

ARTICLE

Received 8 Aug 2014 | Accepted 26 Apr 2015 | Published 15 Jun 2015

DOI: 10.1038/ncomms8283

A network of molecular switches controls the activation of the two-component response regulator NtrC

Dan K. Vanatta¹, Diwakar Shukla^{1,2,3}, Morgan Lawrenz¹ & Vijay S. Pande^{1,2}

Recent successes in simulating protein structure and folding dynamics have demonstrated the power of molecular dynamics to predict the long timescale behaviour of proteins. Here, we extend and improve these methods to predict molecular switches that characterize conformational change pathways between the active and inactive state of nitrogen regulatory protein C (NtrC). By employing unbiased Markov state model-based molecular dynamics simulations, we construct a dynamic picture of the activation pathways of this key bacterial signalling protein that is consistent with experimental observations and predicts new mutants that could be used for validation of the mechanism. Moreover, these results suggest a novel mechanistic paradigm for conformational switching.

¹Department of Chemistry, Stanford University, Stanford, California 94305, USA. ²SIMBIOS NIH Center for Biomedical Computation, Stanford University, Stanford, California 94305, USA. ³Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. Correspondence and requests for materials should be addressed to V.S.P. (email: pande@stanford.edu).

Proteins involved in cellular signalling change their conformation in response to changes in environment (input signal), such as ligand binding or chemical modification, to control downstream cellular processes (output signal)¹. The behaviour of these proteins is controlled by molecular level interactions between specific sets of residues that relay the signal. A detailed map of switches that facilitate conformational change has been obtained for only a few well-studied signalling proteins such as G-protein-coupled receptors and kinases^{2,3}. We propose that atomistic simulations can help identify these key structural elements, and demonstrate this by examining the activation mechanism of nitrogen regulatory protein C (NtrC). This protein belongs to the family of two component regulatory systems ubiquitous in bacteria⁴, in which a phosphate is transferred from a sensor kinase (NtrB) to a response regulator (NtrC)⁵. This transfer activates the protein by triggering a structural rearrangement that exposes a hydrophobic surface and allows the protein to form oligomers essential for ATP hydrolysis⁶. Knowledge of the activation pathways and transition states of this rearrangement would be useful for elucidating new approaches for treating antibiotic-resistant infections.

NtrC activation occurs when a phosphate is transferred from NtrB to D54 in NtrC. The active and inactive states are structurally very similar, with a root-mean-square deviation (RMSD) of only 2.9 Å (ref. 5). The inactive state is more flexible than the active state and could be described as a collection of states that rapidly interconvert⁷. The significant conformational change occurs in the region of residues 82–101, which is spatially removed from the phosphate-binding site at residue 54. The change involves a shift and slight uncoiling of helix 4 and minor rearrangement of the loop region between helix α_3 and beta-sheet β_3 (ref. 6). Recent nuclear magnetic resonance (NMR) studies have shown that both active and inactive states are populated in solution without phosphorylated D54, which demonstrates that phosphorylation is not required for the conformational change to occur⁵. Phosphorylation stabilizes the active state, which results in population inversion from a 2–10% active state population to 99% (ref. 8).

Molecular simulations have been previously used to study conformational changes associated with NtrC. However, because of the large size of the protein and its long activation timescales, these studies have employed biased techniques such as targeted molecular dynamics⁹, separate simulations of the active and inactive states^{7,10,11}, or simulations with coarse grained representations of the protein and solvent^{12–15}. These studies have shed light on various aspects of NtrC activation but have failed to provide a comprehensive view of the conformational landscape of NtrC activation. In particular, kinetics of conformational transitions and detailed molecular description of NtrC activation pathways have not been reported.

Understanding the activation mechanism of unphosphorylated NtrC is the focus of this study. We have employed enhanced sampling algorithms¹⁶ coupled with the Folding@home¹⁷ distributed computing platform to gather unprecedented statistics (1.6 ms) on NtrC dynamics for an atomistic description of the complex network of molecular-switches that drive activation of this key bacterial signalling protein. In this study, we use extensive MD simulations to build a Markov state model (MSM) for estimating the thermodynamics and kinetic properties of NtrC dynamics. MSMs reduce the complex conformational landscape of a protein to a set of metastable conformational states and the rates of transition between them^{18–20}. The MSM of NtrC reveals a network of molecular switches that control NtrC activation and identifies metastable states on the NtrC conformational free energy landscape that could be targeted for inhibitor design.

Results

Simulations reveal molecular switches for NtrC activation.

MSMs for NtrC were built using snapshots along the simulated trajectories, and a 2,300 state MSM was chosen for analysis due to the convergence of its implied timescales, an indicator of Markovian behaviour (see Supplementary Fig. 1 and Methods). We started simulations from the inactive and active NMR structures of NtrC (Fig. 1). A comparison of the MSM states reveals several residues that acquire distinct conformations and switch between different networks of interactions. In particular, we have identified four key ‘molecular switches’ that control NtrC activation as illustrated in Fig. 2a,b. The mechanism includes the F99 and Y94 switches, in which these residues rotate their side chains towards the protein core (Fig. 3a,c), the Y101–T82 switch that breaks its active state hydrogen bond (Fig. 3b) and the K67–Q96–Q95 (KQQ) switch (Fig. 3d), in which the K67–Q96 hydrogen bond breaks, destabilizing the C-terminal turn of helix α_4 , while a new hydrogen bond forms between K67 and Q95. Histograms of the residue pair distances that characterize the molecular switches were weighted by the MSM state equilibrium probability densities (see Methods) and are shown in Fig. 3a–d. The distributions reveal the bimodal nature of these switches and the overall activation process of NtrC. Although the solid line representing the sum of the state probabilities in Fig. 3a–d shows bimodal behaviour, the tail of the inactive state distribution extends into the active region due to greater conformational diversity in this state. This implies that the position of the switches alone is not sufficient to define the states and a kinetic definition of the active and inactive states (described below) must be used for understanding the activation mechanism. The probability density distributions also provide evidence against the commonly accepted theory related to the activation of response regulators such as NtrC and CheY²¹. In two-component systems like these, activation of the response regulator receiver domains has been proposed to rely on a ‘Y-T coupling’ mechanism²² (Y101 and T82 in NtrC), where the interaction between this conserved pair of residues is correlated with the activation of the protein. Figure 3b shows the probability distributions of Y101–T82 distance in the active, inactive and entire conformational ensemble of NtrC. The active ensemble shows two peaks corresponding to conformations with Y-T coupling formed and broken. This figure provides the evidence that the Y101–T82 coupling can be either formed or broken within the active state, but is predominately broken in the inactive state. This result implies that Y-T coupling is not strictly correlated with the activation of NtrC and is in agreement with recent NMR relaxation dispersion experiments¹⁰, which found

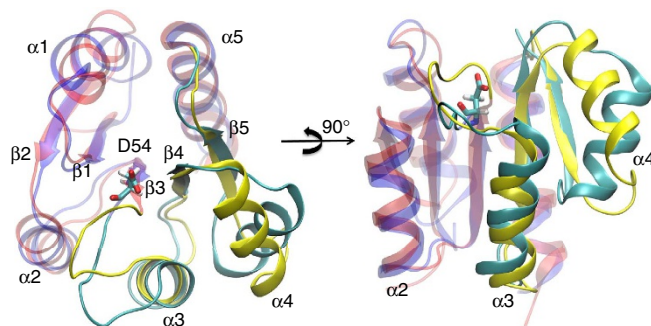


Figure 1 | Conformational changes associated with NtrC activation. The active (blue; cyan) and inactive (red; yellow) crystal structures of NtrC with the regions involved in conformational switching highlighted in cyan (active) and yellow (inactive). The phosphorylation of residue D54 shown above involves large structural changes in and around helices 3 and 4.

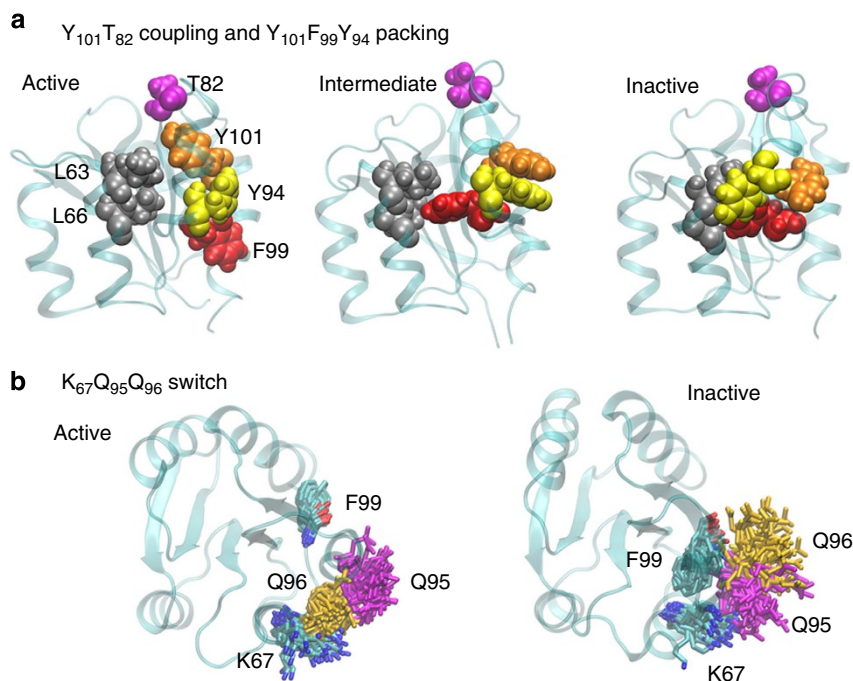


Figure 2 | Molecular switches involved in NtrC activation. (a) Changes in T82–Y101 coupling and changes in hydrophobic packing between residues L63, L66, Y94, F99 and Y101 during activation. (b) Switching of interaction partner of K67 from Q96 in active state to Q95 in inactive state.

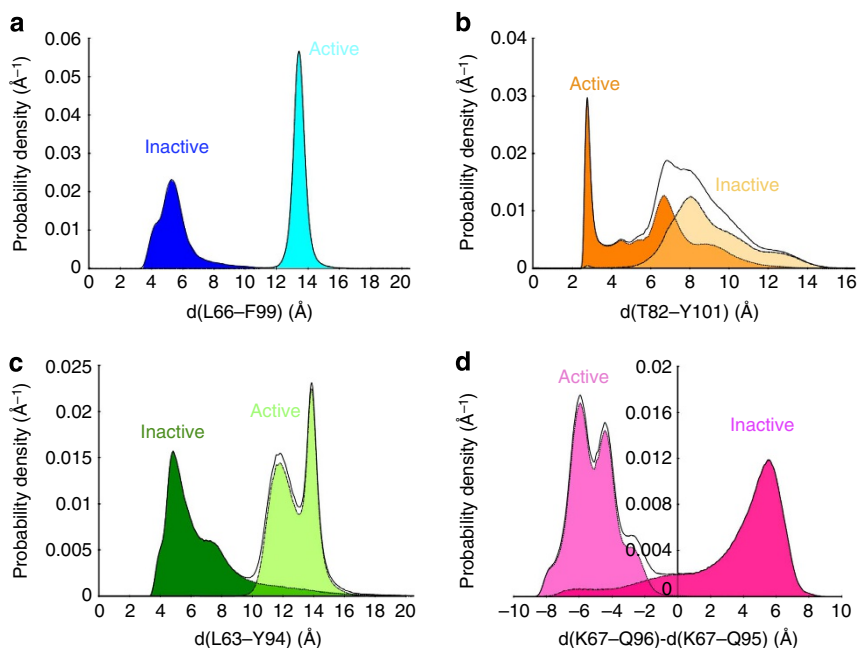


Figure 3 | Thermodynamics of molecular switches. MSM state population weighted probability densities of (a) L66–F99 distance, (b) T82–Y101 distance, (c) L63–Y94 distance and (d) difference between K67–Q96 and K67–Q95 distances (shown separately in Supplementary Fig. 2). The different colour shades correspond to the different kinetically defined states, whereas the solid line represents the entire data set.

the interconversion of the Y101 rotameric state to be faster than the inactive/active conformational transition. Furthermore, our results in Fig. 3a,c,d indicate that the conformational preferences of the three switches identified in this study involving F99, Y94 and the KQQ residues are more strongly correlated with activation than Y–T coupling. In particular, the slowest dynamical process of the MSM (see Methods and Supplementary Figs 2–4) involves the flipping of the KQQ switch identified in this study, which underscores its importance in NtrC activation. These results also show that these couplings or switches could be either

formed or broken within active or inactive ensemble. However, the probability of toggling of individual switches is dependent on the overall state of the protein.

Identifying intermediate states of NtrC. We identify several metastable intermediate states along the activation process that are characterized by the toggling of individual switches between their active and inactive conformations. In the active state of NtrC, hydrophobic residues Y94, F99 and Y101 form a

solvent-exposed hydrophobic surface (Fig. 2a). Breakage of the Y101–T82 hydrogen bond and rotation of the F99 side chain triggers collapse of this exposed hydrophobic surface, and these residues switch to their inactive conformations, directed towards the protein core. These conformational changes are coupled to the exchange of K67 interaction partners, from Q96 to Q95 (Fig. 2b) via switching of the Y94 side chain. The K67–Q96 interaction is disrupted, followed by formation of the K67–Q95 contact and flipping of Y94 towards the protein core. This KQQ conformational switch motif defines several key intermediate states, mapped onto the activation landscape in Fig. 4b, and is enabled by changes in the fold of helix during activation.

Role of transient hydrogen bonds in activation. Our results reveal a hydrogen bond network exchange that allows partial unfolding and refolding of helix α_4 to accommodate the conformational rearrangement of the hydrophobic surface residues. During activation, the helix α_4 C-terminal end gains two backbone hydrogen bonds (between V91–Q95 and S92–Q96 (Supplementary Fig. 5)) and loses two backbone hydrogen bonds (between H84–D88 and S85–A89 (Supplementary Fig. 6)) because of the formation of a helix turn at its N-terminal end. Gardino *et al.*²³ have shown that residues S85 and D86 (Supplementary Fig. 7) at the N-terminal side of helix 4 and residues Q96 and Y101 at the C-terminal side of helix α_4 and beta-sheet β_5 , respectively, form transient hydrogen bonds that compensate for the loss of backbone hydrogen bonds and lower the energy barrier for activation. The MSM-derived thermodynamic populations and exchange kinetics between states with these transient hydrogen bonds reveal that the S85–D86 hydrogen bonds (Supplementary Figs 8 and 9) are formed both in the inactive and active state ensemble in our simulations, not just in the transition state. This indicates that a more complex mechanism is responsible for slow inter-conversion rate of S85D and S85G mutants reported by Gardino *et al.*²⁴, and fast inter-conversion rate of S85N mutant. Similarly, the transient Y101–Q96 hydrogen bond is also found to be well-formed in the inactive state (Supplementary Figs 8 and 9) because of fluctuations of the unfolded, C-terminal half turn of helix α_4 . Gardino *et al.*²³ have shown that both the Q96N and the Y101F mutants of NtrC have a slower inter-conversion rate between active and inactive states ($\approx 3,500$ versus $13,300 \text{ s}^{-1}$ for wild type). The residues Q96 and Y101 interact with different hydrogen bond partners forming T82–Y101, K67–Q96 and Y101–Q96 hydrogen bonds in different conformations (Fig. 2a,b). The substitutions Y101F and Q96N not only affect the transition state stability but also the active/inactive states. Therefore, it is difficult to provide a rational mechanistic

explanation of the slow interconversion rate observed in experiments and further computational investigations are needed to obtain an atomistic perspective of the effect of these substitutions.

Diverse activation pathways due to heterogeneity in basins. The MSM-weighted two-dimensional contour plots of the simulation data are plotted in Fig. 4a as the backbone atom RMSD from both active and inactive states and Fig. 4b as the K67–Q95–Q96 and L66–F99 switch distances. Both plots clearly show two dominant basins, reinforcing our view that both states are populated even without the presence of the activating phosphate group. The plots show that these limited structural features alone are insufficient to characterize the inactive and active states. MSM-derived ‘committor’ values and high-flux pathways between the active and inactive states were computed (see Methods and Supplementary Fig. 3), and states with intermediate committor values are shown as white circles in Fig. 4 to represent transition states with similar probability of transitioning forward to the active state or backward to the inactive state. These states lie around the edges of the inactive basins in both plots, yet they have a 20–80% probability of transitioning into the active basin instead of returning to the inactive basin.

Our MSM approach allows us to define states by their kinetic connections and reveal their role in the activation mechanism, instead of relying on structural differences that can be difficult to elucidate. We can, however, structurally characterize these kinetic intermediate states in terms of the previously described molecular switches and we find that several kinetic intermediate states are present along the activation pathways. For example, the basins in the contour plot for the K67–Q95–Q96 and L66–F99 switch distances (Fig. 4b) are connected by multiple pathways with different kinetic intermediates depending on which switch is toggled first¹². The highest flux pathway involves the kinetic intermediate state, where KQQ switch is toggled before the F99 flip (Supplementary Fig. 10). Similarly, kinetic intermediates along the activation pathways could be identified with differences in the state of molecular switches. The top three activation pathways reveal kinetic intermediates with L63–Y94 switch toggled before KQQ flip (Supplementary Fig. 11) and L63–Y94 switch toggled before the F99 flip.

One of the key debates about the NtrC activation mechanism centres on the requirement for cracking, which is defined as a complete unfolding of helix α_4 for the activation conformational transition^{9,12}. Although we have seen an exchange of hydrogen bond networks involving this helix, as described above, our activation pathway results do not show a cracked helix α_4 . Similarly, ultra-long trajectories generated from a kinetic Monte

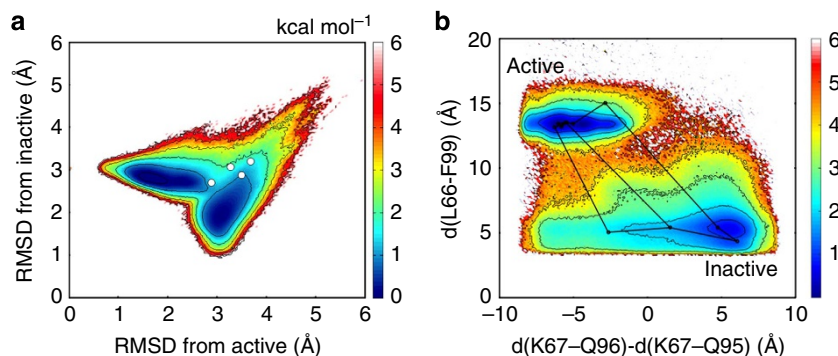


Figure 4 | NtrC activation pathways. Contour plots of (a) RMSD of NtrC from active and inactive states, (b) distance between C α atoms of the L66–F99 residues versus the difference in distance between K67–Q96 and K66–Q95. MSM states with committor values between 0.2 and 0.8 are shown as white circles. The black lines show the three top pathways calculated from transition path theory.

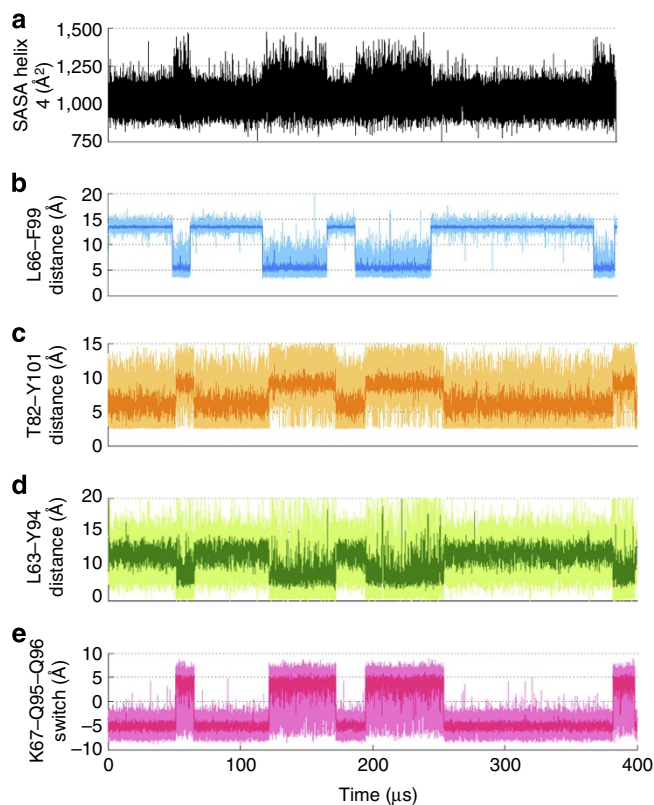


Figure 5 | Kinetics of molecular switches. Variation of the order parameters (a) α_4 -helicity of helix α_4 and distance between (b) L66-F99, (c) T82-Y101, (d) L63-Y94 and (e) the difference of (K67-Q96) and (K67-Q95) residues along a MSM trajectory generated using a kinetic Monte Carlo algorithm and MSM state transition probabilities. The light colour represents the instantaneous value along the trajectory, whereas the dark shade is a moving average over ten frames.

Carlo simulation using the MSM state transition probabilities (Fig. 5a) show that the Solvent Accessible Surface Area of helix α_4 increases slightly in the inactive state because of the rotation of the helix, which is also in agreement with previous NMR results^{23,24}. The increase in the Solvent Accessible Surface Area would be much higher ($\sim 2,600 \text{ \AA}^2$) for the fully unfolded helix α_4 .

Kinetics of NtrC activation. The activation and deactivation timescales (mean first passage time between inactive and active micro states) are found to be 90 and 110 μs , respectively, which are in agreement with the experimental interconversion rate of 77 μs (ref. 24). A visual inspection of the kinetic plots (Fig. 5b–e) suggests that a single thermal fluctuation toggles all the molecular switches via a concerted mechanism, where molecular switches are triggered cooperatively. However, the conformational landscape of molecular switches in Fig. 3 and Supplementary Figs 10–12 indicate a different, more sequential view of NtrC activation, where a particular molecular switch is turned on or off before other switches change their conformation (as described above). Altogether, these results indicate that several switches are required to toggle their states before crossing the final barrier to activation/deactivation and they become locked in that state after the barrier is crossed. For example, the Y101–T82 switch toggles between the on and off state in active basin but it is always broken after crossing the deactivation barrier (Fig. 3b). Similar observations can be made for the L63–Y94 and KQQ switches (Supplementary Figs 10–12).

Figure 5c,d shows that the Y101–T82 and L63–Y94 distances fluctuate between two distinct states but the instantaneous value of these fluctuations (light colour) overlap with the average values (dark colour). This means that these switches are not a strong indicator of state because the instantaneous value of the order parameter does not give any state information. This gives further evidence for our hypothesis that ‘Y-T coupling’ is not strictly correlated with the activation of NtrC. Figure 5b,e shows agreement between the instantaneous value and the average value indicating that the KQQ switch distinguishes the active and inactive states and further implicates this switch as the rate-limiting step.

Finally, the equilibrium population of the active ($\rho_{\text{active}} \sim 0.16$) and inactive state ($\rho_{\text{inactive}} \sim 0.09$) indicate that both states are populated in solution and interconvert rapidly at the simulation temperature (340 K), with a barrier between intermediate and inactive states $\sim 4 \text{ kcal mol}^{-1}$ as estimated from the free energy contour plots (Fig. 4). The predicted barrier height between the intermediates and inactive state is $4\text{--}5 \text{ kcal mol}^{-1}$, which, when allowing for temperature change, is close to the reported 6 kcal mol^{-1} activation barrier measured experimentally using NMR at 300 K (ref. 23).

Discussion

Our results contribute to the prominent debate about conformational change mechanisms in general, comparing a sequential domino brick effect²⁵ or Monod-Wyman-Changeux^{25–27} type of concerted action allostery. For NtrC activation, we find that the molecular switches are all connected (a requirement for the domino effect) and can induce a sequential triggering of switches, such as T82-Y101 hydrogen bond breakage preceding rearrangement of the hydrophobic surface. However, we also observe that the KQQ motif and Y94 switch their conformations cooperatively. The sequential flipping of these residues occurs within the lag time of the model so it appears concerted in the kinetic plots in Fig. 5. Therefore, the mechanism of global conformational change associated with NtrC activation lies between a sequential and cooperative mode of molecular switching so that the system is ‘functionally concerted’. MSMs provide a third and more natural framework for understanding complex conformational changes in proteins. MSMs provide a probabilistic network-switching model where different molecular switch conformations are represented by discrete MSM states and the rate of transition between these states accounts for both the independent switch probability and the conditional probability of a switch toggle given the states of other switches. An important advantage of this discrete probabilistic picture of global conformational change is that it allows characterization of both sequential switching and concerted conformational change, depending on the time resolution needed to understand a particular mechanism.

Comparing this study to recent work in the Kern group²⁸ illustrates the robustness of MSMs for interpreting conformational change. Kern’s work used CHARMM force field instead of amber99sb, which caused an increase in conformational heterogeneity, especially in the inactive state. Also, the Kern study started from several structures along the activation pathway based on a targeted molecular dynamics (TMD) trajectory⁹, whereas the present study used unbiased molecular dynamics (MD) and adaptive sampling. In general, string methods allow sampling of long processes with less computation time than corresponding unbiased methods but increases the risk of discovering unphysical results. In this case, the string method of seeding states along the activation pathway allowed the Kern group to identify an inactive state in better

agreement with recent NMR, where F99 is correctly facing helix 3. Despite these notable differences in methodology, the major conclusions were in agreement with: the inactive is a collection of states interconverting rapidly, there are multiple activation pathways connecting the basins, and a kinetic, rather than structural, state definition is required to differentiate states. A more complete discussion including comparison figures can be found in the Kern study²⁸.

In summary, we have obtained a detailed atomistic view of the conformational transition of the key bacterial signalling protein, NtrC. The mechanistic insights obtained from this study suggest newly recognized molecular switches that play a central role in NtrC activation, such as the KQQ and Y94 switches. These results also suggest that residues far from the activating phosphorylation site and output domain play an important role in activation. The influence of such residues, such as K67, L66 and L63, could be validated using NMR and other spectroscopic methods. The results presented here also support a mechanism in which phosphorylation of D54 would lock the protein in active state, consistent with other work on this system⁶. Hydrogen bonding between phosphate group at D54, which is in proximity to the hydroxyl group of T82, would stabilize the Y101–T82 switch, which we observe to fluctuate without the presence of phosphate, and lock the entire protein in active state.

MSMs have been used extensively to study large-scale conformational changes involved in protein folding as well as more subtle changes involved in receptor activation and ligand binding^{29–36}. This study further supports the utility of MSMs as a natural framework for understanding and quantifying changes in hydrogen bond networks or residue side chain rotations associated with many protein conformational changes in signalling networks. Furthermore, these results indicate that *in silico* approaches²⁰ combined with advances in computer hardware³⁷ can be used successfully and routinely for obtaining mechanistic understanding of cellular signalling phenomena.

Methods

Simulation details. The starting structures for these simulations are the end result of a series of modelling procedures applied to the NMR structures 1DC7 and 1DC8 (ref. 5). In addition to the refinement described by Lei *et al.*⁹, helix 3 was completed and helix 4 was extended by two residues in order to match dihedral and distance constraints. The isomeric state of P104 has also been corrected. The structures were then parameterized using CHARMM27 (ref. 38) with cross-term energy correction map (CMAP) correction³⁹, solvated in TIP3P water⁴⁰, and neutralized with sodium ions. Nanoscale molecular dynamics (NAMD)⁴¹ was used to minimize the structure with a series of conjugate gradients, gradually releasing constraints from the protein. Heating was performed in the NVT (canonical) ensemble, gradually increasing the temperature from 50 to 300 K in increments of 25 K and simulating each step for 50 ps with a time step of 1 fs. Hydrogen atoms were constrained using SHAKE (Gauss-Seidel method which approximates the solution of the linear system of equations)⁴², and a non-bonded cutoff of 14 Å was used. The simulation was continued at 300 K in the NPT ensemble using Langevin dynamics with a time step of 2 fs, a damping coefficient of 5 per ps and a non-bonded cutoff of 12 Å. Pressure was controlled with a Nose-Hoover⁴³ piston with a 200-fs period of oscillation and a damping time of 100 fs. Electrostatics were computed using PME10 with a grid size of $72 \times 72 \times 72 \text{ \AA}^3$, a direct space tolerance of 10^{-6} and an interpolation order of 6.

The two initial structures were placed in a dodecahedron box with radius 10 Å from the proteins (edge length 75 Å) and solvated in 8,000 TIP3P water molecules. The system was then relaxed using a steepest decent algorithm using the Amber99SB⁴⁴ forcefield in the simulation package GROMACS 4.5 (ref. 45) for 500 steps with step size 1 fs. Next, the structures were equilibrated with restrained all protein atom positions for 100 ns with all bonds constrained using linear constraint solver (LINCS)⁴⁶. The equilibration used Berendsen pressure coupling with time constant of 1 ps, compressibility 4.5×10^{-5} and reference pressure 1 bar and the v-rescale temperature coupling to 300 K. Note that in GROMACS, the v-rescale option is velocity rescaling with a stochastic term, which ensures a canonical ensemble⁴⁷. The two equilibrated structures were used to start 30,000 parallel simulations on the distributed computing network Folding@home. These simulations used a 2-fs time step and a grid-based neighbour list was updated every 10 steps with a 10-Å cutoff. All bonds were again constrained using LINCS⁴⁶ with two iterations to fourth order. Production runs were done in the NVT ensemble

using v-rescale⁴⁷ to maintain system temperature of 300 K. Coulombic and van der Waals interactions were switched off between 8 and 8.4 Å. Linear centre of mass motion was removed every step. Random initial velocities were drawn from a Maxwell-Boltzmann distribution at 300 K.

After collecting 250 μs of aggregate data, the data from each conformation were clustered at an RMSD cutoff of 1.25 Å to generate a 6,838 state model for the active state and a 14,232 state model for the inactive state for a round of adaptive sampling¹⁶. Conformations representing the state centres were used to start additional simulations. In addition to the adaptive sampling, 15,000 new simulations were started from the active and the inactive conformations at a slightly increased temperature of 340 K in order to more quickly sample conformational space. In total, 7,917,355 snapshots were stored in increments of 0.2 ns totaling 1,582 μs of data.

All protein structures were rendered in visual molecular dynamics (VMD)⁴⁸ and figures created using Matplotlib⁴⁹ with the Numpy⁵⁰ scientific computing package.

Markov state models. MSMs use discrete time master equations models to map out molecular conformational space^{33,51}. Atomistic molecular dynamics simulations are first clustered based on geometric similarity, then kinetic information is used to group conformations that interconvert rapidly into metastable states. MSMs allow more automated simulation analysis that relies less on human intuition, an advancement over previous techniques. MSMs have been used to make direct connections from simulations to experiments, such as *ab initio* structure prediction for the villin headpiece and characterization of its conformational dynamics^{52–54}. The MSMbuilder 2.5 package⁵⁵ was used to construct a microstate model with 2,314 states. The microstate model was generated by clustering conformations at 2 ns intervals using the RMSD of all backbone heavy atoms as a metric and the k-centres clustering algorithm, followed by ten iterations of the local k-medoids algorithm, with a 2.25-Å RMSD cutoff. The remaining 90% of the data were then assigned to these clusters and used to construct a transition probability matrix. A second clustering based on 25% of the data gave a very similar state decomposition, which suggests the system is well sampled. Maximum Likelihood Estimator¹⁸ was used to generate a transition probability matrix T_{ij} which maps out the probability of transitioning from state i at time t to state j at time $t + \tau$, where τ is the lag time of the model. The Markov lag time is the time interval with demonstrated Markovian behaviour and can be determined by plotting the implied time scales (k) from eigenvalues μ of the transition probability matrix at varied lag times τ as $k = -\tau/\ln(\mu)$. The implied timescales should converge to an unchanged rate when Markovian behaviour is achieved, which corresponds to 4 ns for NtrC (Supplementary Fig. 1)⁵⁶. The first and second eigenvectors of the transition probability matrix T_{ij} provide, respectively, the equilibrium conformational state populations and the slowest dynamical process on the conformational landscape. MSM-weighted probability distributions, both one- and two-dimensional, are obtained by binning the raw data within each MSM state and weighting it by the MSM equilibrium state population. High-flux pathways between active and inactive states and MSM state committor values are also calculated from T_{ij} using the transition matrix definition of transition path theory^{57,58}. Committor values measure the probability of reaching the active state (at values near one) before returning to the inactive state (at values near zero). Intermediate committor values near 0.5 represent transition states capable of transitioning either back to inactive state or forward to the active state.

References

- Grant, B. J., Gorge, A. A. & McCammon, J. A. Large conformational changes in proteins: signaling and other functions. *Curr. Opin. Struc. Bio.* **20**, 142–147 (2010).
- Nygaard, R., Frimurer, T. M., Holst, B., Rosenkilde, M. M. & Schwartz, T. W. Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol. Sci.* **30**, 249–259 (2009).
- Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).
- Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
- Kern, D. *et al.* Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature* **402**, 894–898 (1999).
- Hastings, C. A. *et al.* High-resolution solution structure of the beryll fluoride-activated NtrC receiver domain. *Biochemistry* **42**, 9081–9090 (2003).
- Hu, X. & Wang, Y. Molecular dynamic simulations of the N-terminal receiver domain of NtrC reveal intrinsic conformational flexibility in the inactive state. *J. Biomol. Struct. Dyn.* **23**, 509–517 (2006).
- Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429–2433 (2001).
- Lei, M. *et al.* Segmented transition pathway of the signaling protein nitrogen regulatory protein C. *J. Mol. Bio.* **392**, 823–836 (2009).
- Villali, J., Pontiggia, F., Clarkson, M. W., Hagan, M. F. & Kern, D. Evidence against the ‘Y-T coupling’ mechanism of activation in the response regulator NtrC. *J. Mol. Bio.* **426**, 1554–1567 (2014).

11. Damjanović, A., Garca-Moreno, E. B. & Brooks, B. R. Self-guided Langevin dynamics study of regulatory interactions in NtrC. *Proteins Struct. Funct. Bioinf.* **76**, 1007–1019 (2009).
12. Lätzer, J., Shen, T. & Wolynes, P. G. Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles. *Biochemistry* **47**, 2110–2122 (2008).
13. Pan, A., Sezer, D. & Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **112**, 3432–3440 (2008).
14. Khalili, M. & Wales, D. J. Pathways for conformational change in nitrogen regulatory protein C from discrete path sampling. *J. Phys. Chem. B* **112**, 2456–2465 (2008).
15. Liu, M. S. *et al.* Coarse-grained dynamics of the receiver domain of NtrC: Fluctuations, correlations and implications for allosteric cooperativity. *Proteins Struct. Funct. Bioinf.* **73**, 218–227 (2008).
16. Weber, J. & Pande, V. Characterization and rapid sampling of protein folding Markov State Model topologies. *J. Chem. Theo. Comput.* **7**, 3405–3411 (2011).
17. Shirts, M. & Pande, V. Screen savers of the world unite! *Science* **290**, 1903–1904 (2000).
18. Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).
19. Bowman, G. R., Pande, V. S. & Noe, F. *An Introduction to Markov State Models and their Application to Long-Timescale Molecular Simulation* (Springer, 2014).
20. Shukla, D., Hernandez, C. X., Weber, J. K. & Pande, V. S. Markov state models provide insights into dynamic modulation of protein function. *Acc. Chem. Res.* **48**, 414–422 (2015).
21. Cho, H. S. *et al.* NMR structure of activated CheY. *J. Mol. Bio* **297**, 543–551 (2000).
22. Zhu, X., Rebello, J., Matsumura, P. & Volz, K. Crystal structures of CheY mutants Y106W and T87I/Y106W CheY activation correlates with movement of residue 106. *J. Biol. Chem.* **272**, 5000–5006 (1997).
23. Gardino, A. K. & Kern, D. Functional dynamics of response regulators using NMR relaxation techniques. *Methods Enzymol.* **423**, 149–165 (2007).
24. Gardino, A. K. *et al.* Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell* **139**, 1109–1118 (2009).
25. Koshland, Jr D., Nemethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–385 (1966).
26. Kenakin, T. Principles: receptor theory in pharmacology. *Trends Pharmacol. Sci.* **25**, 186–192 (2004).
27. Changeux, J.-P. & Edelstein, S. J. Allosteric mechanisms of signal transduction. *Science* **308**, 1424–1428 (2005).
28. Pontiggia, F. *et al.* Free energy landscape of activation in a signaling protein at atomic resolution. *Nat. Commun.* **6**, 7284 (2015).
29. Lane, T. J., Shukla, D., Beauchamp, K. A. & Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struc. Bio.* **23**, 58–65 (2013).
30. Kohlhoff, K. *et al.* Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* **6**, 15–21 (2014).
31. Shukla, D., Meng, Y., Roux, B. & Pande, V. S. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* **5**, 3397 (2014).
32. Pan, A. C. & Roux, B. Building markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **129**, 064107 (2008).
33. Noé, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struc. Biol.* **18**, 154–162 (2008).
34. Lawrenz, M., Shukla, D. & Pande, V. S. Cloud computing approaches for prediction of ligand binding poses and pathways. *Scientific Reports* **5**, 7918 (2015).
35. Lapidus, L. J. *et al.* Complex pathways in folding of protein g explored by simulation and experiment. *Biophys. J.* **107**, 947–955 (2014).
36. Harrigan, M. P., Shukla, D. & Pande, V. S. Conserve water: A method for the analysis of solvent in molecular dynamics. *J. Chem. Theo. Comput.* **11**, 1094–1101 (2015).
37. Eastman, P. *et al.* Openmm 4: a reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of chemical theory and computation* **9**, 461–469 (2012).
38. Brooks, B. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217 (2004).
39. MacKerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.* **25**, 1400–1415 (2004).
40. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
41. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comp. Chem.* **26**, 1781–1802 (2005).
42. Ryckaert, J.-P., Cicciotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327–341 (1977).
43. Nosé, S. & Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).
44. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).
45. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theo. Comput.* **4**, 435–447 (2008).
46. Hess, B., Bekker, H., Berendsen, H. & Fraaije, J. *et al.* LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472 (1997).
47. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
48. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
49. Perez, F., Granger, B. E. & Hunter, J. D. Python: an ecosystem for scientific computing. *Comput. Sci. Eng.* **13**, 13–21 (2011).
50. Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).
51. Bowman, G. R., Beauchamp, K. A., Boxer, G. & Pande, V. S. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* **131**, 124101 (2009).
52. Beauchamp, K. A., Ensign, D. L., Das, R. & Pande, V. S. Quantitative comparison of villin headpiece subdomain simulations and triplet-triplet energy transfer experiments. *Proc. Natl Acad. Sci. USA* **108**, 12734–12739 (2011).
53. Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struc. Biol.* **25**, 135–144 (2014).
54. Noé, F. Probability distributions of molecular observables computed from markov models. *J. Chem. Phys.* **128**, 244103 (2008).
55. Beauchamp, K. A. *et al.* MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theo. Comput.* **7**, 3412–3419 (2011).
56. Deuffhard, P., Huisinga, W., Fischer, A. & Schütte, C. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra Appl.* **315**, 39–59 (2000).
57. Weinan, E. & Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **61**, 391–420 (2010).
58. Noé, F., Schütte, C., Vanden-Eijnden, E., Reich, L. & Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl Acad. Sci. USA* **106**, 19011–19016 (2009).

Acknowledgements

D.K.V. is supported by a fellowship from the Center for Molecular Analysis and Design (CMAD) at Stanford University. D.S. and V.S.P. acknowledge support from the Simbios, NIH Center for Biomedical Computation (NIH U54 Roadmap GM072970). V.S.P. acknowledges NIH (R01GM62828). D.S. is supported by the Biomedical Data Science Initiative Fellowship from Stanford University School of Medicine. We thank the users of the Folding@Home distributed computing project who donated compute time used for the simulations. We thank Professor D. Kern for several insightful discussions on the NtrC activation and Francesco Pontiggia for providing initial structures for starting the simulations.

Author contributions

V.S.P. conceived and supervised the project. D.K.V. performed simulations. D.K.V., D.S. and M.L. analysed the data. D.K.V. and D.S. made figures and wrote the manuscript with inputs from M.L. and V.S.P.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npng.nature.com/reprintsandpermissions/>

How to cite this article: Vanatta, D. K. *et al.* A network of molecular switches controls the activation of the two-component response regulator NtrC. *Nat. Commun.* **6**:7283 doi: 10.1038/ncomms8283 (2015).