Published in final edited form as:

Nature. 2022 March ; 603(7901): 528–535. doi:10.1038/s41586-022-04468-9.

Structure determination of high-energy states in a dynamic protein ensemble

John B. Stiller[1], **Renee Otten**[1], **Daniel Häussinger**[2], **Pascal S. Rieder**[2], **Douglas L. Theobald**[3], **Dorothee Kern**[1,✉]

[1]Department of Biochemistry and Howard Hughes Medical Institute, Brandeis University, Waltham, MA, USA

[2]Department of Chemistry, University of Basel, Basel, Switzerland

[3]Department of Biochemistry, Brandeis University, Waltham, MA, USA

Abstract

Macromolecular function frequently requires that proteins change conformation into high-energy states[1–4]. However, methods for solving the structures of these functionally essential, lowly populated states are lacking. Here we develop a method for high-resolution structure determination of minorly populated states by coupling NMR spectroscopy-derived pseudocontact shifts[5] (PCSs) with Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion[6] (PCS–CPMG). Our approach additionally defines the corresponding kinetics and thermodynamics of high-energy excursions, thereby characterizing the entire free-energy landscape. Using a large set of simulated data for adenylate kinase (Adk), calmodulin and Src kinase, we find that high-energy PCSs accurately determine high-energy structures (with a root mean squared deviation of less than 3.5 angström). Applying our methodology to Adk during catalysis, we find that the high-energy excursion involves surprisingly small openings of the AMP and ATP lids. This previously unresolved high-energy structure solves a longstanding controversy about conformational interconversions that are

✉**Correspondence and requests for materials** should be addressed to Dorothee Kern. dkern@brandeis.edu.

Author contributions D.K. conceived the project idea. D.K. and J.B.S. developed the research plan and experimental strategy. J.B.S. purified Adk, performed NMR experiments, and analysed results. D.H. and P.S.R. synthesized lanthanide-binding tags and prepared tagged ubiquitin samples. R.O. purified PPD–SBD proteins and prepared lanthanide-bound samples. R.O. performed the NMR experiments for ubiquitin and PPD–SBD trigger factor. R.O., J.B.S., D.K., P.S.R. and D.H. analysed results from the NMR experiments. D.L.T. designed the expectation maximization algorithm. J.B.S. implemented the algorithm into the XPLOR-NIH software and performed calculations on simulated and real data. All authors discussed results leading to overall scientific findings. J.B.S. and D.K. wrote the manuscript. All authors reviewed and edited the manuscript.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04468-9.

Competing interests The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04468-9.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability

Any relevant code is available by request from the corresponding author. A general script for the Expectation Maximization algorithm and test datasets are made available at https://github.com/kernlab-brandeis/PCS-CPMG.

rate-limiting for catalysis. Primed for either substrate binding or product release, the high-energy structure of Adk suggests a two-step mechanism combining conformational selection to this state, followed by an induced-fit step into a fully closed state for catalysis of the phosphoryl-transfer reaction. Unlike other methods for resolving high-energy states, such as cryo-electron microscopy and X-ray crystallography, our solution PCS–CPMG approach excels in cases involving domain rearrangements of smaller systems (less than 60 kDa) and populations as low as 0.5%, and enables the simultaneous determination of protein structure, kinetics and thermodynamics while proteins perform their function.

Modern structural methods describe macromolecules with atomic resolution, but they primarily capture the lowest-energy, most populated structures. However, macromolecules transiently populate high-energy sub-states that remain unseen in these static snapshots. Excursions into these minorly populated states appear pivotal for macromolecular functions such as catalysis, signalling, membrane transport, folding, protein–protein or DNA–RNA interactions and drug binding[1–4,7–12]. The ability to traverse these specific sub-states is a feature of macromolecules that currently cannot be rationally designed. Therefore, the next phase in structural biology involves expanding our arsenal of methodologies to capture high-energy sub-states. Currently, room-temperature X-ray crystallography and ensemble-refinement approaches excel under conditions in which high-energy states are accessible within the crystal lattice, but they lack energetic information such as sub-state occupancies or interconversion rates[13]. Cryo-electron microscopy shows promise for uncovering minor conformations for large systems, provided that the minor state is sufficiently populated to be observable[14].

Nuclear magnetic resonance (NMR) spectroscopy has championed the exploration of high-energy states, particularly with methods detailing the kinetics and thermodynamics of dynamic processes[6], yet characterizing structures of high-energy states using NMR still remains impossible for most systems. Population-weighted average restraints (for example, nuclear Overhauser effect (NOE), residual dipolar coupling (RDC), paramagnetic relaxation enhancement (PRE) and PCS) have shown promise for describing ensembles[15–17]. Nevertheless, deconvoluting high-resolution minor state structures remains a statistical challenge[18]. Alternatively, backbone chemical shifts or RDC restraints for the high-energy state can be directly measured using NMR relaxation dispersion methods[19]. In some cases, chemical shifts have been used to access high-energy structures in small proteins[19] (less than 150 residues), but this method is hampered by the relatively large uncertainty in relating chemical shifts to structure[20]. Furthermore, the relatively minute size of RDCs (±25 Hz) compared with the uncertainty of high-energy state chemical shifts necessitates exquisite sensitivity[19].

Here we develop a method for high-resolution structure determination of high-energy states by coupling PCSs with CPMG relaxation dispersion. We show that the method is highly accurate, applicable for proteins up to 60 kDa, and sensitive to long-range structural changes such as domain reorientation. This method relies on the use of PCSs, which provide powerful, long-distance restraints during structure determination. PCSs are produced through incorporation of an anisotropic paramagnet and are structurally informative,

possessing both an orientation (φ, θ) and radial dependence (r^{-3}), with consequently extensive effects (5–50 Å). Measured spectroscopically as the difference between a diamagnetic and paramagnetic chemical shift, PCSs are extremely large (±1,600 Hz for ^{1}H at 800 MHz). Whereas chemical shifts alone provide tenuous local structural restraints, PCSs deliver otherwise inaccessible information on tertiary structure and interdomain orientation[21]. In addition, coupling PCSs with relaxation dispersion provides large and tuneable dispersion effects[22,23]. Although this concept has been postulated previously[22,24], rotation of the paramagnet on the millisecond timescale has prevented elucidation of high-energy structures.

PCS–CPMG of Adk during catalysis

We set out to develop and benchmark our method by solving the unknown high-energy state of the enzyme adenylate kinase (Adk) during catalysis. Adk is a model system for multi-domain protein conformational changes that are fundamentally linked to its enzymatic cycle[25]. The active site of Adk, once it is formed through the closure of a rigid-body lid, efficiently catalyses the reversible phosphoryl transfer between ATP and AMP into two ADP molecules[25] (phosphoryl-transfer rate constant ($k_{\text{P-transfer}}$) > 5,000 s^{-1}). Disassembly of the active site, which mediates product release through lid opening, represents the rate-limiting step for catalysis ($k_{\text{cat}} \approx k_{\text{open}} \approx 100\ s^{-1}$, where k_{cat} is the catalytic rate constant and k_{open} is the rate constant for lid opening) (Fig. 1a, b). Whereas the majorly populated closed form has been well-studied using X-ray crystallography[26], the transiently populated, high-energy state–in which product is released from the enzyme–has only been detected indirectly[27–29]. Using the natural zinc-binding site of *Geobacillus stearothermophilus* Adk–exchanged with paramagnetic cobalt–we aimed to produce PCS restraints for high-energy structure determination.

The tetra-cysteine binding site, located in the ATP lid, tightly incorporates either diamagnetic zinc (Zn^{2+}) or paramagnetic cobalt (Co^{2+}) metals, and the similar ionic radii of zinc and cobalt result in equivalent structural and enzymatic features[30]. We collected the [^{1}H-^{15}N]-heteronuclear single quantum coherence (HSQC) spectra of Adk under turnover conditions (20 mM Mg^{2+}-ADP) for both the zinc and cobalt forms; PCSs were recognizable from collinear changes in proton and nitrogen chemical shifts (Fig. 1c). After assigning both species (Extended Data Fig. 1), PCSs were quantified for residues upwards of 35 Å from the metal binding site (Fig. 1d), underscoring their long-range effects. Under turnover conditions, these PCSs reflect a population-weighted average of both the majorly populated closed state and the unknown ‘open’ conformation (Fig. 1a). As Adk is predominately sampling the closed conformation under turnover conditions, the PCSs approximately fit to the known closed crystal structure[26] (Q = 11.6%, where Q is the quality factor[21]; Protein Data Bank (PDB) 4QBH; Fig. 1d, e, Extended Data Fig. 2). Acquisition of spectra in the absence of magnesium (20 mMADP) shifts the open–closed exchange into the slow timescale[25] ($k_{\text{open}} \approx 2\ s^{-1}$; Extended Data Fig. 3) and yields PCSs reporting solely on the closed conformation (Q = 7.6%; Extended Data Fig. 2b). Conversely, spectra in the absence of substrates largely report on an open ensemble with substantially weakened PCSs and sizeable PCS-induced line broadening (Extended Data Figs. 2,4).

To determine the structure of the high-energy state, its corresponding PCSs need to be measured. To this end, we performed $^{1}H_{N}$ CPMG relaxation dispersion experiments for both samples under turnover conditions (Fig. 1f, g). ^{1}H relaxation dispersion was chosen, as the large gyromagnetic ratio of ^{1}H results in a 100-fold greater paramagnetic effect than the commonly used ^{15}N experiments[22]. As anticipated for the global nature of the conformational change in Adk, 103 out of the 186 assigned residues in both the diamagnetic and paramagnetic samples exhibited an exchange contribution ($R_{2,ex}$) greater than 3 Hz. Fitting each of the metallic species globally to the Carver–Richards equation for two-state exchange[31] yielded exchange parameters within the range of experimental error (Zn^{2+} (± s.d.): k_{ex} = 1,355 ± 65 s^{-1}; Co^{2+}: k_{ex} = 1,367 ± 71 s^{-1}). Therefore, both datasets were globally fit together to yield a rate constant, k_{ex}, of 1,428 ± 83 s^{-1} and population of the high-energy state (p_B) of 12.6 ± 2.5% (Supplementary Fig. 1). The equivalent exchange processes seen for the diamagnetic and paramagnetic species indicate that any movement of the metal is contingent on lid opening of Adk.

In the diamagnetic case, each dispersion profile reports solely on the chemical shift difference ($\Delta\delta$) between the closed and open conformations ($|\Delta\delta_{dia}|$). In the paramagnetic case ($\Delta\delta_{para}$), dispersion profiles combine the pre-existing $\Delta\delta_{dia}$ with the change in PCS ($\Delta\delta_{PCS}$) between ground and high-energy states ($|\Delta\delta_{para}| = |\Delta\delta_{dia} + \Delta\delta_{PCS}|$). Consequently, many nuclei that were insensitive to exchange with the diamagnetic metal ($|\Delta\delta_{dia}| \approx 0$) now show marked dispersion profiles ($|\Delta\delta_{PCS}| \approx 0$). In a qualitative indication of opening, we observe that residues in the core and AMP lid display large apparent $\Delta\delta_{PCS}$ (Fig. 1f, g), suggestive of ATP lid and AMP lid opening. Notably, the observed $\Delta\delta_{para}$ is less than would be expected from known open (nucleotide-free) and closed (nucleotide-bound) crystal structures (Extended Data Fig. 4). In accordance, the observed $\Delta\delta_{para}$ values are also substantially lower than measured PCS between the nucleotide-saturated and apo enzyme forms (Extended Data Fig. 4), signifying that the high-energy state must be more closed than previously expected[29,32].

Expectation-maximization method

Ideally, the high-energy structure could be determined by extracting the PCS of each residue in the minor state (PCS_{minor}) from the paired diamagnetic and paramagnetic CPMG experiments; the PCSs would then be used as restraints for structure determination. Relaxation dispersion experiments provide only the magnitude of the $|\Delta\delta|$, with two possible solutions for both the diamagnetic and paramagnetic case. Consequently, four degenerate PCS_{minor} solutions are possible for each residue (Fig. 2a), giving a total of 4^N sets of PCS_{minor} restraints for N residues.

Direct acquisition of the $|\Delta\delta|$ sign can be achieved by comparison of HSQC and HMQC experiments[33]; however, owing to the small difference in the cross-peak positions in these spectra, this route is experimentally impossible for many residues. Alternatively, sign determination can be considered a problem of ambiguity, with the degeneracy in possible PCS_{minor} values being analogous to the challenge faced with ambiguous NOEs. There are four possibilities for each residue, yet only the correct PCS_{minor} set should be self-consistent with the true structure. Therefore, each residue's PCS_{minor} possibilities

can be considered an ambiguous restraint with four weighted spin–metal descriptions. Assuming the PCS data have Gaussian noise, a likelihood function for PCS data can be constructed and the probability of each PCS possibility can be determined via conditional expectation maximization. The probability of each PCS_{minor} is used as a weight during cooling in a simulated annealing protocol. After iteratively updating the probabilities, the restraint may become unambiguous, as the probability of one of the four PCS_{minor} values approaches 100%. Considering the PCS_{minor} ambiguity as a standard classification problem, we therefore developed a stochastic expectation-maximization method to iteratively update likelihood-based restraints during simulated annealing in XPLOR-NIH[34] (Fig. 2b).

Benchmarking using known Adk structures

To benchmark our method, we computationally generated 12 minor states from various Adk crystal structures. Ranging from closed to open, these structures roughly span the possible conformational space sampled by Adk. For each structure, degenerate PCS_{minor} were produced with uncertainties similar to those in our experimental CPMG data. As PCS are long-range restraints, a rigid-body refinement approach was employed. The rigid-body assumption is supported by the fact that individual domains are highly superimposable between all Adk crystal structures despite quaternary conformational changes.

For each simulated minor state, a total of 400 structures were solved over four rounds of structure determination with the XPLOR-NIH software. The first round of structure determination was started from a closed crystal structure[26] (PDB 4QBH). Inferred structures with the highest likelihood were passed to subsequent iterations (Fig. 2b). In all cases, ambiguities in PCS restraints were resolved with the correct PCS found for 90 ± 8% of residues (Fig. 2c, d). The few incorrect PCS identifications were not randomly distributed, but instead clustered at the edges of the rigid bodies where local structural differences most commonly occur (Extended Data Fig. 5). The resultant structures converged excellently to the target structures (Fig. 2e–g). To verify that the method is independent of the initial structure, alternative starting structures were tested and gave similar results upon convergence (Extended Data Fig. 6). Finally, a strong correlation between the likelihood and the root mean square deviation (r.m.s.d.) was found, indicating that the observable likelihood value is a suitable estimate of the relative accuracy of the inferred structures (Extended Data Fig. 7).

Structure of the high-energy state of Adk

Because the simulated cases demonstrated that ambiguous PCSs could accurately solve the minor state of Adk, we applied our methodology to the experimentally determined $^{1}H_{N}$ CPMG data obtained under turnover conditions (Fig. 1f), yielding 93 residues for the refinement procedure of the high-energy structure. Analogous to the simulated cases, a single unambiguous PCS identification was found for each residue after structure determination, with excellent agreement between calculated and observed PCSs (Q = 3.7%; Fig. 3a). In agreement with the $\Delta\delta_{para}$, the PCS differences between the major and minor state are relatively small (Fig. 3b), with the largest changes in the AMP lid.

This high-energy state of Adk is a partially open state with notable changes in both the ATP and AMP lids (Fig. 3c, Supplementary Table 1). Compared with the closed state, the AMP lid opens by about 15°, corresponding to approximately 50% opening (Fig. 3d). Within the ATP lid, a subtle motion of approximately 1.8 Å was found for the Co^{2+} ion, resulting in partial separation of the ATP lid from the bound nucleotides (Fig. 3d). Overall, the structure remains in a partially closed arrangement, with more similarity to the closed state than the open state (r.m.s.d.$_{closed}$ = 2.67 Å versus r.m.s.d.$_{open}$ = 7.03 Å). As the experiments were performed under saturating concentrations of nucleotides ([ADP] = 20 mM, Michaelis constant for ADP($K_{M,ADP}$) ≈ 50μM), the high-energy state must be occupied with substrate or product. Viewing the surface representation of the high-energy-state structure clearly displays how the enzyme is primed for product release (Fig. 3e). Compared with the closed state, in which the nucleotides are buried, the partially opened lid reveals a tunnel for product escape, especially for the AMP lid-bound nucleotide.

The solved high-energy state is a key component in the catalytic cycle that includes conformational-selection and induced-fit steps for substrate binding and product release (Fig. 3f). Beginning with the apo ensemble, partial closing produces a functional substrate-binding pocket. Nucleotide association with this partially closed state (conformational selection) precedes full closure after binding (induced fit). Once fully closed, reversible phosphoryl transfer commences until the overall rate-limiting lid opening occurs stochastically (k_{open} = 180 ± 36 s^{-1}), leading to fast substrate or product release from the newly structurally characterized high-energy state (populated at 12.6 ± 2.5%) and the start of a new catalytic cycle.

PCS–CPMG for large structural changes

The power of PCS restraints lies in their potential to provide long-distance structural information for multi-domain proteins. As a proof of principle for the general use of our approach, we simulated datasets for the large conformational changes observed in calmodulin and Srckinase[35,36] (r.m.s.d. > 30 Å; Fig. 4). Although many proteins lack paramagnetic metal-binding sites, lanthanides can be introduced as replacement for Ca^{2+} (such as in calmodulin[21]), or more generally via a lanthanide-binding tag[37] (simulated here for Src Kinase). The lanthanide metal series exhibit a range of paramagnetic effects and are interchangeable owing to their similar ionic radii. Using multiple paramagnets offers three distinct advantages. First, several PCSs describing the same residue breaks the degeneracy that originates from a single tensor's iso-surface[5]. Second, shared diamagnetic exchange contribution between the two and greater paramagnetic dispersion terms ($|\Delta\delta_{para}| = |\Delta\delta_{dia} + \Delta\delta_{PCS}|$) limit the possible PCS choices (Extended Data Fig. 8). Third, PCSs can be tuned over a wide range due to a wide distribution of lanthanide paramagnetic tensors.

For calmodulin, previous experiments have revealed a large conformational change between an extended and a compact conformation in the absence of ligands, with the compact conformation proposed to be the key conformation for peptide binding[21,38]. Here, the third Ca^{2+} site out of the four binding sites was simulated with either Ho^{3+} or Yb^{3+}, which can be achieved experimentally by mutation of that Ca^{2+} site[21]. PCSs were generated for both states and relaxation dispersion profiles were simulated, assuming that the protein predominately

sampled the extended conformation (major population, p_A = 85%). Using the resultant ambiguous PCS_{minor}, we applied our methodology to infer the transiently accessed compact state. Simulated annealing procedures starting from the extended conformation initially led to incorrect structures and PCS identifications because of an incomplete conformational search, and the incorrectly inferred structure represented a local energy minimum. To overcome this sampling problem, we generated 10,000 structures using the rapidly-exploring random tree (RRT) algorithm from the Integrated Modeling Platform (IMP) software[39]. The likelihood was maximized for each of the 10,000 structures (Fig. 4a). Structures with the highest likelihood, which notably also approach the compact state, were passed to expectation-maximization simulated annealing. After several rounds of refinement, the inferred structure matched the expected compact state extremely well (r.m.s.d. = 1.9 ± 0.1 Å; Fig. 4b), with largely correct PCS identifications (84%; Extended Data Fig. 9).

To test the performance of our methodology with a larger protein and greater domain reorientation, we next tested the 57-kDa Src kinase (Fig. 4c, d). The goal was to further demonstrate the general applicability of PCS–CPMG for non-metal-binding proteins by incorporating the advancement of lanthanide-binding tags. The impressive power of lanthanide-binding tags for ground-state structure determination has been demonstrated previously[37]. Here, we modelled the lanthanide-binding tag on residue 309 of Src (Fig. 4d) and generated PCSs with both Tb^{3+} and Tm^{3+}. The closed conformation was considered the major state, whereas the open conformation was considered the high-energy state. A total of 50,000 structures were generated for Src kinase using the RRT algorithm. Structures approaching the target structure showed the highest likelihood and could be used for additional simulated annealing refinement. Within three rounds of refinement, the inferred structure and the target structure appeared nearly identical (r.m.s.d. = 3.3 ± 0.1 Å; Fig. 4d), with largely accurate PCS probabilities (85%; Extended Data Fig. 9).

PCS–CPMG using a lanthanide-binding tag

As a final proof of concept for our PCS–CPMG method, we tested our approach experimentally with two non-metalloproteins–ubiquitin and the chaperone trigger factor–via the introduction of lanthanide-binding tags. In previous studies where PCS–CPMG was attempted, motion of the paramagnetic metal on the microsecond timescale obscured the detection of protein dynamics[22,24]. We reasoned that the recent improvements in lanthanide-binding tags may have yielded an ideal tag to obtain PCS–CPMG for non-metalloproteins. Ubiquitin was chosen to test PCS–CPMG because it is a classic system that has been studied extensively using NMR spectroscopy, revealing a lack of millisecond motion.

We covalently attached a DOTA-M7PyThiazole lanthanide-binding tag[37] coordinated with either diamagnetic Lu^{3+} or paramagnetic Tm^{3+} to single-cysteine mutants of ubiquitin, K6C and S20C (Fig. 4e, Extended Data Fig. 10a, b). In both mutants, ^1H_N CPMG of the diamagnetic samples yielded flat profiles for nearly all residues of ubiquitin, confirming the lack of millisecond dynamics (Supplementary Figs. 2, 3). By contrast, we observed substantial dispersions in the paramagnetic ^1H_N CPMG profiles for both mutants, albeit with different magnitudes (Fig. 4f). The S20C protein, which is tagged on a loop, displayed large dispersions ranging from 10–20 Hz (Supplementary Fig. 2). Ubiquitin-K6C, which is tagged

on a β-sheet, showed smaller profiles around 1–8 Hz (Supplementary Fig. 3). These results indicate that lanthanide tags should be attached to secondary structure elements rather than loops, as they will report on the local loop movement at the tag coordination site.

To find an optimal lanthanide-binding tag for PCS–CPMG, we then tested the K6C ubiquitin mutant with a different lanthanide-binding tag, DOTA-M8-(4*R*4*S*)-SSPy[40]. Performing $^{1}H_{N}$ CPMG of the paramagnetic bound sample indeed yielded dispersions that were either flat or extremely small (approximately 1 Hz; Fig. 4f), revealing that PCS–CPMG is achievable for non-metalloproteins with this lanthanide tag (Supplementary Fig. 4).

To demonstrate the general applicability of the method, we acquired PCS-enhanced $^{1}H_{N}$ CPMG for the chaperone trigger factor. The full-length trigger factor possesses three domains: a ribosome-binding domain (RBD), a peptidyl-prolyl isomerase domain (PPD), and a substrate-binding domain (SBD). The ribosome-binding domain is bound to the ribosome in vivo but dimerizes in its absence. We thus chose to study the two-domain construct PPD–SBD, which encompasses the functional unit working together to chaperone unfolded proteins to the folded state[41]. $^{1}H_{N}$ CPMG spectra of PPD–SBD showed dispersions at the hinges and loops, indicating domain–domain motion (Supplementary Fig. 5). Guided by our ubiquitin data, we produced a single-cysteine mutant (V270C) and tagged it with the DOTA-M8-(4*R*4*S*)-SSPy bound with Tm^{3+}, which produced PCSs above 8 ppm (Fig. 4g, Extended Data Fig. 10c–e). The tagging location was chosen in an α-helical bundle in the SBD, intended to restrict undesired tag motion. Notably, the Tm^{3+} $^{1}H_{N}$ CPMG dispersions in the SBD were almost unchanged to the dispersion profiles for Lu^{3+}, untagged V270C or wild-type protein, indicating little-to-no tag motion (Fig. 4h). By contrast, the paramagnetic dispersions in PPD were enhanced, indicating a conformational change between PPD and SBD in the high-energy state (Fig. 4h, Supplementary Fig. 5). Unfortunately, the timescale of the motion was in the fast time regime ($k_{ex} \approx 3{,}000\ s^{-1}$), which prevented decoupling of the minor-state population chemical shift and therefore determination of the minor-state pseudocontact shifts.

Discussion

Paramagnetic NMR is at the centre of our approach towards a comprehensive description of protein free-energy landscapes that includes accurate determination of populations of substates, their interconversion rates and, ultimately, their atomic structures. PCSs magnify CPMG relaxation dispersion profiles in a tunable manner through the use of different paramagnetic metals. Standard CPMG relaxation dispersion for interconversion between folded substates often suffers from small chemical shift differences, resulting in minute dispersion amplitudes. Increasing the size of these dispersion profiles with paramagnets both improves detection of the exchange process and has the potential to overcome the shortcoming that populations and chemical shifts cannot be separated in the fast exchange limit ($k_{ex} \gg \Delta\delta$). Increasing $\Delta\delta$ can shift the process into the intermediate time regime ($k_{ex} \approx \Delta\delta_{Para}$), thereby enabling accurate measurement of both populations and chemical shifts. We note that whereas NMR dynamics coupled with structure determination has the reputation of being complicated and inaccessible, our method is straightforward, can perform automated classification of PCSs and is applicable to many systems for solving

their high-energy structures, including non-metalloproteins, as we have shown with ubiquitin and the chaperone trigger factor. These structures of high-energy states can have crucial roles in drug discovery, and this method provides in addition populations and kinetics that are often altered in loss-of-function or gain-of-function disease mutations.

Applying our methodology to Adk during catalysis, we have resolved a mechanistically important controversy: the role of protein dynamics in overall catalytic power and the mechanism of substrate binding and product release. We reveal an unanticipated high-energy structure that is partially open. Solely on the basis of the chemical shifts of the high-energy state, we had previously hypothesized that this minor state involved a more complete opening of both lids[25]. Chemical shifts report on the dissociation of the lids, but not the extent of the domain movement. Our PCS method overcomes this limitation imposed by the local nature of chemical shifts. Furthermore, single-molecule fluorescence resonance energy transfer (FRET) experiments had suggested that the ATP lid undergoes a large conformational change from the ground state to this high-energy state[27,32]. However, as the paramagnetic metal resides in the ATP lid, our data unequivocally reveal that the ATP lid moves only slightly (1.8 Å) between the ground and high-energy state.

Of note, our high-energy structure explains several previously puzzling experimental and computational findings. First, engineered disulfide linkages between the AMP and ATP lids, which would increase the concentration of our partially open state, tightens the affinity for substrates[42]. Second, meta-dynamics simulations of the free energy landscape of Adk when bound with ligands found that opening proceeds through an AMP lid open state[43]. Finally, a recent study detailing the transition state of the conformational change of Adk revealed that the AMP lid movement is the critical energy barrier for the rate-limiting conformational change[44]. Notably, the transition-state ensemble coincides halfway between the fully closed ground state and the high-energy state described here[44]. Our results for Adk show that PCS–CPMG provides the necessary resolution to describe functional transitions within the folded ensemble, including transient states sampled during enzyme catalysis.

While paramagnetic NMR was once limited to metalloproteins, rigid metal-binding tags present the opportunity to generate PCSs in any protein system[45,46]. For conformational changes on the microsecond-to-millisecond timescale, we envision the PCS–CPMG approached outlined here to be an ideal method for high-energy structure determination. We used amide $^{1}H_{N}$ CPMG to probe protein movements, which is applicable for systems smaller than 60 kDa. For larger systems, methyl-based CPMG approaches provide excellent resolution well above 100 kDa. Outside of the millisecond regime, paramagnetic NMR also provides rich parallel structural information for both faster[21] and slower processes[47], with PREs as a complementary approach to obtain structural information on transient states on very fast timescales[17]. Notably, combining PCS restraints with chemical exchange saturation transfer (CEST) experiments could offer an analogous method for high-energy structure determination on the slow timescale in the future. Owing to the long-range nature of PCSs, minor-state structure determination using PCS–CPMG or PCS–CEST should excel in cases involving domain or secondary structure bundle reorientation, which is ubiquitous along the functional landscape of proteins[48]. Combining PCS restraints and diamagnetic chemical

shifts of the high-energy states together with available computational approaches[49] may yield accurate information for any type of conformational change.

Methods

Protein expression and purification

A G-block for *G. stearothermophilus* Adk was synthesized by Integrated DNA Technologies and cloned into the plate-11 vector with a C-terminal TEV cleavage site and His_6-tag. The vector was transformed into BL21(DE3) from New England Biolabs. To prepare protonated samples for NMR peak assignment, cells were grown in ^{15}N NH_4Cl, uniformly labelled ^{13}C glucose, metal-free M9 media until induction ($OD_{600} \approx 1.2$) where $Zn^{2+}Cl_2$ or $Co^{2+}Cl_2$ was added to reach a concentration of 20 μM. Additionally, a final concentration of 1 mM isopropyl β-D-1-thiogalactopyranoside was added and cells were grown for 4 h at 37 °C. For perdeuterated samples necessary for ^1H_N CPMG experiments, cells were grown as described[51]. To prepare pure Zn^{2+} or Co^{2+} samples, no metals were added until induction. Expression was induced with IPTG for 6 h. Protein was purified as described previously[52].

$^{15}N/^2H$ samples of ubiquitin S20C and K6C were gifts from the Grzesiek laboratory and the S. Hiller laboratory, respectively. The lanthanide-binding tags DOTA-M7PyThiazole[37] and DOTA-M8-(4*R*4*S*)-SSPy[40,53] were synthesized according to the described protocols. To attach the lanthanide-binding tags to ubiquitin, a solution of mutant ubiquitin uniformly $^{15}N,^2H$-labelled (300 μM, 0.5 ml) was dialysed in phosphate buffer (10 mM, pH 7.0) with TCEP (2 mM, pH 7.0) overnight. Buffer exchange (phosphate 10 mM, pH 7.0, 0.1 mM TCEP) was performed via ultracentrifugation (Amicon-Ultra 4, 3 kDa cut-off, Merck Millipore). For the thio-ether tag (M7-DOTA-PyThiazole), a solution of the tag in acetonitrile (15 mM) was added in threefold excess to the protein and allowed to conjugate overnight in a shaker. In the case of the disulfide tag (4*R4S*-Ln-M8-DOTA-SPy), a further buffer exchange to non-reducing buffer (phosphate 10 mM, pH 7.0) was achieved via a PD-10 desalting column (GE Healthcare) immediately before the conjugation reaction (2.5-fold excess, 3 h). All conjugation reactions were then quenched by a final buffer exchange to phosphate (10 mM, pH 7.0) using again an Amicon-Ultra 4 and concentration to the desired NMR sample concentration. All operations were carried out at room temperature.

The plasmid of the trigger factor construct without the ribosome-binding domain (PPD–SBD) was a gift from S. Hiller. The single-Cys mutation V270C (numbering based on full-length protein) was generated using the QuickChange Lightning site-directed mutagenesis kit (Agilent Technologies) using the forward (5′-gctccatgtttttacggcattcagcgcgcagaccttctacgg-3′) and reverse (5′-ccgtagaaggtctgcgcgctgaatgccgtaaaaacatggagc-3′) primers obtained from Genewiz. The protein was essentially expressed using the growth media (except for the addition of MEM) and outlined protocol[51]. Protein expression was induced overnight with 0.4 mM IPTG at 25 °C.

The wild-type PPD–SBD and V270C mutant were purified as described[41], but without the unfolding–refolding step, and 2 mM TCEP was used in all purification buffers for the V270C mutant. The tagging reaction with the DOTA-M8-4*R*4*S* tag was performed similarly

as described above for ubiquitin: first a buffer-exchange step into 25 mM potassium phosphate, pH 6.5, 100 mM KCl, 0.5 mM EDTA, 100 μM TCEP was performed using a 10 kDa concentrator and the remaining TCEP was removed using Zeba desalting columns (Thermo Scientific) immediately prior to the conjugation reaction. A three-fold excess of DOTA-M8-(4*R*4*S*)-SSPy-Tm or 1.6-fold excess of DOTA-M8-(4*R*4*S*)-SSPy-Lu tags was used in the conjugation mixture, respectively, and the reaction was allowed to proceed overnight at 4°C. The samples were buffer exchanged using a 10 kDa concentrator into 25 mM potassium phosphate, pH 6.5, 100 mM KCl, 0.5 mM EDTA and concentrated to about 500 μM.

NMR spectroscopy

Adk samples were prepared at 2 mM enzyme in 50 mM MOPS, 50 mM NaCl, 2 mM TCEP and 0.02% sodium azide at pH 7.0. Samples tested under turnover conditions included 20 mM $MgCl_2$ and 20 mM ADP. All experiments were acquired on an Agilent DD2 600 MHz four-channel spectrometer equipped with a triple-resonance cryogenically cooled probe head. For assignment, a combination of HNCO, HNCACB and CBCA(CO)NH experiments were performed on ^{13}C, ^{15}N Adk (Co^{2+} or Zn^{2+}) at 40 °C. Triple-resonance experiments yielded backbone assignment of 92% for Zn^{2+} and 83% for Co^{2+} states.

Ubiquitin samples contained between 400–800 μM protein in 10 mM phosphate, pH 7.0, 10% D_2O. All PPD–SBD samples contained 500 μM protein in 25 mM potassium phosphate, pH 6.5, 100 mM KCl, 0.5 mM EDTA, and 10% D_2O (the untagged V270C sample additionally contained 2 mM TCEP). Assignments for wild-type PPD–SBD and ubiquitin were taken from the BioMagResBank[54] entries 27239 and 15410, respectively; the single-Cys mutation in PPD–SBD showed only minor chemical shift perturbations compared to wild-type (Extended Data Fig. 10). Assignments for the protein samples with the paramagnetic lanthanide-binding tags were obtained iteratively using a combination of a 3D ^{15}N-edited NOESY spectrum, [^1H,^{15}N]-HSQC spectra recorded at different temperature between 20–30 degrees, and initial ‘guesses’ for isolated cross peaks in the untagged sample. The tensor was calculated from the PCSs using these initial assignments and used to ‘calculate’ the peak positions in the paramagnetic sample using NUMBAT[50] or Paramagpy[55]. For ubiquitin, we obtained complete assignments for the paramagnetic sample, except for severely exchange broadened residues (approximately 10–15 for each sample) near the metal position.

For PCS acquisition, a modified version of the ^1H_N CPMG relaxation dispersion experiment[56] was performed at 25 °C. The pulse program included a modified P-element–which reduced antiphase signal refocused over the CPMG period–designed by L. E. Kay. Relaxation dispersion experiments were acquired using a 24 ms constant-time relaxation period with 16 ν_{CPMG} frequencies ranging from 83 to 1,500 Hz for Adk and ubiquitin; for PPD–SBD a 12 ms constant-time relaxation period was used with 14 ν_{CPMG} values, where ν_{CPMG} is the frequency of the refocusing pulse train. Data were processed with NMRPipe/NMRDraw software[57] and analysed with CCPNMR[58] or POKY[59]. Relaxation dispersion intensities were determined using the PINT peak fitting software[60] and analysed using the general Carver–Richards equation for two-site exchange with an in-house script

built with LmFit[31,61]. Zn^{2+} and Co^{2+} datasets were globally fit first individually and then simultaneously. Uncertainties on individual data points were calculated from the standard pooled deviation in duplicate data points ($n = 3$). A minimum of 2% error was applied to all residues.

Simulated annealing protocol

Having developed a method for resolving PCS ambiguity (see Supplementary Methods for the description of our stochastic likelihood expectation-maximization method for ambiguous PCS data), we set up rigid-body simulated annealing procedures in the XPLOR-NIH structure determination software[34]. The expectation-maximization algorithm has the advantage of being easily incorporated into the simulated-annealing procedure of the XPLOR-NIH software package and can rapidly calculate the probability of each PCSs choice at each step of the simulated-annealing process. Each PCS_{minor} is treated as a separate restraint and weighted as follows:

$$w_{ij} = \frac{1}{q_{ij}}$$

where q_{ij} is the probability of each PCS_{minor}. The weights are then used to calibrate the restraint energy function in XPLOR-NIH refinement programs as follows, where $PCS_{i,Calc}$ is the calculated PCS for each residue:

$$E = (\mathrm{PCS}_{i,\,\mathrm{Calc}} - \mathrm{PCS}_{ij}) \times \frac{s}{w_{ij}^2}$$

At the onset of the simulated annealing protocol, each q_{ij} is initialized to 0.25. The parameter s is an exponential scaling energy factor that increased from 0.04 at the beginning of the annealing protocol to a final value of 1.0, 5.0, 7.0 and 7.0 for first, second, third and fourth simulated annealing iterations, respectively. Rigid-body refinement was performed with an annealing protocol from an initial temperature of 2,000 decreasing to a final temperature of 20 °C over 1,770 cooling steps.

At each cooling step, the q_{ij} probability of each PCS was updated iteratively until the likelihood converged. The energy functions were reweighted and the simulated annealing protocol continued. As PCSs provide only long-distance information, rigid-body refinement included two stages. For Adk calculations, the first 1,720 cooling steps were performed (2,000 to 25 °C) with four bodies: a core domain (amino acids 1–28, 62–113 and 165–217), an AMP lid (32–55), a hinge (116–125) and an ATP lid (128–157). All residues between the four sections were given complete freedom. Molecular dynamics were performed in torsion space. Second, 50 cooling steps were performed (25 to 20 °C) with Adk defined by secondary structure bundles: (1) a reduced core domain (amino acids 1–28, 81–113 and 181–217), AMP lid 1 (amino acids 32–48), AMP lid 2 (amino acids 44–55), AMP lid 3 (amino acids 62–79), ATP lid (amino acids 128–157), hinge 1 (amino acids 116–125) and hinge 2 (amino acids 165–178). The bundling was defined by known crystal structures of Adk.

For determination of either the simulated or real minor structures, a total of 3–4 rounds of refinement was performed and between 100–250 structures were solved in each pass. In the first pass, the closed crystal structure[26] (PDB 4QBH) was used as a starting point for simulated annealing. At the end of this pass, the total likelihood of the structure was calculated. In subsequent refinement procedures, the highest likelihood structures, and associated probabilities, were passed, with the chance of each structure X_i being chosen defined by:

$$P(X_i) = \frac{\mathscr{L}_i}{\sum_i \mathscr{L}_i}$$

To determine structural uncertainty in the high-energy state of Adk, a leave-one-out cross-validation analysis was performed on the final refinement procedure. Simply, the last refinement step was repeated N times, where N is the number of residues in the dataset (N = 93). On each iteration, one residue was removed from the dataset and 100 structures were calculated. The structural parameters (for example, AMP lid angle, etc.) were calculated for the top 10 highest likelihood structures. The uncertainty in high-energy state structure was then determined by calculating the standard deviation of each structural parameter over all N repeats (Fig 3d).

Benchmarking PCS–CPMG with computer-generated data

To theoretically benchmark the capacity of PCS–CPMG to solve high-energy structures, we generated datasets for 12 deposited Adk structures ranging from open to closed (PDB_{chain}, where 'PDB' is the accession code and 'chain' is the molecule used for simulations from the deposited structure: $2EU8_B$ (ref. [62]), $2AKY_A$ (ref. [63]), $1ZIP_A$ (ref. [64]), $2BBW_B$, $2AK3_A$ (ref. [65]), $2BBW_A$, $2AR7_B$, $1DVR_A$ (ref. [66]), $2RH5_A$ (ref. [67]), $4AKE_B$ (ref. [68]) and $4AKE_A$ (ref. [68])). For simulated structures of Adk treated here, homology models of the known crystal structure were produced with the SWISS-MODEL software[69] to incorporate the *G. stearothermophilus* Adk sequence. PCSs for the target structure were produced by modelling a Co^{2+} ion into the ATP lid, and then, generating a magnetic susceptibility tensor with X_{ax} and X_{rh} equal to 16.8 and 0.65 × 10^{-32} m^3, respectively. The Euler angles for the susceptibility tensor were chosen randomly. Protons were added with the 'reduce' software[70] and PCS was calculated for the $^1H_N{}^5$. Similarly, PCSs were generated for the closed state using the tensor calculated from PCS collected in the 20 mM ADP condition. From there, $|\Delta\delta_{Para}|$ defined as:

$$|\Delta\delta_{Para}| = |(PCS_{Target} - PCS_{Closed}) \pm \Delta\delta_{Dia}|$$

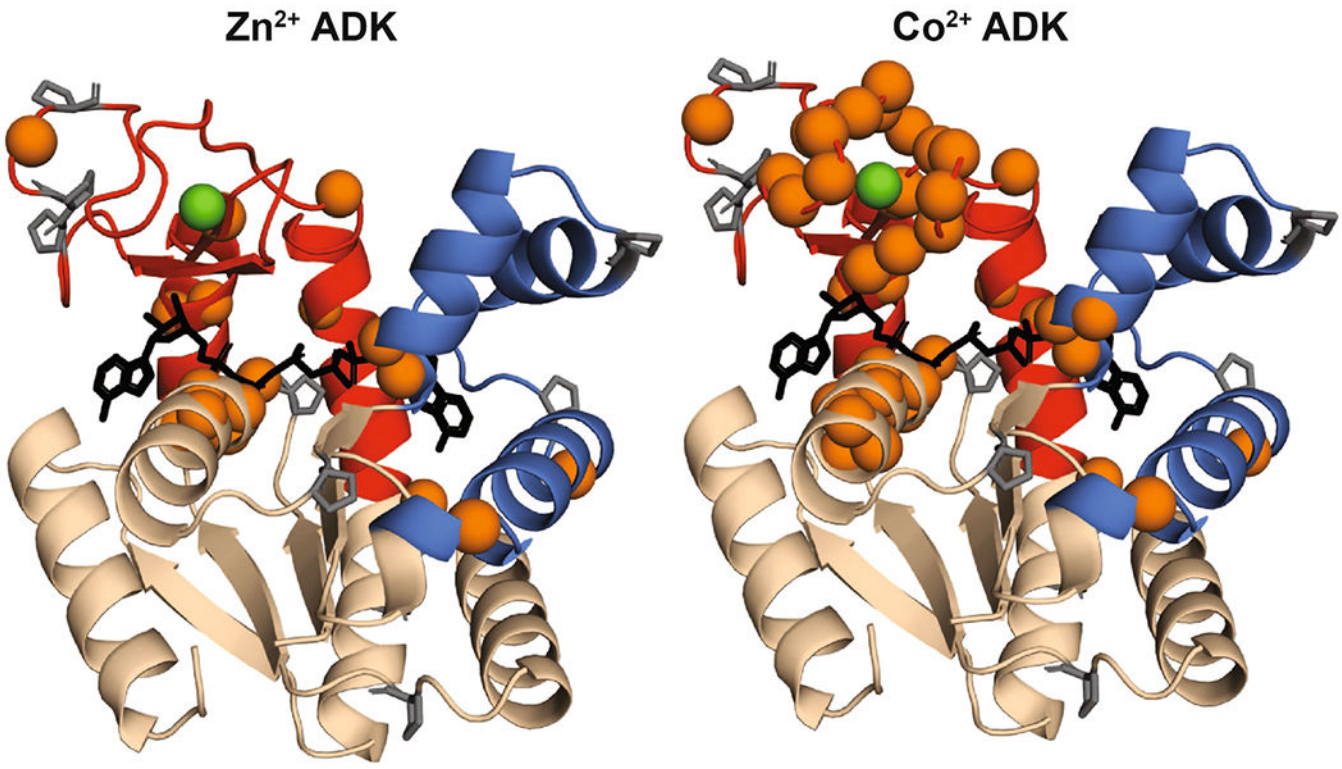
where $\Delta\delta_{Dia}$ came from the diamagnetic relaxation dispersion experiments and the sign was chosen randomly for each residue. A 2% relative error was added to each residue's $\Delta\delta_{Para}$.

For Src kinase and calmodulin, an analogous routine was applied. For calmodulin, the extended conformation was considered the major state (PDB 1CLL[71]) and the compact structure was considered the minor state (PDB 1PRW[35]). Similarly, the closed conformation of Src was considered the major state (PDB 2SRC[72]) and the open structure was considered

the minor state (PDB 1Y57[36]). Paramagnetic metals were added to each structure and then magnetic susceptibility tensors were calculated. For Src, the metals were placed 9 Å from residue 309, mimicking a metal-binding tag. For calmodulin, the metals were placed in the third Ca^{2+} binding site, between residues 130–137. Axial and rhombic components of the susceptibility tensor were chosen based on those seen for lanthanides incorporated into calbindin[73]. Euler angles were chosen randomly. For the minor state, up to a 5° random Euler angle change was randomly added between the major and minor state. The PCSs of both states were used to calculate a |ΔPCS|. Diamagnetic chemical shift differences were sampled from an exponential distribution with a mean of 0.05 ppm, which matches what is observed for Adk. For both systems, CPMG data was generated assuming a major state population of 85% and an exchange rate of 1,500 s^{-1}. Ambiguous PCSs were then calculated as described in the Supplementary Methods.

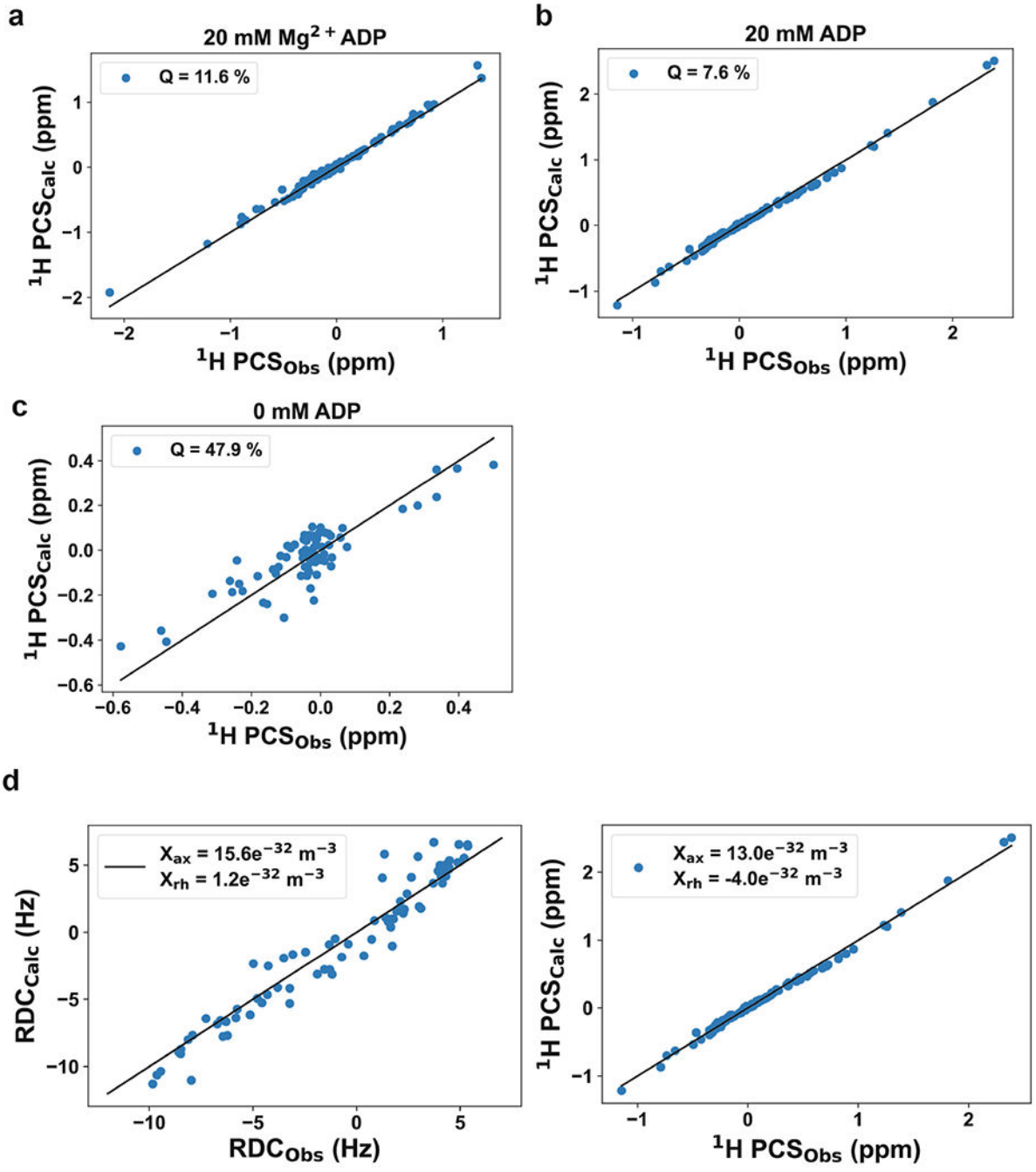
To sample the conformational space of Src and calmodulin, the RRT algorithm from the IMP software was used[39]. For Src, a linker between the SH3 binding site and the catalytic domain was used (residues 254–260) and 50,000 structures were generated. For calmodulin, two points of flexibility were added (residues 64–65 and residues 77–81) and 10,000 structures were generated. The likelihood of each structure was maximized using the expectation-maximization algorithm. Following, three rounds of simulated annealing in XPLOR-NIH were performed. In the first iteration, started structures were chosen from the IMP dataset based upon their likelihood as described above.

Extended Data

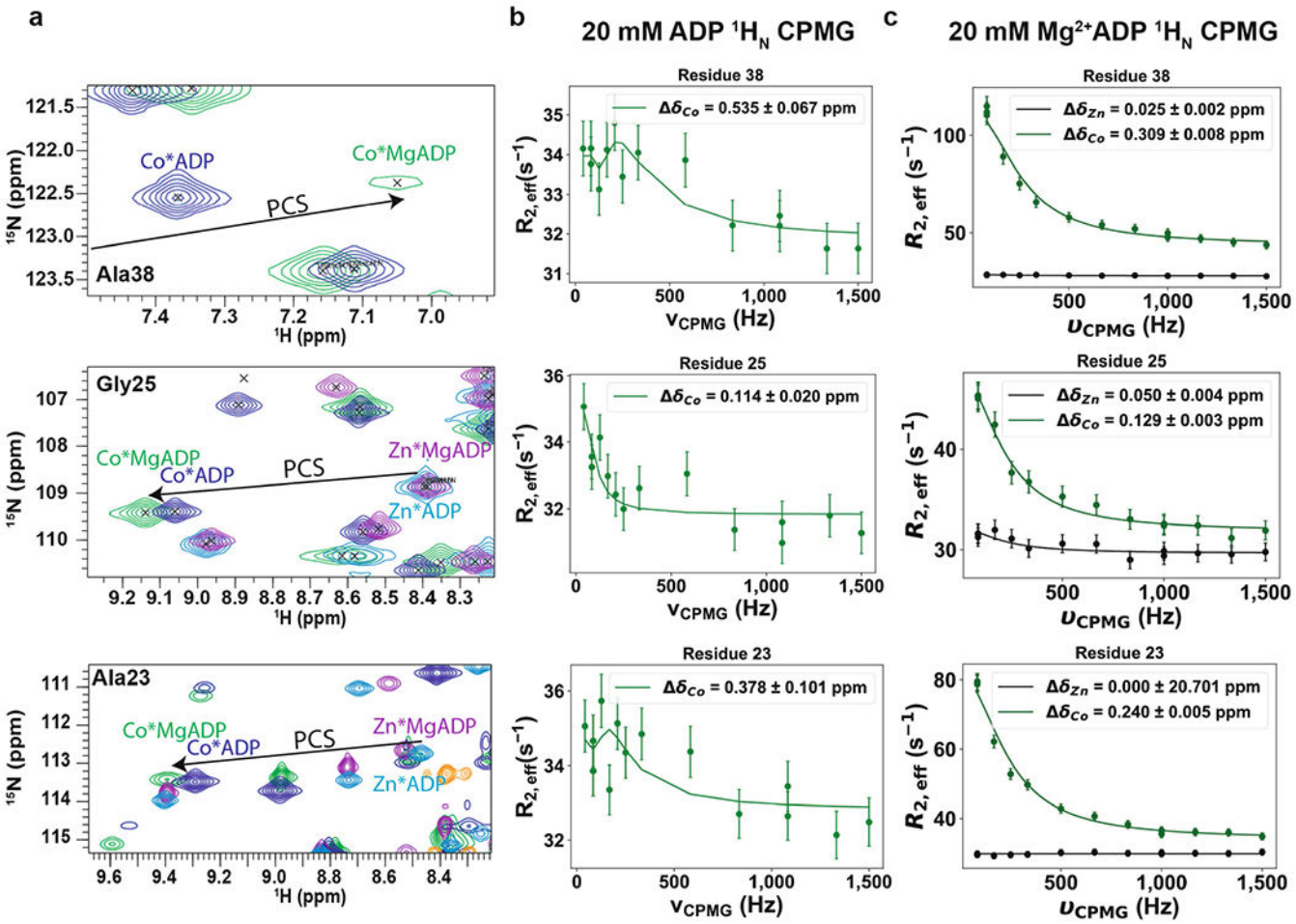


Extended Data Fig. 1 |. Diamagnetic and paramagnetic samples were assigned by triple resonance experiments.

Backbone assignment of Zn^{2+} (left) and Co^{2+} (right) Adk proteins under saturating nucleotide conditions plotted onto Ap5A bound crystal structure, PDB 4QBH[26]. Orange spheres depict residues where no assignment was determined. Prolines are shown as gray sticks. All other amides are assigned. For the cobalt species, many residues surrounding the metal binding site are lost due to either Curie relaxation or exchange on the intermediate timescale.

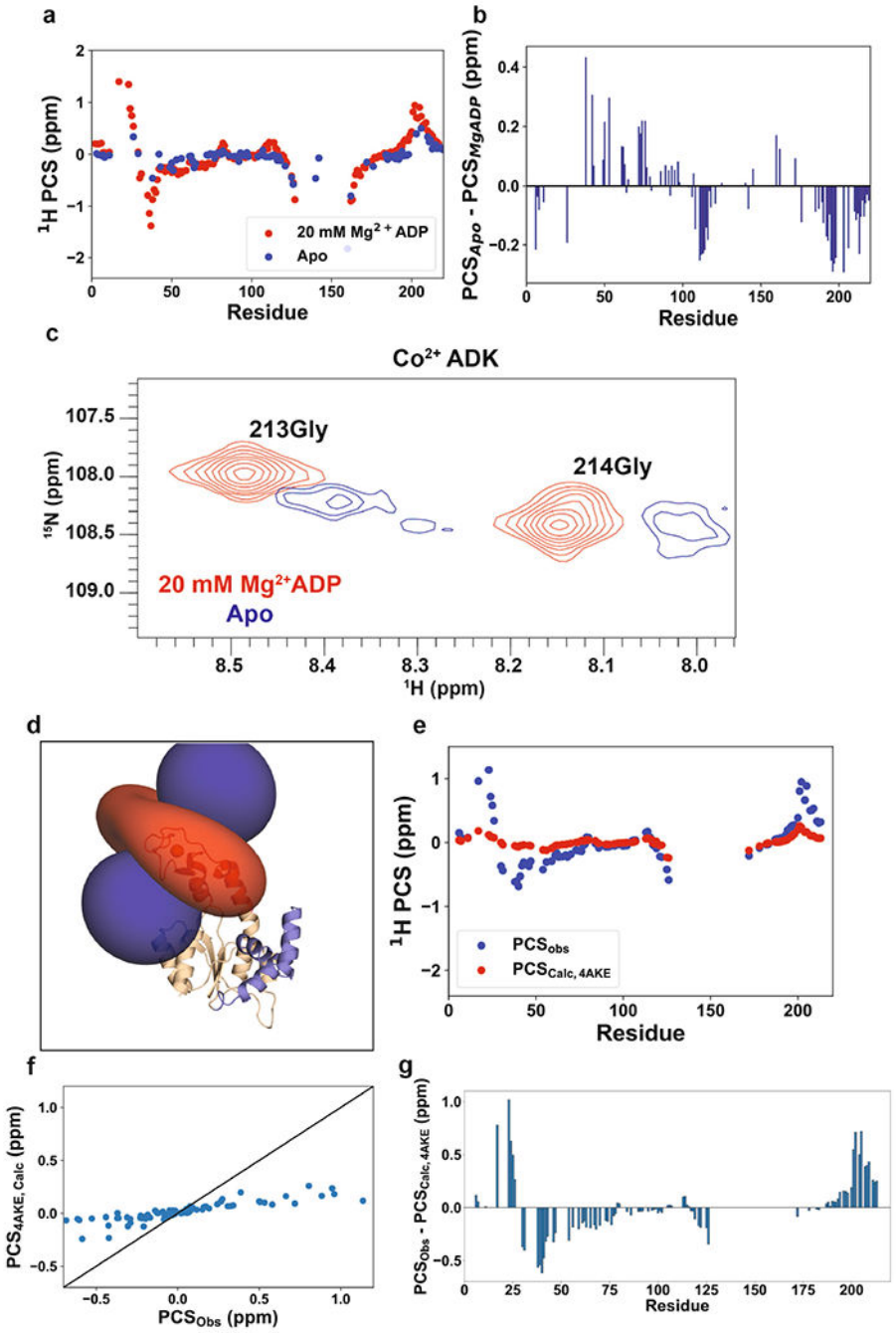


Extended Data Fig. 2 |. PCSs in different ligand states reflect the conformational state of Adk. Fits of PCSs extracted from [^{1}H-^{15}N]-HSQC spectra in either 20 mM Mg^{2+} ADP (a), 20mM ADP (b), and apo (c) to the closed crystal structure (PDB 4QBH[26]). Notably, the Mg^{2+}/ADP data fits reasonably well (Q = 11.6%), but reports on PCSs from both the closed crystal structure and the minor state (i.e., population averaged). In the absence of magnesium, the open/closing exchange rate is in the slow time regime, leading to PCSs that better fit the closed state (Q = 7.6%). The *apo* PCSs report on a more open state, and, therefore, fit significantly worse (Q = 47.9%) to the closed state. (d) Calculating tensors using either paramagnetic-induced RDCs (left) or PCSs (right) provide similar tensor values, indicating little ps-ns motion of the paramagnet.



Extended Data Fig. 3 |. In the absence of magnesium, the open/closed rate exchange rate is in the slow exchange regime, in agreement with earlier reports[25].
(a) ^{1}H, ^{15}N]-HSQC spectra for Zn^{2+} and Co^{2+} samples with either 20 mM ADP or 20mM Mg^{2+} ADP. Residues that were broadened in the Mg^{2+} ADP sample show reduced linewidths in the ADP sample, indicating a shift in exchange timescales from intermediate to slow exchange. Black lines trace the PCSs between diamagnetic and paramagnetic samples. For residue 38, the corresponding diamagnetic peaks are at 8.46 ppm (^{1}H) and 123.5 ppm (^{15}N) and are not shown for illustration purposes. (b) $^{1}H_N$ CPMG dispersion profiles for

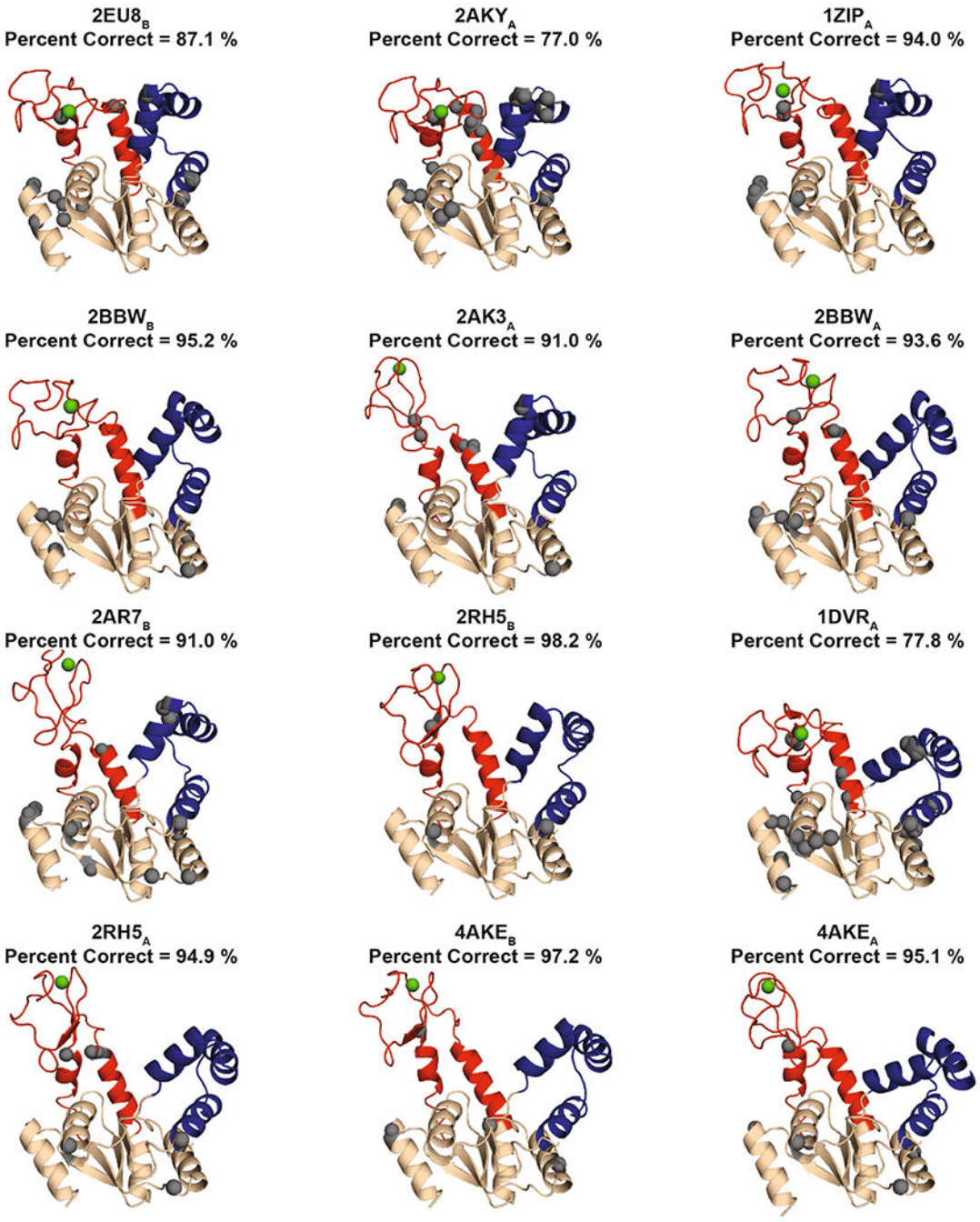
Co^{2+} Adk with 20 mM ADP and 0 mM Mg^{2+}. Representative traces show slow exchange that is fit to the Tollinger equation[75] (*F*-statistics were used to determine whether the description by a slow exchange model compared to a "no-exchange" model was justified at the 95% confidence interval; p>0.05). (c) Representative CPMG relaxation dispersion profiles for residues in the presence on 20 mM Mg^{2+} ADP. Notably, the paramagnetic chemical shift differences in the absence and presence of magnesium are similar, whereas the timescale is significantly altered ($k_{open,\ ADP}$ = 2.6 ± 0.3 s^{-1} vs $k_{open,\ ADP}$ = 180 ± 36 s^{-1}). Uncertainties (s.d.) in $R_{2,eff}$ are determined from the rmsd in the intensities of duplicate points (n = 3) according to the definition of pooled relative standard deviation. Uncertainties (s.d.) in chemical shift differences were calculated from the covariance matrix.



Extended Data Fig. 4 |. Comparison of PCS values determined during *apo* and turnover conditions and correlation of PCSs during turnover conditions with the open structure of Adk.

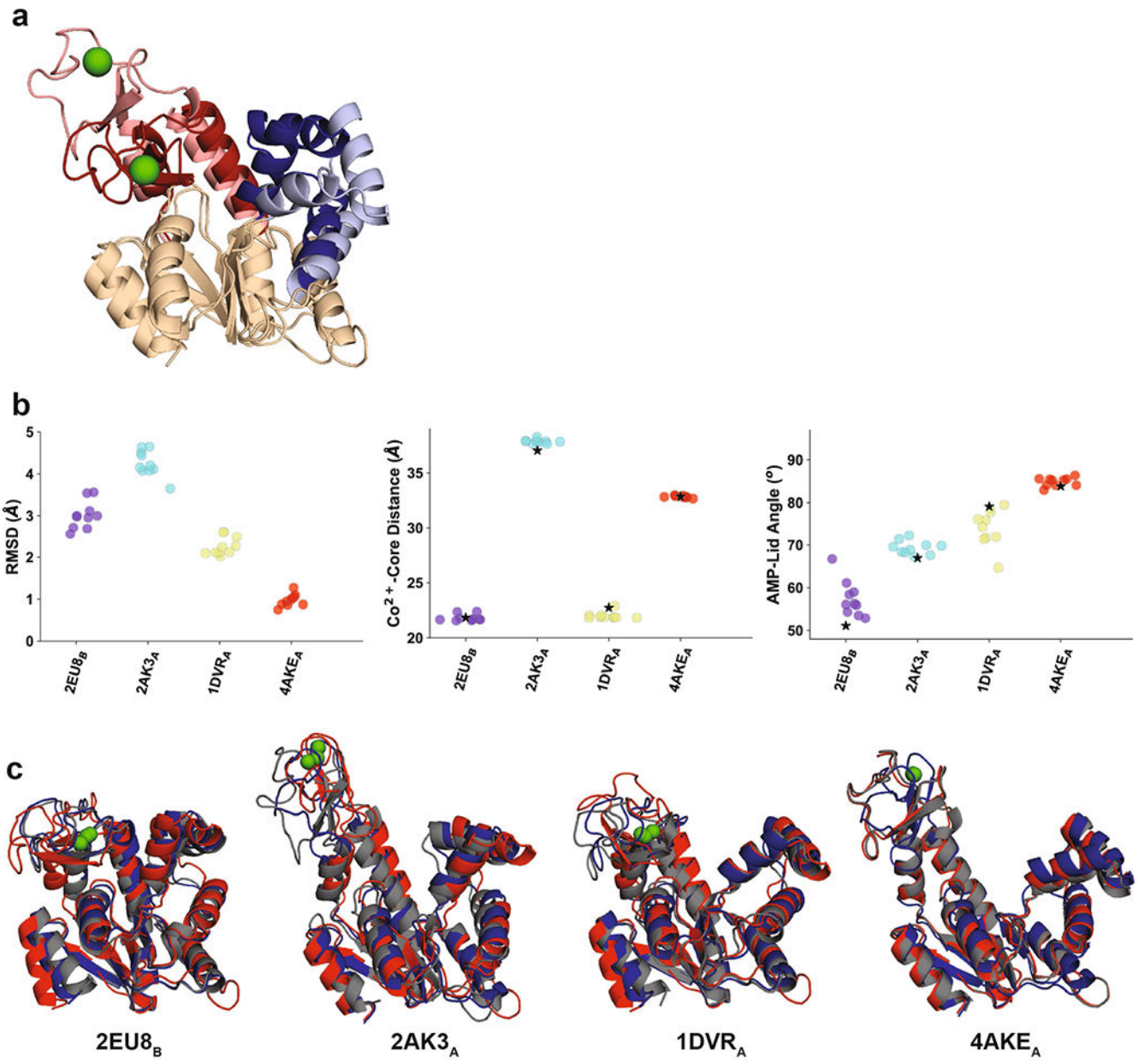
(a) Overlay of PCS values obtained for *apo* and Mg^{2+} ADP conditions. Values were determined from [^{1}H-^{15}N]-HSQC comparison in Zn^{2+} and Co^{2+} states. Note the sizable loss of PCS in the *apo* state compared to the closed state, indicating a more open structure in the absence of ligand. (b) Difference in PCS values for *apo* and turnover conditions. Large absolute differences of > 0.1 ppm are observed for many residues. (c) Zoom in of [^{1}H-^{15}N]-HSQC spectra in either 20 mM Mg^{2+} ADP or apo conditions of Co^{2+} Adk. Noticeable line broadening is observed for *apo* conditions. (d) Fit of open state (4AKE[68]) to

observed PCS shift data during catalytic turnover. (e) Best-fit tensor for PCS to open state structure. (f) Calculated PCSs for open state structure when fit with observed PCSs. A poor fit is found as the observed PCSs do not report on the open state structure. (g) The PCS difference expected between the open and closed state structures. Differences of |0.5 ppm| or greater would be expected for residues in AMP lid and core domain near the ATP lid.



Extended Data Fig. 5 |. Expectation-maximization during simulated annealing leads to correct PCS identifications for nearly all residues.

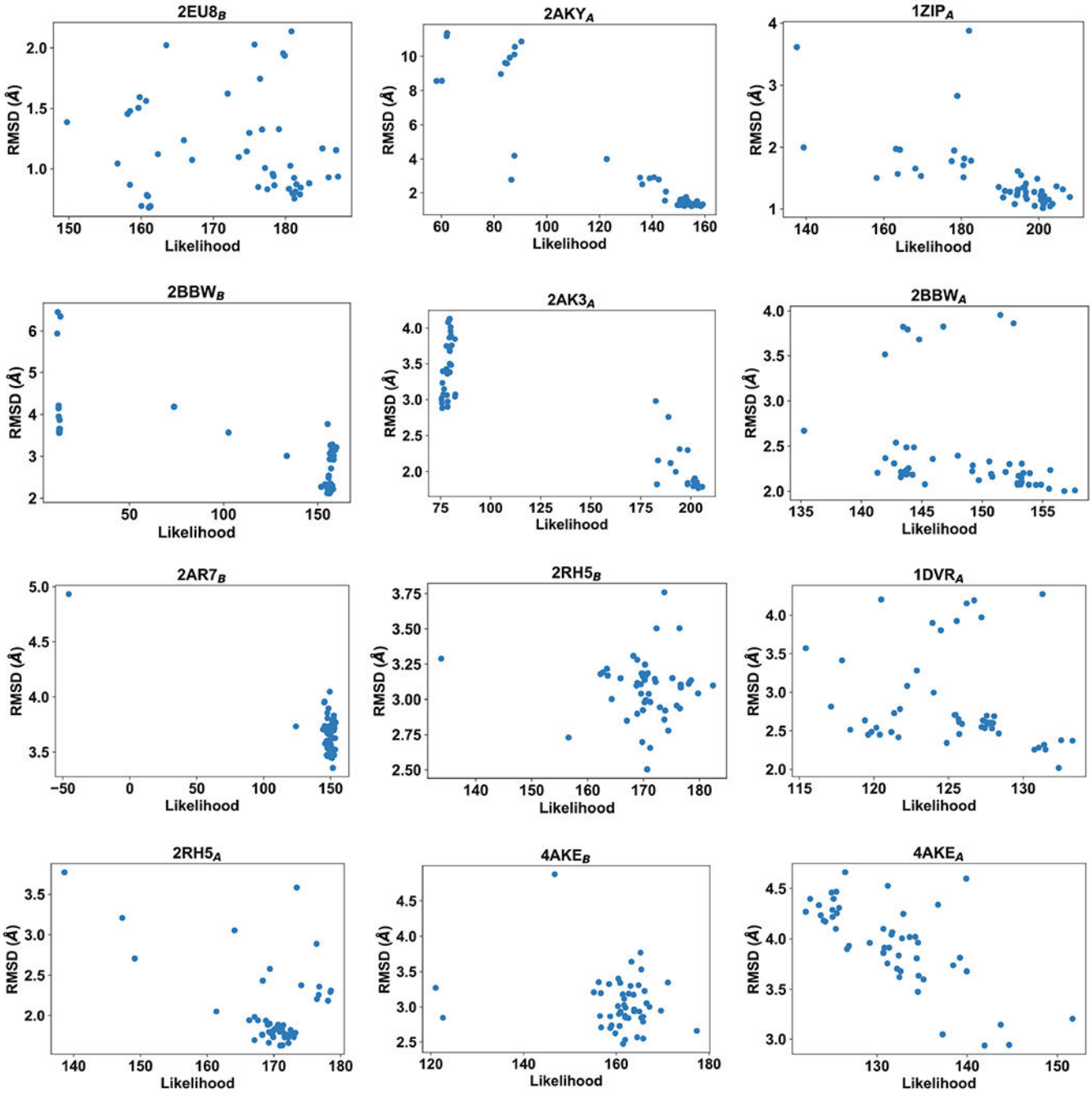
All 12 structures are shown with incorrectly chosen PCS as gray spheres. Mistakes usually occur near the end of secondary structure units, where local differences between crystal structures are most prevalent. For each structure, the core domain, ATP-lid, and AMP-lid are colored in wheat, red, and blue, respectively. The cobalt metal is shown as a green sphere.



Extended Data Fig. 6 |. Starting from an open or closed starting structure of Adk in the PCS-CPMG maximum-likelihood calculations results in equivalent final Adk structures.

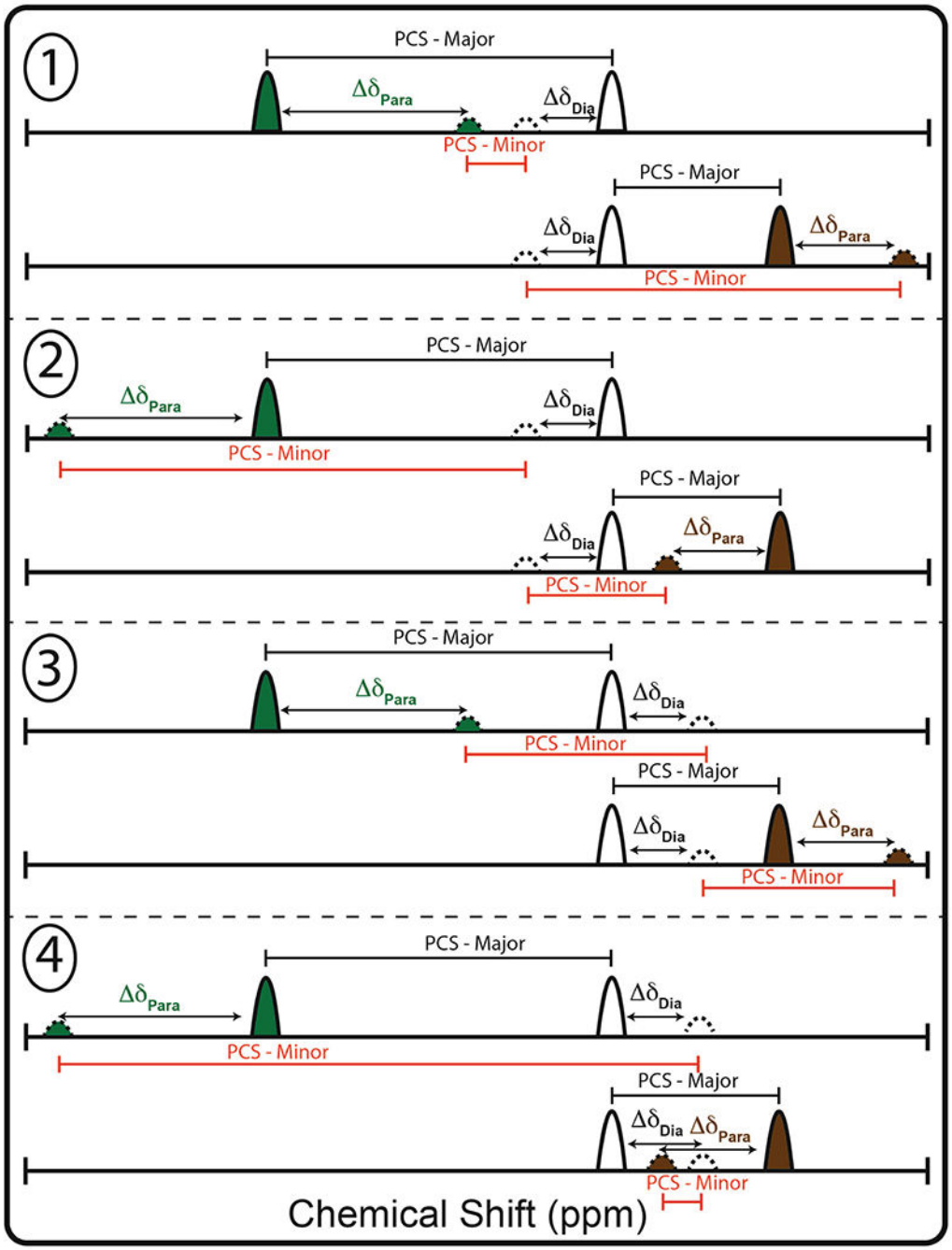
(a) Open (lighter colors) and closed (darker colors) crystal structures, PDBs 4AKE[68] and 4QBH[26], respectively. (b) RMSD, Co^{2+}-Core distance, and AMP-lid angle results for four calculated structures, all started from the open state (PDB 4AKE[68]). Similar to starting from the closed state (Fig. 2f), starting from an open state result in excellent collective

variables, and excellent agreement with the experimental structures (shown as black stars). (c) Alignment of inferred structures starting from a closed state (blue) and open state (red) compared to the target structure (gray), highlighting that the new method results in converged and accurate structures independent of the starting model.



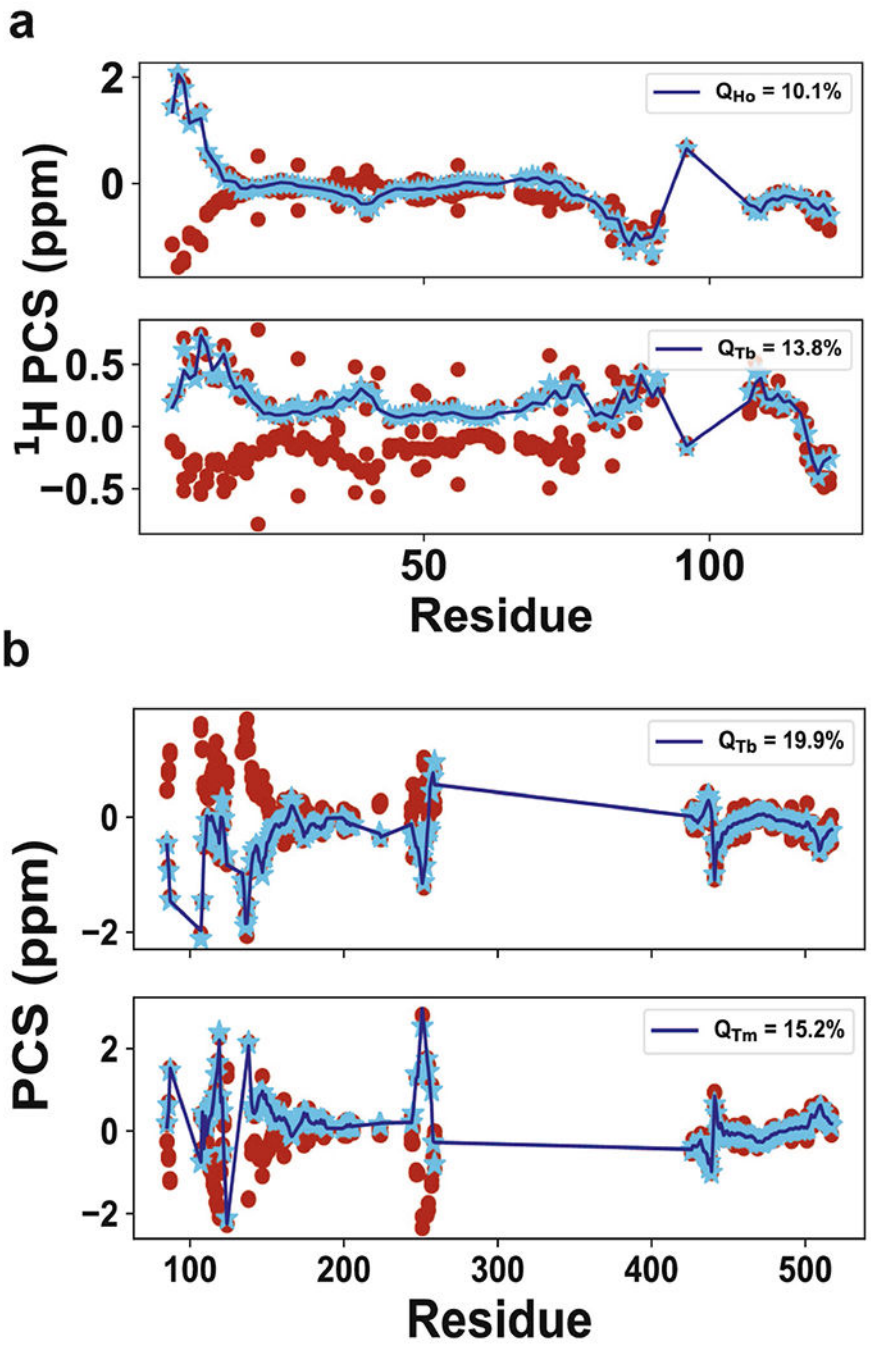
Extended Data Fig. 7 |. Plots of likelihood versus RMSD to the target structures for the first simulated annealing run for all 12 simulated Adk structures.

Structures which possess relatively low likelihood also have higher RMSD.



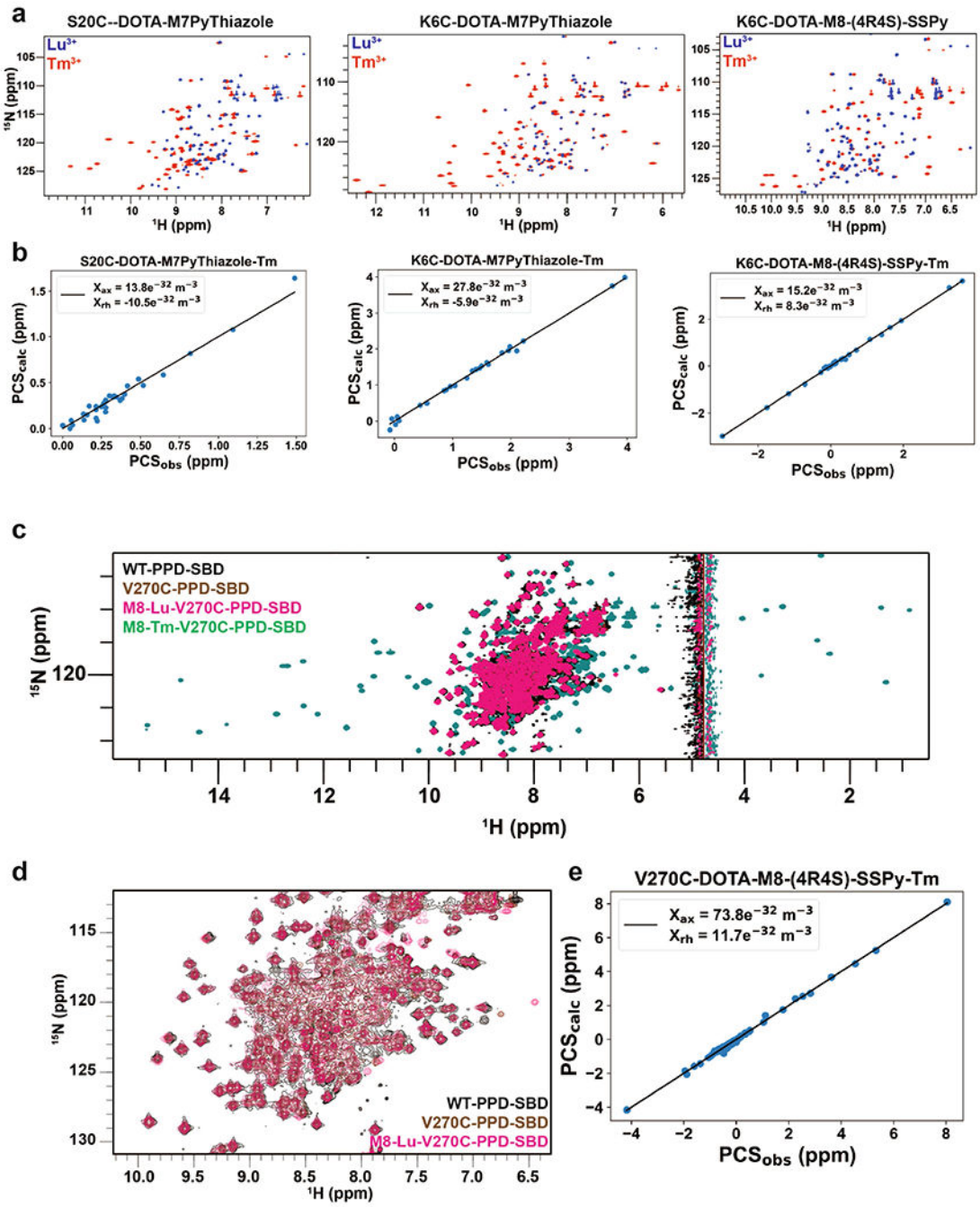
Extended Data Fig. 8 |. Eight PCS choices possible in the case where CPMG is performed in the presence of one diamagnetic and two paramagnetic metals.

As the diamagnetic sign is shared between the two paramagnetic dispersions, choices 1/3 and 2/4 are linked. This reduces the possible PCS choice for a specific residue.



Extended Data Fig. 9 |. Expectation-maximization of ambiguous PCSs are solved during structural calculations for calmodulin and src kinase.

PCSs of the final structures for calmodulin (a) and src kinase (b) have low *Q* values for both metals as well as accurate PCS identifications (blue stars for the correct PCS identification, red spheres are incorrect PCS identifications, dark blue line indicates the calculated PCS).



Extended Data Fig. 10 |. Ubiquitin and the chaperone trigger factor support lanthanide-binding tag coordination and produce substantial paramagnetic tensors.

(a) [^{1}H-^{15}N]-HSQC spectra of ubiquitin mutants S20C (left) and K6C (middle, right) bound with either the DOTA-M7PyThiazole (left, middle) or DOTA-M8-(4*R*4*S*)-SSPy(right) lanthanide-binding tags. Each spectrum shows large PCSs induced by Tm^{3+} bound tags. (b) Correlation plots between each ubiquitin variant's PCSs and the calculated PCSs. (c) [^{1}H-^{15}N]-HSQC spectra of WT PPD-SBD, V270C PPD-SBD, M8-Lu-V270C PPD-SBD, and M8-Tm-V270C PPD-SBD. (d) Zoom-in of center section of the [^{1}H-^{15}N]-HSQC spectra

of (c) showing nearly identical spectra for each sample. (e) Correlation plots between each M8-Tm-V270C PPD-SBD PCSs and the calculated PCSs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank I. Hertel-Hering and M. Rogowski for expression of the K6C and S20C ubiquitin constructs, respectively; and S. Hiller for the trigger factor expression plasmid. This work was supported by the Howard Hughes Medical Institute (HHMI) to D.K., the NIH (R01GM121384 and R01GM132499) to D.L.T., and the Department of Chemistry at University of Basel to D.H. Computational resources were provided by NSF XSEDE computing resources. We acknowledge computational support from the Brandeis HPCC which is partially supported by the NSF through DMR-MRSEC 2011846 and OAC-1920147.

Data availability

The NMR assignments of *G. stearothermophilus* adenylate kinase in its Zn^{2+}- and Co^{2+}-bound states have been deposited in the BioMagResBank[74] under accession codes 51232 and 51233, respectively. Peak lists for the tagged ubiquitin variants can be obtained from https://github.com/kernlab-brandeis/PCS-CPMG. All relevant data are available from the corresponding author upon request.

References

1. Sekhar A & Kay LE. NMR paves the way for atomic Level descriptions of sparsely populated, transiently formed biomolecular conformers. Proc. Natl Acad. Sci. USA 110, 12867–12874 (2013). [PubMed: 23868852]
2. Orellana L. Large-scale conformational changes and protein function: breaking the in silico barrier. Front. Mol. Biosci 6, 117 (2019). [PubMed: 31750315]
3. Nussinov R. Introduction to protein ensembles and allostery. Chem. Rev 116, 6263–6266 (2016). [PubMed: 27268255]
4. Haliloglu T. & Bahar I. Adaptability of protein structures to enable functional interactions and evolutionary implications. Curr. Opin. Struct. Biol 35, 17–23 (2015). [PubMed: 26254902]
5. Bertini I, Luchinat C. & Parigi G. Magnetic susceptibility in paramagnetic NMR. Prog. Nucl. Magn. Reson. Spectrosc 40, 249–273 (2002).
6. Kleckner IR. & Foster MP. An introduction to NMR-based approaches for measuring protein dynamics. Biochim. Biophys. Acta 1814, 942–968 (2011). [PubMed: 21059410]
7. Boehr DD, McElheny D, Dyson HJ & Wright PE The dynamic energy Landscape of dihydrofolate reductase catalysis. Science 313, 1638–1642 (2006). [PubMed: 16973882]
8. Cianfrocco MA et al. Human TFIID binds to core promoter DNA in a reorganized structural state. Cell 152, 120–131 (2013). [PubMed: 23332750]
9. Zhao J, Benlekbir S. & Rubinstein JL. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. Nature 521, 241–245 (2015). [PubMed: 25971514]
10. Neudecker P et al. Structure of an intermediate state in protein folding and aggregation. Science 336, 362–366 (2012). [PubMed: 22517863]
11. Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A. & Al-Hashimi HM. Visualizing transient Low-populated structures of RNA. Nature 491, 724–728 (2012). [PubMed: 23041928]
12. Zhao B, Guffy SL, Williams B. & Zhang Q. An excited state underlies gene regulation of a transcriptional riboswitch. Nat. Chem. Biol 13, 968–974 (2017). [PubMed: 28719589]

13. Fraser JS. et al. Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc. Natl Acad. Sci. USA 108, 16247–16252 (2011). [PubMed: 21918110]
14. Bonomi M. & Vendruscolo M Determination of protein structural ensembles using cryo-electron microscopy. Curr. Opin. Struct. Biol 56, 37–45 (2019). [PubMed: 30502729]
15. Vogeli B, Olsson S, Guntert P. & Riek R. The exact NOE as an alternative in ensemble structure determination. Biophys. J 110, 113–126 (2016). [PubMed: 26745415]
16. Leung HT. et al. A rigorous and efficient method to reweight very Large conformational ensembles using average experimental data and to determine their relative information content. J. Chem. Theory Comput 12, 383–394 (2016). [PubMed: 26632648]
17. Clore GM & Iwahara J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient Low-population states of biological macromolecules and their complexes. Chem. Rev 109, 4108–4139 (2009). [PubMed: 19522502]
18. Maltsev AS, Grishaev A, Roche J, Zasloff M & Bax A. Improved cross validation of a static ubiquitin structure derived from high precision residual dipolar couplings measured in a drug-based Liquid crystalline phase. J. Am. Chem. Soc 136, 3752–3755 (2014). [PubMed: 24568736]
19. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR & Kay LE A transient and Low-populated protein-folding intermediate at atomic resolution. Science 329, 1312–1316 (2010). [PubMed: 20829478]
20. Nerli S, McShan AC & Sgourakis NG Chemical shift-based methods in NMR structure determination. Prog. Nucl. Magn. Reson. Spectrosc 106-107, 1–25 (2018). [PubMed: 31047599]
21. Bertini I. et al. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. Proc. Natl Acad. Sci. USA 101, 6841–6846 (2004). [PubMed: 15100408]
22. Hass MAS et al. A minor conformation of a lanthanide tag on adenylate kinase characterized by paramagnetic relaxation dispersion NMR spectroscopy. J. Biomol. NMR 61, 123–136 (2015). [PubMed: 25563704]
23. Xu D. et al. Ligand proton pseudocontact shifts determined from paramagnetic relaxation dispersion in the limit of NMR intermediate exchange. J. Phys. Chem. Lett 9, 3361–3367 (2018). [PubMed: 29864276]
24. Eichmuller C & Skrynnikov NR Observation of microsecond time-scale protein dynamics in the presence of Ln^{3+} ions: application to the N-terminal domain of cardiac troponin C. J. Biomol. NMR 37, 79–95 (2007). [PubMed: 17180551]
25. Kerns SJ et al. The energy landscape of adenylate kinase during catalysis. Nat. Struct. Mol. Biol 22, 124–131 (2015). [PubMed: 25580578]
26. Moon S, Bannen RM, Rutkoski TJ, Phillips GN Jr & Bae E. Effectiveness and limitations of local structural entropy optimization in the thermal stabilization of mesophilic and thermophilic adenylate kinases. Proteins 82, 2631–2642 (2014). [PubMed: 24931334]
27. Hanson JA et al. Illuminating the mechanistic roles of enzyme conformational dynamics. Proc. Natl Acad. Sci. USA 104, 18055–18060 (2007). [PubMed: 17989222]
28. Aden J & Wolf-Watz M NMR identification of transient complexes critical to adenylate kinase catalysis. J. Am. Chem. Soc 129, 14003–14012 (2007). [PubMed: 17935333]
29. Pelz B, Zoldak G, Zeller F, Zacharias M & Rief M Subnanometre enzyme mechanics probed by single-molecule force spectroscopy. Nat. Commun 7, 10848 (2016). [PubMed: 26906294]
30. Mukhopadhyay A et al. Crystal structure of the zinc-, cobalt-, and iron-containing adenylate kinase from *Desulfovibrio gigas*: a novel metal-containing adenylate kinase from Gram-negative bacteria. J. Biol. Inorg. Chem 16, 51–61 (2011). [PubMed: 20821240]
31. Carver JP & Richards RE General 2-site solution for chemical exchange produced dependence of T2 upon Carr–Purcell pulse separation. J. Mag. Res 6, 89–105 (1972).
32. Aviram HY et al. Direct observation of ultrafast Large-scale dynamics of an enzyme under turnover conditions. Proc. Natl Acad. Sci. USA 115, 3243–3248 (2018). [PubMed: 29531052]
33. Skrynnikov NR, Dahlquist FW & Kay LE Reconstructing NMR spectra of “invisible” excited protein states using HSQC and HMQC experiments. J. Am. Chem. Soc 124, 12352–12360 (2002). [PubMed: 12371879]

34. Schwieters CD, Kuszewski JJ, Tjandra N & Clore GM The Xplor-NIH NMR molecular structure determination package. J. Mag. Res 160, 65–73 (2003).
35. Fallon JL & Quiocho FA A closed compact structure of native Ca^{2+}-calmodulin. Structure 11, 1303–1307 (2003). [PubMed: 14527397]
36. Cowan-Jacob SW et al. The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. Structure 13, 861–871 (2005). [PubMed: 15939018]
37. Müntener T, Kottelat J, Huber A & Häussinger D New lanthanide chelating tags for PCS NMR spectroscopy with reduction stable, rigid linkers for fast and irreversible conjugation to proteins. Bioconjugate Chem. 29, 3344–3351 (2018).
38. Chou JJ, Li S, Klee CB & Bax A Solution structure of Ca^{2+}-calmodulin reveals flexible hand-like properties of its domains. Nat. Struct. Biol 8, 990–997 (2001). [PubMed: 11685248]
39. Russel D et al. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol. 10, e1001244 (2012). [PubMed: 22272186]
40. Häussinger D, Huang JR & Grzesiek S DOTA-M8: an extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy. J. Am. Chem. Soc 131, 14761–14767 (2009). [PubMed: 19785413]
41. Morgado L, Burmann BM, Sharpe T, Mazur A & Hiller S The dynamic dimer structure of the chaperone Trigger Factor. Nat. Commun 8, 1992 (2017). [PubMed: 29222465]
42. Kovermann M, Grundstrom C, Sauer-Eriksson AE, Sauer UH & Wolf-Watz M Structural basis for ligand binding to an enzyme by a conformational selection pathway. Proc. Natl Acad. Sci. USA 114, 6298–6303 (2017). [PubMed: 28559350]
43. Li D, Liu MS & Ji B Mapping the dynamics landscape of conformational transitions in enzyme: the adenylate kinase case. Biophys. J 109, 647–660 (2015). [PubMed: 26244746]
44. Stiller JB et al. Probing the transition state in enzyme catalysis by high-pressure NMR dynamics. Nat. Catal 2, 726–734 (2019). [PubMed: 32159076]
45. Saio T & Ishimori K Accelerating structural life science by paramagnetic lanthanide probe methods. Biochim. Biophys. Acta 1864, 129332 (2019).
46. Nitsche C & Otting G Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. Prog. Nucl. Magn. Reson. Spectrosc 98-99, 20–49 (2017). [PubMed: 28283085]
47. Ma RS et al. Determination of pseudocontact shifts of low-populated excited states by NMR chemical exchange saturation transfer. Phys. Chem. Chem. Phys 18, 13794–13798 (2016). [PubMed: 27001533]
48. Gerstein M, Lesk AM & Chothia C Structural mechanisms for domain movements in proteins. Biochemistry 33, 6739–6749 (1994). [PubMed: 8204609]
49. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). [PubMed: 34265844]
50. Schmitz C, Stanton-Cook MJ, Su XC, Otting G & Huber T Numbat: an interactive software tool for fitting Delta chi-tensors to molecular coordinates using pseudocontact shifts. J. Biomol. NMR 41, 179–189 (2008). [PubMed: 18574699]
51. Cai M, Huang Y, Craigie R & Clore GM A simple protocol for expression of isotope-labeled proteins in *Escherichia coli* grown in shaker flasks at high cell density. J. Biomol. NMR 73, 743–748 (2019). [PubMed: 31677040]
52. Otting G, Ruckert M, Levitt MH & Moshref A NMR experiments for the sign determination of homonuclear scalar and residual dipolar couplings. J. Biomol. NMR 16, 343–346 (2000). [PubMed: 10826886]
53. Joss D, Walliser RM, Zimmermann K & Häussinger D Conformationally locked lanthanide chelating tags for convenient pseudocontact shift protein nuclear magnetic resonance spectroscopy. J. Biomol. NMR 72, 29–38 (2018). [PubMed: 30117038]
54. Romero PR et al. BioMagResBank (BMRB) as a resource for structural biology. Methods Mol. Biol 2112, 187–218 (2020). [PubMed: 32006287]
55. Orton HW, Huber T & Otting G Paramagpy: software for fitting magnetic susceptibility tensors using paramagnetic effects measured in NMR spectra. Magn. Reson 1, 1–12 (2020).

56. Ishima R & Torchia DA Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. J. Biomol. NMR 25, 243–248 (2003). [PubMed: 12652136]
57. Delaglio F et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277–293 (1995). [PubMed: 8520220]
58. Vranken WF et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59, 687–696 (2005). [PubMed: 15815974]
59. Lee W, Rahimi M, Lee Y & Chiu A POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. Bioinformatics. 37, 3041–3042 (2021).
60. Niklasson M et al. Comprehensive analysis of NMR data using advanced line shape fitting. J. Biomol. NMR 69, 93–99 (2017). [PubMed: 29043470]
61. Newville M, Stensitzki T, Allen DB & Ingargiola A LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python https://lmfit.github.io/lmfit-py/ (2014).
62. Counago R, Chen S & Shamoo Y In vivo molecular evolution reveals biophysical origins of organismal fitness. Mol. Cell 22, 441–449 (2006). [PubMed: 16713575]
63. Abele U & Schulz GE High-resolution structures of adenylate kinase from yeast ligated with inhibitor Ap5A, showing the pathway of phosphoryl transfer. Protein Sci. 4, 1262–1271 (1995). [PubMed: 7670369]
64. Berry MB & Phillips GN Jr. Crystal structures of *Bacillus stearothermophilus* adenylate kinase with bound Ap5A, Mg^{2+} Ap5A, and Mn^{2+} Ap5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg^{2+} and Mn^{2+}. Proteins 32, 276–288 (1998). [PubMed: 9715904]
65. Diederichs K & Schulz GE The refined structure of the complex between adenylate kinase from beef heart mitochondrial matrix and its substrate AMP at 1.85 Å resolution. J. Mol. Biol 217, 541–549 (1991). [PubMed: 1994037]
66. Schlauderer GJ, Proba K & Schulz GE Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP. J. Mol. Biol 256, 223–227 (1996). [PubMed: 8594191]
67. Henzler-Wildman KA. et al. Intrinsic motions along an enzymatic reaction trajectory. Nature 450, 838–844 (2007). [PubMed: 18026086]
68. Muller CW, Schlauderer GJ, Reinstein J. & Schulz GE. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure 4, 147–156 (1996). [PubMed: 8805521]
69. Arnold K, Bordoli L, Kopp J. & Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195–201 (2006). [PubMed: 16301204]
70. Word JM, Lovell SC, Richardson JS. & Richardson DC Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J. Mol. Biol 285, 1735–1747 (1999). [PubMed: 9917408]
71. Chattopadhyaya R, Meador WE, Means AR & Quiocho FA Calmodulin structure refined at 1.7 Å resolution. J. Mol. Biol 228, 1177–1192 (1992). [PubMed: 1474585]
72. Xu W, Doshi A, Lei M, Eck MJ. & Harrison SC Crystal structures of c-Src reveal features of its autoinhibitory mechanism. Mol. Cell 3, 629–638 (1999). [PubMed: 10360179]
73. Bertini I, Janik MB, Lee YM, Luchinat C & Rosato A Magnetic susceptibility tensor anisotropies for a lanthanide ion series in a fixed protein matrix. J. Am. Chem. Soc 123, 4181–4188 (2001). [PubMed: 11457182]
74. Ulrich EL et al. BioMagResBank. Nucleic Acids Res. 36, D402–D408 (2007). [PubMed: 17984079]
75. Tollinger M, Skrynnikov NR, Mulder FA, Forman-Kay JD & Kay LE Slow dynamics in folded and unfolded states of an SH3 domain. J. Am. Chem. Soc 123, 11341–11352 (2001). [PubMed: 11707108]

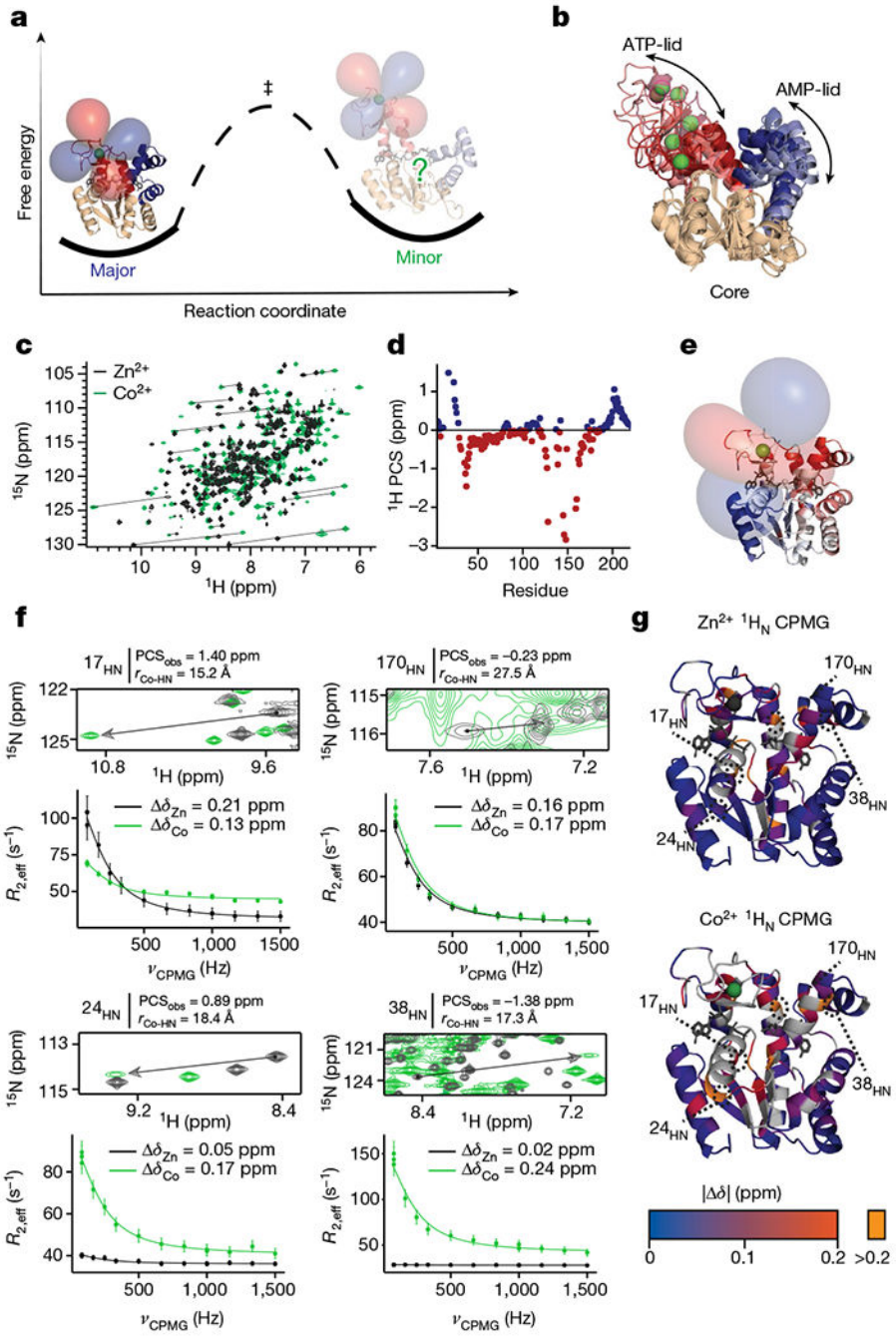


Fig. 1 |. Paramagnetic enhanced NMR for structure determination of a high-energy state during catalysis.

a, Simplified free-energy landscape for the rate-limiting step in the catalytic cycle of Adk, underscoring the unknown minorly populated structure (green). The core (tan), AMP lid (blue), and ATP lid (red) are shown. The paramagnetic iso-surfaces resulting from the bound paramagnetic metal (green sphere) illustrate the concept of using the difference in paramagnetic restraints for the minor state structure determination. The transition state for Adk's conformational change is indicated with a double dagger. **b**, Structural flexibility

of Adk shown by crystal structures indicating closed and open conformations. **c**, [^{1}H-^{15}N]-HSQC of diamagnetic (Zn^{2+}, black) and paramagnetic (Co^{2+}, green) Adk during catalysis (substrate saturation, 20 mM Mg^{2+}-ADP). Black lines indicate the co-linear PCSs seen for almost all residues. **d, e,** Quantifiable PCSs are observed globally (**d**) and approximate the closed conformation of Adk (**e**) (see Extended Data Fig. 2). The paramagnetic iso-surface was produced using the NUMBAT software[50]. PCS values are coloured onto the structure with non-assigned residues in white. **f,** Representative dispersion profiles of $^{1}H_{N}$ CPMG of diamagnetic and paramagnetic Adk during catalysis, with observed PCS in the [^{1}H-^{15}N]-HSQC indicated above the graph. Uncertainties (s.d.) in $R_{2,\mathrm{eff}}$ are determined from the r.m.s.d. in the intensities of duplicate points ($n = 3$) according to the definition of pooled relative standard deviation. $R_{2,\mathrm{eff}}$ is the effective transverse relaxation rate at each refocusing pulse train frequency (ν_{CPMG}. Uncertainties (s.d.) in $\Delta\delta$ were calculated from the covariance matrix of a global fit of diamagnetic and paramagnetic data. **g**, $\Delta\delta$ extracted from fits of **f** are plotted as colours on the structure.

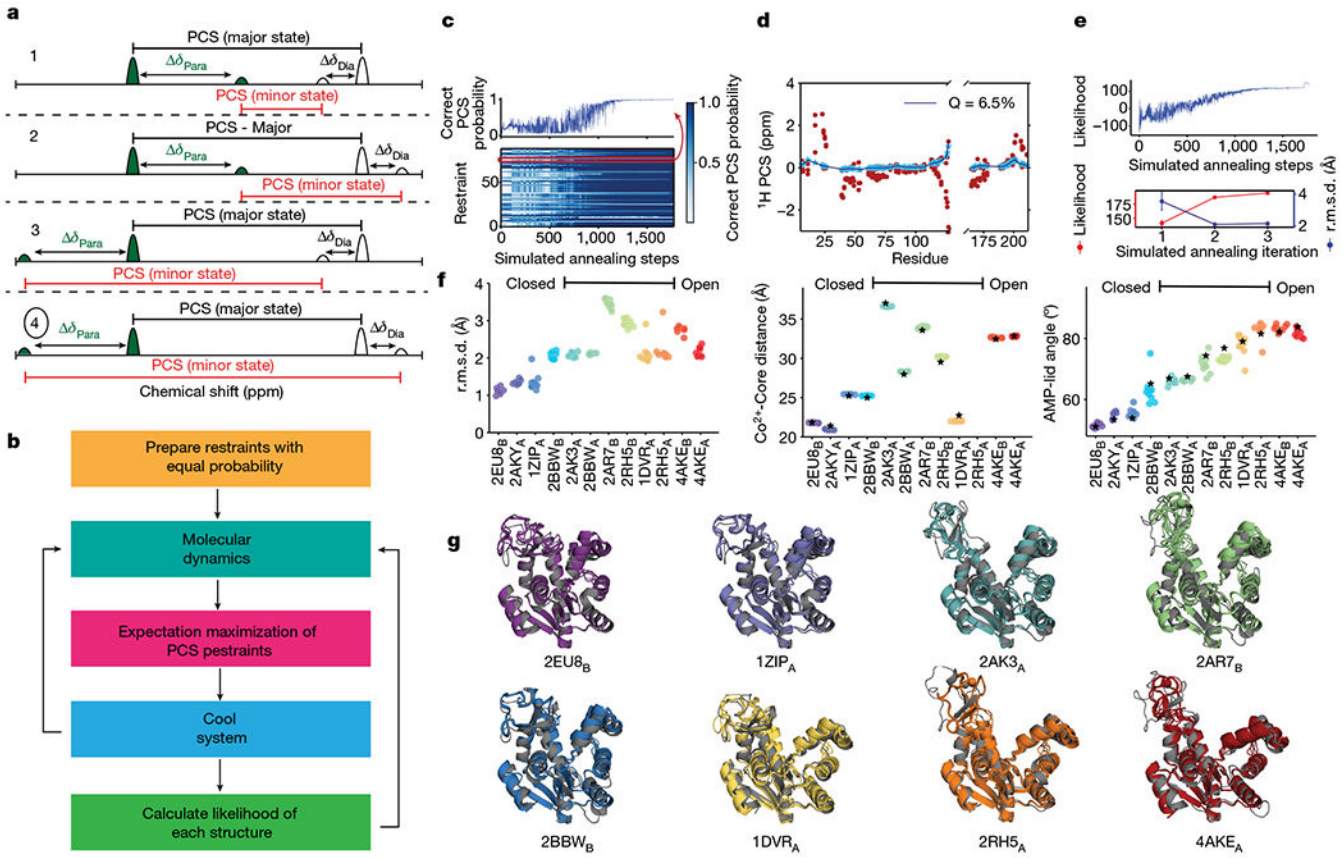


Fig. 2 |. Maximum-likelihood classification method for high-energy structure determination.

a, One-dimensional representation of the four possible PCS_{minor} restraints for each residue based on diamagnetic–paramagnetic relaxation dispersion. **b**, Flow chart of our approach to determine the high-energy structure including an expectation-maximization, simulated annealing scheme. **c–g**, Benchmarking of the iterative PCS–CPMG method. **c–e**, Representative results for the open Adk structure. **c,** PCS probability map for each of the 93 restraints over the course a simulated annealing run. A representative trace for residue 211 illustrates convergence of PCS identification. **d,** PCSs of the final structure have low Q values as well as accurate PCS identifications (blue stars indicate correct PCS identification, red spheres are incorrect PCS identifications, dark blue line indicates the calculated PCS). **e,** The likelihood increases over the course of the run (top), and multiple iterations show an increased likelihood with a concomitantly decreased r.m.s.d. (bottom; $n = 400$ structures; mean ± s.d.). Convergence of the likelihood and r.m.s.d. is observed between the second and fourth run. **f,** The same procedure is performed for a series of Adk structures from closed to open conformations. All are well-determined based on r.m.s.d. (left), metal coordinates (middle) and AMP lid position (right). Each sphere represents one of the top ten solutions, and stars indicate the correct metal location or AMP lid position based on the target X-ray structure. **g,** Alignment of the top five inferred structures in colour with target structures in grey, showing excellent agreement with small structural differences being predominately local.

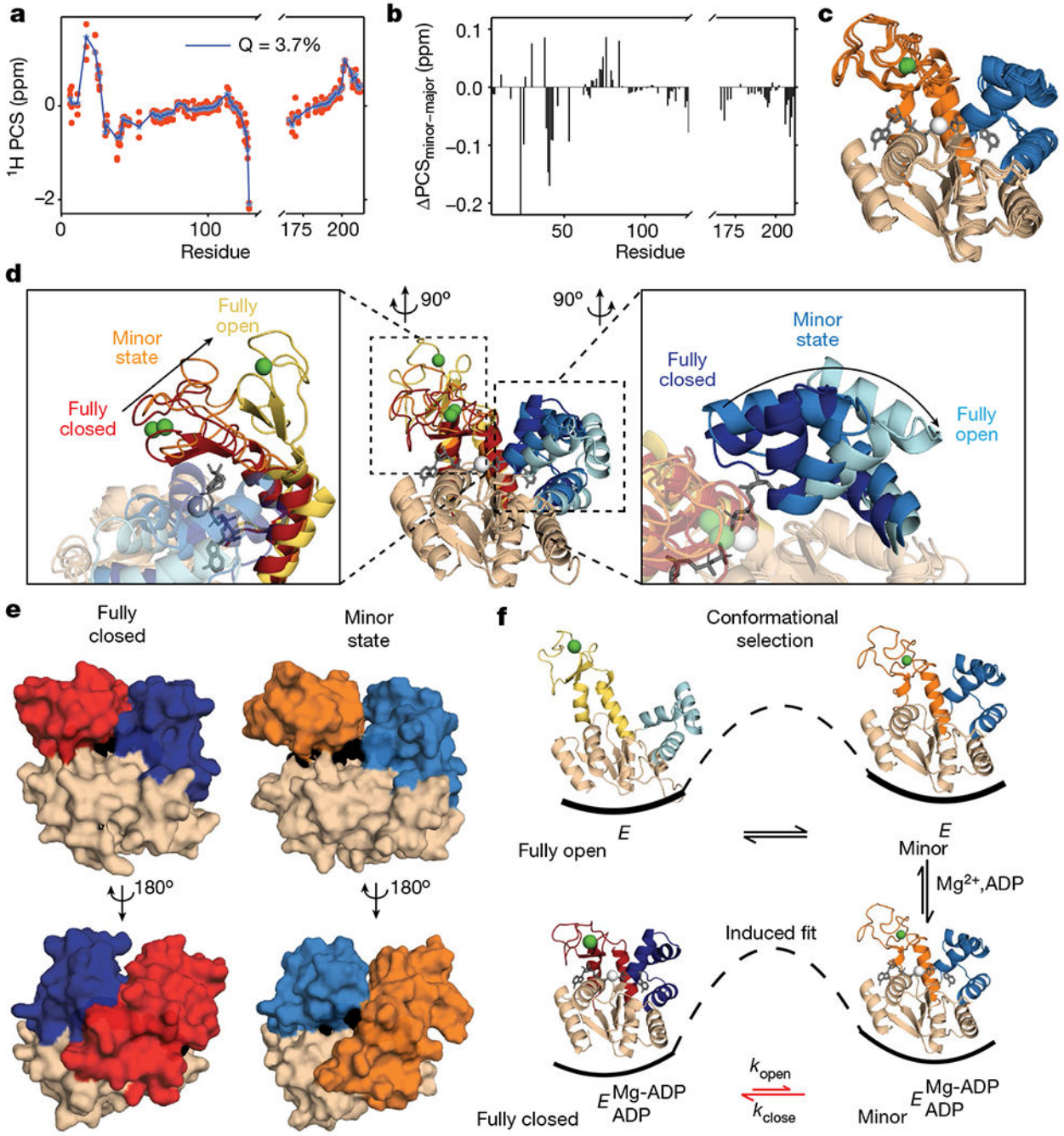


Fig. 3 |. The high-energy state of Adk during enzyme catalysis.

a, PCS identifications for minor state after maximum-likelihood calculation from experimental data (Fig. 1f). Low Q values indicate a good correspondence between the back-calculated and identified PCSs. **b,** The difference in PCS between major and minor state. **c,** The top five highest likelihood structures of the minor state solved by the expectation-maximization simulated annealing protocol. The position of Zn^{2+} or Co^{2+} is shown as a green sphere; the catalytic Mg^{2+} coordinating two ADP substrates is shown as a white sphere. **d,** Structural comparison of the major closed state (blue and red), average minor

state (light blue and orange) and open structure (cyan and yellow). The ATP lid shows slight opening in the minor state with a 1.8 Å change in metal position (left). The AMP lid domain shows greater movement with approximately 50% transition from a closed to open conformation (right). **e**, Surface representation of the closed crystal structure (left) and minor state (right) in front and side view (nucleotides are shown in black and Mg^{2+} as a white sphere). **f,** Proposed mechanism for substrate binding and product release, including conformational selection for substrate binding to the minor state, followed by an induced-fit step of full closure over the nucleotides.

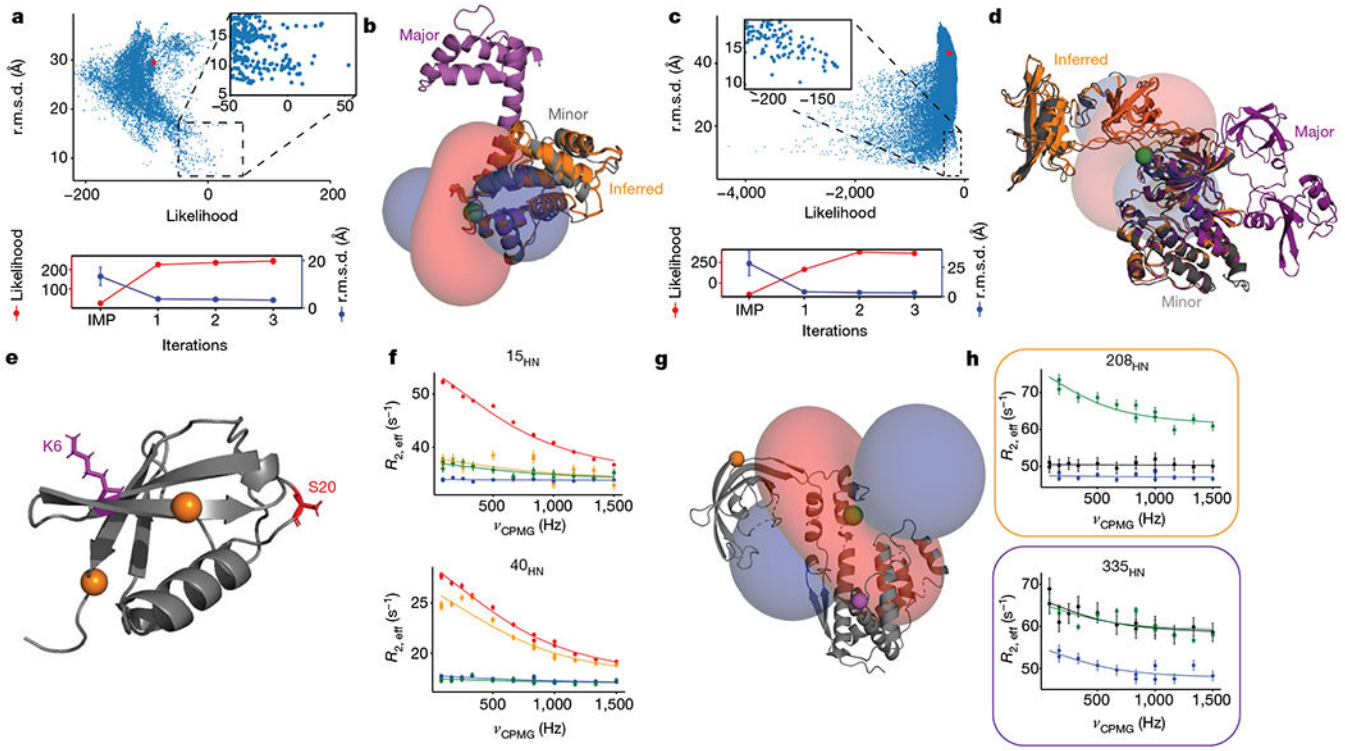


Fig. 4 |. General applicability of PCS–CPMG methodology.

a–h, PCS–CPMG data for calmodulin (**a, b**), Src kinase (**c, d**), ubiquitin (**e, f**) and trigger factor (**g, h**). **a, c,** Red stars indicate the starting point–the major state structure, **b, d,** Large conformational changes are solved with paramagnetic dispersion restraints to high accuracy for calmodulin (**b;** 2 Å r.m.s.d.) and Src kinase (**d;** 2.5 Å r.m.s.d.) as seen from the close superposition between the calculated (orange) and target conformation (grey). Simulated paramagnetic metals, shown in green, are either added via replacement of the known Ca^{2+} site (**b**) or using a paramagnetic binding tag (**d**). r.m.s.d. values in **a,** bottom and **c,** bottom represent mean ± s.d. of the top 10 structures based on likelihood. **e,** Locations of ubiquitin tagging locations are shown in purple (K6) and red (S20).**f,** $^{1}H_{N}$ CPMG profiles for four representative residues, shown as orange spheres in **e**, for S20C-M7-Tm^{3+} (red), K6C-M7-Tm^{3+} (orange), K6C-M8-Tm^{3+} (green) and K6C-M8-Lu^{3+} (blue). For visualization, all profiles were normalized to the same fitted $R_{2,inf}$ value, where $R_{2,inf}$ is the exchange-free transverse relaxation rate. **g,** Tensor of PPD–SBD V270C with metal (green) located at the centre of the tensor. **h,** Representative $^{1}H_{N}$ CPMG profiles for wild-type (black), V270C M8-Lu^{3+} (blue) and V270C M8-Tm^{3+} (green) tagged protein for residues in the PPD (orange) and SBD (purple), which are shown as spheres in **g.** Uncertainties in $R_{2,eff}$ (s.d.) (**f, h**) are determined from the r.m.s.d. in the intensities of duplicate points ($n = 3$) according to the definition of pooled relative standard deviation.