Dynamic design: manipulation of millisecond timescale motions on the energy landscape of Cyclophilin A

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.

[¬] Present Addresses: 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 authors. E-mail: jordi.juarez@ed.ac.uk, mail@julienmichel.net

1

Abstract

Proteins need to interconvert between many conformations in order to function, many of which are formed transiently, and sparsely populated. Particularly when the lifetimes of these states approach the millisecond timescale, identifying the relevant structures and the mechanism by which they inter-convert remains a tremendous challenge. Here we introduce a novel combination of accelerated MD (aMD) simulations and Markov State modelling (MSM) to explore these 'excited' conformational states. Applying this to the highly dynamic protein CypA, a protein involved in immune response and associated with HIV infection, we identify five principally populated conformational states and the atomistic mechanism by which they interconvert. A rational design strategy predicted that the mutant D66A should stabilise the minor conformations and substantially alter the dynamics whereas the similar mutant H70A should leave the landscape broadly unchanged. These predictions are confirmed using CPMG and R1ρ solution state NMR measurements. By accurately and reliably exploring functionally relevant, but sparsely populated conformations with milli-second lifetimes *in silico*, our aMD/MSM method has tremendous promise for the design of dynamic protein free energy landscapes for both protein engineering and drug discovery.

Introduction

Capturing the complete dynamics of proteins to understand their many functions has long been a goal of structural biology. Experimentally, trapping proteins into relevant conformational states and characterising these at atomic resolution provides tremendous insight[1–3]. Such static 'snapshots' are inherently incomplete and especially in cases where proteins are particularly dynamic, understanding function can require characterisations of conformational states that are formed only transiently, and populated as sparsely as a few per cent[4–6]. These conformers are typically challenging to characterise experimentally. Solution-state NMR has proven unique in its ability to detect and structurally characterise these functionally relevant states, whose lifetimes can be on the order of a few milliseconds and are otherwise 'invisible' to experimental measures[6–8]. Such states can play important roles in processes as diverse as protein folding, molecular recognition and catalysis, but remain challenging to characterise[9–16]. To engineer proteins with new functions, and for drug discovery, there is a clear need to be able to characterise, explore and manipulate the full set of conformational states accessed by proteins, and the mechanism by which they interconvert.

Molecular dynamics (MD) simulations provide an attractive means to accomplishing this goal *in silico*. To characterise low amplitude structural fluctuations in local minima, simulations on timescales of a few hundred nano-seconds may be sufficient[17,18]. Substantially longer simulations are required in order to sample sparsely populated 'excited' conformational states with millisecond lifetimes. Such a 'brute force' calculation is possible at present only for relatively small proteins using dedicated hardware that is not widely available[19]. To make progress, two broad categories of enhanced sampling methods have been developed with the goal of increasing total

3

effective time covered by a simulation. Predefined collective variables can be exhaustively sampled (Metadynamics[20–22], Umbrella Sampling[23–25], Steered MD[26–28]), or the potential energy function of the forcefield can be explicitly altered (scaled MD[29–32] accelerated MD (aMD)[33,34]). The former class of methods require the degrees of freedom underpinning the motion of interest to be supplied in advance, which pre-supposes that the relevant motions are already known. The latter class of methods can overestimate the population of high energy regions of the free energy surface (FES) and so explore unphysical regions of conformational space.

Aiming to improve upon these methods, Markov State Modelling (MSM) based sampling approaches have emerged as an exciting method for sampling protein dynamics on millisecond timescales[2,35–41]. These techniques may be used to sample conformational space using a series of relatively short MD trajectories. These trajectories are analysed with the goal of obtaining a discrete set of interconverting conformational 'states', where each has no knowledge of any previously visited state. This method allows a series of independently generated MD trajectories to be combined into a complete description of the energy landscape of the protein. MSM have been shown to reproduce experimental observables in model systems where the structures of the relevant states were known in advance[35,42,43], and in cases where the timescales for conformational transitions were sufficiently short to be validated with 'brute force' MD methods[19,37,44].

In this work, we generate a protocol suitable for sampling transitions between structurally ambiguous states separated by millisecond kinetic barriers (Fig. 1b). This is achieved by harnessing the ability of aMD to efficiently explore conformational space,

4

while relying on the MSM procedures to recover a more accurate thermodynamic and kinetic information description of the system. Because this approach provides atomically detailed structural ensembles and detailed information about how they interconvert, it is well suited to both generate experimentally testable mechanistic hypotheses and enable the rational design of conformational ensembles.

We test the novel aMD/MSM protocol by both characterising and rationally modifying the free energy surface of human cyclophilin A (CypA). This enzyme is a human prolyl isomerase whose incorporation into new virus assemblies is essential for HIV-1 infectivity[45–48] and HCV replication[49], making it a major drug target[50–57]. This protein has been widely studied[58–67] and possesses two functional loops, residues 65-77 (70s loop, Fig. 1a) and 100-110 (100s loop, Fig. 1a) that have been established by NMR to undergo substantial dynamics on the millisecond timescale[68–70]. An atomistic picture of the conformational changes has proven elusive. Our aMD/MSM procedure reveals a range of conformational sub-states of CypA where both the 100s and 70s loops exchange between open and closed states (Fig. 1c/d).

To manipulate the energy surface, we implemented a design strategy based on hydrogen bonding patterns to stabilise specifically sparsely populated conformers with millisecond lifetimes (Fig. 2). This procedure led us to a previously unidentified mutation D66A. Whereas WT has the 70s loop predominantly in the closed state, consistent with the majority of X-ray structures of CypA solved to date[52,55,70–75], D66A was predicted to predominantly occupy the open state of the 70s loop. Similarly, as a negative control, a second similar mutation H70A was predicted by aMD/MSM to closely resemble the WT (Fig. 3).

5

Using a combination of $R_{1\rho}$ and CPMG NMR experiments, with X-ray crystallography we tested these predictions (Fig. 4,5) and verified that exactly as predicted, the 70s loop is substantially disordered and open in D66A, and adopts a conformation very similar to the minor conformations populated in WT ensemble. Also consistent with the predictions of aMD/MSM, the free energy landscape of H70A was largely the same as for WT.

Overall, we demonstrate that the aMD/MSM method accurately predicts the conformational fluctuations of the challenging and dynamic protein CypA including states with millisecond lifetimes. The resulting free energy surface is sufficiently accurate to enable the design of mutants that stabilise specific conformations. aMD/MSM provides a powerful tool for not only the atomistic characterisation, but also the manipulation of the free energy surface of dynamic proteins, including controlling states with millisecond lifetimes.
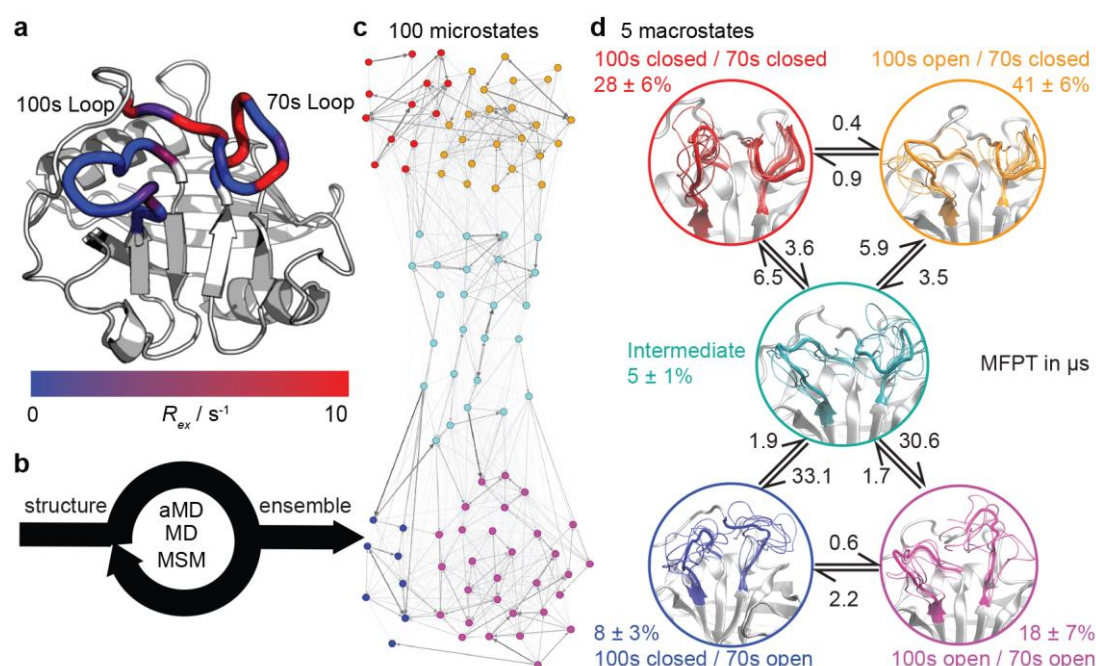
Figure 1. Calculation the dynamic ensemble of WT CypA. **a** an X-ray diffraction structure of CypA, with the 100s and 70s loops indicated. A number of residues have been previously determined to undergo conformational exchange[68,69] which is evidence by a large value of R_{ex} for these residues, measured by NMR. These values are consistent with the formation of transiently sampled, but sparsely populated alternative loop conformations. **b** It is desirable to determine the nature of the interconverting structures, their populations and mechanism by which they exchange. We accomplish this for CypA using a combination of aMD, MD and MSM methodologies (detailed methods). **c** A 100 microstates MSM for wild-type CypA that describes 70s and 100s loop motions. The width of the arrows is proportional to the inter-conversion rates. **d** The microstates were clustered into a minimal set of sub-states that together contain the relevant amplitudes of motion and time-scales of state-to-state inter-conversion present in the full ensemble. The calculated rates and populations are indicated. Error bars on reported populations were obtained by bootstrapping of the MD trajectories assigned to the individual microstates.
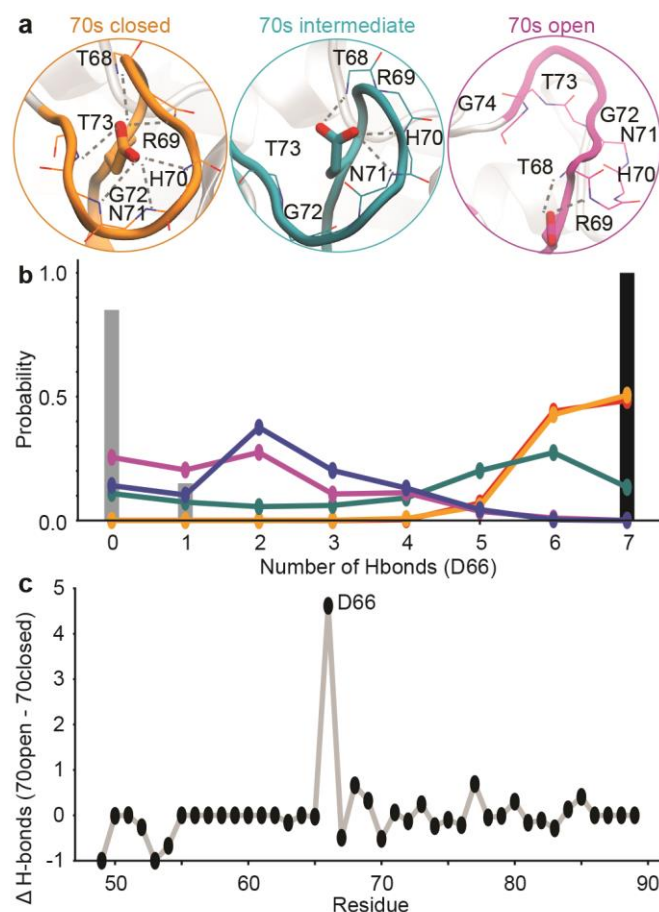
Figure 2 Design of ensemble disrupting mutation. a Specific interactions made by D66 with 70s loop residues in representative MD snapshots of the closed, intermediate and open states. **b** The probability distribution of the numbers of intra-molecular H-bonds between D66 and 70s loop residues in the 5 sub-states. These are compared to X-ray structure PDB 1AK4[71] (black), and NMR ensemble PDB 2N0T (grey)[61]. **c** The difference in the average number of intra-molecular H-bonds in the 70s closed (orange/red) and 70s open (blue/purple) states. Residue D66 stands out as 'designable' as it has substantially more H-bonds stabilising the closed state than the open state of the 70s loop.

Results

Atomic-level insights in millisecond timescale CypA loop motion mechanisms

All-atom accelerated molecular dynamics (aMD) simulations[33,34] were initiated from an X-ray structure of wild type CypA (PDB ID 1AK4[71]). In aMD conformational sampling is accelerated by adding a biasing potential to the molecular mechanics potential energy

function when it falls below a threshold (supplementary methods). Structures sampled from the aMD trajectory were used to seed sets of equilibrium MD trajectories, which were used for Markov state model (MSM) analysis. In a MSM, the protein dynamics is modelled as a set of Markov jump processes between discrete conformational states. These Markov processes can be thought of as jumps between nodes in a network, with the edges representing transition probabilities between these nodes, and where no knowledge of previously visited nodes is required to estimate the probability for a particular transition. The procedure enables reconstruction of a conformational ensemble from sets of independently run short MD trajectories that have each individually explored only a small portion of that conformational ensemble. Three rounds of MD simulations and MSM estimations were required to produce a 100 microstate MSM (Fig 1c) that was numerically stable, corresponding to a cumulative MD sampling time of 187.5 µs. Details about model validation against experimental observables are given in the Supporting Information (Figure S1-S3).

To obtain physical insight into the calculated ensemble, the 100 MSM microstates were further lumped into a coarser 5-macrostate model (see SI methods). This five state conformational landscape reveals that both the 100s and 70s loops can adopt 'open' and 'closed' conformations. While the inter-conversion of both loops were largely independent, the inter-conversion of the 100s loop was one order of magnitude faster than the 70s loop (Fig 1d, Fig S4), suggesting that the slower loop dynamics of the protein are largely associated with the opening and closing of the 70s loop. The most populated state for WT CypA, (100s open/70s closed, Fig. 1c/d orange, 41±6%) closely resembles the conformation typically observed in X-ray structures of CypA (Fig. S4), and fluctuations within the state were small (Cα RMSF values of 2.7 Å and 1.1 Å for

9

the 100s and 70s loops, respectively). From this state, the 100s loop can rapidly open to adopt the 100s closed/70s closed state (Fig 1c/d, red, RMSF 2.6 Å and 1.1 Å).

These two conformational states must overcome a substantially larger kinetic barrier to adopt open conformations of the 70s loop (100s open/70s open, magenta) and (100s closed/70s open, blue). In contrast to the 70s closed states (red/orange), the 70s loop in its open conformations is relatively disordered, as evidenced by RMSF values >4 Å. These two 70s open states can interconvert, with the 100s loop opening/closing in an analogous fashion to the inter-conversions observed in the 70s closed states. While fluctuations of the 100s loop are rapid and largely independent of the 70s loop conformations, the opening/closing transition of the 70s loop was more complex, passing via an intermediate fifth state (teal) where both 70s and 100s loops display positional fluctuations that are intermediate between open and closed (RMSF ca. 2.9 Å). Qualitatively, it is important for the 100s loop to simply 'not be in the way', as some conformations of the open 100s state can effectively block the opening of the 70s loop.

To validate these findings we sought a mutation that would stabilise the group of minor, 70s open, conformations (teal/blue/purple). The number of hydrogen bonds each residue makes to the 70s loop was assessed in both the open and closed forms (Fig. 2b). Notably, D66 was found to adopt a large number of H-bonds in the open form, and substantially fewer in the closed (Fig. 2c), making this residue stand out as 'designable' and provide a target for altering the free energy landscape of CypA. The critical role of D66 in the proposed 70s loop motions mechanism is not supported by a recent NMR structural ensemble that has been proposed to describe the CypA 70s loop conformational changes measured by CPMG experiments.[61] Analysis of the structures in that NMR bundle shows little changes in hydrogen bonding interactions with the 70s-

loop (Fig. 2b). Further none of the structures in the NMR bundle contain a 70s-loop conformation similar to that commonly observed in high-resolution X-ray structures of CypA (Fig. S5). aMD/MSM simulations of D66A were performed as described for WT CypA (Fig 1). The resulting ensemble predicted that the population of the previous ground state was substantially reduced (from ca. 40% to ca. 15%, Fig. 3a), and the intermediate (teal) state was increased in a compensatory fashion to become the new ground state (from ca. 5% to ca. 40%). Overall with respect to CypA the population of states in which the 70s-loop is closed/open transitioned from ca. 70:30% to 25:75%. Moreover, the mean first passage times between the conformational states were predicted to decrease by 10-15 fold with respect to wild type (Fig. 3a and 3c). Finally, as a negative control, we performed aMD/MSM simulations on the mutation H70A[69]. The hydrogen bonding pattern for this residue was predicted to be broadly similar when the 70s loop is either open or closed (Fig. 2c). Consistent with this, aMD/MSM simulations determined that the conformational ensemble for H70A should resemble WT (Fig. 3b).
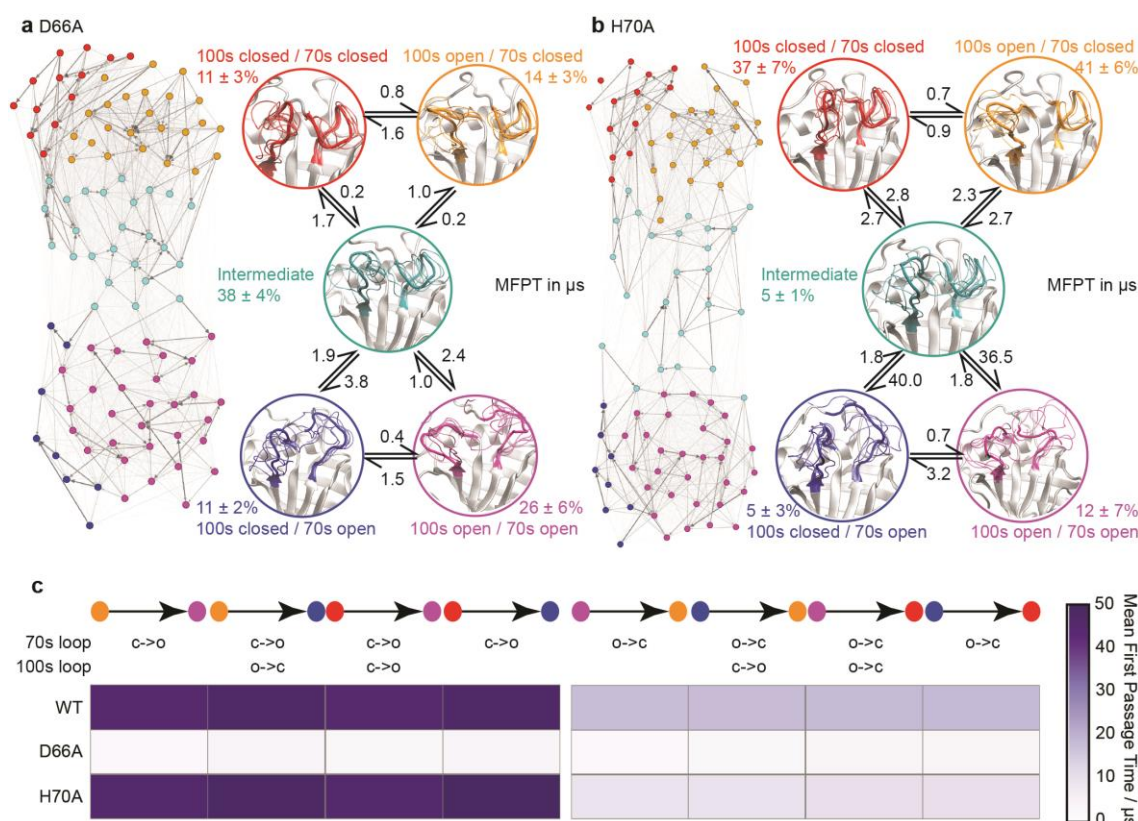
Figure 3 Ensembles of mutant CypA proteins D66A and H70A . a MSM model for D66A. **b** MSM model for H70A. Other details as in Fig 1c. **c** Heatmap of calculated mean first passage times (MFPT) between non-adjacent macro-states. The symbols 'c' and 'o' denote transitions between closed and open loop conformations.

Experimental validation of the calculated ensembles

To establish the validity of mechanistic hypotheses originating from the aMD/MSM models, we sought to compare our results to experimental data. Initially, we examined binding of an inhibitor against both WT and D66A. The K_d for D66A was 12-fold weaker than for WT (Fig. 4a), indicating a substantial change in the underlying protein conformations. Crystallisation of D66A proved challenging as we might expect for a protein with increased disorder, although both WT and D66A were successfully crystallised in the presence of the inhibitor[51] and their structures were solved (Table S2) Comparison of electron density profiles revealed that while the 70s-loop adopts as

expected a closed conformation in the WT complex (Fig 4b orange), electron density in the 70s loop for the D66A complex was largely absent (Fig. 4b blue), consistent with the notion that this loop occupies a predominantly disordered conformation in D66A.

To probe the conformational dynamics more directly, isotopically enriched samples of WT and D66A CypA were prepared for NMR analysis (Fig. S6-S7). ^{1}H-^{15}N HSQC spectra of both proteins revealed that while both were predominantly well structured (Fig. 4c, Fig. S8), there are substantial differences in the ground-state ensembles, exactly as predicted from the simulations. On a per-residue basis, the largest deviations in observed chemical shift positions between D66A and WT were found in the 70s loop region (Fig. 4e). The changes in chemical shift were quantitatively estimated by calculating the chemical shifts of individual conformers in the aMD/MSM ensembles using the program shiftx2[76]. The calculated CSP values are consistent with the experimental observation of significant changes in the 70s loop region, though the magnitude of the CSPs is underestimated (Fig 4f).

To determine if there are any changes in fast timescale motions of the 70s loop, the random coil index (RCI) was calculated for WT and D66A from the observed chemical shifts (Fig. 4d). These confirmed that the 70s loop is substantially more disordered in D66A, as expected from the simulations. This finding was further corroborated by measurement of the H-N steady-state heteronuclear NOE, where substantially lower values were recorded in the 70s loop for the D66A versus the WT. The heteronuclear NOEs were calculated from the MD ensembles, and also found to be in reasonable qualitative agreement with the experimental measurements (Fig. 4g/h).

To further characterise the fast timescale dynamics, R_2 and R_1 values were additionally measured at two magnetic field strengths, and a model-free analysis was performed. Such an analysis requires successively more complex motional models to be fitted to

13

individual residues, in order to achieve an overall self-consistent picture of the dynamics. The procedure was unable to provide a self-consistent picture of relaxation for either CypA or D66A, with large variations in parameters between adjacent residues and 'fast' local fluctuations measured to be comparable to the global tumbling rate (Fig. S9). This is likely because of the substantial contributions of chemical exchange to the individual measurements (*vide infra*). Of the measurements obtained, the heteronuclear NOE nevertheless strikes a reasonable compromise between being relatively resistant to the effects of chemical exchange, whilst still being sensitive to relatively rapid local fluctuations within the molecule (Fig. 4g/h).
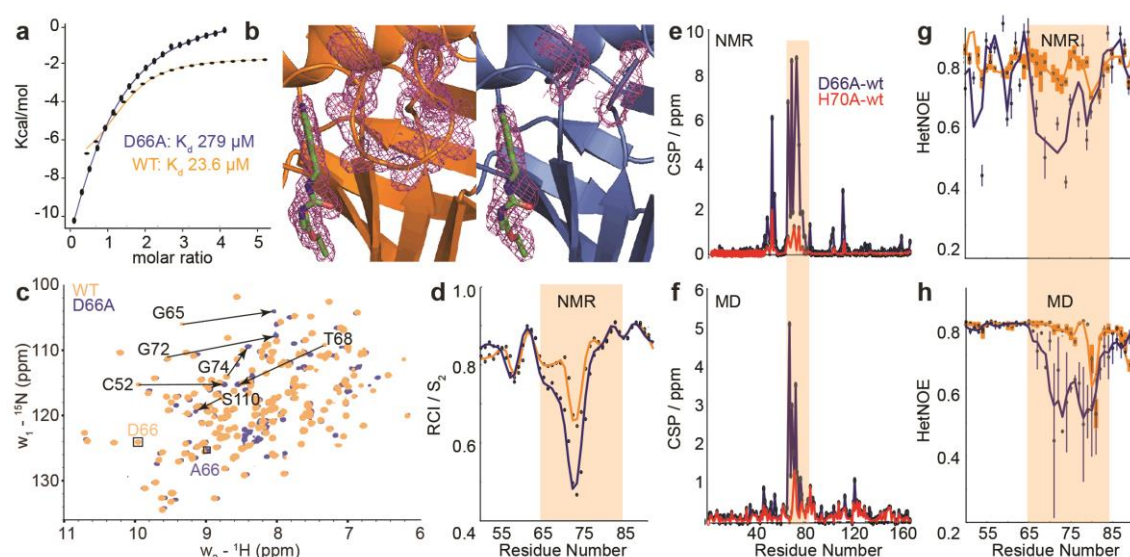

Figure 4. **Comparison of biophysical characterisation and MD/MSM conformational ensembles for WT CypA, D66A and H70A. a** Binding isotherms for WT and D66A measured by ITC against an inhibitor. **b** X-ray structure of WT and D66A bound to inhibitor. Magenta meshes indicate $2Fo$-Fc electron density map at an isocontour of 1.0σ for the inhibitor and protein residues 65 to 75. **c** ^{1}H-^{15}N HSQC correlation spectra of WT and D66A at (10 ^{o}C), residues showing significant CSPs are highlighted. **d** Random coil index S_2 values calculated from Talos+ for WT and D66A revealing increased disorder in the vicinity of the 70s loop for D66A. **e** Measured CSP per residue with respect to WT for D66A (blue) and H70A (red)[69]. **f** Calculated CSPs per residue from the computed ensembles. **g** Measured steady-

state ^{15}N-{^{1}H} heteronuclear NOE transfers for WT and D66A in the vicinity of the 70s loop. **h** Calculated ^{15}N-{^{1}H} heteronuclear NOE transfers in the vicinity of the 70s loop from the computed ensembles.

While the fast ps/ns time-scale dynamics indicate substantial differences in the 70s loop, the major goal was to determine if the structure of D66A resembles the excited state of WT CypA, thus validating the design strategy. A combination of $R_{1\rho}$ and CPMG experiments were employed to study these conformational fluctuations on the millisecond timescale (Fig. 5a,b). The CPMG data for WT strongly resembles that obtained previously[68,69]. The combined data were analysed globally within a modified version of the software CATIA[77]. Such an analysis can ascertain a minimally complex binding model, and provide experimental measures of the kinetic and thermodynamic parameters that govern the exchange. Moreover, the analysis also provides chemical shifts (Fig. 5c,d), which characterise the structural changes that accompany the millisecond conformational rearrangements. For WT, a two-state model was found to explain well the combined dataset to within experimental error. The exchange rate k_{ex} was determined to be approximately 2200 s^{-1} and a population of excited state was determined to be approximately 2% (Table S3).

Remarkably, the chemical shift differences directly measured from WT and D66A HSQC spectra were in good agreement with the chemical shift differences between ground and excited states of WT obtained by CPMG/R1ρ measurements (Fig. 5e). This correlation strongly supports the *in silico* design strategy demonstrating that the ground state of D66A is similar to the 'excited' minor conformational state adopted by the WT, and that the design process effectively swapped the major and minor conformers.

15

Having established that the ground state of D66A resembles the 'excited' state of WT, we acquired CPMG and R1ρ data to further characterise its millisecond dynamics. The raw exchange rates R_{ex} for D66A in the 70s loop R_{ex} were found substantially smaller than those detected for the WT protein, and residues that had large R_{ex} values in WT were reduced to zero in D66A (e.g. A66, G74 Fig. 5) indicating the success of the design protocol. Consistent with this, the pattern of specific residues involved with the conformational dynamics was very different (Fig. 5, Table S3) to WT. As with WT, the data could be well explained by fitting to a 2-state model, where the exchange rate was similar to WT (k_{ex} ca. 2,000 s^{-1}), but the population of the excited state was much smaller, very close to our detection threshold (p_B ca. 0.5%). Analysis of the excited state chemical shifts shows that this new state is not moving towards a particularly disordered state, nor the ground state of CypA. Instead, since a number of residues in the 80s loop were found to have relatively large R_{ex} values in D66A this suggests that the excited state involves rearrangements of the 80s loop, outwith the scope of the present simulations that were designed to resolve millisecond timescale motions of the 70s and 100s loops. Overall the key finding is that the NMR data shows that the 'ground' conformational state of D66A has the 70s loop in an 'open' conformation, which from the chemical shifts, is similar to the 'excited' state adopted by the WT protein.
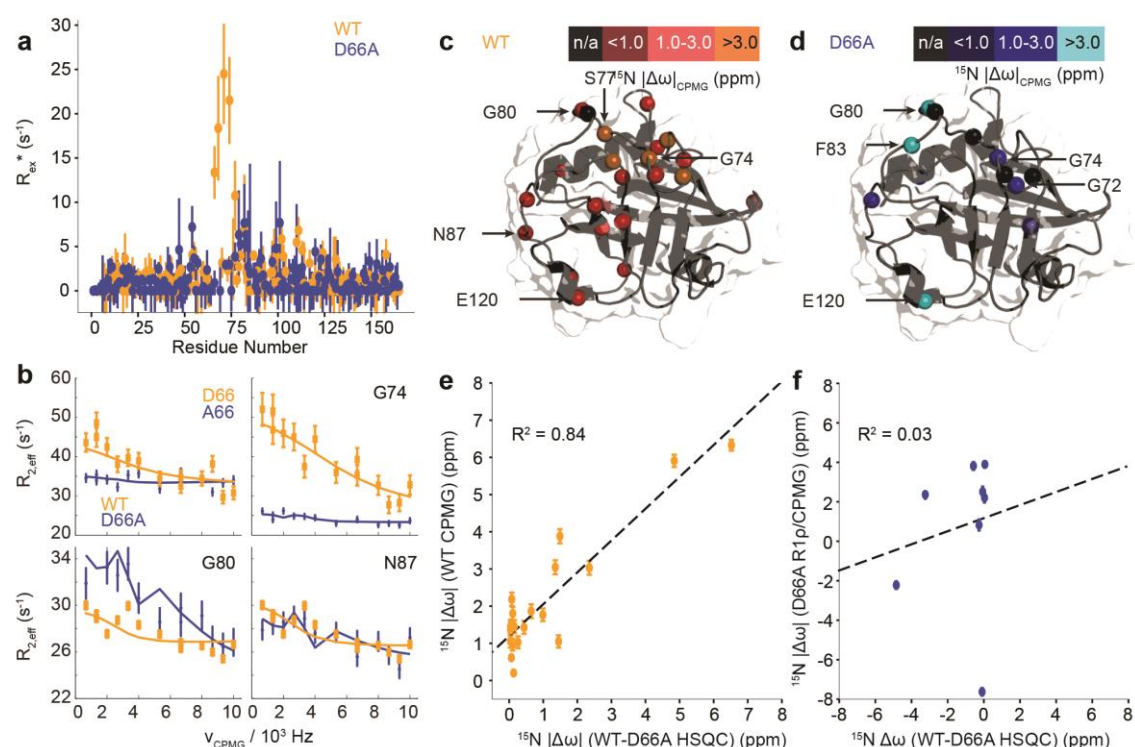
Figure 5. Characterisation of µs-ms dynamics in CypA and D66A by CPMG and R1ρ NMR measurements. a estimated contribution to relaxation from exchange, taken as the average of the difference between the lowest and highest frequency measurements from CPMG dispersion curves. **b** Comparison of selected dispersion profiles measured for CypA (orange) and D66A (blue) at 283 K and 600 MHz. **c** The experimentally determined chemical shift changes, Δω, from analysis of the CPMG/R1ρ curves for WT and (**d**) D66A. **e** Comparison of magnitude of Δω between CypA and D66A from HSQC measurements (x-axis) and from CPMG/R1ρ measurements on the WT (y-axis). The agreement between the two is remarkably good. **f** Comparison of magnitude of Δω between CypA and D66A from HSQC measurements (x-axis) and from CPMG/R1ρ measurements on D66A (y-axis). The agreement is poor revealing that the 'excited state' in D66A is not the ground state of WT CypA.

Discussion

The work demonstrates that our new hybrid method aMD/MSM can accurately and reliably provide a comprehensive description of protein free energy surfaces, including a characterisation of sparsely populated conformational states with millisecond lifetimes. The simulations are hypothesis generating, and are sufficiently accurate to be amenable to rational design strategies whose goal is to adjust the relative populations of the different conformational states (Fig. 3).

17

It is interesting to compare different ensembles of WT CypA available in the literature. A notable ensemble of WT CypA [61] was calculated recently from solution-state NMR data. Our aMD/MSM ensemble is also in good agreement with NOE, J-coupling and RDC solution-state NMR data (Figure S3). However while both ensembles suggest large amplitude motions of the 70s loop, the NMR ensemble does not support the critical role of D66 in a conformational change between open and closed loop conformations, and is also inconsistent with the closed 70s loop conformation observed by X-ray diffraction experiments.

To further test our method, we constructed an ensemble of CypA purely from our aMD trajectories. Unlike the aMD/MSM ensemble, the resulting aMD ensemble predicts that CypA adopts predominantly a 100sopen/70sopen conformation (Fig. S11), a prediction inconsistent with experimental data. Taken together, the aMD/MSM ensemble is the only method that characterises the conformational exchange, revealing the presence of both open and closed conformations of 70s loop in the WT with sufficient precision to provide the means to redesign the conformational landscape.

The resulting conformational states of |CypA also provide insight into function. The open conformational state of the 70s loop has been linked to restriction of the HIV-2 virus. A double mutant (D66N/R69H) of the rhesus macaque homologue TRIMCyp adopts a 70s open conformation, which has been linked to potent HIV-2 restriction.[78] These observations lead to the interesting speculation that the sparsely populated open 70s conformation in human CypA plays an important and unrealised role in the pathway of HIV-2 infectivity, suggesting stabilisation of this conformation could lead to novel therapeutics.

Approaches have been developed based on using antibodies to stabilise transiently populated conformation states to facilitate ligand discovery efforts[79–81]. Our results open an exciting opportunity for determining this *in silico* as part of rational drug design endeavours. In the specific case of CypA, current treatments methods have insufficient isoform selectivity[53,54]. We envision that aMD/MSM can be used to selectively target transient states that are differently populated among different protein isoforms, hence paving the way for the rational drug design of isoform selective inhibitors.

There has been substantial discussion on the relationship between millisecond timescale motions in CypA and its catalytic cycle.[82] Initial measurements suggested that the two could be linked, although this has been disputed by later more detailed measurements [60,62]. Our results indicate that the opening and closing of the 70s loop is largely independent of the rest of the protein, supporting the notion that its fluctuations are disconnected from the active site.

There remain technical challenges when comparing aMD/MSM simulations to experimental data. To achieve this, it is necessary to calculate chemical shifts from the structural ensembles and combine them appropriately to encapsulate the effects of intermediate exchange in order to accurately capture the effects of millisecond dynamics in NMR data. Chemical shift calculation algorithms such as shiftX2 are trained on static structures and not MD trajectories, and so their accuracy in such applications has not been extensively validated. Nevertheless, the agreement between the predicted and measured CSPs between WT/D66A and H70A/WT are very

encouraging. However agreement between R_{ex} values calculated for one protein calculated in isolation is less impressive, perhaps suggesting that systematic deficiencies in the chemical shift calculations are removed when one takes the difference between two simulated states. Moreover, when using the expectation values for the CSPs from shiftX2 with the rate constants predicted by the MSM leads to a large overestimation of R_{ex} values, on the order of hundreds per second. This suggests that a combination of both the populations and the kinetics together with the estimated chemical shifts are together not in perfect agreement. This can be partially explained by noting that while the CPMG and R1ρ data was acquired at 10 ^{o}C, the simulations were performed at 25 ^{o}C. These technical challenges aside, it is clear that the main predictions of aMD/MSM are fundamentally correct, and the key conformational ensembles predicted are sufficiently accurate to enable rational design efforts and to generate and test structural hypotheses.

Overall we demonstrate that aMD/MSM can provide atomic-level insights into the thermodynamics and kinetics of independent conformational sub-ensembles that characterise the free energy surface of CypA. The picture is chemically intuitive, and sufficiently accurate for rational design efforts aimed at stabilising particular conformational states over others, even when the lifetimes of the individual states reaches the millisecond timescale. We anticipate that this protocol can greatly facilitate future protein structure-dynamics-function relationships studies and will unlock new opportunities in bioengineering or drug discovery.

Methods

Molecular dynamic simulations were performed using AMBER16[83] and GROMACS5.0[84] software packages and processed using MDTRAJ.[85] MSMs were built and analysed using PyEMMA 2.3.0.[86] PALES[87] and BME[88] were used to compute RDCs, chemical shifts were calculated with ShiftX2[76] and MultiEx[89]. NMRPipe[90] and Sparky3.1[91] were used to process all spectra. The CCP4 suite[92] was used for the structure refinement of crystallographic data. Detailed simulation and experimental methods are provided in the Supporting Information.

Supporting information

Detailed Materials and Methods section, MSM validation procedures, LC-MS analysis for CypA and D66A, additional NMR experiments, CypA assignment, X-ray Refinement statistics, exchange parameters for CypA and D66A, measured R_{ex} values. Instructions to download the available datasets. Figures S1-S12. Tables S1-S5.

Acknowledgements

References

1.   Lee, G. M. & Craik, C. S. Trapping Moving Targets with Small Molecules. *Science (80-.).* **324**, 213–215 (2009).

2.   Chen, S. *et al.* The dynamic conformational landscape of the protein methyltransferase SETD8. *Elife* **8**, (2019).

3.   Lee, G. M., Shahian, T., Baharuddin, A., Gable, J. E. & Craik, C. S. Enzyme Inhibition by Allosteric Capture of an Inactive Conformation. *J. Mol. Biol.* **411**, 999–1016 (2011).

4.   Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **8**, 932–935 (2001).

5.   Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying 'invisible' excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012).

6.   Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nature Chemical Biology* **5**, 808–814 (2009).

7.   Tzeng, S.-R. & Kalodimos, C. G. Allosteric inhibition through suppression of

transient conformational states. *Nat. Chem. Biol.* **9**, 462–465 (2013).

8.   Yuwen, T., Sekhar, A., Baldwin, A. J., Vallurupalli, P. & Kay, L. E. Measuring Diffusion Constants of Invisible Protein Conformers by Triple-Quantum 1H CPMG Relaxation Dispersion. *Angew. Chemie - Int. Ed.* **57**, 16777–16780 (2018).

9.   Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: Signaling and other functions. *Current Opinion in Structural Biology* **20**, 142–147 (2010).

10.   Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).

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

12.   Tzeng, S.-R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488**, 236–240 (2012).

13.   Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci.* **108**, 16247–16252 (2011).

14.   Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

15.   Vallurupalli, P., Chakrabarti, N., Pomès, R. & Kay, L. E. Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. *Chem. Sci.* **7**, 3602–3613 (2016).

16.   Wang, Y., Papaleo, E. & Lindorff-Larsen, K. Mapping transiently formed and sparsely populated conformations on a complex energy landscape. *Elife* **5**, (2016).

17. Maximova, T., Moffatt, R., Ma, B., Nussinov, R. & Shehu, A. Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. *PLOS Comput. Biol.* **12**, e1004619 (2016).

18. Huggins, D. J. *et al.* Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **9**, e1393 (2019).

19. Shaw, D. E. *et al.* Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (80-.).* **330**, 341–346 (2010).

20. Laio, A. & Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562–12566 (2002).

21. Leone, V., Marinelli, F., Carloni, P. & Parrinello, M. Targeting biomolecular flexibility with metadynamics. *Current Opinion in Structural Biology* **20**, 148–154 (2010).

22. Camilloni, C., Cavalli, A. & Vendruscolo, M. Replica-averaged metadynamics. *J. Chem. Theory Comput.* **9**, 5610–5617 (2013).

23. Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932–942 (2011).

24. Souaille, M. & Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **135**, 40–57 (2001).

25. Bartels, C. & Karplus, M. Probability Distributions for Complex Systems: Adaptive Umbrella Sampling of the Potential Energy. *J. Phys. Chem. B* **102**, 865–880 (1998).

26. Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11**, 224–230 (2001).

27. Park, S. & Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120**, 5946–5961 (2004).

28. Xiong, H., Crespo, A., Marti, M., Estrin, D. & Roitberg, A. E. Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship. *Theor. Chem. Acc.* **116**, 338–346 (2006).

29. Sinko, W., Miao, Y., de Oliveira, C. A. F. & McCammon, J. A. Population based reweighting of scaled molecular dynamics. *J. Phys. Chem. B* **117**, 12759–12768 (2013).

30. Miao, Y. & McCammon, J. A. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol. Simul.* **42**, 1046–1055 (2016).

31. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).

32. Tsujishita, H., Moriguchi, I. & Hirono, S. Potential-scaled molecular dynamics and potential annealing: effective conformational search techniques for biomolecules. *J. Phys. Chem.* **97**, 4416–4420 (1993).

33. 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).

34. 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).

35. Noé, F., Wu, H., Prinz, J. H. & Plattner, N. Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. *J. Chem. Phys.* **139**, 184114 (2013).

36. Husic, B. E. & Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society* **140**, 2386–2396 (2018).

37. Olsson, S. & Noé, F. Mechanistic models of chemical exchange induced relaxation in protein NMR. *J. Am. Chem. Soc.* **139**, 200–210 (2017).

38. Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).

39. Plattner, N. & Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **6**, 7653 (2015).

40. Plattner, N., Doerr, S., De Fabritiis, G. & Noé, F. Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **9**, 1005–1011 (2017).

41. Malmstrom, R. D., Lee, C. T., Van Wart, A. T. & Amaro, R. E. Application of Molecular-Dynamics Based Markov State Models to Functional Proteins. *J. Chem. Theory Comput.* **10**, 2648–2657 (2014).

42. Prinz, J. H., Keller, B. & Noé, F. Probing molecular kinetics with Markov models: Metastable states, transition pathways and spectroscopic observables. *Physical Chemistry Chemical Physics* **13**, 16912–16927 (2011).

43. 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).

44. 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).

45. Thali, M. *et al.* Functional association of cyclophilin A with HIV-1 virions.

Nature **372**, 363–365 (1994).

46.   Franke, E. K., Yuan, H. E. H. & Luban, J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **372**, 359–362 (1994).

47.   BonHomme, M., Wong, S., Carter, C. & Scarlata, S. The pH dependence of HIV-1 capsid assembly and its interaction with cyclophilin A. *Biophys. Chem.* **105**, 67–77 (2003).

48.   Towers, G. J. *et al.* Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. *Nat. Med.* **9**, 1138–1143 (2003).

49.   Chatterji, U. *et al.* The Isomerase Active Site of Cyclophilin A Is Critical for Hepatitis C Virus Replication. *J. Biol. Chem.* **284**, 16998–17005 (2009).

50.   Guichou, J.-F. *et al.* Structure-based design, synthesis, and biological evaluation of novel inhibitors of human cyclophilin A. *J. Med. Chem.* **49**, 900–910 (2006).

51.   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**, 12777 (2016).

52.   Steadman, V. A. *et al.* Discovery of Potent Cyclophilin Inhibitors Based on the Structural Simplification of Sanglifehrin A. *J. Med. Chem.* **60**, 1000–1017 (2017).

53.   Sweeney, Z. K., Fu, J. & Wiedmann, B. From chemical tools to clinical medicines: Nonimmunosuppressive cyclophilin inhibitors derived from the cyclosporin and sanglifehrin scaffolds. *J. Med. Chem.* **57**, 7145–7159 (2014).

54.   De Simone, A. *et al.* A computationally designed binding mode flip leads to a novel class of potent tri-vector cyclophilin inhibitors. *Chem. Sci.* **10**, 542–547 (2019).

55.   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).

56. Tang, H. Cyclophilin Inhibitors as a Novel HCV Therapy. *Viruses* **2**, 1621–1634 (2010).

57. Nigro, P., Pompilio, G. & Capogrossi, M. C. Cyclophilin A: a key player for human disease. *Cell Death Dis.* **4**, e888 (2013).

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

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

60. Otten, R. *et al.* Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* **9**, 1314 (2018).

61. 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).

62. Wapeesittipan, P., Mey, A. S. J. S., Walkinshaw, M. D. & Michel, J. Allosteric effects in cyclophilin mutants may be explained by changes in nano-microsecond time scale motions. *Commun. Chem.* **2**, 41 (2019).

63. Camilloni, C. *et al.* Cyclophilin A catalyzes proline isomerization by an electrostatic handle mechanism. *Proc. Natl. Acad. Sci.* **111**, 10203–10208 (2014).

64. Papaleo, E., Sutto, L., Gervasio, F. L. & Lindorff-Larsen, K. Conformational Changes and Free Energies in a Proline Isomerase. *J. Chem. Theory Comput.* **10**, 4169–4174 (2014).

65. McGowan, L. C. & Hamelberg, D. Conformational Plasticity of an Enzyme during Catalysis: Intricate Coupling between Cyclophilin A Dynamics and Substrate Turnover. *Biophys. J.* **104**, 216–226 (2013).

66.   Agarwal, P. K., Geist, A. & Gorin, A. Protein Dynamics and Enzymatic Catalysis: Investigating the Peptidyl−Prolyl Cis−Trans Isomerization Activity of Cyclophilin A †. *Biochemistry* **43**, 10605–10618 (2004).

67.   Schlegel, J., Armstrong, G. S., Redzic, J. S., Zhang, F. & Eisenmesser, E. Z. Characterizing and controlling the inherent dynamics of cyclophilin-A. *Protein Sci.* **18**, 811–824 (2009).

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

69.   Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).

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

71.   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).

72.   Kontopidis, G., Taylor, P. & Walkinshaw, M. D. Enzymatic and structural characterization of non-peptide ligand–cyclophilin complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 479–485 (2004).

73.   Howard, B. R., Vajdos, F. F., Li, S., Sundquist, W. I. & Hill, C. P. Structural insights into the catalytic mechanism of cyclophilin A. *Nat. Struct. Mol. Biol.* **10**, 475–481 (2003).

74.   Keedy, D. A. *et al.* Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *Elife* **4**, (2015).

75.   Liu, C. *et al.* Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.* **7**, 10714 (2016).

76.   Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: Significantly

improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).

77.   Hansen, D. F., Vallurupalli, P., Lundström, P., Neudecker, P. & Kay, L. E. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? *J. Am. Chem. Soc.* **130**, 2667–2675 (2008).

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

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

80.   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).

81.   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).

82.   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).

83.   Case, D. A. *et al.* Amber 16. *University of California, San Francisco* (2016). doi:10.1002/jcc.23031

84.   Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

85.   McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).

86.   Scherer, M. K. *et al.* PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).

87.   Zweckstetter, M. NMR: Prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **3**, 679–690 (2008).

88.   Bottaro, S., Bengtsen, T. & Lindorff-Larsen, K. Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. *bioRxiv* (2018). doi:10.1101/457952

89.   Baldwin, A. J. An exact solution for R2,eff in CPMG experiments in the case of two site chemical exchange. *J. Magn. Reson.* **244**, 114–124 (2014).

90.   Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

91.   Goddard, Td. & Kneller, D. G. Sparky 3. *Univ. California, San Fr.* **14**, 15 (2004).

92.   Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235–242 (2011).

Graphical TOC