Informing NMR experiments with molecular dynamics simulations to characterize the dominant activated state of the KcsA ion channel

Sergio Pérez-Conesa,[1, a)] Eric G. Keeler,[2, a)] Dongyu Zhang,[2, a)] Lucie Delemotte,[1, b)] and Ann E. McDermott[2, b)]

[1)] *KTH Royal Institute of Technology, Science for Life Laboratory, Stockholm, Sweden*

[2)] *Department of Chemistry, Columbia University, New York, New York 10027, United States*

(Dated: 15 December 2020)

As the first potassium channel with a X-ray structure determined, and given its homology to eukaryotic channels, the pH-gated prokaryotic channel KcsA has been extensively studied. Nevertheless, questions related in particular to the allosteric coupling between its gates remain open. The many currently available X-ray crystallography structures appear to correspond to various stages of activation and inactivation, offering insights into the molecular basis of these mechanisms. Since these studies have required mutations, complexation with antibodies, and substitution of detergents for lipids, examining the channel under more native conditions is desirable. Solid-state NMR (SSNMR) can be used to study the wild-type protein under activating conditions (low pH), at room temperature, and in bacteriomimetic liposomes. In this work, we sought to structurally assign the activated state present in SSNMR experiments. We used a combination of molecular dynamics (MD) simulations, chemical shift prediction algorithms, and Bayesian inference techniques to determine which of the most plausible X-ray structures resolved to date best represents the activated state captured in SSNMR. We first identified specific nuclei with simulated NMR chemical shifts that differed significantly when comparing partially open vs. fully open ensembles from MD simulations. The simulated NMR chemical shifts for those specific nuclei were then compared to experimental ones, revealing that the simulation of the partially open state was in good agreement with the SSNMR data. Nuclei that discriminate effectively between partially and fully open states belong to residues spread over the sequence and provide a molecular level description of the conformational change.

[a)] These three authors contributed equally

[b)] Correspondance to: lucied@kth.se, aem5@columbia.edu

I. INTRODUCTION

Transmembrane allostery is at the heart of many signalling events in human biochemistry.[1] KcsA has proved to be an excellent model system for studying this intriguing phenomenon, and has also served as a prototype for ion channel function.[2] Activation coupled inactivation, a hallmark functional feature of essentially all ion channels, has been studied in KcsA by numerous biophysical and biochemical tools[3–11]. The literature in aggregate has resulted in a dominant hypothesis in which manipulation of an intracellular gate results in opening to a metastable activated state that lives for some milliseconds, during which time it directly triggers a slow allosteric inactivation involving loss of ions in the selectivity filter[3,5,12]. The details of the opening and coupled inactivation processes remain unclear and of great interest. Identifying the stable intermediates of these processes is central to clarifying the molecular basis for inactivation coupling in channels.

Solid-state nuclear magnetic resonance (SSNMR) studies of KcsA have provided unique insights regarding dynamics and thermodynamics of full length wild-type channels in hydrated lipid bilayers. These studies have the advantage of being carried out in near-native conditions, without use of antibodies and with freely varied ligand concentrations ([K^+] and [H^+]). In our NMR studies, we have stabilized three states, which we have referred to as deactivated (stabilized at neutral pH and high [K^+]), activated (stabilized at low pH and high [K^+]) and inactivated (stabilized at low [K^+]), with a naming convention selected because of the functional consequences of varying pH and [K^+] (Fig. 1.a). An important prerequisite for interpretation of the NMR data is identifying (for each of these experimental conditions) the atomic arrangement that best represents the majority species in the sample. Identification of the deactivated and inactivated states is largely straightforward given the sample conditions used and the stability of these states: both X-ray diffraction (XRD) and solid-state NMR data suggest that the samples are homogeneous and robustly reproducible. Here the deactivated state corresponding to a channel with a closed activation gate (also called hydrophobic bundle crossing or inner helix bundle) and a conductive selectivity filter fully loaded with potassium ions (PDB ID 1K4C)[13–17]. The inactivated state corresponds to a channel with a wide open inner gate and a pinched selectivity filter depleted of ions (PDB ID 5VKE, 5VKH, 3F7V, 3F7Y, 3F5W)[4,9,18,19] (Fig. 1.a).

The activated state, the only state that conducts ions, on the other hand, presents a particular challenge. Indeed, the seminal crystallography work by Cuello et al. showed that the development of a wide open activation gate was coupled to the appearance an inactive selectivity filter, and many

other studies showed that the selectivity filter tended to collapse when the gate was held open, an observation compatible with the metastable short-lived nature of the activated state.[3,6,18] Although the activated state has proven difficult to stabilize, X-ray crystallography studies involving a number of mutations and antibody binding succeeded in stabilizing a family of configurations with a conductive selectivity filter and the activation gate having a range of degrees opening, suggesting an almost continuous opening process[18]. A mutant channel (where all but one of the proton sensor charges were neutralized) was captured together with a conductive selectivity filter/Partially Open (PO) pore, with degrees of opening varying between 14 Å and 16 Å cross bundle distances at the narrowest site in the pore (measured as the distance between T112 C_α of opposing subunits) (PDB IDs 3FB5 3FB6)[18]. Recently a non-inactivating mutant E71A structure was engineered with cysteines at positions 28 and 118 forming disulfide bonds to capture the channel in a Fully Open (FO) state, revealing a wider pore, of 23 Å cross bundle distance at the inner gate (PDB ID 5VK6). In this structure, interestingly, the gate was as wide open as in the inactivated state structure[4].

These results raise a question about the degree of opening in the low pH near-native condition SSNMR activated samples. *De novo* structure determination of such a system from NMR restraints has not yet been reported. In this work, we have used an alternative approach to compare the compatibility of configurational ensembles obtained using MD simulation initiated from XRD structures with the NMR spectra. Chemical shifts (CS) have been shown to be very sensitive to biopolymer conformation, and in some cases contain sufficient information to correctly identify 3-dimensional structures.[20–23] The use of chemical shift prediction tools has also been refined to account for averaging of the experimental chemical shifts due to internal dynamics.[24–26] Therefore we leverage this information to identify the most likely structure for the NMR activated state. To do so, we ran MD simulations based on the two most likely representative XRD structures of the activated state (Partially Open PO, 3FB5 and Fully Open FO 5VK6) and the deactivated state (Closed C, 5VKH) and predicted the ensemble of CS sampled by the trajectories of the NMR-assigned nuclei in the different residues. We then calculated relative chemical shifts ($\Delta CS^{X}_{sim} = CS^{X}_{sim} - CS^{Closed}_{sim}$, where X represents the FO or PO state), which are expected to be less affected by systematic errors than absolute chemical shifts. Using Bayesian inference, we determined signals that distinguished the PO and FO ensembles. We then calculated the deviation of these 49 ΔCS^{X}_{sim} from the ΔCS_{exp} derived from the SSNMR experiments, leading us to conclude that the activated state probed in the SSNMR is closer to the PO than the FO state.

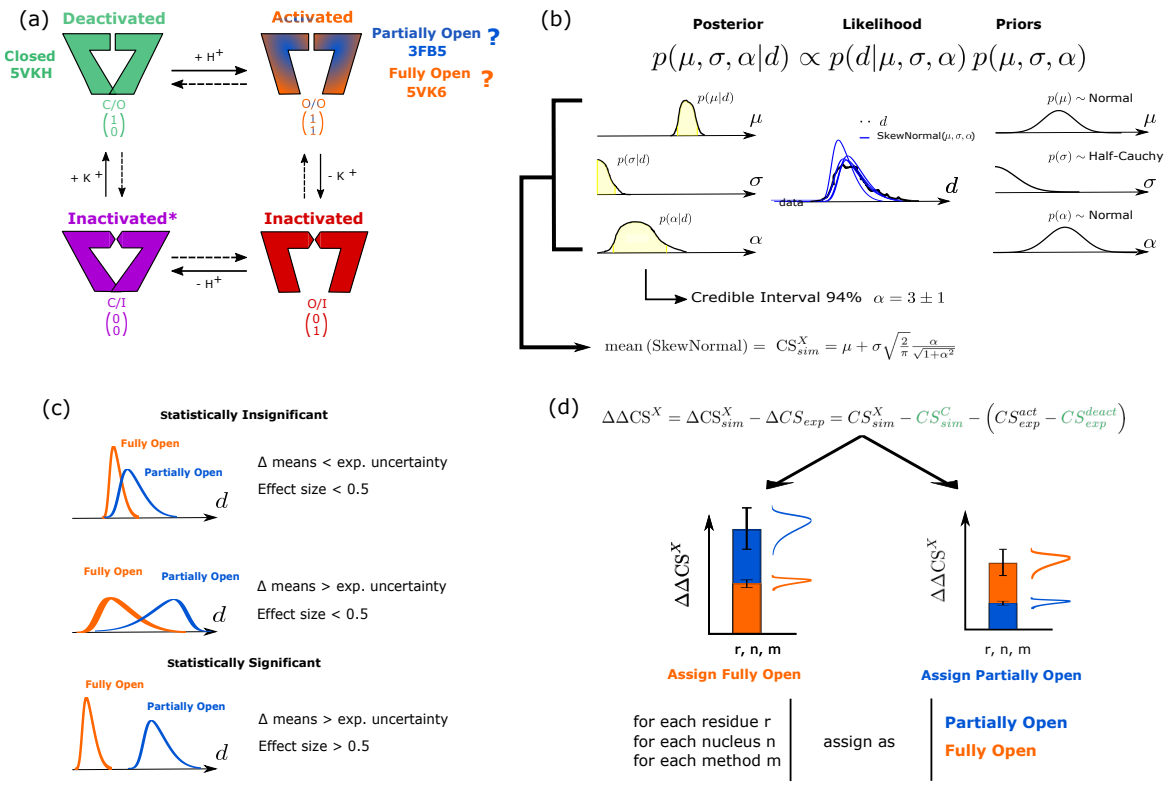

FIG. 1. Summary scheme of the various computational steps carried out in this work. a) Illustration of the gating cycle of KcsA and the corresponding X-ray crystallography structures with the nomenclature used for the different states. b) Bayes law (top equation) and its terms: the posterior distribution ($p(\mu,\sigma,\alpha|d)$) is proportional to the product of the likelihood ($p(d|\mu,\sigma,\alpha)$) and the priors ($p(\mu,\sigma,\alpha)$). The posterior probability of the parameters of a skew-normal distribution representative of the CS calculated from the MD simulation ensemble is proportional to the product of the likelihood of the data (predicted chemical shifts) by the prior distribution of the parameters (chosen here as uninformative). The definition of the mean of this skew-normal distribution that is reported as the $\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$ with its 94% credible interval. c) Criteria imposed to consider a given residue's CS as statistically significant. This is illustrated using two idealized CS-likelihoods obtained from the posteriors inferred for the two possible activated states, showing how their overlap and location affects statistical significance. This was done by assessing the difference in means between the two states ($\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}} - \mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}}$) and the effect size of this difference $\left(\frac{\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}} - \mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}}}{\mathrm{std}_{pool}}\right)$ where $\mathrm{std}_{pool} = \sqrt{\frac{var(\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}}) - var(\mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}})}{2}}$ and *var* is the variance function). d) Assignment of nuclei indicating compatibility of the activated state NMR sample with the PO or FO MD simulation ensembles. The method we use involves calculating the CS difference between the C and the PO/FO states (labeled X) measured using SSNMR ($\Delta\mathrm{CS}_{\mathrm{exp}}$) and predicted from the MD simulations ensemble ($\Delta\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$) for the nuclei determined in c). If the difference between the ΔCS measured experimentally and using the simulation ensemble ($\Delta\Delta\mathrm{CS}^{\mathrm{X}}$) is smaller for the PO than the FO state, this nucleus supports the conclusions that the experimentally observed activated state is the PO state seen in crystal structures. The reverse is true if the difference is smaller for the FO state.

II. METHODS

A. Expression and purification of U-^{13}C,^{15}N wild-type KcsA

Uniformly ^{13}C,^{15}N labeled KcsA was overexpressed and purified as described previously from *Escherichia coli* PASK90/JM83 and PASK90/BL21(DE3) cells in M9 minimal media.[2,12,14] Purified, uniformly ^{13}C-^{15}N labeled, KcsA in detergent dodecyl β-D-maltoside (DM) (Anatrace,

Maumee, OH) was reconstituted into liposomes by reducing the detergent concentration through dialysis; the liposomes consisted of 3:1 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) : 1,2-di-(9Z-octadecenoyl)-sn-glycero-3-phospho-(10-rac-glycerol) (DOPG) (Avanti Polar Lipids, Alabaster, AL) at pH 4.0 and 50 mM KCl with a lipid to protein (*w/w*) ratio of 1:1. The total ionic strength was kept at 100 mM by compensation with NaCl.

B. NMR Spectroscopy

25 (or 9) mg of hydrated proteoliposome sample was centrifuged into a Bruker 3.2 (or 1.9) mm rotor after 3-5 freeze-thaw cycles. SSNMR experiments were performed on either a Bruker AVANCE II or NEO spectrometer operating at 21.1 T ($\omega_{0,\mathrm{H}}/2\pi$ = 900 MHz) or 17.6 T ($\omega_{0,\mathrm{H}}/2\pi$ = 750 MHz). Experiments were collected with either a 3.2 mm HCN E-free probe (900 MHz) or a 1.9 mm HCN probe (750 MHz) (Bruker BioSpin, Billerica, MA), as indicated.

Experiments at 900 MHz were collected with a spinning frequency, $\omega_r/2\pi$, of 16666 ± 10 Hz and a set point temperature of 267 ± 1 K. The sample temperature is estimated to be 290 ± 5 K due to sample heating from MAS and RF pulses.[27] Standard $\pi/2$ pulse lengths of 2.55 µs, 4 µs, and 5.1 µs were used for the ^{1}H, ^{13}C, and ^{15}N channels, respectively, corresponding to $\omega_1/2\pi$ = 98 kHz (1H), 62.5 kHz (^{13}C), and 49 kHz(^{15}N). Two-dimensional (2D) ^{13}C-^{13}C, 2D ^{13}C-^{15}N (NcaCX, NcoCX), and three-dimensional (3D) ^{13}C-^{13}C-^{15}N (NCOCX, NCACX) correlation experiments were used to determine the chemical shift assignments.[28–30] The ^{13}C-^{13}C mixing was performed using dipolar assisted rotational resonance (DARR) mixing of 50 ms.[31] The heteronuclear polarization transfer used SPECtrally Induced Filtering In Combination with Cross Polarization (SPECIFIC-CP).[32] SPECIFIC-CP standard RF conditions of $\omega_{1,\mathrm{C}} = 1.5\omega_r$ and $\omega_{1,\mathrm{N}} = 2.5\omega_r$ for NCA, and $\omega_{1,\mathrm{C}} = 2.5\omega_r$ and $\omega_{1,\mathrm{N}} = 1.5\omega_r$ for NCO and contact times between 4.5 and 5.0 ms were used for SPECIFIC-CP experiments. ^{1}H decoupling of $\omega_{1,\mathrm{H}}/2\pi > 90$ kHz during acquisition was performed using SPINAL-64 decoupling.[33] More detailed experimental parameters for the 3D experiments are given in Table S1.

Experiments at 750 MHz were collected with spinning frequencies of 16000 ± 10 or 33333 ± 10 Hz and a set point temperature of 230 ($\omega_r/2\pi$ = 33 333 Hz) or 240 ($\omega_r/2\pi$ = 16 000 Hz) $\pm$ 1 K. The sample temperature is estimated to be 290 ± 5 K due to sample heating from MAS and RF pulses.[27] Standard $\pi/2$ pulse lengths of 2.55 µs, 3.3 µs, and 4.3 µs were used for the ^{1}H, ^{13}C, and ^{15}N channels, respectively, corresponding to $\omega_1/2\pi$ = 98 kHz (^{1}H), 76 kHz(^{13}C),

and 58 kHz (^{15}N). The same 3D experiments for chemical shift assignments were performed at 750 MHz. The ^{13}C-^{13}C mixing at 750 MHz was performed using COmbined $R2_n^v$-Driven (CORD) mixing of 50 ms.[34] SPECIFIC-CP RF conditions of $\omega_{1,C} = 0.5\omega_r$ and $\omega_{1,N} = 1.5\omega_r$ for NCA and NCO at $\omega_r/2\pi = 33\,333$ Hz and $\omega_{1,C} = 1.5\omega_r$ and $\omega_{1,N} = 2.5\omega_r$ for NCA, and $\omega_{1,C} = 0.5\omega_r$ and $\omega_{1,N} = 1.5\omega_r$ for NCO at $\omega_r/2\pi = 16\,000$ Hz. SPECIFIC-CP contact times of between 4.5 and 5.0 ms were used. A 2D ^{13}C-^{15}N ZF-TEDOR experiment[35] was collected with 1.5 ms mixing and $\omega_r/2\pi = 16000$ Hz for identification of the proline chemical shifts. ^{1}H decoupling of $\omega_{1,H}/2\pi > 90$ kHz during acquisition was performed using swept-frequency two-pulse phase-modulated (SW$_f$-TPPM) decoupling.[36] More detailed experimental parameters for the 3D experiments are given in Table S2.

All data were processed in Topspin (Bruker BioSpin) or NMRPipe.[37] The ^{13}C and ^{15}N chemical shifts were referenced via the substitution method to the methylene peak of adamantane at 40.48 ppm (0.5% DSS in D_2O scale), and to the peak of $^{15}NH_4Cl$ to 39.3 ppm (at 298 K, NH_3 (liquid) scale), respectively.[38–41] Peak picking and assignments were made in CcpNmr.[42] All the spin systems were identified first in each 3D spectrum and then the residue type and their secondary structure were predicted using PLUQin.[43] Backbone resonance assignments were determined based on correlations in the 3D NCACX, NCOCX, and CANCO spectra, beginning from unique spin systems (for which the residue type was identified using PLUQin). Sidechain residue assignments were determined mainly based on 3D NCACX, NCOCX, and 2D ^{13}C-^{13}C DARR experiments. The following tolerance thresholds were applied for the agreement of chemical shifts when comparing the various spectra of the activated state for assignments: 0.2 ppm for ^{13}C, and 0.5 ppm for ^{15}N. Monte Carlo/simulated annealing (MC/SA) algorithms in the software packages NSGA-II and MC/SA were used to facilitate assignments.[44]

C. Molecular dynamics simulations

Coordinates for PO, FO and C KcsA structures (PDB ID: 3FB5, 5VK6, 5VKH)[4,18] were prepared using CHARMM-GUI[45]. The constructs included residues 26-121. For 3FB5 Rosetta[46] was used to add missing residues 114-121 in alpha-helical conformation. The N- and C- termini were capped by methylation (NME) and acetylation (ACE). We note that the pH sensor was not included in the simulations to keep consistent the chain-lengths across states. Using PYMOL[47] the following mutations were reverted: for the FO state (5VK6), A28C, E71A, L90C, R117Q, E118C,

E120Q, R121Q; for the PO state (3FB5), L90C, R117Q, E120Q, R121Q and for the closed state (5VKH), Y82A, L90C, F103A. None of the residues that were mutated in the activated states (3FB5 and 5VK6) were later found to be important to discriminate between the PO and FO states. For all states, residues E71, E118 and E120 were protonated since H-NMR spectra showed these residues to be protonated in the activated state.[15,48] We note that, the simulated protein contains no histidines due to the truncation of the sequence. The selectivity filter was initially fully loaded with K^+ in agreement with the X-ray diffraction and crystallization waters bound to the channel were included. The protein was embedded in a DOPE:DOPG (3:1) membrane bilayer containing 150 lipids and surrounded by a neutralizing solution of 50 mM NaCl and 50 mM KCl at a ratio of 75 TIP3P[49] rigid water-molecules for each lipid molecule. All states were modelled including the co-purifying DOPG lipids, as identified in the X-ray diffraction structures.

The systems were equilibrated using the CHARMM36 force field[50] in an extended CHARMM-GUI protocol. 1µs-long simulations were performed using GROMACS2019[51]. The Parrinello-Rahman barostat[52] ($P = 1\,\text{bar}$ $\tau = 5\,\text{ps}$) and Bussi thermostat[53] ($T = 290\,\text{K}$ $\tau = 1\,\text{ps}$) were employed to mimic experimental conditions. The parts of the simulation after a water molecule enters the selectivity filter were excluded since such is a symptom of inactivation (see Fig S1)[6]. Additionally, the initial parts of the simulation where the inner-gate distance was unstable were also excluded from analysis (see Fig S2). Additional technical details of the simulation can be obtained from the simulation files freely available at `https://github.com/delemottelab/Informing_NMR_experiments_w_MD` and `https://osf.io/6h2z5/`.

D. Statistical Inference of Chemical Shifts from MD simulations

N, C, C_α , and C_β CS of the simulation snapshots were calculated using the SPARTA+[54] and SHIFTX2[55] chemical shift prediction tools, with their default options. Only nuclei experimentally assigned for both states were considered. Since KcsA is a homotetramer, we were able to collect four times the data per residue under the assumption that the trajectories of the monomers were independent. Nevertheless, since there are contact surfaces between the monomers that can affect the CS prediction, the prediction software was given the full tetramer and the data obtained for the four equivalent residues was concatenated *a posteriori*. The CS-prediction error for each MD simulation snapshot was not modelled. The overall agreement of the conclusions drawn for both methods highlights the general robustness of the methods.

The statistical distributions of the CS were modelled using Bayesian inference with the python libraries PYMC3[56] and Arviz[57]. Bayesian inference assumes that the probability distribution of a random variable has a given functional form and enables to infer the probability of the parameters of the distribution given the data sampled (d). A schematic description of the procedure is found in Fig. 1b . The CS of each particular nucleus, state, residue, and prediction method was considered as an independent variable and was therefore modelled separately in a non-hierarchical framework. For every variable, two thousand data points were selected randomly out of all the time frames obtained from each MD trajectory.

The likelihood function chosen was a skew-normal distribution as implemented by PMC3:

$$L(d|\,\mu,\sigma,\alpha) = \frac{2}{\sigma}\cdot\phi\left(\frac{d-\mu}{\sigma}\right)\cdot\Phi\left((d-\mu)\,\frac{\alpha}{\sigma}\right)\cdot \tag{1}$$

where d is the data sampled for the $\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$ of a particular nucleus, and for a particular residue in that particular state X. ϕ and Φ are the probability distribution function and cumulative distribution function of the Gaussian distribution:

$$\phi(x) = \frac{1}{\sqrt{2}}e^{-\frac{1}{2}x^2} \tag{2}$$

$$\Phi(x) = \frac{1}{2}\left[1+\frac{2}{\sqrt{\pi}}\int_0^{x/\sqrt{2}} t^2 dt\right] \tag{3}$$

The skew-normal function has three parameters: the location (μ), the scale parameter (σ), and skewness parameter (α). In order to conduct Bayesian inference, previous assumptions of the parameters are included by assigning probability distribution functions, *priors*, to them. For the μ parameters, normal distributions are assigned:

$$N(\mu) = \frac{1}{10\,\mathrm{ppm}\sqrt{2}}e^{-\frac{1}{2}\left(\frac{\mu-\bar{d}}{10\,\mathrm{ppm}}\right)^2} \tag{4}$$

Where $\bar{d}$ is the average value of the sampled $\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$ data (d) and 10 ppm the scale of the normal distribution. For σ, the priors assigned are half-cauchy distributions, assigned with a beta parameter of 10 ppm:

$$C(\sigma) = \frac{1}{\pi 10\,\mathrm{ppm}\left[1+\left(\frac{\sigma}{10\,\mathrm{ppm}}\right)^2\right]}\forall\sigma\geq 0 \tag{5}$$

For α, normal distributions are also assigned:

$$N(\alpha) = \frac{1}{10\,\mathrm{ppm}\sqrt{2}}e^{-\frac{1}{2}\left(\frac{\alpha}{10\,\mathrm{ppm}}\right)^2} \tag{6}$$

The posterior distribution was sampled using the NUTS[58] algorithm with 4 chains. It is important to remark that the mean of a skew-normal distribution is not the location parameter μ but rather $mean = \mu + \sigma\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}$ .This mean value of the distribution is what we will consider the chemical shift $\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$ of state X predicted by our MD simulation, or in other words the simulated position of the peak in the NMR spectrum. This corresponds to a physical assumption of rapid population weighted average of the chemical shifts sampled in the protein dynamics. In similar fashion, the distribution variance was calculated as $var = \sigma^2\left(1 - \frac{2\alpha^2}{(\alpha^2+1)\pi}\right)$. Posterior predictive checks were carried out after the inference to test the adequacy of the model (Fig S3). The main features of the data distributions are generally in good agreement with the model. Additional tests and inference information can be found in the SI and `https://github.com/delemottelab/Informing_NMR_experiments_w_MD`.

All quantities obtained with Bayesian inference are not assigned a single value, but rather a probability distribution. As a consequence, for simplicity, the distributions of these random variables can be summarized by the interval that contains 94% of its probability. This interval is known as the 94% credible interval (CI) and is written as its center plus or minus the distance to the bounds of the interval. A graphical representation can be found in Fig 1b.

E. Selecting nuclei whose CS differs statistically between states

The next step was to determine which nuclei were statistically different in the PO and FO trajectories. For this, the difference between the simulated chemical shifts for both states modeled as described above was calculated ($\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}} - \mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}}$), as well as the effect size of this difference $\left(\frac{\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}} - \mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}}}{\mathrm{std}_{pool}}\right)$ where $\mathrm{std}_{pool} = \sqrt{\frac{var(\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}}) - var(\mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}})}{2}}$ and var is the variance function. The first expression measures if there is a difference between the CS predicted for both states and the second if this difference is large compared to the variance of the distributions. The importance and graphical interpretation of the difference in means and effect size can be seen in Figure 1c.

A CS variable for a given nucleus in a given residue generated using a given chemical shift prediction method is considered to be have discrimination power if the center of the difference in means credible interval is greater in absolute value than the experimental tolerance (0.2 ppm for C, C_α , C_β and 0.5 for N) and the center of the effect size credible interval is larger than 0.5 in absolute value. An effect size greater than 0.5 is considered medium as a general rule in statistics[59]. This procedure is repeated independently for each studied nuclei, for each chemical

shift prediction method, and for each residue.

F. Comparing CS differences from MD simulations and NMR

Site specific CS for the PO and FO states were contrasted with experimentally measured CS to assess which state is present in the experiment. A schematic representation of the procedure can be found in Fig. 1d. 94% credible intervals were calculated for both the FO and PO states (designated "X" since it is to be determined which is correct), using the expressions $\Delta CS^{X}_{sim} = CS^{X}_{sim} - CS^{C}_{sim}$, $\Delta CS_{exp} = CS^{pH=4,act}_{exp} - CS^{ph=7.5,deact}_{exp}$, and $\Delta\Delta CS^{X} = \Delta CS_{exp} - \Delta CS^{X}_{sim}$. The state, X, for which $\Delta\Delta CS^{X}$ is lower in magnitude has a predicted ΔCS^{X}_{sim} closer to the experimentally measured ΔCS_{exp} and we conclude it to be likely that this state is dominant in the experiment at that nucleus. This procedure was repeated for all nuclei in all residues for which experimental shifts were available, using both CS prediction software packages.

III. RESULTS

A. NMR CS Assignments of the activated state sample

Chemical shift assignments of the activated state (pH 4.0, 50 mM [K^+]) of KcsA were determined using three-dimensional spectra (NCACX, NCOCX, and CANCO) acquired at two different external magnetic fields (900 and 750 MHz), (see Table S3). Representative strip plots of the backbone walk for the assignments of residues T74 to G79 using these spectra are given in Figure S5. The completeness of the backbone and sidechain assignments, along with the experimental connectivities of each resonance in each experiment, are presented in a schematic representation in Figure S6. A number of the assigned resonances, including some that are used in the analysis presented here, are well resolved in various 2D spectra, as illustrated in Figures 2 and S7. Chemical shifts for the deactivated state at pH 7.5, 50 mM [K^+] (previously reported,[15] Table S4) were re-referenced using the protocol described in this paper so as to be comparable to the shifts determined here for the activated state (pH 4.0, 50 mM [K^+]). Distributions of the chemical shift differences (comparing activated and deactivated states) indicate that the chemical shift referencing is consistent for the two datasets (Figure S8). In fact, the site-specific chemical shifts were found to be similar for the activated and deactivated states, with few notable exceptions, suggesting that the two states adopt similar structural folds, as has been indicated from X-ray crystallography

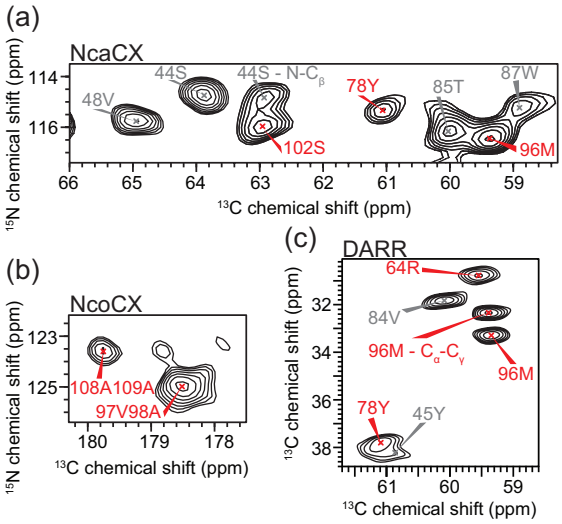

FIG. 2. Representative experimental NMR spectra of KcsA in the activated state (3:1 DOPE:DOPG, 50 mM KCl, pH 4.0). (a) NCA region of 2D ^{15}N-^{13}C NcaCX spectrum with assigned peaks shown. All peaks in the NcaCX spectrum are N^i-C^i_α unless noted otherwise. (b) NCO region of 2D ^{15}N-^{13}C NcoCX spectrum with assigned peaks shown. All peaks in the NcoCX spectrum are N^i-C'^{i-1} unless noted otherwise. (c) C_α -C_β region of 2D ^{13}C-^{13}C DARR spectrum with assigned peaks labeled. All peaks in the DARR spectrum are C^i_α-C^i_β unless noted otherwise. Peaks used to distinguish between fully open vs partially open (identified by statistical inference analysis of the CS predictions of MD snapshots) are labeled in red, and those identified (by the same methods) as spectators are labeled in gray.

studies (Table S5). Resonances were considered to be experimental markers if the chemical shift difference (activated vs. deactivated, $CS_{exp}^{pH\ 4,\ act} - CS_{exp}^{pH\ 7.5,\ deact}$) was greater than the experimental tolerance (0.2 ppm for ^{13}C and 0.5 ppm for ^{15}N). If this difference does not reach this threshold the assigned resonance was considered to be an experimental spectator.

B. Sampling the deactivated and activated KcsA conformational ensemble using MD simulations

In order to run MD simulations to characterize the activated state, the most plausible XRD structures were selected as starting coordinates. There is a rich set of such structures for KcsA, which mostly differ in their inner gate opening (measured as the distance between T112 C_α from opposing subunits): 3FB5 (14 Å), 3FB6 (16 Å), 5VK6 (22 Å) and 3F5W (31 Å)[4,18]. In this work, the 3FB5 structure was used instead of 3FB6 because of its higher resolution and therefore fewer missing side chains. Nevertheless, the two structures are very similar and the MD simulation initiated from 3FB5 coordinates explores openings compatible with 3FB6 (Fig. S3). 3F5W was found to have an unstable inner-gate that moved towards openings compatible with 5VK6 and therefore was not included in this study (Fig. S4). Therefore, the initial configurations were prepared based

on the 3FB5 PO structure, and from the FO 5VK6 structure. We also ran simulations of the C state (conductive selectivity filter, closed inner gate), to compare the activated state to, preparing a system based on the 5VKH high resolution crystal structure.

The simulation conditions were chosen to match the experimental conditions as closely as possible. The lipid and solution composition in the computational system were identical to the NMR experiment: 3:1 DOPE:DOPG in a symmetrical bilayer and a concentration of 50 mM of NaCl and 50 mM of KCl in the solution. The simulation temperature and pressure were set to 290 K and 1 atm, respectively, thus close to experimental conditions. To mimic the pH of the experiments, acidic residues E71, E118, and E120 were protonated and the protonation states were kept fixed due to limitations of classical MD simulations. The main difference between the simulated and experimental constructs is the truncation of the sequence at residues 25 and 121 due to the lack of experimental coordinates to model the intracellular domains. In the three states, the models were overall stable over the hundreds of nanoseconds timescale, as characterized by the RMSD (Fig. S9).

The opening of the inner gate was found to be fairly stable for both the C and FO (Fig S3). In contrast, the PO state inner gate experienced some initial instability, followed by asymmetrization, and finally a drift towards closing. For this reason data collected between 400 ns and 1000 ns, corresponding to a stable opening, was used in the analysis. This observation is consistent with observations from previous MD simulation studies[3]. The conductive selectivity filter of all three states was fully loaded with K^+ at the start of the simulation and was found to rapidly evolve to the WKK0KW occupancy (Fig. S1). In the FO state simulation, a water molecule entered the selectivity at t=400 ns. Water penetration in the selectivity filter has been linked to inactivation, particularly in this inactivation-prone wide-gate state[6]. Therefore the rest of the simulation data was discarded. In contrast, consistent with previous observations[3], for the C state, no instabilities in the structure or the selectivity filter occupancy were noticed such that the full trajectory length was considered for the analysis. The simulation data for the different states that were used for CS analysis is summarized in figure (Fig. S1).

C. Chemical Shift Prediction from the MD Results

We used SPARTA+[54] and SHIFTX2,[55] two well established tools, to predict the chemical shifts expected for the ensemble of structures observed in the MD trajectories. Predictions were made

based on numerous frames from the MD trajectory, and subsequently the average and the distribution in predicted shifts was analyzed, following a procedure analogous to Robustelli et al.[24] The two tools (SPARTA+ and SHIFTX2) produced chemical shifts in qualitative agreement; however, the predicted chemical shifts were not identical for the two tools (Figure S10). The results from both methods were thus analyzed separately to ensure robustness of the conclusion. We note that direct comparison of the CS predicted directly from the XRD structures does not give a clear picture of which state is closer to experiment (Fig. SS11)

The distribution of the predicted CS based on the MD simulation trajectory is strikingly non-Gaussian as can be seen in Figures S12 - S19. This fact introduces difficulty in assessing the statistical significance of differences, and a need for a robust method to assess the statistical significance of changes in CS between simulated states. Therefore, we chose to model the CS distributions of the simulation using Bayesian inference (Fig. 1b).

The goal of Bayesian inference is to numerically calculate the *posterior* distribution of the parameters of the distribution of CS shifts (chosen here as a skew-normal distribution, see Methods, Figure 1b), $\mathrm{p}(\mu, \sigma, \alpha|\mathrm{d})$, given our data (which is here the CS predictions from MD snapshots). The posterior distribution is thus a measure of the uncertainty of the CS shifts distribution parameters considering the data we have collected. Note that we assign uncertainty to our knowledge of the parameters and not to the data directly, contrary to a classical statistics approach in which parameters that we assume to be certain are estimated, and then the probability to obtain the data sampled calculated by a p-value.

To calculate the posterior probability we must choose the terms of the right hand side of Bayes law. The *likelihood* term or probability of the data given the parameters, $\mathrm{p}(d|\mu, \sigma, \alpha)$, is specified by choosing the distribution with which we model our data presented in the central column of Figure 1b. After the inference has been carried out, the posterior distribution of the parameters can be sampled to obtain the ensemble of skew-normal distributions inferred. This ensemble can be compared to the data to see the adequateness of the model. This procedure is known as *posterior predictive check*. As Figures S12 - S19 shows, our data is adequately characterized by skew-normal distributions. A skew-normal distribution can be defined by three parameters: the location (μ), the scale parameter (σ), and skewness parameters (α).

Finally the posterior distributions are estimated numerically using the NUTS[58] algorithm. This algorithm also calculates the normalization factor $p(d)$ or the probability of the data. Once these posterior distributions are known, we can calculate the probability distributions of the mean of the

skew-normal distribution $\left(\mu+\sigma\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}\right)$, which we ascribe to the NMR-peak position and call $\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$, where X is the state simulated. $\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$, like all quantities obtained with Bayesian inference, is thus not assigned a single number but rather a probability distribution. As a consequence, for simplicity, the distributions of these random variables can be summarized by the interval that contains 94% of its probability. This interval is known as the credible interval (CI) and is written as its center plus or minus the distance to the bounds of the interval. This is analogous to the confidence intervals in classical statistics.

D. Activated State Characterization Through Statistical Analysis

We approached the question of what structure best represents the activated samples studied by NMR spectroscopy by asking two questions. First, we asked which chemical shifts are predicted to distinguish the FO and the PO states. Then we asked whether the NMR shifts corresponding to the activated state suggest better agreement with the FO or the PO configuration.

Many of the predicted $\mathrm{CS}^{\mathrm{X}}_{\mathrm{sim}}$ shifts do not change sufficiently between the PO and FO states to offer any statistical power of discrimination (Figure S20,S21). We thus developed the following protocol to identify nuclei capable of discriminating between the two states.

The difference between the simulated chemical shifts of both states ($\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}}-\mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}}$) and the effect size of this difference $\left(\frac{\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}}-\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}}}{\mathrm{std}_{pool}}\right)$, where $\mathrm{std}_{pool}=\sqrt{\frac{var(\mathrm{CS}^{\mathrm{FO}}_{\mathrm{sim}})-var(\mathrm{CS}^{\mathrm{PO}}_{\mathrm{sim}})}{2}}$ and var is the variance function, were calculated. The first expression estimated whether there was a difference between the CS predicted for both states and the second if this difference was large compared to the variance of the CS distributions. A graphical interpretation of this procedure can be found in Figure 1c.

An example of these quantities for C_α and the chemical shift prediction method SPARTA+ is presented in Figure 3. In this figure, we present the difference in means (left graph) and the effect size (right graph) for the different amino-acids whose C_α CS have been assigned in the experiment corresponding to activated and deactivated states.

A CS variable of a given nucleus in a given residue and generated by a given chemical shift prediction method is considered as having discrimination power if the center of the CS difference CI is greater in absolute value than the experimental tolerance (0.2 ppm for C, C_α , C_β and 0.5 for N) and the center of the effect size CI is larger than 0.5 in absolute value. This procedure is repeated independently for each studied nucleus, for each chemical shift prediction method, and

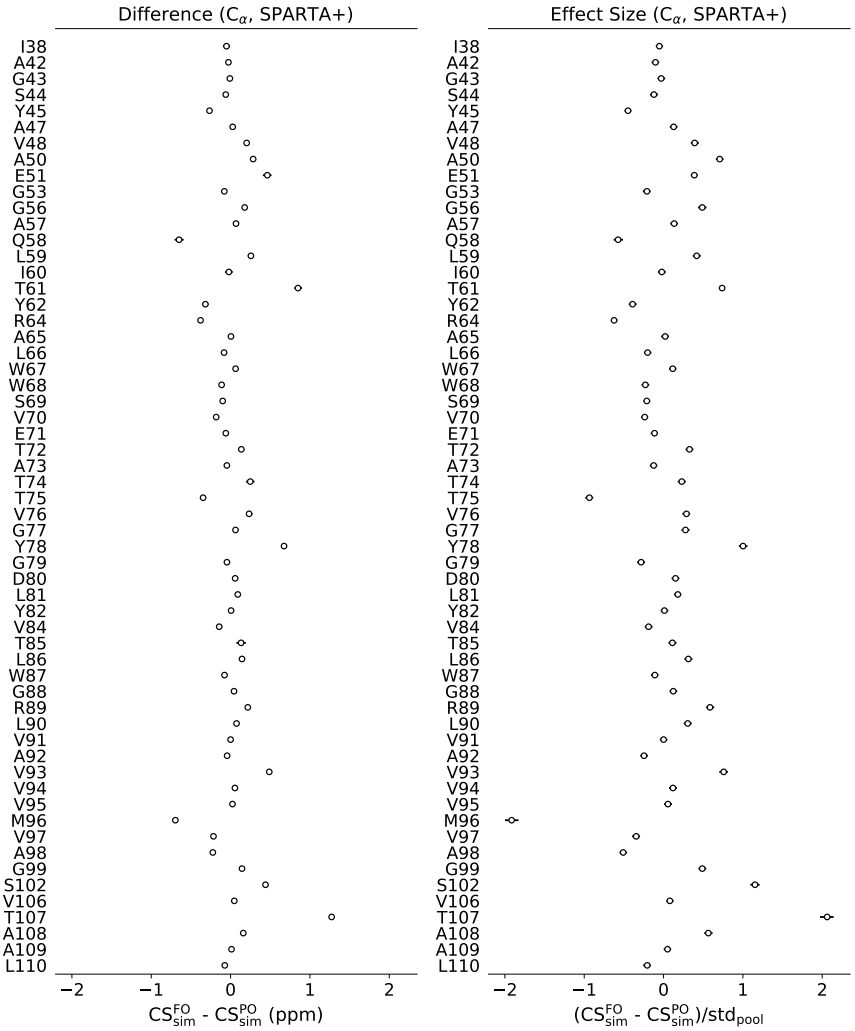

FIG. 3. Statistical filtering of simulated CS. Horizontal bars represent the 94% credible interval of the variable distribution and circles represent its center. The vertical axis represents the residues that are part of the experimental dataset in both pH conditions. We show here CS differences for the C_α nuclei using the SPARTA+ chemical shift prediction method. Left: credible interval of CS^{X}_{sim} difference in means between the Fully Open (FO) and Partially Open (PO) states. If a credible interval center is outside the experimental tolerance range (green shade), the difference is considered significant. Right: credible interval of the effect size (difference in means scaled by the pooled standard deviation) between the FO and PO states. If the credible interval center is greater than 0.5 in absolute value (green shade), the effect size is considered medium-large[59]. A residue must have a significant difference in means and a large effect size to be considered as having discrimination power. The same procedure is done with the rest of the nuclei and chemical shift prediction methods (Figs. S23,S24)

for each residue (Figures S23, S24). 14 C_α nuclei, 11 C_β , 11 C, and 13 N nuclei were found to fit these criteria, yielding a total of 49 nuclei with potential to distinguish the two open states.

Our strategy for comparing the experimental chemical shifts to the predicted chemical shifts from the structural ensemble sampled during MD simulations involved analysis of the difference of the chemical shift changes between the FO/PO states and the C state computed using the simulations ensemble (ΔCS^{X}_{sim}), and the experimental chemical shift change between the activated state

and deactivated state (ΔCS_{exp}). An overview of the method can be found in Figure 1d. Compared to comparing directly the FO/PO state CS_{sim} with the activated state CS_{exp}, such a strategy is likely to be less prone to errors, both in CS referencing and systematic errors in the prediction, namely consistent effects that occur in the experimental data but are not captured by the prediction algorithm (e.g. effects of the lipid environment, or tertiary interactions). Indeed, the use of $\Delta\Delta CS^X$ (ΔCS^X_{sim} - ΔCS_{exp}) has allowed us to make stronger predictions than absolute CS (CS_{sim} - CS_{exp}) would have (Figure S25). Additionally, the error distribution of the $\Delta\Delta CS^X$ is narrower and more centered at zero than the analogous distribution of absolute chemical shifts (Figure S26) implying that the use of $\Delta\Delta CS^X$ has smaller random and systematic errors and is thus is more suited for state prediction than absolute chemical shifts. Considering absolute chemical shift predictions may be enough when the system undergoes major conformational transitions. The approach we propose should instead be considered when the structure of a reference state is available and changes in the system are relatively subtle, with many structural features preserved between the various conformational states of the system.

Figure 4a shows the $\Delta\Delta CS^X$ for the PO (blue bars) and FO states (orange bars) for the different nuclei and predicted using the two different prediction softwares. Missing bars correspond to an absence of experimental CS signal or to a statistical non-significance of the simulated value difference between states (Figure S22). The lower the $\Delta\Delta CS^X$, the closer the simulation prediction is to the experiment. Therefore if, for a given nucleus, a given method, and a given residue, the bar corresponding to the PO state is lower than the bar corresponding to the FO state, we consider the CS^{PO}_{sim} of the PO state to agree better with the $CS^{pH\,4,\,act}_{exp}$ of the activated state. If both signals are below the green shaded area, which indicates the experimental tolerance, both signals are indistinguishable. Most of the nuclei identified in this study indicate a better agreement with the PO state (61%) than with the FO state (25%), suggesting that the activated samples prepared for NMR analysis closely agree with a partial opening of the activation gate.

The nuclei displaying better agreement with the PO state (highlighted in blue in Figure 4b,c) are located in the turret region linking TM1 and the pore helix (A50, E51, G56, Q58, L59, I60), in the selectivity filter (T75, Y78) and in the transmembrane part of TM1 (M96, V97, S102, F103, T106 and T112). Additional nuclei spread all over the sequence were inconclusive (grey highlights in Figure 4c - Y62, R64, R88, V93, A98, A108 and A109). Only two nuclei displayed better agreement with the FO state (orange highlights in Figure 4b,c - T61 and G99).

Within the approach of analyzing the difference in predicted and experimental chemical shifts,

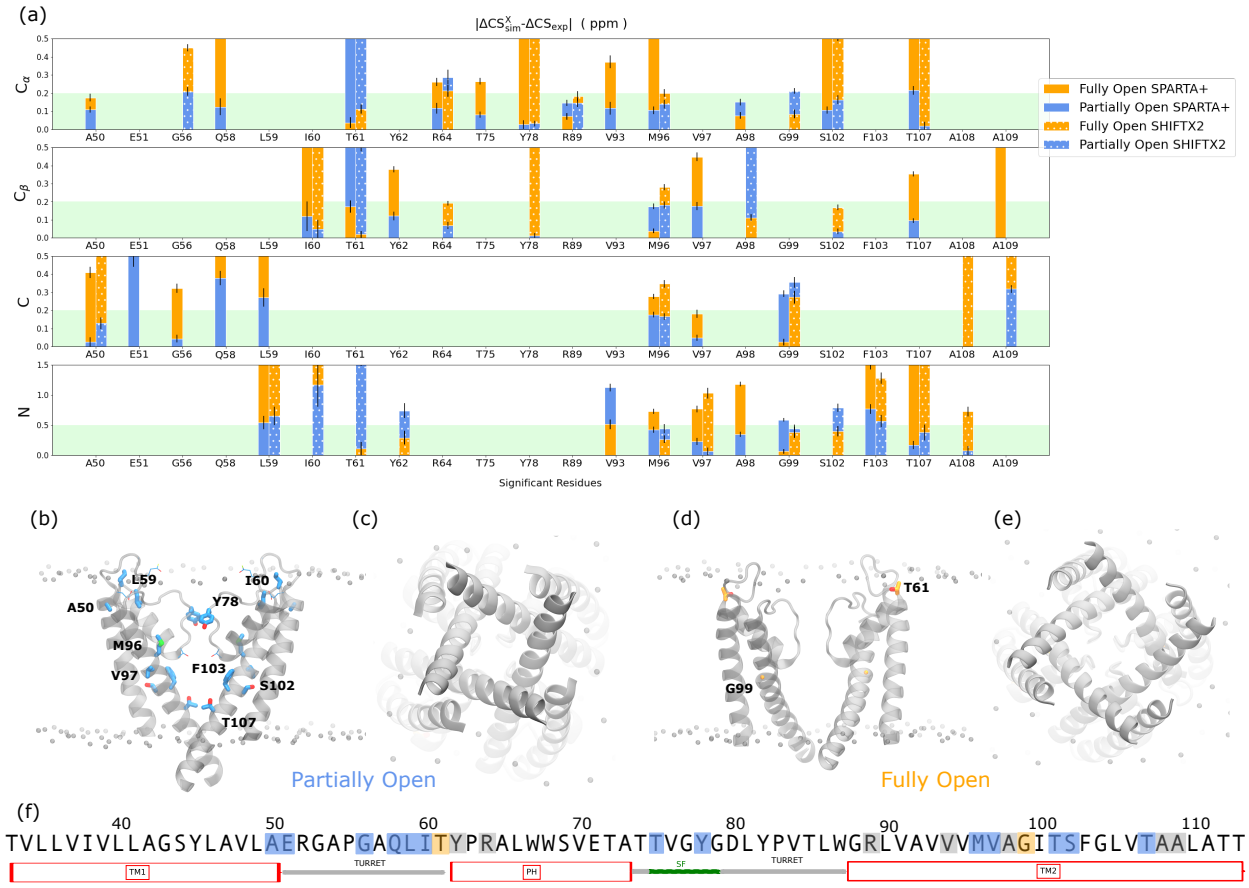

FIG. 4. a) Centers of 94% credible intervals of the difference in relative chemical shifts between experiment and simulation in absolute value. The limits of the credible interval are shown as error bars. Both experiment and simulation use as reference the closed state chemical shifts: $\Delta CS^{X}_{sim} = CS^{X}_{sim} - CS^{C}_{sim}$ and $\Delta CS_{exp} = CS^{pH\ 4}_{exp} - CS^{pH\ 7.5}_{exp}$, where X represents the FO or PO sate. The residues found to be capable of discriminating between PO and FO using our statistical criteria are represented on the x-axis for the different nuclei and chemical shift prediction methods. The agreement between MD simulations and the NMR experiments is higher the closer the value is to zero. The PO state (blue) has in general a better agreement with experiment than FO state (orange). The green shade depicts the typical experimental uncertainty. If two methods produce statistically identical CS or the signal is missing in the spectrum, the bar is not represented. b) Side view of the molecular structure of the PO state. Only two subunits are shown. Residues identified using one chemical shift prediction method are shown with a thin "licorice representation". Residues identified using two chemical shift prediction methods are shown with a thick "licorice representation". c) Molecular structure of the PO state viewed from the intracellular side. d) and e) show the same representations as b) and c) but for the FO state. f) KcsA sequence, truncated to region of the protein used in the MD simulations. Blue highlights represent residues whose experimental chemical shift agrees with the simulated PO state, orange highlights those in agreement with the FO state and grey highlights inconclusive residues.

one can imagine a number of logical ways to compare the various structural states or MD ensembles (FO, PO, and C) to the two experimental states (activated and deactivated). We developed an *ad hoc* classification system summarized briefly in Figure 5 and in more detail in Figure S27. Individual nuclei are first grouped in 5 classes based on the predicted shifts for the three MD ensembles, as follows: (A) a very well populated group for which the three structural states are

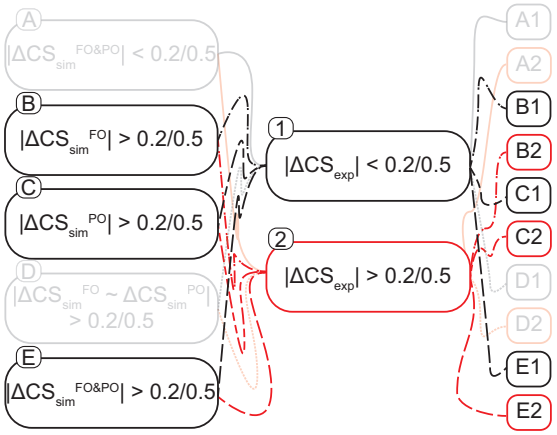

FIG. 5. Flow chart demonstrating the combinations of the ΔCS^{X}_{sim} and ΔCS_{exp} that yield various classifications of the state markers identified in this study. A threshold of tolerance of 0.2 ppm for ^{13}C and 0.5 ppm for ^{15}N was used to determine if the CS change was significant. FO = fully open, PO = partially open, C = closed, $\Delta CS^{X}_{sim} = CS^{X}_{sim} - CS^{Closed}_{sim}$, and $\Delta CS_{exp} = CS^{pH\ 4,\ act}_{exp} - CS^{pH\ 7.5,\ deact}_{exp}$.

predicted to give rise to indistinguishable shifts (akin to experimental spectators), (B) a well populated group for which the FO state is predicted to be distinct from the C and the PO states (which are themselves not distinguishable from each other); (C) a poorly populated group for which, curiously, PO is distinct from FO and C (which are themselves not distinguishable from each other); (D) an unpopulated hypothetical group for which the C state is distinct but PO and FO are indistinguishable ($\Delta CS^{FO}_{sim} \sim \Delta CS^{PO}_{sim}$), which has been eliminated by our statistical filtering analysis; and (E) a fairly well populated group for which all three computed states are expected to give rise to distinct chemical shifts. Each nucleus is further classified in terms of whether (1) the two experimental states have indistinguishable chemical shifts, or (2) if a change in the chemical shift is observed. Each nucleus is then identified as being category B1, B2,... up to E2 (and in some cases is identified as belonging to two classes because the two chemical shift prediction tools that were used were not in agreement).

Two classes of nuclei are of particular note in this study. A somewhat common type, **E2**, where the predicted CS is different for each state and the experimental data for the nucleus also indicates a shift, has strong ability to address the questions posed in this work. (This class includes A50C, E51C, Q58C, T107C_{α} , T107N, A108C, A108N, A109C_{β}). Many of these markers clearly show a better agreement with the conclusion that the sample is PO rather than FO (Figure 4a). Nuclei classified as **B1**, for which the FO state (but not the PO state) is predicted to show a deviation from C, and for which the experimental data obtained at both pH conditions are indistinguishable, are the most common (apart from the spectators **A1**, which are not discussed here). Resonances that

fall into the **B1** classification in aggregate offer strong support for the conclusion that the activated samples studied experimentally are closer to PO than to FO (A50C_α , G56C_α , G56C, Q58C_α , I60C_β , T75C_β , Y78C_α , Y78C_β , R89C_α , M96C_α , M96C_β , M96C, M96N (SHIFTX2), V97C, V97N (SHIFTX2), A98C_β , G99C_α , S102C_α , S102C_β).

Another group of CS are classified as **B2**, when the FO state is predicted to differ from the PO and C states, and the experimental data do show a shift (V97C_β , G99C, S102N, T107C_β , A109C) and yet another as **E1** for which the predicted CS differs for each structural state while the data do not reflect a change in CS. These two types of markers could be inconclusive although in some cases there is a much better match for one hypothesis than the other (typically the PO state being in better agreement).

A handful are observed as class **C1** if the predicted PO shifts are different from the FO and C states while the experimental data show no shift. At first sight, these would appear to indicate that the sample is FO. The pattern is difficult to interpret mechanistically and the phenomenon is mainly observed in the turret region for which the connection to opening is unclear and where antibody binding the crystals structure may have complicated the MD sampling (T61C_β (SHIFTX2), R64C_β , and G99N (SPARTA+)).

A single resonance is classified into **C2** (A98C_α), where the PO shift is distinct from the other states and there is an observed experimental chemical shift change between the activated and deactivated states. This particular class would suggest that the resonances that are identified within it would favor the sample to be in the PO state; however, because these classifications do not take into account the directionality of the CS change, the resulting marker cannot necessarily be said to indicate one structural state over another. Notably, some classifications in this study are populated by no resonances, namely **D1** and **D2**, due to the initial conditions of the analysis that require a resonance be distinguishable between the FO and PO states.

E. Molecular Basis for KcsA activation

The chemical shift markers identified by the SSNMR experiments reflect conformational gating of the corresponding residue during opening. These sites exhibit a variety of conformational differences in the MD simulations, including sidechain rotamer hops and backbone rearrangements (especially at the TM2 hinge) (Figure 6).

Several of the residues that enabled the distinction between the PO and FO states (A50, A98,

S102, T107, A108 and A109) also exhibit experimental chemical shift changes when comparing deactivated and activated datasets (Figure 5). S102 and T107, which undergo rotameric changes in their χ_1 dihedral angles, in particular, offer particularly strong support to the conclusion of partial opening (Figure 6d,h).

Scrutinizing additional dihedral angle plots revealed other substantial rearrangements at residues identified to discriminate the PO and the FO states in the selectivity filter (Y78), in the hinge region (M96, V97, A98, and S102) and in the residues facing the central pore cavity (F103 and T107) (Figure 6). The hinge region residues displayed subtle changes in the alpha helical basin, as expected from the hinge motion that is key for gate opening described in numerous studies (Figure 6d-i).[3,4] F103 has indeed been reported as a key player in transmitting the allosteric signal coupling the activation gate and the inactivation gate at the selectivity filter.[3] In the C and PO states, several rotameric states are accessible to the χ_1 angle of this residue, while in the FO state the residue is locked in the state found in the XRD structure avoiding exposure to the hydrated cavity. Several studies have found that its clash with T74 leads to SF pinching linked to inactivation[3,60]. Our work seems to indicate that in high [K^+] and low pH conditions, locking of this residue in the rotameric state compatible with inactivation is not favored. T107 assumes a similar rotameric state in the C, PO and FO state XRD structures (Figure 6h). Intriguingly, our MD simulations show that another χ_1 dihedral conformation is more prevalent in the FO state. This FO rotameric state appears unfavorable in the activated state condition SSNMR. While the molecular basis for this observation remains unclear, we note that a CS prediction based on a pure XRD structure diffraction (in the absence of MD simulations) would have missed this observation.

The only SF residue with discriminating power which shows an interesting rotameric change is Y78 (Figure 6b). Indeed the other SF residue, T75, is a weak marker, with the C_α nucleus CS pointing towards the PO state only when predicted by SPARTA+. Y78 is at the center of a hydrogen bonding hub which connects the selectivity filter and the pore helix of adjacent subunits, via hydrogen bonding to the side chains of T72 and W68 and the backbone carbonyl of E71 (Figure 6c). The rotameric state of Y78 χ_2 differs between the XRD structures of the C and PO state, on one hand, and of the FO state, on the other (Figure 6b). The slight change in Y78 orientation in the FO state appears to lead to a motion of the pore helix. Interestingly no pore helix residues, including T72 and W68, showed a change in predicted CS between the PO and the FO states.

Only two residues indicated a better agreement with the FO state, T61 and G99. It would be intriguigh if a portion of the channel were in a partially open like state and another portion were

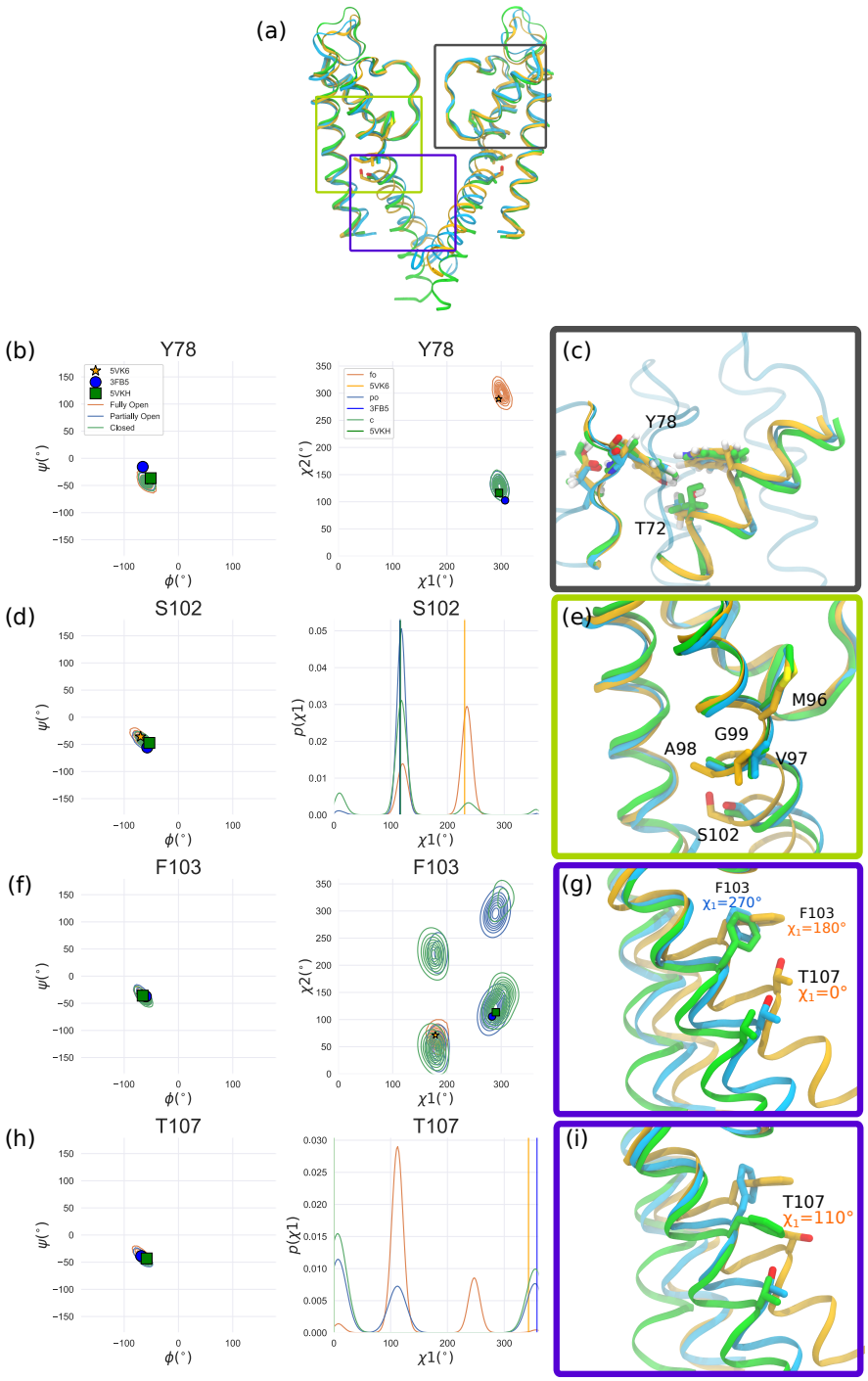

FIG. 6. Structural comparison between C, PO and FO states (a) Side view of 2 opposite subunits in the C (green ribbon), PO (blue) and FO states (orange). The selectivity filter region is highlighted by a grey box, the hinge region by a green box and the cavity facing helices by a purple box. (b) Dihedral (Ramachandran and χ_1-χ_2 plots) angle plots showing the MD simulation distributions (surface contours, coloring is the same as in panels a) and (c) corresponding molecular models showing details of the environment around Y78 in the three states. Snapshots were aligned on the selectivity filter residues backbone (75-80) of the subunit Y78 belongs to. (d-e) Same as (c) for S102 (f-i) same as (c) for F103 and T107. The molecular views in the purple boxes (g,i) show the rearrangement of the rotameric state of T107 from the crystal structure to the most prevalent rotameric state observed in MD simulations, in the FO state.

fully open, but this seems not likely to be quite correct. For example T61 is a more complex situation without a clear assignment to either state. A comparison of the Ramachandran distributions of T61 in both states reveals that both states explore a common conformer but the PO state samples an additional conformer (Fig. S28). This conformer, which is only present in the PO state simulations, is only explored by one of four subunits. If the chemical shifts produced from this subunit are removed then the difference in CS between PO and FO states becomes statistically insignificant (Fig. S29). We attribute this conformational state to a fluctuation in the dynamics of the flexible turret region which may not be representative of the entire configurational ensemble. As mentioned above, this region was in contact with the antibodies used to stabilized the sample in the XRD structure and numerous indications in the literature suggest dynamic plasticity for these residues.[18] G99 is located in the center of the hinge region, surrounded by residues that exhibit good agreement with the PO state. The fact that its chemical shift agrees better with the FO state appears intriguing. Indeed, the dihedral angles explored by G99 in both states are fairly similar and correspond to the initial angles of the X-ray structure (Fig. S28). Why it is the only residue in better agreement with the FO state in this region remains to be explained.

IV. CONCLUSIONS

The open conductive state of KcsA has remained elusive to structural studies due to its short lifetime. Mutagenesis and other engineering strategies have led to the determination of several XRD structures, but since they were obtained under somewhat artificial conditions, the physiologically relevant state remains of intertest to be fully characterized. Spectroscopy studies have revealed that there is considerable conformational heterogeneity within the open states of KcsA of the degree of opening, making a static characterization not only challenging but also limited in its physiological relevance. SSNMR allows the study of a sample at near-physiological conditions, but interpreting the spectra in terms of structure is challenging. MD simulations can assist with this interpretation since the conditions of the MD simulation and the SSNMR experiment can be nearly equivalent, and the time-scales probed are compatible. This procedure rests on the use of chemical shift prediction software which have been refined over the years to yield robust predictions.

Here, we have designed a computational pipeline to characterize the activated state present in the SSNMR activated state sample. The two candidates from X-ray crystallography that we

compared it to were a partially open PO (PDB ID 3FB5) and a fully open FO (PDB ID 5VK6) structure. While a direct comparison between the experimental and computationally predicted CS was noisy, implementing a Bayesian inference scheme in which nuclei were selected when they had the predictive power to statistically discriminate the PO and the FO states, and the use of difference chemical shifts (relative to the deactivated/closed state) for both the experimental and the predicted shifts, proved more successful. With this approach were able to conclude that the majority state in the SSNMR resembles the PO state.

Why does the system stall in the partially open state, instead of advancing to form the inactivated state as it does in electrophysiology experiments? Formation of the fully open state may linked to ion expulsion, and so may be prevented in these SSNMR experiments by the relatively high concentration of ambient potassium, i.e. perhaps the FO state is unstable in the presence of bound potassium ions. Indeed, under some conditions the FO state characterized by X-ray crystallography exhibited reduction in the ion occupancy in the selectivity filter.[18] MD simulations of the opening also appear to indicate a slow process from partial opening to full opening, with an unknown inherent bottleneck, of a possibly entropic nature, or it is possible that the system does not fully open absent other coupled processes such as ion loss. Alternatively, perhaps the various differences between the sample conditions (protein concentration, symmetric vs. asymmetric lipids and pH, etc.) cause the system to stall in one case and advance to inactivation in the other. In this sense, whether the state captured in the SSNMR experiment under low pH and high [K^+] truly represents the long sought-after conductive state transiently responsible for conduction awaits further study.

The conclusion that the activated state studied by SSNMR is a partially open state seems compatible a number of earlier studies or hypotheses. It has been speculated in prior studies that the state stabilized for NMR studies by cardiolipin lipids at low pH is partially open.[61] It is also consistent with the fact that preparation of the FO state for X-ray crystallography experiments required a number of mutations and cross links, without which the partially open state resulted. The results collectively suggest a mechanism in which opening, once triggered by protonation on the pH gate, proceeds to an intermediate point which can be stabilized under appropriate experimental conditions such as high K^+ concentration. The degree of opening is possibly a balancing act between repulsions in the protonated pH sensor that tend to push the system open, and conformational clashes in the hinge and base of the selectivity filter, which tend to keep the system closed.

This work demonstrates an approach to combining SSNMR and MD simulations to provide

a molecular level description of a physiologically important allosteric process in a membrane protein.

V. SUPPLEMENTARY MATERIALS

See Supplementary Materials for additional data figures and tables.

VI. ACKNOWLEDGEMENTS

E.G.K. was supported by a postdoctoral fellowship from the National Institutes of Health (NIH) (Grant No. GM135350). A.E.M. was supported by Grant No. GM088724 from the NIH and Grant No. 1412253 from the National Science Foundation. The NMR data were collected at the New York Structural Biology Center (NYSBC) with support from the Center on Macromolecular Dynamics by NMR Spectroscopy, a Biomedical Technology Research Resource (NIH Grant No. GM118302). The NYSBC is also supported by the Empire State Division of Science Technology and Innovation and the Office of Research Infrastructure Program (NIH Grant No. CO6RR015495). A.E.M. is a member of the NYSBC. L. D. would like to thank the support of Science for Life Laboratory, the Göran Gustafsson Foundation, and the Swedish research council (Grant No. VR-2018-04905). Simulaitons were performed on computational resources provided by the Swedish National Infrastructure for Computing (SNIC) at the PDC Centre for High Performance Computing (PDC-HPC). The authors thank Zhiyu Sun for synthesis of one of the KcsA samples used in this work.

VII. DATA AVAILABILITY

The data that support the findings of this study are openly available in `https://github.com/delemottelab/Informing_NMR_experiments_w_MD` and `https://osf.io/6h2z5/` with the raw NMR data deposited in BMRbig (`https://bmrbig.org`) with under deposition ID XXXXX.

REFERENCES

[1]Z. Cournia and A. Chatzigoulas, "Allostery in membrane proteins," Current Opinion in Structural Biology **62**, 197–204 (2020).

[2]Y. Xu, M. P. Bhate, and A. E. McDermott, "Transmembrane allosteric energetics characterization for strong coupling between proton and potassium ion binding in the KcsA channel," Proceedings of the National Academy of Sciences of the United States of America **114**, 8788–8793 (2017).

[3]J. Li, J. Ostmeyer, L. G. Cuello, E. Perozo, and B. Roux, "Rapid constriction of the selectivity filter underlies C-type inactivation in the KcsA potassium channel," J. Gen. Physiol. **215**, 1408–1420 (2018).

[4]L. G. Cuello, D. M. Cortes, and E. Perozo, "The gating cycle of a K+ channel at atomic resolution," Elife **6**, 1–17 (2017).

[5]Y. Xu, D. Zhang, R. Rogawski, C. M. Nimigean, and A. E. McDermott, "Identifying coupled clusters of allostery participants through chemical shift perturbations," Proc. Natl. Acad. Sci. U. S. A. **116**, 2078–2085 (2019).

[6]S. Furini and C. Domene, "Critical assessment of common force fields for molecular dynamics simulations of potassium channels," Journal of Chemical Theory and Computation **16**, 7148–7159 (2020).

[7]J. Ostmeyer, S. Chakrapani, A. C. Pan, E. Perozo, and B. Roux, "Recovery from slow inactivation in K+ channels is controlled by water molecules," Nature **501**, 121–124 (2013).

[8]Z. Sun, Y. Xu, D. Zhang, and A. E. McDermott, "Probing allosteric coupling in a constitutively open mutant of the ion channel KcsA using solid-state NMR," Proceedings of the National Academy of Sciences of the United States of America **117**, 7171–7175 (2020).

[9]C. Ader, R. Schneider, S. Hornig, P. Velisetty, V. Vardanyan, K. Giller, I. Ohmert, S. Becker, O. Pongs, and M. Baldus, "Coupling of activation and inactivation gate in a K-channel: Potassium and ligand sensitivity," EMBO Journal **28**, 2825–2834 (2009).

[10]A. J. Labro, D. Marien Cortes, C. Tilegenova, and L. G. Cuello, "Inverted allosteric coupling between activation and inactivation gates in K+ channels," Proceedings of the National Academy of Sciences of the United States of America **115**, 5426–5431 (2018).

[11]L. Delemotte, "Opening leads to closing: Allosteric crosstalk between the activation and inactivation gates in KcsA," Journal of General Physiology **215**, 1356–1359 (2018).

[12]D. Zhang, G. S. Howarth, L. A. Parkin, and A. E. McDermott, "NMR studies of lipid regulation of the K+ channel KcsA," Biochim. Biophys. Acta - Biomembr. , 183491 (2020).

[13]Y. Zhou, J. H. Morais-Cabral, A. Kaufman, and R. Mackinnon, "Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 Å resolution," Nature **414**, 43–48

(2001).

[14]M. P. Bhate, B. J. Wylie, A. Thompson, L. Tian, C. Nimigean, and A. E. McDermott, "Preparation of uniformly isotope labeled KcsA for solid state NMR: Expression, purification, reconstitution into liposomes and functional assay," Protein Expression and Purification **91**, 119–124 (2013).

[15]B. J. Wylie, M. P. Bhate, and A. E. McDermott, "Transmembrane allosteric coupling of the gates in a potassium channel," Proceedings of the National Academy of Sciences of the United States of America **111**, 185–190 (2014).

[16]J. Medeiros-Silva, D. Mance, M. Daniëls, S. Jekhmane, K. Houben, M. Baldus, and M. Weingarth, "^{1}H-Detected Solid-State NMR Studies of Water-Inaccessible Proteins In Vitro and In Situ," Angewandte Chemie - International Edition **55**, 13606–13610 (2016).

[17]S. Jekhmane, J. Medeiros-Silva, J. Li, F. Kümmerer, C. Müller-Hermes, M. Baldus, B. Roux, and M. Weingarth, "Shifts in the selectivity filter dynamics cause modal gating in channels," Nature Communications **10**, 1–12 (2019).

[18]L. G. Cuello, V. Jogini, D. M. Cortes, A. C. Pan, D. G. Gagnon, O. Dalmas, J. F. Cordero-Morales, S. Chakrapani, B. Roux, and E. Perozo, "Structural basis for the coupling between activation and inactivation gates in K+ channels," Nature **466**, 272–275 (2010).

[19]M. P. Bhate, B. J. Wylie, L. Tian, and A. E. McDermott, "Conformational dynamics in the selectivity filter of KcsA in response to potassium ion concentration," Journal of Molecular Biology **401**, 155–166 (2010).

[20]A. C. De Dios, J. G. Pearson, and E. Oldfield, "Secondary and tertiary structural effects on protein NMR chemical shifts: An ab initio approach," Science **260**, 1491–1496 (1993).

[21]E. Oldfield, "Chemical shifts and three-dimensional protein structures," Journal of Biomolecular NMR **5**, 217–225 (1995).

[22]D. S. Wishart and D. A. Case, *Methods in Enzymology*, Vol. 338 (Elsevier Masson SAS, 2002) pp. 3–34.

[23]M. Niklasson, A. Ahlner, C. Andresen, J. A. Marsh, and P. Lundström, "Fast and Accurate Resonance Assignment of Small-to-Large Proteins by Combining Automated and Manual Approaches," PLoS Computational Biology **11** (2015), 10.1371/journal.pcbi.1004022.

[24]P. Robustelli, K. A. Stafford, and A. G. Palmer, "Interpreting protein structural dynamics from NMR chemical shifts," Journal of the American Chemical Society **134**, 6365–6374 (2012).

[25]D. W. Li and R. Brüschweiler, "PPM: A side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles," Journal of Biomolecular NMR **54**, 257–265 (2012).

[26]M. Berjanskii, D. Arndt, Y. Liang, and D. S. Wishart, "A robust algorithm for optimizing protein structures with NMR chemical shifts," Journal of Biomolecular NMR **63**, 255–264 (2015).

[27]D. Zhang, B. Itin, and A. E. McDermott, "TmDOTP: An NMR-based thermometer for magic angle spinning NMR experiments," Journal of Magnetic Resonance **308**, 106574 (2019).

[28]J. Pauli, M. Baldus, B. Van Rossum, H. De Groot, and H. Oschkinat, "Backbone and side-chain ^{13}C and ^{15}N signal assignments of the α-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla," ChemBioChem **2**, 272–281 (2001).

[29]T. I. Igumenova, A. J. Wand, and A. E. McDermott, "Assignment of the Backbone Resonances for Microcrystalline Ubiquitin," Journal of the American Chemical Society **126**, 5323–5331 (2004).

[30]Y. Li, D. A. Berthold, H. L. Frericks, R. B. Gennis, and C. M. Rienstra, "Partial ^{13}C and ^{15}N chemical-shift assignments of the disulfide-bond-forming enzyme DsbB by 3D magic-angle spinning NMR spectroscopy," ChemBioChem **8**, 434–442 (2007).

[31]K. Takegoshi, S. Nakamura, and T. Terao, "^{13}C-^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR," Chemical Physics Letters **344**, 631–637 (2001).

[32]M. Baldus, A. T. Petkova, J. Herzfeld, and R. G. Griffin, "Cross polarization in the tilted frame: Assignment and spectral simplification in heteronuclear spin systems," Molecular Physics **95**, 1197–1207 (1998).

[33]B. M. Fung, A. K. Khitrin, and K. Ermolaev, "An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids," Journal of Magnetic Resonance **142**, 97–101 (2000).

[34]G. Hou, S. Yan, J. Trébosc, J. P. Amoureux, and T. Polenova, "Broadband homonuclear correlation spectroscopy driven by combined $R2_n^v$ sequences under fast magic angle spinning for NMR structural analysis of organic and biological solids," Journal of Magnetic Resonance **232**, 18–30 (2013).

[35]C. P. Jaroniec, C. Filip, and R. G. Griffin, "3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon-nitrogen distances in uniformly ^{13}C,^{15}N-labeled solids," Journal of the American Chemical Society **124**, 10728–10742 (2002).

[36]R. S. Thakur, N. D. Kurur, and P. K. Madhu, "Swept-frequency two-pulse phase modulation for heteronuclear dipolar decoupling in solid-state NMR," Chemical Physics Letters **426**, 459–463

(2006).

[37]F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, "NMRPipe: A multidimensional spectral processing system based on UNIX pipes," Journal of Biomolecular NMR **6**, 277–293 (1995).

[38]C. R. Morcombe and K. W. Zilm, "Chemical shift referencing in MAS solid state NMR," Journal of Magnetic Resonance **162**, 479–486 (2003).

[39]A. McDermott and Z. Gu, "Carbon and Nitrogen Chemical Shifts of Solid State Enzymes," Encyclopedia of Magnetic Resonance (2007), 10.1002/9780470034590.emrstm0053.

[40]R. K. Harris, E. D. Becker, S. M. De Cabral Menezes, P. Granger, R. E. Hoffman, and K. W. Zilm, "Further Conventions for NMR shielding and chemical shifts (IUPAC recommendations 2008)," Magnetic Resonance in Chemistry **46**, 582–598 (2008).

[41]P. Bertani, J. Raya, and B. Bechinger, "^{15}N chemical shift referencing in solid state NMR," Solid State Nuclear Magnetic Resonance **61-62**, 15–18 (2014).

[42]W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue, "The CCPN data model for NMR spectroscopy: Development of a software pipeline," Proteins: Structure, Function and Genetics **59**, 687–696 (2005).

[43]K. J. Fritzsching, M. Hong, and K. Schmidt-Rohr, "Conformationally selective multidimensional chemical shift ranges in proteins from a PACSY database purged using intrinsic quality criteria," Journal of Biomolecular NMR **64**, 115–130 (2016).

[44]Y. Yang, K. J. Fritzsching, and M. Hong, "Resonance assignment of the NMR spectra of disordered proteins using a multi-objective non-dominated sorting genetic algorithm," Journal of Biomolecular NMR **57**, 281–296 (2013).

[45]S. Jo, X. Cheng, J. Lee, S. Kim, S. J. Park, D. S. Patel, A. H. Beaven, K. I. Lee, H. Rui, S. Park, H. S. Lee, B. Roux, A. D. MacKerell, J. B. Klauda, Y. Qi, and W. Im, "CHARMM-GUI 10 years for biomolecular modeling and simulation," J. Comput. Chem. **38**, 1114–1124 (2017).

[46]F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker, and I. André, "Modeling symmetric Macromolecular structures in Rosetta3," PLoS One **6** (2011), 10.1371/journal.pone.0020450.

[47]Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.8," (2015).

[48]M. P. Bhate and A. E. McDermott, "Protonation state of E71 in KcsA and its role for channel collapse and inactivation," Proceedings of the National Academy of Sciences of the United States of America **109**, 15265–15270 (2012).

[49]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926–935 (1983).

[50]R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ Dihedral Angles," J. Chem. Theory Comput. **8**, 3257–3273 (2012).

[51]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindah, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1-2**, 19–25 (2015), arXiv:arXiv:1503.05249v1.

[52]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**, 7182–7190 (1981).

[53]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126** (2007), 10.1063/1.2408420, arXiv:arXiv:0803.4060v1.

[54]Y. Shen and A. Bax, "SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network," Journal of Biomolecular NMR **48**, 13–22 (2010).

[55]B. Han, Y. Liu, S. W. Ginzinger, and D. S. Wishart, "SHIFTX2: Significantly improved protein chemical shift prediction," Journal of Biomolecular NMR **50**, 43–57 (2011).

[56]J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," PeerJ Comput. Sci. **2016**, 1–24 (2016), arXiv:1507.08050.

[57]R. Kumar, C. Carroll, A. Hartikainen, and O. A. Martin, "{ArviZ} a unified library for exploratory analysis of {Bayesian} models in {Python}," J. Open Source Softw. (2019), 10.21105/joss.01143.

[58]P. Sanjeevi and P. Viswanathan, "NUTS scheduling approach for cloud data centers to optimize energy consumption," Computing **99**, 1179–1205 (2017).

[59]S. S. Sawilowsky, "Very large and huge effect sizes," J. Mod. Appl. Stat. Methods **8**, 597–599 (2009).

[60]W. Kopec, B. S. Rothberg, and B. L. de Groot, "Molecular mechanism of a potassium channel gating through activation gate-selectivity filter coupling," Nat. Commun. **10**, 1–15 (2019).

[61]E. A. van der Cruijsen, A. V. Prokofyev, O. Pongs, and M. Baldus, "Probing Conformational Changes during the Gating Cycle of a Potassium Channel in Lipid Bilayers," Biophysical Journal **112**, 99–108 (2017).