bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# Atomic Level Characterisation of Millisecond-Time Scale Protein Motions through a Combined Molecular Simulations and NMR Approach

Jordi Juárez-Jiménez[†], Arun A. Gupta[†¬], Gogulan Karunanithy[$], Antonia S. J. S. Mey[†], Charis Georgiou[†¬], Harris Ioannidis[†], Alessio De Simone[†¬], Paul N. Barlow[†], Alison N. Hulme[†], Malcolm D. Walkinshaw[o], Andrew J. Baldwin[$] and Julien Michel[†*]

[†] EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh, EH9 3FJ, United Kingdom.

[$] Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QZ, United Kingdom.

[o] School of Biological Sciences, Michael Swann Building, Max Born Crescent, Edinburgh, EH9 3BF, United Kingdom.

[¬] Current affiliations: A. G: Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom; C.G: Department of Biomedicine, University of Bergen, 5020 Bergen, Norway; A.D.S: Sygnature discovery, Biocity, Nottingham NG1 1GR, United Kingdom.

[*] Corresponding author. E-mail: mail@julienmichel.net

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Abstract**

Advances in biomolecular sciences are closely linked to our ability to chart the energy landscapes of biomolecules with atomic details. Here we validate a new paradigm to characterise thermodynamics and kinetics of millisecond timescale conformational transitions between ground state and transient excited states in the enzyme cyclophilin A (CypA). We describe a novel methodology that combines molecular dynamics simulations and Markov State modelling with NMR measurements to provide atomic-level insights into the nature of CypA transient conformational states. The computed conformational ensembles also enabled the predictive design and experimental validation of a single-site mutant that dramatically perturbs millisecond timescale loop motions, converting a CypA excited state into the ground state. The resulting models open up new horizons for targeting CypA with inhibitors and pave the way towards rational design of protein energy landscapes for protein engineering and drug discovery purposes.

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The structural characterization of conformational changes between long-lived low-energy (ground) states and short-lived sparsely populated (excited) states in biological macromolecules remains a major unsolved challenge in structural biology. Despite advances in integrating data from multiple biophysical techniques,[1–12] atomistic descriptions of the conformational changes that link the ground and excited states of a protein remain scarce. New approaches to provide such descriptions could transform the fields of drug discovery and protein engineering.[13–15]

While X-ray diffraction provides atomic resolution structures, it struggles to resolve lowly populated states (ca. <20%), and time-resolution of conformational motions has been possible in only a handful of special cases.[16,17] Molecular dynamics (MD) simulations provide atomic resolution, but their accuracy is limited by the approximate nature of potential energy functions. MD also only readily characterizes transient states that exchange with a ground state within few microseconds, and efforts to surpass this timescale are generally accompanied by loss of temporal resolution.[18–21] NMR relaxation experiments provide temporal resolution on a broad range of timescales but lack spatial resolution in comparison with X-ray diffraction or molecular modelling techniques.[22–25]

Here we implement and validate an approach that combines MD, NMR and X-ray crystallography, to characterize in atomic detail a millisecond timescale conformational change in the human enzyme cyclophilin A (CypA).[26–28] Extensive biophysical studies have previously sought to characterize its diverse millisecond timescale motions using X-ray and NMR, but a detailed atomistic description of the conformational changes involved has remained elusive.[29–33] CypA is a major drug target, but atomic level description of its ground state has proven insufficient to develop isoform selective CypA inhibitors.[34–41] The structural ensembles reported herein give key insights into the

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

nature of a millisecond conformational change in CypA, and these guided the design and experimental validation of a single-site mutant that converts a transient state into a ground state, thereby unlocking new opportunities for CypA inhibitors design efforts.

## Results and discussion

### An experimentally validated ensemble of Cyclophilin A conformations

The workflow outlined in Fig. 1a was applied to CypA. Previous NMR results had identified a set of residues displaying significant chemical-exchange contribution ($R_{ex}$ > 4 $s^{-1}$) to the transverse relaxation rate $R_2$.[29] The majority of those residues clustered in the region between Asp66 and Gly75, herein referred to as the 70s-loop. Accelerated molecular dynamics (aMD) simulations were initiated from the X-ray structure of wild type CypA that entailed enhanced motions specifically in the vicinity of the 70s-loop while maintaining the integrity of the overall fold (Fig 1a, step I and SI for details.)[42,43]

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

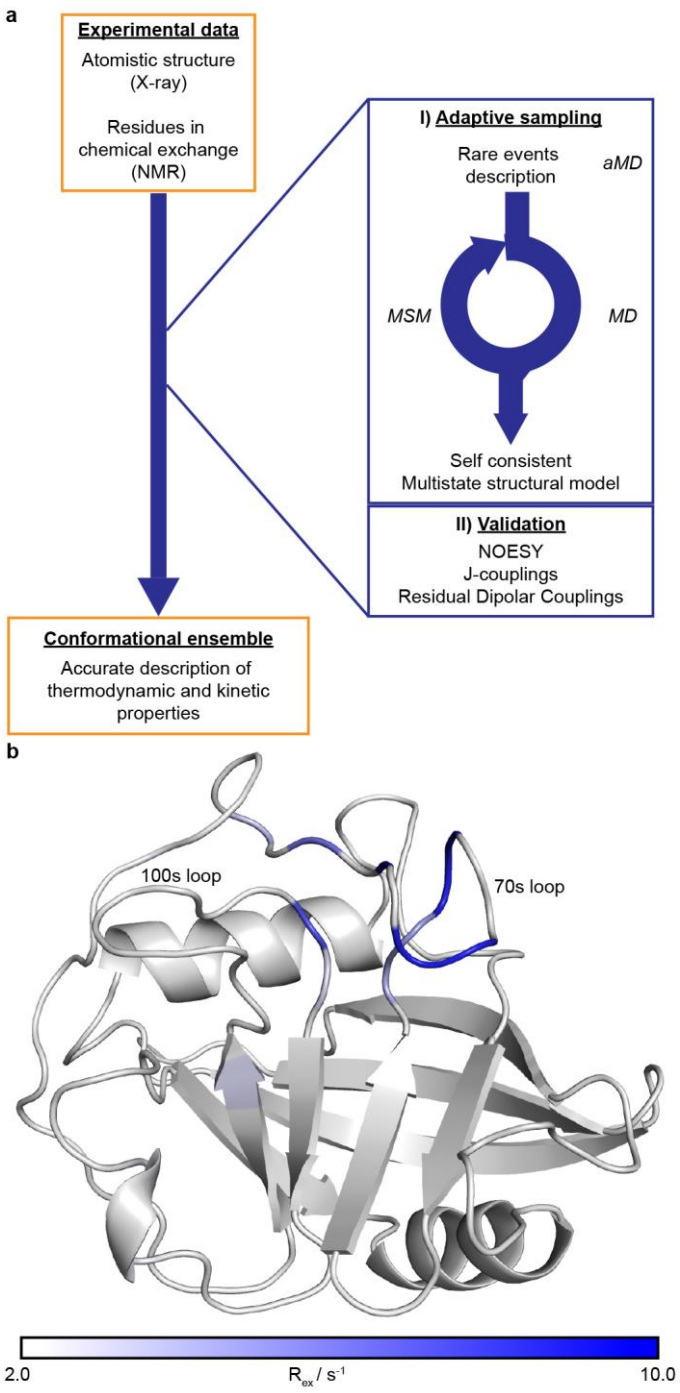


**Figure 1 Workflow for generating a conformational ensemble describing millisecond timescale protein motions**. **a** I) Combination of accelerated MD (aMD) simulations to generate rare event loop motions and iterative rounds of MD simulations and Markov States Models (MSM) construction to recover equilibrium properties. II) Model validation against a range of NMR observables. **b** Structure of the ground-state of CypA. Residues displaying significant $R_{ex}$ contributions are highlighted in blue.

aMD-biased simulations successfully generated large-scale rearrangements of the 70s-loop as well as of the loop between residues Ala101 and Gln111 (hereafter referred

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to as the 100s-loop) on timescales of a few hundred ns. To recover the equilibrium properties of the aMD-generated conformational changes, CypA snapshots that were representative of the conformational transition were pooled with snapshots from the initial and final conformational states sampled during the aMD simulations. These structures were used as inputs for a follow-up stage that featured rounds of equilibrium MD simulations and Markov state model (MSM) construction. This step was iterated until a self-consistent MSM was obtained. Three rounds amounting to a cumulative sampling time of 187.5 µs were sufficient to produce a numerically stable MSM (Fig. S1-S2).

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

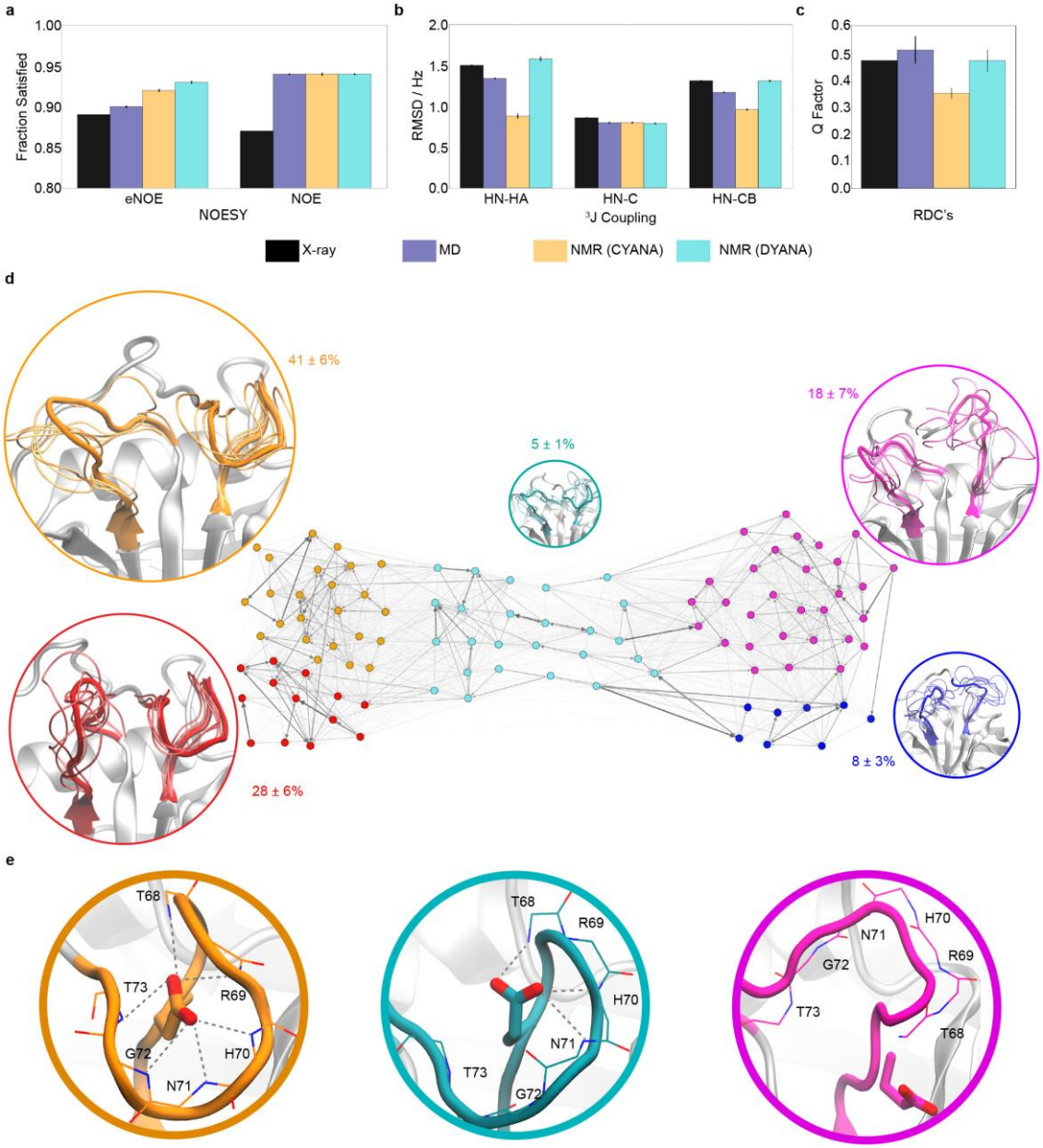


**Figure 2 .CypA loop dynamics, model validation, description and interpration. a** Agreement between measured and back-calculated eNOE and NOE derived distances for the MD ensemble, two previously published NMR ensembles, and an X-ray diffracted structure of CypA. Bar heights indicate the fraction fulfilled allowing a 0.5 Å deviation tolerance. **b** RMSD between measured and back-calculated $^{3}$J couplings**. c** Q-factor values for measured and back-calculated Residual Dipolar Couplings. **d** MSM describing thermodynamics and kinetics of 70s- and 100s-loop motions in CypA. **e** Details of Asp66 interactions with CypA in the orange, teal and magenta macrostates. Error bars on all reported quantities were obtained by bootstrapping.

The conformational ensemble was cross-validated by back-calculation of a set of NMR observables that included standard and exact NOEs, $^{3}$J-coupling measurements

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and Residual Dipolar Couplings (RDCs).[6,44] Observables were also back-calculated from a 40-structures CYANA NMR ensemble,[31] a 20 structures DYANA NMR ensemble,[45] and a representative CypA X-ray structure.[46] The MD ensemble was found to reproduce about 90% and 95% of derived proton-proton distances from eNOE and NOE measurements, which was comparable to both NMR ensembles and slightly better than the X-ray structure. $^{3}$J-coupling values were also computed for backbone HN-HA, HN-C, and HN-CB pairs (Fig 2b). HN-HA RMSDs values for the MD ensemble are slightly above those reported in benchmark studies of smaller proteins,[47,48] and above those obtained with the CYANA ensemble. The DYANA ensemble that was not refined against the particular set of $^{3}$J coupling values used here shows poorer agreement. The RMSD for $^{3}$J coupling values derived from the X-ray structure is similar to the one obtained from the MD ensemble. There is little difference in accuracy for HN-C and HN-CB pairs between the MD and NMR ensembles, whereas the X-ray structure is slightly worse. Finally, the RDCs are better reproduced with the CYANA ensemble, whereas the X-ray structure, MD and DYANA ensembles are of similar accuracy (Fig 2c). Overall the accuracy of the MD ensemble, which does not include any refinement step against NOEs, $^{3}$J-coupling or RDCs, was deemed encouraging, and further insights into loop motion mechanisms were sought by detailed inspection of the MSM.

**Atomic-level insights in millisecond timescale CypA loop motion mechanisms**

A MSM description of CypA 70s- and 100s-loop dynamics is presented in Fig. 2d. For ease of interpretation, the MSM states were further grouped into a coarser, five-state description (see SI for details). The most populated macro-state (orange in Fig. 2d) resembles the conformation observed in most X-ray structures reported for CypA, with the 70s- and 100s-loops in 'closed' and 'open' conformations, respectively. In this 70s-

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

loop closed/100s-loop open (closed/open, orange in Fig. 2d) macro-state both loops undergo low positional fluctuations (average Cα RMSF values of 1.1 Å and 2.7 Å for the 70s and 100s loops respectively). This closed/open macro-state interchanges with a similarly flexible closed/closed macro-state (red, RMSF 1.1 Å and 2.6 Å). The 100s-loop is as flexible in the open/open (magenta, RMSF 2.49 Å) and open/closed (blue, RMSF 3.33 Å) states as in the orange and red macro-states. However the 70s-loop displays greater positional fluctuations (RMSF 4.4 Å and 4.1 Å). This suggests that the open 70s-loop is disordered in these states, in contrast with the rigid structure observed in both the X-ray structure and in the closed/open macro-state. The transition between closed and open 70s-loop conformations proceeds *via* an intermediate (teal) semi-open 70s-loop macro-state (teal) that features positional fluctuations intermediate between the other macro-states (RMSF ca. 2.9 Å).

The order-disorder transition of the 70s-loop is critically controlled by a network of hydrogen bonds established between the backbone nitrogen atoms of residues Thr68 to Gly74 and the side-chain of Asp66. Progression from the well-populated closed/open or closed/closed macro-states to the intermediate semi-open macro-state requires partial breaking of this array of hydrogen-bonds (Gly72-Gly75), to allow partial unwinding of the 70s-loop, before progression to completely disordered 70s-loop macro-states, where no interaction between the side chain of Asp66 and backbone amide N of Thr68-Gly74 is observed (Fig. 2e). Structures of rhesus macaque TRIMCyp (RhTC) showing that a D66N-R69H double substitution promotes the opening of the 70s-loop strengthens the proposed mechanism.[49,50] However, whereas RhTC adopts a structured open 70s-loop conformation that is resolved in X-ray experiments, the present calculations suggest that the closed 70s-loop observed in the X-ray structure of human CypA exchanges with a previously uncharacterized open and disordered 70s loop conformation.

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**A rationally designed single-site mutant converts a transient CypA state into the ground state and disrupts millisecond time scale motions of the 70s loop**

This led to the hypothesis that mutation of Asp66 to Alanine (D66A) would shift the 70s-loop towards a conformation resembling the intermediate semi-open state. The hypothesis was assessed by applying the same MD/MSM protocol to CypA structures featuring the D66A mutation (Fig 3a). Calculations were also carried out for the H70A mutant as a negative control. Although His70 is spatially close to Asp66, this mutant has previously been reported to only cause a minor change in the 70s-loop motions.[29]

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

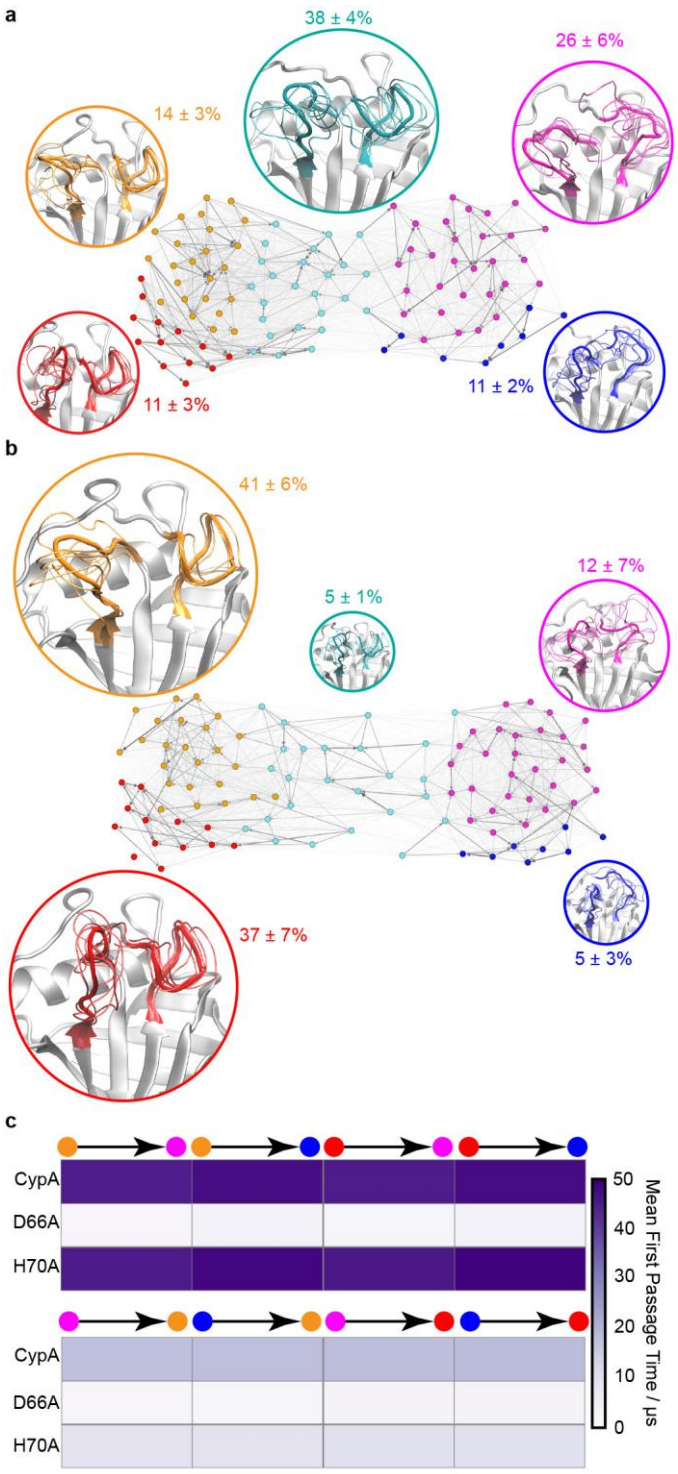


**Figure 3. Rational modulation of 70s-loop order-disorder transition via single site mutations in CypA. a** MSM model for D66A. **b** MSM model for H70A. Other details as in Fig 2d. **c** Heatmap of calculated mean first passage times (MFPT) between non-adjacent macro-states.

The computations predicted that the open/closed (orange) state is significantly destabilized in the D66A mutant with a population decreasing from ca. 40% to less than

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

15% (Fig. 3a). Meanwhile, the intermediate (teal) state seen in CypA is considerably stabilized, becoming the most abundant (40%). Overall with respect to CypA the population of states in which the 70s-loop is ordered has decreased from about 70% to 25%. In addition, while Mean First Passage Times (MFPT) among closed and open macro-states are similar between CypA and H70A, they have decreased by 10-15 fold (Fig. 3c) in D66A, thus suggesting that ordered to disordered transitions in the 70s loop happen faster in D66A than in CypA or H70A.

Experimental evidence for the dramatically altered energy landscape of D66A was sought by carrying out NMR experiments. Thus single-labelled and double-labelled CypA and D66A were expressed in and purified from BL21(DE3)pLysS *E. coli* cells (Fig. S4-S5).

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

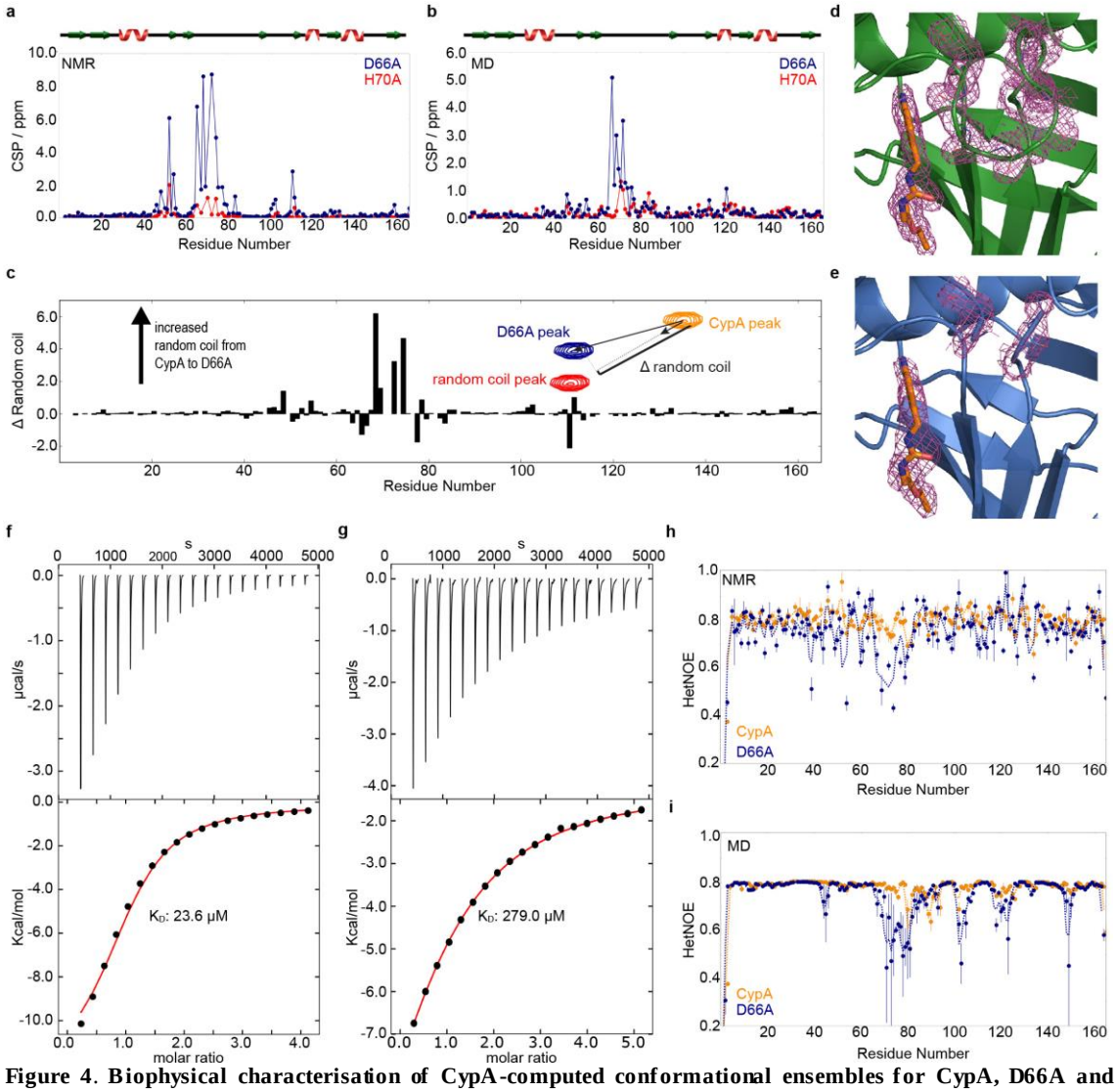


**Figure 4**. **Biophysical characterisation of CypA-computed conformational ensembles for CypA, D66A and H70A. a** Measured CSPs per residue (w.r,t CypA) for D66A, and for H70A as reported elsewhere.[29] **b** Calculated CSPs per residue (w.r.t to CypA) for D66A and H70A from the MSMs. **c** Per residue projection of the vector joining D66A and CypA peaks in HSQC spectra on the unit vector joining the CypA peak and predicted random coil peak position. A large positive value indicates a significant shift towards a random-coil chemical shift in the mutant. **d** X-ray structure of CypA in complex with ligand **1**. Magenta meshes indicate the 2Fo – Fc electron density map at an isocontour of 1.0σ for **1** and protein residues 65 to 75. **e** X-ray structure of D66A:**1** complex. Same details as in **d**. **f** Binding isotherm of compound **1** to CypA. **g** Binding isotherm of compound **1** to D66A. **h** Measured $^{15}N$ -$^{1}H$ hetereonuclear NOE values for CypA and D66A. **i** Calculated $^{15}N$ -$^{1}H$ hetereonuclear NOE values for CypA and D66A.

The well dispersed HSQC spectra recorded for the D66A mutant indicated the protein was natively folded (Fig. S6). A chemical shift perturbation (CSP) analysis, with respect to CypA, indicated strikingly larger changes for the D66A mutant compared to

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the previously reported H70A mutant (Fig. 4a, Table S1). The largest deviations are observed in the 70s-loop region with CSP values of up to 9.0 ppm. Additional perturbations occur around residues 50 and 110 that flank the 70s-loop region. These measurements are in good agreement with the predicted CSP patterns obtained through back-calculation of chemical shifts from the computed ensembles for CypA, D66A and H70A (Fig. 4b). Moreover, the neighbour corrected intrinsically disordered protein (ncIDP) web server was used to compare chemical shift changes with respect to random coil values (Fig 4c). The collective displacement of chemical shifts in the 70s-loop towards random-coil values supported the MD predicted order-disorder transition.

Further corroboration was provided by co-crystallization and X-ray diffraction experiments on CypA and D66A in complex with inhibitor **1** (Table S2).[36] Comparison of electron density profiles revealed that the 70s-loop adopts an ordered and closed conformation in the CypA:**1** complex (Fig 4d) while in the D66A:**1** complex electron density in the 70s-loop region is largely absent (Fig. 4e), consistent with a predominantly disordered conformation. The electron density for other parts of the protein and for **1** was similar, suggesting that the conformational change is restricted to the 70s-loop region. Moreover, the ITC estimated $K_d$ value of **1** is ca. 12-fold weaker for D66A than for CypA (Fig. 4f and 4g) suggesting that the order-disorder transition of the 70s-loop weakens binding of compounds to the Abu and Pro pockets without completely disrupting the CypA active site. This is in line with the conservation of binding mode between CypA and D66A observed for **1** by X-ray crystallography.

The narrow difference between effective correlation time for internal motions of a large number of residues and global tumbling times for both CypA and D66A precluded the obtention of generalized $S^2$ order parameters to characterize differences in ps-ns timescale motions between CypA and D66A (Fig. S7). Instead $\{^1H\}$-$^{15}N$ heteronuclear

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

NOE values were back-calculated by combining MD trajectories with NMR measurements of global tumbling times. The measured heteronuclear NOE values for D66A are consistently lower than for CypA in the 70s-loop region, even though the global tumbling times are similar (9.1 and 9.2 ns respectively). This suggests increased ps-ns time scale motions for D66A (Fig. 4h), an observation that is in line with the back-calculated NOE values, which show a significant dip in the 70s-loop region for D66A (Fig. 4i).

CPMG measurements turned out to be insufficient to reliably fit intermediate exchange parameters for D66A. However additional $R_{1\rho}$ experiments provided sufficient data to enable a combined fit of the $R_{1\rho}$ and CPMG measurements (Fig. S7). The data for CypA and D66A could be fitted in both cases to a two-state exchange model with similar $k_{ex}$ values (ca. 2000 $s^{-1}$) for CypA and D66A, but with significant differences in excited state populations ($p_b$ ca. 0.5% for D66A and ca. 2% for CypA). The fit also enabled determination of the magnitude of the chemical shift changes $\Delta\omega$ for residues showing significant dispersions for CypA, and the magnitude and sign of chemical shift changes for D66A.

Remarkably the pattern and magnitude of $\Delta\omega$ between ground state and excited state is dramatically different between CypA and D66A (Fig. 5a, 5b, Table S3). Major differences are also apparent in the comparison of measured $R_{ex}$ values for CypA and D66A, which shows that the large dispersions measured for residues in the 70s-loop in CypA have been completely quenched in D66A (Fig. 5c, Table S4) while $R_{ex}$ contributions can be measured in D66A for residues outside the 70s region. This suggests that D66A may undergo different millisecond exchange processes. CPMG dispersion profiles for individual residues (Fig. 5d) further support this hypothesis. Residues in the 70s-loop region (e.g. Asp/Ala66; Gly74) that exhibited strong

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

dispersion in CypA are no longer sensitive to CPMG pulses in D66A, whereas other residues show more pronounced or similar dispersions elsewhere in the structure (e.g. Gly80, Asn87). Furthermore, the difference in chemical shifts measured between CypA and D66A proteins in their ground states correlates well with the difference in chemical shifts between CypA ground and excited states (Fig. 5e). This provides additional evidence that the exchange process involving the 70s-loop in CypA is an order-disorder transition. By contrast the differences in chemical shifts measured between CypA and D66A ground states do not correlate with the differences in chemical shifts between D66A ground and excited states (Fig. 5f). Altogether the data suggests that the exchange process measured for D66A is unrelated to an order-disorder transition of the 70s-loop region.

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

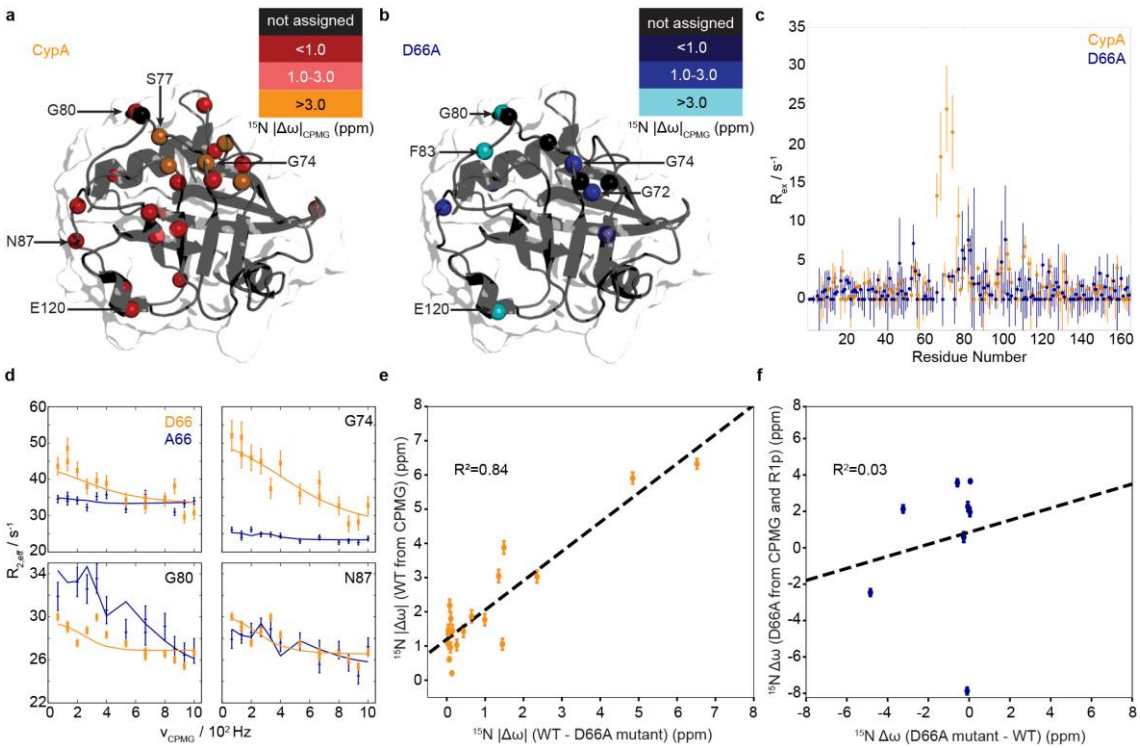


**Figure 5. Characterisation of µs-ms dynamics in CypA and D66A. a** Location of residues showing significant changes in chemical shifts between ground and excited states in CypA as determined from exchange parameters. **b** Same as **a** for D66A. **c** estimated contribution to relaxation from exchange, taken as the average of the difference between the two lowest and highest frequency measurements from CPMG dispersion curves. **d** Comparison of selected dispersion profiles measured for CypA (orange) and D66A (blue) at 283 K and 600 MHz. **e** Comparison of magnitude of differences in chemical shifts between CypA and D66A as observed by HSQC measurements and between CypA ground and excited states as determined from exchange parameters. **f** Same as **e** but for differences in chemical shifts between D66A and CypA as observed by HSQC measurements, and between ground and excited states of D66A as determined from exchange parameters.

## Discussion

The present results demonstrate comprehensive description of specific millisecond-timescale processes in proteins by combining molecular simulations with NMR measurements. The resulting conformational ensemble for CypA describes a range of NMR observables with accuracy comparable to that of NMR structure-refinement protocols. Yet there are remarkable structural differences. The DYANA ensemble does not display large amplitude conformational changes of the 70s-loop or the 100s-loop.[45] Conformational variability in the CYANA ensemble is mostly in the 70s-loop region,[31]

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

but the closed state structures in that ensemble do not resemble that observed in the X-ray structure. By contrast the present MD-derived ensemble suggests that both the 70s-loop and 100s-loop undergo large amplitude motions. The source of these discrepancies may be linked to the nature of the NMR measurements used to validate the ensembles. Indeed, back-calculated NOE, J-coupling and RDC values from a single X-ray structure are only slightly less accurate than those obtained with the various NMR and MD ensembles (Fig. 2a-c). Most of the available NMR observables report on the position of residues in structured regions of the protein, and experimental data to characterize the flexible loop regions are relatively scarce. Indeed reanalysis of the MD ensemble using only states from the closed/open (orange) macro-state yields observables of similar accuracy to the full ensemble (Fig. S8). This stresses the need to validate conformational ensembles with experimentally testable hypotheses.

Our results open the door to *in silico* generation of such hypotheses for manipulating thermodynamics and kinetic parameters in a protein ensemble, information that is lacking in traditional workflows for NMR-based structure calculations. There has been much debate about the relationship between millisecond timescale motions in CypA and catalysis.[51] The present results suggest that the previously reported millisecond timescale process involving the 70s-loop region in CypA is a local order-disorder transition that does not dramatically perturb the shape of the active site, as evidenced by X-ray and ITC measurements on a CypA-ligand complex. These results are in line with previously reported observations that millisecond time scale motions of the 70s-loop do not appear to be linked with the catalytic cycle.[52,53] Previous work has shown that while human CypA only weakly restricts the HIV-2 virus, mutations in the 70s-loop region of RhTC confer potent HIV-2 restriction.[49] The proposed mechanism involved alternative conformations of the 70s-loop thought to be inaccessible to CypA.[50] Our results suggest

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

instead that CypA's energy landscape readily permits exploration of large scale rearrangement of the 70s-loop; thus detailed characterisation of transient conformational states emerge as a powerful strategy to understand how mutations in protein sequence may evolve new function.

Alternative conformational states of the 70s-loop in CypA also offers prospects for the design of next-generation inhibitors that could potentially overcome isoform-selectivity concerns.[54,55] The D66A mutant offers a template for such efforts and further work to validate the mechanism of order-disorder transition in the 70s-loop of other cyclophilin isoforms is warranted. Engineering methodologies based on antibodies have been developed to trap transient conformational states and facilitate ligand discovery efforts.[56–58] Our results suggest that molecular simulations may be used enhance the effectiveness of conformational trapping technologies.

The ability to describe millisecond timescale conformational changes in proteins by MD simulations is shedding new light on NMR relaxation measurements. The interpretation of such experiments is traditionally performed by fitting data to two or three states phenomenological models. By contrast MSMs constructed from MD datasets typically use a much larger number of states to describe protein dynamics. In the present case there is an apparent discrepancy between CPMG-derived populations for the CypA excited state (ca. 2%, Table S3) and those for the CypA open 70s-loop macro-states from the MSM (ca. 25%, Fig. 2e). Quantitative agreement is not expected owing to technical limitations of the simulations. However other factors may contribute to this discrepancy. Inspection of the distributions of chemical shifts of snapshots sampled from the various macro-states show that in many cases there is a significant overlap in chemical shift values between distinct macro-states (Fig. S10). Re-analysis of the MSMs, under the assumption that open 70s-loop conformations with chemical shifts

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

similar to closed 70s-loop conformations cannot be distinguished from the ground state lead to populations for the ground state and excited state (ca. 91/9% respectively, Table S5) that are in closer agreement with the NMR measurements. This suggests that two-state CPMG analyses may underestimate the populations of minor states owing to a degree of degeneracy of chemical shifts with respect to different protein conformations.

Overall this work has demonstrated that molecular simulation methodologies may now be combined with NMR to offer atomic-level details about millisecond-timescale conformational changes in proteins. Such ability is expected to be widely enabling for protein structure-dynamics-function relationships studies, and for unlocking new opportunities in bioengineering or drug discovery.

## Methods

Detailed simulation and experimental methods are provided in the SI.

## Contributions

M.D.W., P.N.B., A.N.H., A.J.B., J.J-J. and J.M. conceived the overall strategy of the study. J. J-J., A.S.J.S.M. and J.M. conceived and implemented the aMD/MD/MSM approach, carried out and analysed the computational work. A.G., G.K. and A.J.B. performed and analysed NMR experiments. A.D.S. and A.N.H. synthesized and purified compound **1**. C.G., H.I., P.N.B. and M.D.W. performed X-ray crystallography. A.G., H.I. and P.N.B. performed Iso-Thermal Calorimetry experiments. All authors discussed the results, designed experiments and J. J-J, G.K., A.J.B. and J.M. wrote the manuscript.

## Acknowledgements

Authors thank Prof. Beat Vogeli for kindly providing the set of NMR observables of Cyp A and Prof. Xavier Hanoulle and Prof. Guy Lippens for kindly providing the NMR

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

residue assignment of wild type CypA. J. M. is supported by a Royal Society University Research Fellowship. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 655667 awarded to J. J-J., from the European Research Council under the European Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement No. 336289 and from EPSRC (grant no. EP/P011330/1). This project made use of time on the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk) granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (grant no. EP/L000253/1).

## References

1. Fenwick, R. B., van den Bedem, H., Fraser, J. S. & Wright, P. E. Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc. Natl. Acad. Sci.* **111,** E445–E454 (2014).
2. Fraser, J. S. *et al*. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci.* **108,** 16247–16252 (2011).
3. Basudhar, D. *et al*. Analysis of cytochrome P450 CYP119 ligand-dependent conformational dynamics by two-dimensional NMR and x-ray crystallography. *J. Biol. Chem.* **290,** 10000–10017 (2015).
4. Olsson, S., Wu, H., Paul, F., Clementi, C. & Noé, F. Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Natl. Acad. Sci.* **114,** 8265–8270 (2017).
5. Fenwick, R. B., Schwieters, C. D. & Vögeli, B. Direct Investigation of Slow

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Correlated Dynamics in Proteins via Dipolar Interactions. *J. Am. Chem. Soc.* **138,** 8412–8421 (2016).

6. Vögeli, B., Kazemi, S., Güntert, P. & Riek, R. Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. *Nat. Struct. Mol. Biol.* **19,** 1053–1058 (2012).
7. Meng, Y. *et al*. Predicting the Conformational Variability of Abl tyrosine kinase using Molecular Dynamics Simulations and Markov State Models. *J. Chem. Theory Comput.* acs.jctc.7b01170 (2018). doi:10.1021/acs.jctc.7b01170
8. Baldwin, A. J., Hansen, D. F., Vallurupalli, P. & Kay, L. E. Measurement of methyl axis orientations in invisible, excited states of proteins by relaxation dispersion NMR spectroscopy. *J. Am. Chem. Soc.* **131,** 11939–11948 (2009).
9. Shaw, D. E. *et al*. Atomic-level characterization of the structural dynamics of proteins. *Science* **330,** 341–346 (2010).
10. Ban, D., Smith, C. A., de Groot, B. L., Griesinger, C. & Lee, D. Recent advances in measuring the kinetics of biomolecules by NMR relaxation dispersion spectroscopy. *Arch. Biochem. Biophys.* **628,** 81–91 (2017).
11. Robustelli, P., Stafford, K. A. & Palmer, A. G. Interpreting protein structural dynamics from NMR chemical shifts. *J. Am. Chem. Soc.* **134,** 6365–6374 (2012).
12. Persson, F. & Halle, B. Transient Access to the Protein Interior: Simulation versus NMR. *J. Am. Chem. Soc.* **135,** 8735–8748 (2013).
13. van den Bedem, H. & Fraser, J. S. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nat. Methods* **12,** 307–318 (2015).
14. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nature Chemical Biology* **5,** 808–814 (2009).
15. Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

proteins: Signaling and other functions. *Current Opinion in Structural Biology* **20,** 142–147 (2010).

16. Mizohata, E., Nakane, T., Fukuda, Y., Nango, E. & Iwata, S. Serial femtosecond crystallography at the SACLA: breakthrough to dynamic structural biology. *Biophys. Rev.* **10,** 209–218 (2018).
17. Stagno, J. R. *et al.* Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541,** 242–246 (2017).
18. Laio, A. & Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12562–12566 (2002).
19. De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* acs.jmedchem.5b01684 (2016). doi:10.1021/acs.jmedchem.5b01684
20. Camilloni, C., Cavalli, A. & Vendruscolo, M. Replica-averaged metadynamics. *J. Chem. Theory Comput.* **9,** 5610–5617 (2013).
21. Bernardi, R. C., Melo, M. C. R. & Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* **1850,** 872–877 (2015).
22. Otten, R., Villali, J., Kern, D. & Mulder, F. A. A. Probing microsecond time scale dynamics in proteins by methyl 1H Carr-Purcell-Meiboom-Gill relaxation dispersion NMR measurements. Application to activation of the signaling protein NtrCr. *J. Am. Chem. Soc.* **132,** 17004–17014 (2010).
23. Baldwin, A. J., Religa, T. L., Hansen, D. F., Bouvignies, G. & Kay, L. E. 13CHD2 Methyl group probes of millisecond time scale exchange in proteins by 1H relaxation dispersion: An application to proteasome gating residue dynamics. *J. Am. Chem. Soc.* **132,** 10992–10995 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24. Tzeng, S. R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462,** 368–372 (2009).

25. Berjanskii, M. V. & Wishart, D. S. Unraveling the meaning of chemical shifts in protein NMR. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1865,** 1564–1576 (2017).

26. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* **337,** 476–478 (1989).

27. Wang, P. & Heitman, J. The cyclophilins. *Genome Biol* **6,** 226 (2005).

28. Davis, T. L. *et al*. Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. *PLoS Biol*. **8,** e1000439 (2010).

29. Eisenmesser, E. Z. *et al*. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438,** 117–21 (2005).

30. Fraser, J. S. *et al*. Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462,** 669–673 (2009).

31. Chi, C. N. *et al*. A Structural Ensemble for the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution. *Angew. Chemie - Int. Ed*. **54,** 11657–11661 (2015).

32. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science (80-. )*. **295,** 1520–1523 (2002).

33. Holliday, M. J. *et al*. Structure and dynamics of GeoCyp: A thermophilic cyclophilin with a novel substrate binding mechanism that functions efficiently at low temperatures. *Biochemistry* **54,** 3207–3217 (2015).

34. Ni, S. *et al*. Discovering potent small molecule inhibitors of cyclophilin A using de novo drug design approach. *J. Med. Chem.* **52,** 5295–8 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

35. Sweeney, Z. K., Fu, J. & Wiedmann, B. From Chemical Tools to Clinical Medicines: Nonimmunosuppressive Cyclophilin Inhibitors Derived from the Cyclosporin and Sanglifehrin Scaffolds. (2014).

36. Ahmed-Belkacem, A. *et al*. Fragment-based discovery of a new family of non-peptidic small-molecule cyclophilin inhibitors with potent antiviral activities. *Nat. Commun.* **7,** (2016).

37. Daum, S. *et al*. Isoform-specific inhibition of cyclophilins. *Biochemistry* **48,** 6268–77 (2009).

38. Nagaraju, M., McGowan, L. C. & Hamelberg, D. Cyclophilin a inhibition: Targeting transition-state-bound enzyme conformations for structure-based drug design. *J. Chem. Inf. Model.* **53,** 403–410 (2013).

39. Georgiou, C. *et al*. Pushing the Limits of Detection of Weak Binding Using Fragment-Based Drug Discovery: Identification of New Cyclophilin Binders. *J. Mol. Biol.* **429,** 2556–2570 (2017).

40. Guichou, J.-F. *et al*. Structure-based design, synthesis, and biological evaluation of novel inhibitors of human cyclophilin A. *J. Med. Chem.* **49,** 900–10 (2006).

41. Dunsmore, C. J. *et al*. Design and Synthesis of Conformationally Constrained Cyclophilin Inhibitors Showing a Cyclosporin-A Phenotype in C. elegans. *ChemBioChem* **12,** 802–810 (2011).

42. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120,** 11919–11929 (2004).

43. Pierce, L. C. T., Salomon-Ferrer, R., Augusto F. De Oliveira, C., McCammon, J. A. & Walker, R. C. Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory Comput.* **8,** 2997–3002 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

44. Vögeli, B., Olsson, S., Güntert, P. & Riek, R. The Exact NOE as an Alternative in Ensemble Structure Determination. *Biophys. J.* **110,** 113–126 (2016).

45. Ottiger, M., Zerbe, O., Güntert, P. & Wüthrich, K. The NMR solution conformation of unligated human cyclophilin A. *J. Mol. Biol.* **272,** 64–81 (1997).

46. Gamble, T. R. *et al.* Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87,** 1285–1294 (1996).

47. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PLoS One* **7,** (2012).

48. Lindorff-Larsen, K., Maragakis, P., Piana, S. & Shaw, D. E. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J. Phys. Chem. B* **120,** 8313–8320 (2016).

49. Price, A. J. *et al.* Active site remodeling switches HIV specificity of antiretroviral TRIMCyp. *Nat. Struct. Mol. Biol.* **16,** 1036–1042 (2009).

50. Caines, M. E. C. *et al.* Diverse HIV viruses are targeted by a conformationally dynamic antiviral. *Nat. Struct. Mol. Biol.* **19,** 411–416 (2012).

51. Kamerlin, S. C. L. & Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis. *Proteins: Structure, Function and Bioinformatics* **78,** 1339–1375 (2010).

52. Otten, R. *et al.* Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* **9,** (2018).

53. Wapeesittipan, P., Mey, A., walkinshaw, malcolm & Michel, J. Allosteric effects in catalytic impaired variants of the enzyme cyclophilin A may be explained by changes in nano-microsecond time scale motions. *bioRxiv* (2018).

54. Sweeney, Z. K., Fu, J. & Wiedmann, B. From chemical tools to clinical medicines: Nonimmunosuppressive cyclophilin inhibitors derived from the

bioRxiv preprint doi: https://doi.org/10.1101/490987; this version posted December 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cyclosporin and sanglifehrin scaffolds. *J. Med. Chem.* **57,** 7145–7159 (2014).

55. De Simone, A. *et al*. A computationally designed binding mode flip leads to a novel class of potent tri-vector cyclophilin inhibitors. *Chem. Sci*. (2019). doi:10.1039/C8SC03831G

56. Lawson, A. D. G. Antibody-enabled small-molecule drug discovery. *Nature Reviews Drug Discovery* **11,** 519–525 (2012).

57. Hutchings, C. J., Koglin, M., Olson, W. C. & Marshall, F. H. Opportunities for therapeutic antibodies directed at G-protein-coupled receptors. *Nature Reviews Drug Discovery* **16,** 787–810 (2017).

58. Tehan, B. G. & Christopher, J. A. The use of conformationally thermostabiliseded GPCRs in drug discovery: application to fragment, structure and biophysical techniques. *Current Opinion in Pharmacology* **30,** 8–13 (2016).