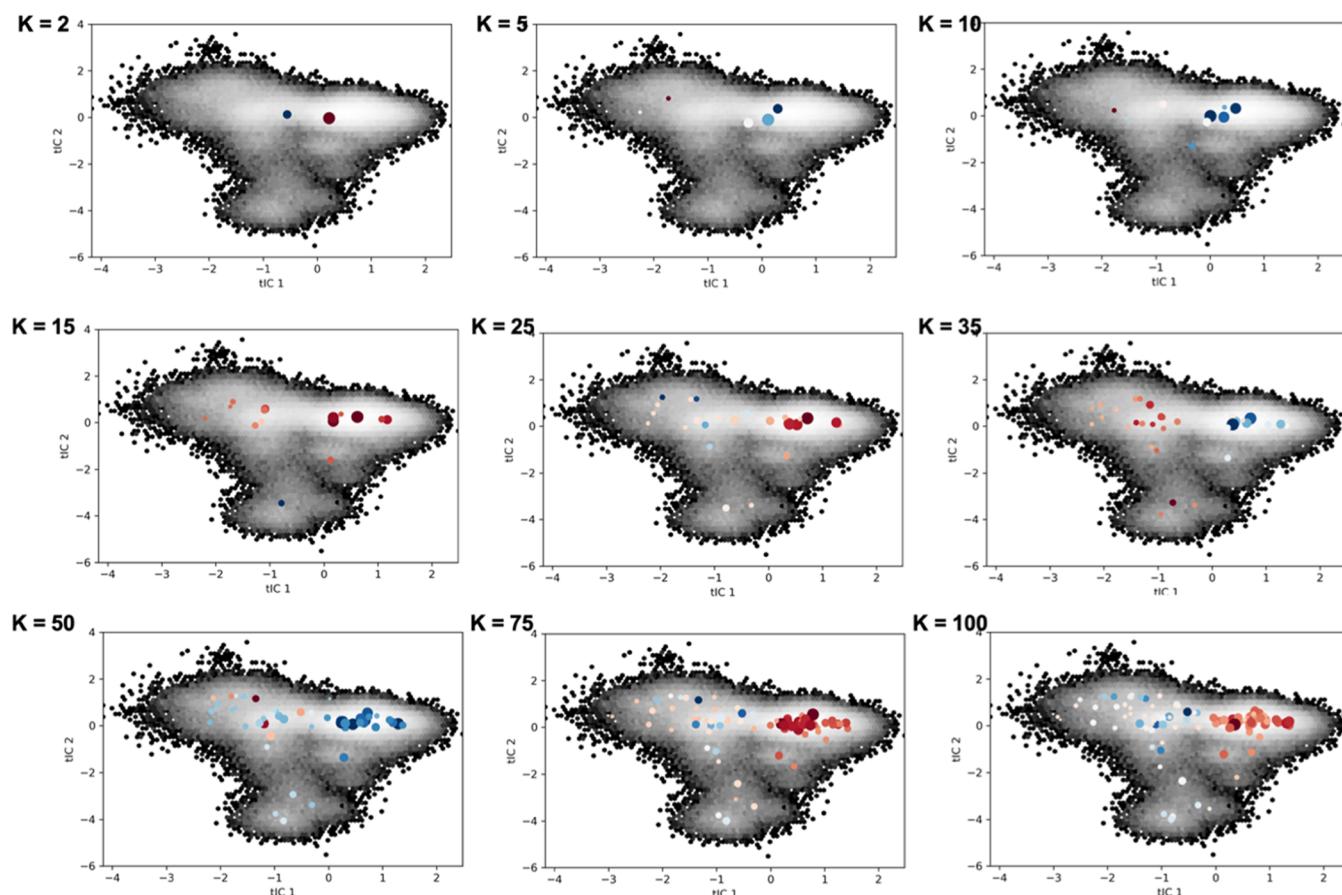


Figure 4. MSM plots on the first two tICA components of the tICA projection of dihedral angles with different cluster sizes for the *k*-means algorithm. The circles indicate microstate centers, the size of the circle shows the population size, and the slowest processes are shown as transitions from red states to blue states.

captures the slowest processes of the NTSR1 dynamics, we calculated the implied time scales (ITS) for the MSM model of dominant tICs. The tIC1 and tIC2 of dihedral angles showed convergence of the slowest processes near 25 ns (Figure 3E), and tIC1 and tIC2 showed convergence near 75 ns for contact distances (Figure 3F). Furthermore, we also generated MSMs with different cluster sizes of *k*-means to determine the impact of different cluster sizes for MSM generation (Figure 4 and S7).

The density projections of tIC1 and tIC2 of dihedral angles and contact distances show that the tIC1 separates the input features into two distinct populations, which correspond to the active and inactive conformations. However, to ensure the conformational landscape sampled by all the four simulation setups is indeed unique, we calculated the cluster centers in the dihedral projection (Figure S8). There is a distance of 0.345 between inactive and inactive–agonist cluster centers and a distance of 0.209 between active and active–agonist cluster centers. The active conformations in the case of dihedral angles have a tIC1 coordinate <-0.324 (Figure 3B) and in the case of contact distances have a tIC1 coordinate >0.363 (Figure 3E). Along tIC2, in the case of both dihedral angles and contact distances, we observe that the active conformation of NTSR1 is clustered into two alternative states. In the case of dihedral angles, tIC2 > 0.0126 corresponds to one state and tIC2 < -3.171 corresponds to the other state (Figure 3B); similarly, in the case of contact distances, tIC2 > 2.055 is one state and tIC2 < -0.100 is the other state (Figure 3E). This finding is consistent with the literature, which showed that there exist two

alternative states of active NTSR1, the canonical and non-canonical activation states.¹⁷ Furthermore, we performed free energy analysis of the tICA projection of dihedral angles to determine the existence of a free energy path connecting the active and inactive conformations (Figure S11). We projected the free energy landscape of NTSR1 on to tIC1 and tIC2. We again observe that tIC1 separates the active (blue) and inactive (red) states, whereas tIC2 separates the two alternative conformations of the active state. Furthermore, there exists a region of stability (tIC1: -0.663 and tIC2: 0.941), with the free energy <3.1 kcal/mol, where the intermediate conformations are clustered together. This region also links the active and inactive states along tIC1 and represents a path of minimum free energy that connects the active and inactive state basins.

To interpret the tICs (tIC1 and tIC2) and identify the microswitches, we calculated the contribution of individual input features (dihedral angles and contact distances) and ranked them. This task can be conceptualized as an ML problem, with the high-dimensional dihedral angle and contact distance features as the inputs to the ML model and the dominant tICs (tIC1 and tIC2) as the labels. We used RF (an ML algorithm³⁰) to learn the functional mapping of input features (dihedral angles and contact distances) to the labels (dominant tICs). The important features learnt by the fully trained RF model correspond to the most important features for generating the dominant tICs from the input features (Tables 1, 2, S2, and S3). By using this method, we ranked the important dihedral angles and contact distances and analyzed their role in governing the

Table 1. Feature Importance of Top 10 Features in the Computation of Sparse-tICA Projection for Dihedral Angles

tIC 1	contribution	tIC 2	contribution
GLU123 ^{2.60} — TYR125 ^{2.62}	0.004807	SER111 ^{2.49} — CYS151 ^{3.35}	0.00367
LEU166 ^{3.50} — ASN256 ^{5.58}	0.003628	VAL239 ^{5.41} — PHE247 ^{5.49}	0.003332
SER161 ^{3.45} — TYR167 ^{3.51}	0.003426	ALA139 ^{3.23} — CYS141 ^{3.25}	0.003138
ILE115 ^{2.53} — ALA154 ^{3.38}	0.00335	TYR104 ^{4.42} — SER161 ^{3.45}	0.003027
ASP149 ^{3.33} — LEU204 ^{4.62}	0.003202	ARG322 ^{6.54} — TYR342 ^{7.31}	0.002941
TYR103 ^{2.41} — THR185 ^{4.42}	0.003061	ASP149 ^{3.33} — THR155 ^{3.39}	0.00282
SER111 ^{2.49} — LEU147 ^{3.31}	0.002926	ASP112 ^{2.50} — ILE310 ^{6.42}	0.002792
ARG142 ^{3.26} — PHE205 ^{4.63}	0.002827	LEU105 ^{2.43} — ASN158 ^{3.42}	0.002785
LEU73 ^{1.42} —VAL79 ^{1.48}	0.002762	VAL101 ^{2.39} — ILE189 ^{4.46}	0.00276
VAL159 ^{3.43} — LEU162 ^{3.46}	0.00268	LEU246 ^{5.48} — VAL314 ^{6.46}	0.002684

Table 2. Feature Importance of Top 10 Features in the Computation of Sparse-tICA Projection for Contact Distances

tIC 1	contribution	tIC 2	contribution
PHE146 ^{3.30} — LEU147 ^{3.31}	0.008514	LEU366 ^{7.55} — ASN365 ^{7.54}	0.006207
LEU105 ^{2.43} — TYR104 ^{2.42}	0.007275	ALA139 ^{3.23} — ASP138 ^{3.22}	0.006123
LEU162 ^{3.46} — SER161 ^{3.45}	0.006737	PHE86 ^{1.55} —LEU85 ^{1.54}	0.006004
ALA139 ^{3.23} — ASP138 ^{3.22}	0.006294	PHE146 ^{3.30} — LEU147 ^{3.31}	0.005714
PRO248 ^{5.50} — PHE247 ^{5.49}	0.00561	GLY106 ^{2.44} — SER107 ^{2.45}	0.005676
THR155 ^{3.39} — ALA154 ^{3.38}	0.005287	MET207 ^{4.65} — GLY208 ^{4.66}	0.005474
VAL160 ^{3.44} — SER161 ^{3.45}	0.00516	THR67 ^{1.36} —ALA68 ^{1.37}	0.005383
LEU97 ^{2.35} —SER96 ^{2.34}	0.005157	PHE146 ^{3.30} — TYR145 ^{3.29}	0.00538
THR349 ^{7.38} — ASN350 ^{7.39}	0.004906	THR349 ^{7.38} — LEU348 ^{7.37}	0.00503
VAL355 ^{7.44} — TYR354 ^{7.43}	0.004868	PRO121 ^{2.58} — VAL122 ^{2.59}	0.004932

transition between the clusters on the density projection of tIC1 and tIC2. From this list of the top ranked features, we identified the microswitches by using the inclusion criteria such that only the residues that are oriented toward the binding site of the agonist or are involved in the intrahelix interactions were considered to be potential microswitches. The inclusion criteria helped eliminate the residues that are facing outward toward the lipid membrane, are not involved in the activation mechanism, and hence are not microswitches. These inclusion criteria identified 10 residues (L105^{2.43}, S111^{2.49}, E123^{2.60}, Y144^{3.28}, R148^{3.32}, C151^{3.35}, L162^{3.46}, F312^{6.44}, Y354^{7.43}, and Y364^{7.53}) that are most likely to be microswitches governing the transitions between the discrete perturbations on the density projection of dihedral angles and contact distances. We also investigated the sequence conservation of the 10 residues identified as microswitches. We used the GPCRdb server to calculate the residue identity and similarity of the microswitch residues

(Table S1).^{37,38} We considered amino acids with Sneath's index < 15 to be similar.³⁹ We identified that S111^{2.49}, E123^{2.60}, Y144^{3.28}, R148^{3.32}, and C151^{3.35} are the least conserved. The hydrophobic residues (L105^{2.43} and L162^{3.46}) have low identity but high similarity, indicating their positional importance. Finally, aromatic residues (F312^{6.44}, Y354^{7.43}, and Y364^{7.53}) have a high sequence similarity and high sequence identity.

To determine the order of occurrence of the identified changes in the important residues, we calculated the transfer entropy between the featurized trajectory (contact distance pairs) of the residues that were identified as microswitches (Table S4). The transfer entropy calculations allowed us to determine the precedence in conformational changes of residues of interest and construct a mechanism of activation for NTSR1.^{40,41} The transfer entropy calculations showed that the contact distance features between E123^{2.60}—R148^{3.32} and R148^{3.32}—Y354^{7.43} have the highest transfer entropy (0.1283), which is followed by transfer entropy between other binding pocket residues. This indicates that the reorganization of residues closer to the binding pocket of the agonist occurs first during the NTSR1 activation mechanism. The C151^{3.35}—F312^{6.44} and S111^{2.49}—C151^{3.35} transfer entropies are ranked next (0.0971) in Table S4, indicating that the conformational changes in F312^{6.44} occur before the interaction between S111^{2.49}—C151^{3.35} is formed. Finally, the contact distances between L105^{2.43}, L162^{3.46}, and Y364^{7.53} of the NPxxY region are ranked (0.0935), showing that these changes are the next to follow. However, this sequence of events only leads us to an intermediate state in the GPCR activation as the simulation data showed that the TM3–TM6 distance was only 9.0 Å. The transfer entropy between Y364^{7.53} of the NPxxY region and neighboring residues is followed by that of the residues of the binding pocket and the S111^{2.49}—C151^{3.35} pair, indicating that further rearrangements in these residues occur before the TM3–TM6 distances is increased to 12.2 Å (Cα distance of cytosolic terminal residues) and that NTSR1 acquires the active conformation. The transfer entropy-based ordering of microswitch residues into a sequence motivated us to analyze conformational dynamics of the microswitch residues in the various simulation trajectories and elucidate the mechanism of activation of NTSR1.

From the analysis of the states of the MSM of the density projection of tIC1 and tIC2 (dihedral angles and contact distances), we identified a distinct intermediate state in the NTSR1 activation mechanism [dihedral angles tIC1 and tIC2 coordinates (−0.032, 0.013), Figure 3A and contact distances tIC1 and tIC2 coordinates (0.005, 0.368), Figure 3D]. By analyzing the corresponding structure, we determined that the TM3–TM6 distance in this intermediate state is ~9 Å (Cα distance between the terminal residues of TM3–TM6). We then explored what sequence of steps drives the transition from the active state (pink, Figure 5) to this intermediate state and what steps drive the transition from the inactive state (green, Figure 5) to the intermediate state (cyan, Figure 5). Upon analysis of the trajectory corresponding to the cluster center of the MSM state corresponding to the reorientation of R148^{3.32}. We observed dynamic interaction between NTS and R148^{3.32} (Supporting Information Movie 1), where NTS forms double hydrogen bonds with the terminal −NH groups of R148^{3.32}, and these interactions are broken as R148^{3.32} begins its reorientation. The transition from the inactive conformation to the intermediate occurs, and then, active conformation begins with the reorientation of R148^{3.32} away from the terminal

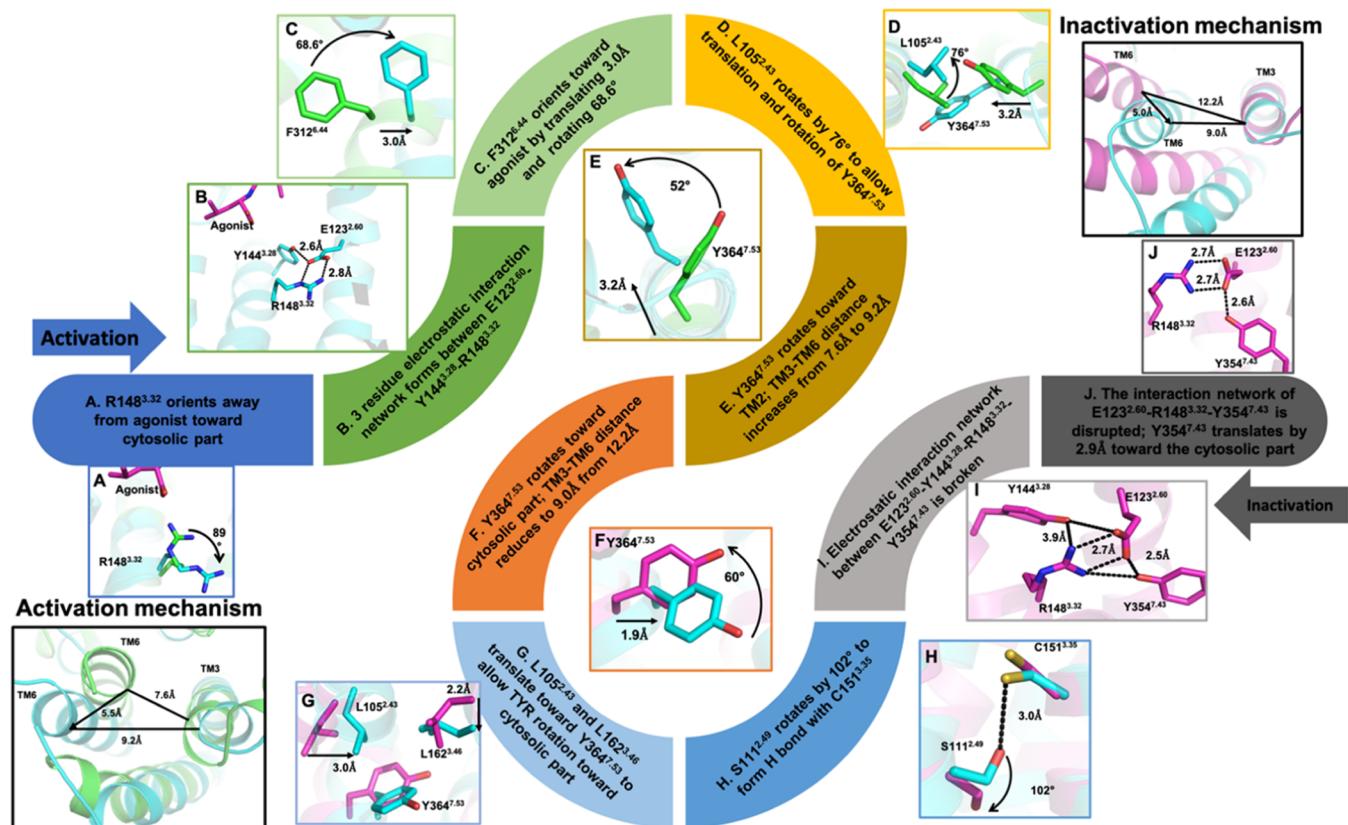


Figure 5. Proposed pathway of NTSR1 activation and inactivation dynamics determined by the dominant tICs and their feature importance (green: inactive state, cyan: intermediate state, and pink: active state). Each cell in the pathway and the associated figure box shows the residue level changes that happen in NTSR1 trajectories. The black boxes show the shift in TM6 with respect to TM3 for activation or inactivation of NTSR1. (A) Rotation in R148^{3.32}, which begins the activation mechanism. (B) Three-residue network in the binding pocket of the agonist. (C) Rotation of F312^{6.44}, which lies between the binding pocket and Y364^{7.53}. (D) Rotation of L105^{2.43}. (E) Rotation and translation of Y364^{7.53} and increase in the TM3–TM6 distance, which lead to the activation of NTSR1. (F) Rotation of Y364^{7.53} and reduction of the TM3–TM6 distance, which lead to the inactivation of NTSR1. (G) Translation in L105^{2.43} and L162^{3.46} to allow translocation of Y364^{7.53}. (H) H-bond formation between S111^{2.49} and C151^{3.35}. (I) Breaking of the four-residue network. (J) Translation of Y354^{7.43}.

residue (L13) of the NTS peptide (Figure 5A). This reorientation happens with an 89° rotation of R148^{3.32} (the angle between C α and N ϵ atoms) toward the cytosolic part and allows R148^{3.32} to form an intrahelix three-residue network of E123^{2.60}–Y144^{3.28}–R148^{3.32}. This three-residue network forms directly below the agonist and is faced toward the cytosolic part (Figure 5B). The formation of this interaction is followed by a 3.0 Å translation (C α distance) and a 68.6° rotation (angle between C α and the aromatic ring) in F312^{6.44} away from TM3 (Figure 5C). Based on the transfer entropy rankings, the movement of F312^{6.44} is one of the first residue rearrangements in TM6 outward movement. Furthermore, to accommodate for the reorientation of Y364^{7.53} in the NPxxY region, L105^{2.43} rotates by 76° (angle between C α and C γ) toward TM3 and away from Y364^{7.53} (Figure 5D). This allows the Y364^{7.53} of the NPxxY region to reorient, without steric hindrance, by translating 3.2 Å (C α distance) and rotating 52° (angle between C α and the hydroxyl group) toward TM3 (Figure 5E). Cumulatively, this set of conformational changes leads to the transition from the inactive state (green, Figure 5) to the intermediate state (cyan, Figure 5).

Similarly, we investigated the set of conformational changes that drive the transition from the intermediate state (cyan, Figure 5) to the active state (pink, Figure 5). These changes are analyzed from a trajectory where the NTSR1 protein transitions from the active conformation to the intermediate conformation.

In an inverse chronological order, these changes can be interpreted as the reorientation of Y364^{7.53} of the NPxxY region by rotating 60° (angle between C α and the hydroxyl group) and translating 1.9 Å (C α distance, Figure 5F). This leads to the 3.0 and 2.2 Å (C α distance) translation in L105^{2.43} and L162^{3.46} toward TM7, respectively, which fills the intrahelix cavity formed by the reorientation of Y364^{7.53} (Figure 5G). The next step is the rotation of S111^{2.49}, by 102° (angle between C α and the hydroxyl group), which leads to intrahelical H-bond formation between –OH of S111^{2.49}, and –SH of C151^{3.35} (Figure 5H). The intrahelix four-residue interaction network is formed as E123^{2.60}–Y144^{3.28}–R148^{3.32}–Y354^{7.43}, which lies just below the agonist binding site (Figure 5I). One of the final conformational changes for NTSR1 to transition from the intermediate state to the active state is the disruption of the network E123^{2.60}–R148^{3.32}–Y354^{7.43} due to the translation of Y354^{7.43} by 2.9 Å (C α distance) toward the cytosolic part (Figure 5J). Conclusively, this set of conformational changes leads to the transition from the intermediate state (cyan, Figure 5) to the active state (pink, Figure 5). Through the analysis of the conformational changes of the microswitches (identified by the interpretation of tIC1 and tIC2), we were able to elucidate one of the parallel pathways of NTSR1 activation.

CONCLUSIONS

In this work, we simulated the NTSR1 protein (in the absence and presence of its agonist NTS) to unravel its mechanism of activation and identify the kinetically stable intermediate states. We interpreted the simulation data by generating the density projections along the reaction coordinates. To find out the reaction coordinates, tICA was used to project the slowest relaxing degrees of freedom in NTSR1. The key motifs and residues with the slowest relaxing mode were identified by clustering states and building the MSM. The network of these states demonstrated that there exist multiple parallel paths of activation for NTSR1. We then used RF regression and ranked the contribution of individual residues for the calculation of dominant tICs. This allowed us to determine the high-ranking features as the microswitches driving the transition of the intermediate states of NTSR1. We identified the four-residue electrostatic network of E123^{2,60}–R148^{3,32} and R148^{3,32}–Y354^{7,43}, in the inactive conformation, which undergoes reorientation and gets disrupted. This is followed by the rotation and translation of F312^{6,44} away from TM3. The Y364^{7,53} of the NPxxY region is also reoriented to lead to an intermediate conformation where L105^{2,43} and L162^{3,46} are translated toward TM7. In this intermediate conformation, the –OH of S111^{2,49} and –SH of C151^{3,35} form a hydrogen bond, and the four-residue network of E123^{2,60}–R148^{3,32} and R148^{3,32}–Y354^{7,43} is formed again. Finally, Y354^{7,43} is translated toward the cytosolic end of the protein to lead to the active conformation. The proposed sequence of conformational changes can be one of the parallel pathways driving the transition between the inactive and active states of the NTSR1 protein. The further investigation of the intermediate states and their stability can be carried out through free energy analysis of the pathway proposed in this work, for instance, through the string method with a swarm of trajectories. The activation dynamics elucidated in this work furthers the understanding of molecular pathways of NTSR1, which can potentially facilitate the development of highly specific therapeutics targeting NTSR1.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acschemneuro.2c00154>.

Visualization of the simulation system of NTSR1, sequence alignment of NTSR1 structures used for simulation, TM region identification in NTSR1, lag time analysis of sparse-tICA projection of dihedral angles and contact distances, MSM with different cluster sizes on sparse-tICA projection of contact distances, tICA and Kernel tICA projection of dihedral angles and contact distances, top 20 contributing dihedral angles of the first four tICs of dihedral angles and contact distances in the sparse-tICA projection, top 30 transfer entropy pairs of contact distances, and count of simulations ([PDF](#))

The interactions of the agonist (L13) with Y145^{3,29} and R322^{6,55} from the active conformation ([MP4](#))

AUTHOR INFORMATION

Corresponding Author

Amir Barati Farimani — Department of Mechanical Engineering, Biomedical Engineering, Chemical Engineering and Machine Learning, Carnegie Mellon University,

Pittsburgh, Pennsylvania 15213, United States; [orcid.org/0000-0002-2952-8576](#); Email: barati@cmu.edu; <https://www.baratilab.com>

Author

Prakhar Yadav — Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States; [orcid.org/0000-0002-1170-6290](#)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acschemneuro.2c00154>

Author Contributions

A.B.F. conceptualized the project. P.Y. and A.B.F. contributed to the overall design of the project, data interpretation, and preparation of the manuscript. P.Y. performed the experiments and analyzed the data.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the use of the supercomputing resource Arjuna provided by the Pittsburgh Supercomputing Center (PSC). This work was supported by the Center for Machine Learning in Health (CMLH) at Carnegie Mellon University and the start-up fund from the Mechanical Engineering Department at CMU. The authors also acknowledge the comments and feedback provided by Parisa Mollaei, Rishikesh Magar, and Zhonglin Cao.

REFERENCES

- (1) Carraway, R.; Leeman, S. E. The Isolation of a New Hypotensive Peptide, Neurotensin, from Bovine Hypothalamus. *J. Biol. Chem.* **1973**, *248*, 6854–6861.
- (2) Vincent, J.-P.; Mazella, J.; Kitabgi, P. Neurotensin and Neurotensin Receptors. *Trends Pharmacol. Sci.* **1999**, *20*, 302–309.
- (3) Boules, M.; Li, Z.; Smith, K.; Fredrickson, P.; Richelson, E. Diverse Roles of Neurotensin Agonists in the Central Nervous System. *Front. Endocrinol.* **2013**, *4*, 36.
- (4) Wu, Z.; Martinez-Fong, D.; Trédaniel, J.; Forgez, P. Neurotensin and Its High Affinity Receptor 1 as a Potential Pharmacological Target in Cancer Therapy. *Front. Endocrinol.* **2013**, *3*, 184.
- (5) Mustain, W. C.; Rychahou, P. G.; Evers, B. M. The Role of Neurotensin in Physiologic and Pathologic Processes. *Curr. Opin. Endocrinol., Diabetes Obes.* **2011**, *18*, 75–82.
- (6) Bissette, G.; Nemeroff, C. B.; Loosen, P. T.; Prange, A. J.; Lipton, M. A. Hypothermia and Intolerance to Cold Induced by Intracisternal Administration of the Hypothalamic Peptide Neurotensin. *Nature* **1976**, *262*, 607–609.
- (7) Schimpff, R.-M.; Avard, C.; Fénelon, G.; Lhiaubet, A.-M.; Tennezé, L.; Vidailhet, M.; Rostène, W. Increased Plasma Neurotensin Concentrations in Patients with Parkinson's Disease. *J. Neurol., Neurosurg. Psychiatry* **2001**, *70*, 784–786.
- (8) Alifano, M.; Souazé, F.; Dupouy, S.; Camilleri-Broët, S.; Younes, M.; Ahmed-Zaid, S.-M.; Takahashi, T.; Cancellieri, A.; Damiani, S.; Boaron, M.; Broët, P.; Miller, L. D.; Gespach, C.; Regnard, J. F.; Forgez, P. Neurotensin Receptor 1 Determines the Outcome of Non-Small Cell Lung Cancer. *Clin. Cancer Res.* **2010**, *16*, 4401–4410.
- (9) Griebel, G.; Holsboer, F. Neuropeptide Receptor Ligands as Drugs for Psychiatric Diseases: The End of the Beginning? *Nat. Rev. Drug Discovery* **2012**, *11*, 462–478.
- (10) Tanaka, K.; Masu, M.; Nakanishi, S. Structure and Functional Expression of the Cloned Rat Neurotensin Receptor. *Neuron* **1990**, *4*, 847–854.
- (11) Chalon, P.; Vita, N.; Kaghad, M.; Guillemot, M.; Bonnin, J.; Delpech, B.; Le Fur, G.; Ferrara, P.; Caput, D. Molecular Cloning of a

- Levocabastine-Sensitive Neuropeptide Binding Site. *FEBS Lett.* **1996**, *386*, 91–94.
- (12) Savdie, C.; Ferguson, S. S. G.; Vincent, J. P.; Beaudet, A.; Stroh, T. Cell-Type-Specific Pathways of Neuropeptide Endocytosis. *Cell Tissue Res.* **2006**, *324*, 69–85.
- (13) Ferguson, S. S. G. Evolving Concepts in G Protein-Coupled Receptor Endocytosis: The Role in Receptor Desensitization and Signaling. *Pharmacol. Rev.* **2001**, *53*, 1–24.
- (14) Besserer-Offroy, É.; Brouillette, R. L.; Lavenus, S.; Froehlich, U.; Brumwell, A.; Murza, A.; Longpré, J.-M.; Marsault, E.; Grandbois, M.; Sarret, P.; Leduc, R. The Signaling Signature of the Neuropeptide Type 1 Receptor with Endogenous Ligands. *Eur. J. Pharmacol.* **2017**, *805*, 1–13.
- (15) White, J. F.; Noinaj, N.; Shibata, Y.; Love, J.; Kloss, B.; Xu, F.; Gvozdenovic-Jeremic, J.; Shah, P.; Shiloach, J.; Tate, C. G.; Grisshammer, R. Structure of the Agonist-Bound Neuropeptide Receptor. *Nature* **2012**, *490*, 508–513.
- (16) Egloff, P.; Hillenbrand, M.; Klenk, C.; Batyuk, A.; Heine, P.; Balada, S.; Schlinkmann, K. M.; Scott, D. J.; Schütz, M.; Plückthun, A. Structure of Signaling-Competent Neuropeptide Receptor 1 Obtained by Directed Evolution in Escherichia Coli. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E655–E662.
- (17) Kato, H. E.; Zhang, Y.; Hu, H.; Suomivuori, C.-M.; Kadji, F. M. N.; Aoki, J.; Krishna Kumar, K.; Fonseca, R.; Hilger, D.; Huang, W.; Latorraca, N. R.; Inoue, A.; Dror, R. O.; Kobilka, B. K.; Skiniotis, G. Conformational Transitions of a Neuropeptide Receptor 1–G II Complex. *Nature* **2019**, *572*, 80–85.
- (18) Huang, W.; Masureel, M.; Qu, Q.; Janetzko, J.; Inoue, A.; Kato, H. E.; Robertson, M. J.; Nguyen, K. C.; Glenn, J. S.; Skiniotis, G.; Kobilka, B. K. Structure of the Neuropeptide Receptor 1 in Complex with β -Arrestin 1. *Nature* **2020**, *579*, 303–308.
- (19) Krumm, B. E.; White, J. F.; Shah, P.; Grisshammer, R. Structural Prerequisites for G-Protein Activation by the Neuropeptide Receptor. *Nat. Commun.* **2015**, *6*, 7895.
- (20) Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem. Rev.* **2017**, *117*, 139–155.
- (21) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. Activation Mechanism of the B2-Adrenergic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 18684–18689.
- (22) Huang, W.; Manglik, A.; Venkatakrishnan, A. J.; Laeremans, T.; Feinberg, E. N.; Sanborn, A. L.; Kato, H. E.; Livingston, K. E.; Thorsen, T. S.; Kling, R. C.; Granier, S.; Gmeiner, P.; Husbands, S. M.; Traynor, J. R.; Weis, W. I.; Steyaert, J.; Dror, R. O.; Kobilka, B. K. Structural Insights into M-Opioid Receptor Activation. *Nature* **2015**, *524*, 315–321.
- (23) McGibbon, R. T.; Pande, V. S. Variational Cross-Validation of Slow Dynamical Modes in Molecular Kinetics. *J. Chem. Phys.* **2015**, *142*, 124105.
- (24) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139*, 015102.
- (25) Manglik, A.; Kim, T. H.; Masureel, M.; Altenbach, C.; Yang, Z.; Hilger, D.; Lerch, M. T.; Kobilka, T. S.; Thian, F. S.; Hubbell, W. L.; Prosser, R. S.; Kobilka, B. K. Structural Insights into the Dynamic Process of B2-Adrenergic Receptor Signaling. *Cell* **2015**, *161*, 1101–1111.
- (26) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (27) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985**, *83*, 4069–4074.
- (28) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.
- (29) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.
- (30) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
- (31) Kanungo, T.; Mount, D. M.; Netanyahu, N. S.; Piatko, C. D.; Silverman, R.; Wu, A. Y. An Efficient K-Means Clustering Algorithm: Analysis and Implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892.
- (32) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010**, *52*, 99–105.
- (33) Hernández, C. *Mdentropy: V0.2*; Zenodo, 2015.
- (34) Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.
- (35) Sultan, M.; Pande, V. S. TICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017**, *13*, 2440–2447.
- (36) Schult, D. A. Exploring Network Structure, Dynamics, and Function Using NetworkX. *Proceedings of the 7th Python in Science Conference*; SciPy, 2008; pp 11–15.
- (37) Pády-Szekeres, G.; Esguerra, M.; Hauser, A. S.; Caroli, J.; Munk, C.; Pilger, S.; Keserű, G. M.; Kooistra, A. J.; Gloriam, D. E. The G Protein Database, GproteinDb. *Nucleic Acids Res.* **2022**, *50*, D518–D525.
- (38) Kooistra, A. J.; Mordalski, S.; Pády-Szekeres, G.; Esguerra, M.; Mamyrbekov, A.; Munk, C.; Keserű, G. M.; Gloriam, D. E. GPCRdb in 2021: Integrating GPCR Sequence, Structure and Function. *Nucleic Acids Res.* **2021**, *49*, D335–D343.
- (39) Sneath, P. H. A. Relations between Chemical Structure and Biological Activity in Peptides. *J. Theor. Biol.* **1966**, *12*, 157–195.
- (40) E, W.; Vanden-Eijnden, E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391–420.
- (41) Meng, Y.; Shukla, D.; Pande, V. S.; Roux, B. Transition Path Theory Analysis of C-Src Kinase Activation. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 9193–9198.

Recommended by ACS

Molecular Dynamics Refinement of Open State Serotonin 5-HT_{3A} Receptor Structures

Zoe Li, Xiaolin Cheng, et al.

FEBRUARY 09, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ

Structural and Molecular Determinants for Isoform Bias at Human Histamine H₃ Receptor Isoforms

Sabrina N. Rahman, Christopher J. Langmead, et al.

JANUARY 26, 2023

ACS CHEMICAL NEUROSCIENCE

READ

Neuropeptide Receptor Allosterism Revealed in Complex with a Biased Allosteric Modulator

Brian E. Krumm, Bryan L. Roth, et al.

MARCH 14, 2023

BIOCHEMISTRY

READ

5-HT₃ Receptor MX Helix Contributes to Receptor Function

James Mocatta, Sarah C. R. Lummis, et al.

JULY 22, 2022

ACS CHEMICAL NEUROSCIENCE

READ

Get More Suggestions >