

PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Analysis of the solution conformations of T4 lysozyme by paramagnetic NMR spectroscopy

Jia-Liang Chen,^a Yin Yang,^a Lin-Lin Zhang,^a Haobo Liang,^b Thomas Huber,^{*b} Xun-Cheng Su^{*a,b} and Gottfried Otting^{*b}

^a State Key Laboratory of Elemento-organic Chemistry, The Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Nankai University, Tianjin 300071, China

^b Research School of Chemistry, Australian National University, Canberra, ACT 0200

* Corresponding authors

e-mail: t.huber@anu.edu.au, xunchengsu@nankai.edu.cn and gottfried.otting@anu.edu.au

Keywords: domain movement, T4 lysozyme, paramagnetic NMR spectroscopy, lanthanide ion, pseudocontact shift, residual dipolar coupling, paramagnetic relaxation enhancement

Abstract: A large number of crystal structures of bacteriophage T4 lysozyme (T4-L) have shown that it contains two subdomains, which can arrange in a compact conformation (closed state) or, in mutants of T4-L, more extended structures (open state). In solution, wild-type T4-L displays only a single set of nuclear magnetic resonance (NMR) signals, masking any conformational heterogeneity. To probe the conformational space of T4-L, we generated a site-specific lanthanide binding site by attaching 4-mercaptomethyl dipicolinic acid via a disulfide bond to Cys44 in the triple-mutant C54T/C97A/S44C of T4-L and measured pseudocontact shifts (PCS) and magnetically induced residual dipolar couplings (RDC). The data indicate that, in solution and in the absence of substrate, the structure of T4-L is on average more open than suggested by the closed conformation of the crystal structure of wild-type T4-L. A slightly improved fit was obtained by assuming a population-weighted two-state model involving an even more open conformation and the closed state, but paramagnetic relaxation enhancements measured with Gd^{3+} argue against such a conformational equilibrium. The fit could not be improved by including a third conformation picked from the hundreds of crystal structures available for T4-L mutants.

1 Introduction

Domain motions are of fundamental importance for the function of enzymes.¹⁻⁷ While experiments often reveal evidence for multiple conformations, detailed analyses are difficult when the exchange between different conformations is fast. Fitting the data by an ensemble of conformations is an ill-posed problem, which typically has multiple non-unique solutions. In addition, it is difficult to determine the structures in the ensemble without relying on model building. Detailed models can be obtained by molecular dynamics simulations, but arguably the most accurate models come from X-ray crystallography, where snapshots of different individual conformations can be frozen in single crystals.³

T4 lysozyme (T4-L) is the protein, for which the protein data bank holds more crystal structures of wild-type and mutant forms⁸ than for any other protein. The extraordinary number of mutants offers a rich set of experimentally determined conformations to choose from for ensemble calculations. The structure of T4-L comprises two subdomains, an N-terminal domain (N-domain) and a C-terminal domain (C-domain), which are connected by a long helix.⁹ The function of T4-L is to break down the peptidoglycan wall of the bacterial host of the T4 bacteriophage. The peptidoglycan substrate binds in a cleft between the N- and C-domain. The single-crystal structures of different T4-L mutants showed that the N- and C-domains move essentially as rigid bodies but their relative orientations can be quite variable, mediated by a hinge-bending motion.^{8,10} For example, the crystal structure of the I3P mutant of T4-L (PDB code 1L97)¹¹ detected the enzyme in a wide open conformation, whereas the wild-type protein crystallized in a closed conformation (PDB code 2LZM¹²) similar to the closed conformation observed for the active-site mutant T26E in the presence of substrate (PDB code 148L;¹³ Fig. 1).

It is a fundamental drawback of single-crystal environments that the influence of intermolecular contacts in the crystal lattice on the final conformation is difficult to assess and that mutations are required to stabilize different conformational states. In principle, this limitation can be overcome by fusion of the target protein to an independently crystallising scaffold. This approach was explored by a construct of T4-L fused to the polymerizing protein module 2TEL that generated crystals in which only the C-terminal, but not the N-terminal domain of T4-L made crystal contacts.¹⁴ Unfortunately,

the C-domain and, even more so, the N-domain showed very large B -factors, requiring manual docking of the T4-L domains into low-resolution electron density. Interestingly, the reported average conformation was quite open (PDB code 2QAR).¹⁴ To investigate the hinge-bending motion in the absence of any crystal packing constraints, however, alternative methods must be sought.

Comparison of electron paramagnetic resonance (EPR) spectra of free and substrate-bound T4-L with nitroxide tags confirmed that hinge-bending domain movement is associated with substrate binding and provided evidence for multiple conformations in vitrified solution.¹⁵ Single-molecule fluorescence studies in the presence of substrate indicated that such movements are part of the enzymatic cycle of T4-L.¹⁶ Movements during enzymatic turnover were also detected in single-molecule experiments using single-walled carbon nanotube (SWNT) field-effect transistor (FET) devices.^{17,18} These experiments indicated little movement in the absence of substrate on time scales between 20 μ s and hundreds of seconds, whereas fluorescence-correlation spectroscopy suggested that the hinge-bending motion exists also in the absence of substrate on a 15 μ s time scale.¹⁹ Molecular dynamics simulations confirmed the existence of some hinge-bending motions even on the sub-nanosecond time scale^{20,21} and transitions between different conformations were observed in an essential dynamics analysis.²² None of these methods, however, provides detailed structural information about the conformations involved and their populations in solution. An NMR structure analysis of the cysteine-free double-mutant C54T/C97A (referred to as WT*) by residual dipolar couplings (RDC) determined an average conformation in solution, revealing a mostly open conformation, but could not discern this structure from a 2-state model.²³

To gain more insight into the conformational space occupied by T4-L in solution, we used paramagnetic NMR spectroscopy to collect long-range structural restraints by pseudocontact shift (PCS), residual dipolar coupling (RDC) and paramagnetic relaxation enhancement (PRE) measurements. By using the triple-mutant C54T/C97A/S44C, the single cysteine residue at position 44 allowed site-specific attachment of the small lanthanide binding tag 4-mercaptomethylene dipicolinic acid (4MMDPA, Fig. S1).²⁴ The activity and stability of the C54T/C97A mutant (WT*) has previously been shown to be essentially the same as for the wild-type enzyme. WT* also crystallizes isomorphously

with wild-type and shows virtually the same structure except in the vicinity of the mutations.²⁵ Ser44 resides in the second α -helix of the N-domain, facing away from the substrate binding cleft (Fig. 1). Residues 54 and 97 are equally located in structurally conserved regions of the protein. None of these mutations are thus expected to perturb hinge-bending motions between the N- and C-domain. The position of the lanthanide tag in the T4-L S44C-4MMDPA construct allows the observation of PCSs in both domains with the paramagnetic lanthanides Yb³⁺, Tm³⁺ or Tb³⁺, while Gd³⁺ can be used to observe pure PREs.

PCSs, RDCs and PREs are sensitive to conformational fluctuations in different ways. PREs are proportional to $1/r^6$, where r is the distance between the paramagnetic centre and a nuclear spin. Therefore, PREs induced by Gd³⁺ bound to T4-L S44C-4MMDPA are potentially sensitive reporters of even small populations of the closed state. In contrast, PCSs are less sensitive to the distance from the paramagnetic centre. The PCS, $\Delta\delta^{\text{PCS}}$, of a nuclear spin can be described by²⁶

$$\Delta\delta^{\text{PCS}} = 1/(12\pi r^3)[\Delta\chi_{\text{ax}}(3\cos^2\theta - 1) + 1.5\Delta\chi_{\text{rh}} \sin^2\theta \cos 2\phi] \quad (1)$$

where r , θ , and ϕ are the polar coordinates of the nuclear spin relative to the principal axes of the magnetic susceptibility anisotropy ($\Delta\chi$) tensor, $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ are the axial and rhombic components of the $\Delta\chi$ tensor, and the PCS is the difference in chemical shifts (measured in ppm) between samples with paramagnetic and diamagnetic tag. The coordinate system defined by the $\Delta\chi$ tensor delivers not only distance restraints but also information about relative domain orientations.

Any paramagnetic tag that produces PCSs also generates weak alignment in a magnetic field and, therefore, RDCs.²⁷ RDCs are independent of the distance from the paramagnetic centre, depending only on the bond orientations relative to the alignment tensor. In an entirely rigid molecule, the axial and rhombic components of the alignment tensor A are proportional to the corresponding components of the $\Delta\chi$ tensor, with the same orientation:

$$A_{\text{ax,rh}} = B_0^2/(15\mu_0kT) \Delta\chi_{\text{ax,rh}} \quad (8)$$

where B_0 is the magnetic field strength, μ_0 is the induction constant, k is the Boltzmann constant and T is the temperature.²⁶ In practice, RDCs are very sensitive to structural noise arising either from inaccuracies in the 3D structure coordinates or from bond movements, leading to smaller alignment tensors than expected.^{28,29} Systematic underestimation of the alignment tensor magnitude results in particular, if only $^1D_{\text{HN}}$ couplings are available and the alignment tensor fit is to a crystal structure with $> 1.5 \text{ \AA}$ resolution.²⁸ In a first approximation, we took these effects into account by an order parameter S of about 0.9.³⁰ In general, magnetic alignment of a two-domain protein by a paramagnetic lanthanide located in one of the domains provides a powerful way to assess its orientation relative to the other domain, as has been demonstrated in an exemplary manner with calmodulin.^{29,31-35}

In the following we show that the combined PCS and RDC data of T4-L WT* indicate a more open structure than the crystal structure of wild-type T4-L. This average conformation is similar to the states identified for T4-L fused to the crystallization module 2TEL¹⁴ and related to the structure identified by RDCs measured in multiple alignment media.²³ The fit was slightly improved by assuming a weighted average between an even more open state and the closed conformation. PREs were used to distinguish between the single-state and two-state models.

2 Experiments

2.1 Protein expression and purification

The C54T/C97A/S44C triple mutant of T4-L (in this work referred to as T4-L S44C) was cloned into the PET3a vector and expressed in *E. coli*. ^{15}N -labelled protein was prepared by growing cells in M9 medium with $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source. The protein was first purified using a DEAE column, and the low-salt fractions containing target protein were collected and concentrated. Pure protein was obtained using a SP-Sepharose column, followed by gel-filtration. The protein yield was about 20 mg of purified protein per litre M9 medium.

2.2 Synthesis of the 4MMDPA tag and ligation

The 4MMDPA tag (Fig. S1) was synthesized according to the published protocol.²⁴ The tag was ligated to the protein as described previously³⁶ with minor modification. First, the mutant T4-L S44C was activated with ten equivalents of Ellman's reagent, 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB), in 20 mM Tris-HCl at pH 7.2. Next, the resulting mixture was incubated at room temperature for about two hours, after which the excess of reagent was removed by a PD-10 column. 4MMDPA was added in three-fold excess to a 0.2 mM solution of T4-L S44C-TNB in 20 mM Tris-HCl and the pH was adjusted to about 7.2. After incubation of the above mixture for three hours, a cation-exchange SP-Sepharose column was used to remove unligated protein and any free tag. The pure ligation product was obtained in about 65% yield.

2.3 Protein NMR measurements

All NMR experiments were performed at 25 °C in 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer (pH 6.5) on a Bruker AV 600 NMR spectrometer equipped with a QCI cryoprobe. A 3D NOESY-¹⁵N-HSQC spectrum (80 ms mixing time) was recorded of a 0.7 mM solution of T4-L S44C in 90% H₂O and 10% D₂O. ¹⁵N-HSQC spectra were recorded of solutions of 0.1 mM T4-L S44C-4MMDPA loaded with one equivalent of diamagnetic Y³⁺ or paramagnetic lanthanide (Tb³⁺, Tm³⁺ or Yb³⁺).

¹D_{HN} RDCs induced by Tb³⁺ and Tm³⁺ were measured on a 600 MHz NMR spectrometer using the IPAP pulse scheme,³⁷ using $t_{1\max} = 95$ ms and $t_{2\max} = 130$ ms. The spectra were recorded of a 0.1 mM solution of T4-L S44C-4MMDPA in complex with one equivalent of paramagnetic Tm³⁺ or diamagnetic Y³⁺, respectively.

2.4 Determination of the $\Delta\chi$ tensors

$\Delta\chi$ tensors were determined using the PCSs of the backbone amide protons observed with Tb³⁺, Tm³⁺ and Yb³⁺ by fitting to the N-domain (residues 14-65) of the crystal structure of wild-type T4-L (2LZM), using a common metal ion position for all three sets of PCSs. At the same time, the fit took the RDCs into account by assuming that the alignment tensor is directly proportional to the $\Delta\chi$ tensor except for an order parameter $S = 0.9$ to

allow for disorder due to N-H bond librations. The fit was performed using the program PyParaTools.³⁸ The $\Delta\chi$ tensor parameters and metal position determined for the N-domain were then used to back-calculate PCSs and RDCs of all spins in the set of 572 previously reported T4-L conformations (Table S3), which were aligned to the structure 2LZM by superimposing the N-domains. The same metal position was used to calculate PREs. The quality of the tensor fits were assessed by Q -factors calculated as

$$Q = \sqrt{\frac{\sum(PCS^{obs} - PCS^{calc})^2}{\sum(PCS^{obs})^2}} \quad (2)$$

where PCS^{obs} and PCS^{calc} are the observed and back-calculated PCS values, respectively.

2.5 Ensemble analysis

A mixture model from pairs and triplets of structures was applied to simulate PCSs experienced by T4-L S44C in solution. The PCSs of T4-L conformations in solution were estimated by a simple n -component mixture model

$$P_{mix}^j = \sum_{i=1}^n w_i P_i^j$$

where P_i^j is the PCS value of spin j in state i , and w_i denotes the populations of states i ,

which are constrained such that the total population $\sum_{i=1}^n w_i = 1$.

2.6 PRE analysis

PRE measurements were performed following a published protocol.³⁹ ^{15}N -HSQC spectra were recorded of a 0.1 mM solution of T4-L S44C-4MMDPA in complex with one equivalent of paramagnetic Gd^{3+} or diamagnetic Y^{3+} , respectively. Peak intensities were measured as peak heights obtained by line fitting. The intensity ratios of paramagnetic versus diamagnetic cross-peaks were normalized by comparison with the peak intensities of amide protons located further than 43 Å from the paramagnetic centre. The enhancement of the transverse relaxation rate of an amide proton, Γ_2 , can be described by

$$\Gamma_2 = R_2^{para} - R_2^{dia} \quad (4)$$

$$\frac{I_{para}}{I_{dia}} = \frac{R_2^{dia} \exp(-\Gamma_2 t)}{R_2^{para}} \quad (5)$$

where R_2^{para} and R_2^{dia} are the transverse relaxation rates of the amide proton in the presence of Gd^{3+} and Y^{3+} , respectively, and I_{para} and I_{dia} are the cross-peak heights observed for the paramagnetic and diamagnetic protein samples, respectively. The total duration t of the INEPT delays was 9 ms in the ^{15}N -HSQC experiments. R_2^{dia} values were estimated from the line widths observed in the ^{15}N -HSQC spectrum of the diamagnetic sample. In measuring the line widths, $^3J_{HN}$ coupling constants were neglected as all line widths were at least 13 Hz.

The metal position found in the N-domain by fitting $\Delta\chi$ tensors using PCSs was used to back-calculate PREs induced by Gd^{3+} using

$$\Gamma_2 = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{15r^6} (4\tau_c) \quad (6)$$

where μ_0 is the permeability of vacuum, γ_I the proton gyromagnetic ratio, μ_B the electron Bohr magneton, g_e the electron g factor ($g_e = 2$ for Gd^{3+}), S the electron spin quantum number ($S = 7/2$ for Gd^{3+}), r the distance between the paramagnetic centre and the proton, and τ_c the effective correlation time ($\tau_c^{-1} = \tau_s^{-1} + \tau_r^{-1}$, where τ_r is the estimated rotational correlation time of the protein, and τ_s is the estimated electron relaxation time for Gd^{3+}).⁴⁰ While the full equation describing Γ_2 contains additional dispersive terms that depend on the Larmor frequencies of the nuclear and electronic spins,⁴¹ these are neglected in equation 6 because, for $\tau_s \gg \tau_r$, these terms are at least 40 times smaller than the term shown. τ_r has been measured for a T4-L mutant to be 10.8 ns at 25 °C⁴² and τ_s of Gd^{3+} is assumed to be similar⁴³ or longer (tens or hundreds of nanoseconds) in high-field NMR magnets.⁴⁴ Using $\tau_r = 10.8$ ns and assuming $\tau_s \gg \tau_r$, equation 6 can thus be written as

$$\Gamma_2 = K/r^6 \quad (7)$$

with $K = 11.1 \times 10^9 \text{ \AA}^6/\text{s}$.

3 Results

3.1 Protein ligation

The T4-L S44C-4MMDPA construct produced NMR spectra of high quality (Fig. 2).

Compared to the unligated protein, no significant chemical shift changes were observed for the protein residues except for those close to residue 44. The NOEs observed in a 3D NOESY- ^{15}N -HSQC spectrum confirmed that the introduction of the lanthanide tag did not significantly perturb the structure of the protein.

3.2 $\Delta\chi$ -tensor determinations

Titration of paramagnetic ions (Tb^{3+} , Tm^{3+} or Yb^{3+}) into solutions of T4-L S44C-4MMDPA generated significant PCSs for many of the amide cross-peaks (Fig. 2). The cross-peaks of the paramagnetic species increased with increasing concentration of lanthanide ion, indicating slow exchange between the metal in the protein-bound and free states.

Large PCSs were observed for the residues of the N-domain, whereas the residues of the C-domain experienced smaller PCSs (Table S1 and Fig. S2). Smaller PCSs are expected due to the longer distance of the C-domain from the tagging site (Fig. 1). PRE-induced line broadening prevented the measurement of RDCs for many amide resonances in the N-domain in the presence of Tb^{3+} , while adding large uncertainties to RDCs measured with Tm^{3+} (Table S2). Nonetheless, the RDCs measured for the C-domain were clearly of comparable size as those of the N-domain (Table S2). Similarly, using the PCSs to fit $\Delta\chi$ tensors to the individual N- or C-domains yielded tensors of similar magnitude (Table 1). If the motion of the C-domain relative to the N-domain were much less restricted, akin to the situation in, e.g., calmodulin,³¹ the averaging would have substantially reduced the $\Delta\chi$ tensors and the RDCs measured for the C-domain. The apparent absence of such scaling indicates that any hinge-bending motion between N- and C-domains is limited in amplitude, suggesting that the limited conformational space sampled by the crystal structures of different T4-L mutants is representative of the situation in solution.

$\Delta\chi$ tensors were fitted first to the N-domain excluding the data from the N-terminal α -helix, which, in the crystal structures, tends to follow the movement of the C-domain (Fig. 1). The $\Delta\chi$ -tensor fit to the N-domain used the PCSs measured with Tb^{3+} , Tm^{3+} and Yb^{3+} with a single common metal position, thus determining the coordinates of the paramagnetic centre. The fits simultaneously took the RDCs into account by assuming direct proportionality between the $\Delta\chi$ and alignment tensors, except that the alignment

tensor was assumed to be scaled by an order parameter $S = 0.9$ to account for structural noise in amide bond orientations.

The $\Delta\chi$ tensors determined for the N-domain (Table 1) were similar in magnitude to those determined previously for ArgN-4MMDPA, for which the lanthanide position is restrained by additional coordination to a carboxyl group of the protein.²⁴ Immobilization of the lanthanide by additional coordination seems to occur also in T4-L S44C-4MMDPA, as the $\Delta\chi$ tensor fit positioned the lanthanide ion within 2.4 Å of the side-chain carboxyl group of Glu45. Immobilization of the lanthanide in the T4-L S44C-4MMDPA construct is important, as flexibility of a paramagnetic tag can compromise the prediction of PCSs close to the paramagnetic centre, if the motions change the coordinates of the metal ion.⁴⁶

3.3 Fitting of PCSs and RDCs to different crystal structures of T4-L

The tensors fitted to the N-domain allowed the prediction of PCSs and RDCs for the entire protein. Among 572 conformations in crystal structures of wild-type and mutant T4-L in the Protein Data Bank (Table S3), the best-fitting structure proved to be the multi-site mutant G28A/I29A/G30A/C54T/C97A (PDB code 1SSY, chain B),⁴⁷ with a Q -factor of 5.6% for the fit over all PCS data from the three lanthanide ions. Apart from the standard WT* mutations to replace the cysteine residues, the mutations in this structure are located in the hinge between the two domains. This result indicates that the average structure of WT* is more open in solution than the crystal structure of the wild-type protein (2LZM; Fig. 3a) and also slightly more open than the structure of WT* determined earlier to be in best agreement with RDCs (150L, conformer C; Fig. 3b).²³ The structure comparison with the T4-L module fused to 2TEL¹⁴ is more difficult, because the scarcity of crystal contacts blurred the electron density, resulting in a distorted model of the N-domain. Overall, however, the degree of opening of the active site is comparable to the structure of 1SSY (Fig. 3b).

3.4 Ensemble structural analysis

The fit between experimental and back-calculated PCSs can be improved by assuming equilibrium between different conformations. To find the best-fitting pair of structures among all available crystal structures of wild-type and mutant T4-L, we repeated the

fitting using a 2-state model for all 163306 possible pair-wise combinations of 572 T4-L structures in the PDB (Table S3). The best-fitting pair of structures comprised the I3P mutant (PDB ID 1L97, conformer B; Fig. 1) and the A98V/V149I/T152S mutant (PDB ID 1L51⁴⁸), with a population weighting of 55% to 45%. 1L97 is in a wide-open conformation, whereas 1L51 is in the closed conformation of the wild-type protein 2LZM (r.m.s.d. <0.2 Å). Fig. 1 shows that the structures 1L97 and 2LZM are more open and closed, respectively, than the structure 1SSY identified as the best-fitting single conformation (Fig. 3a).

While the 2-state model yielded a better fit of the paramagnetic data of the C-domain than any of the crystal structures alone, the Q -factor improved only marginally over the representation by the structure 1SSY (4.7% versus 5.7%; Fig. 4). Likewise, the 2-state model did not produce a significantly improved fit of back-calculated versus experimental RDCs compared with the structure 1SSY (Fig. 5). Finally, no significant improvement in the fit was obtained for a 3-state model that included all possible (over 46 million) combinations of three out of the 572 T4-L coordinate sets of Table S3 with variable populations scanned in steps of 0.1 (10%). The PCS and RDC data thus indicate that the open conformation of 1SSY is a good representation of the average structure in solution, with the possibility of fast exchange between an even more open conformation (1L97) and the closed state (1L51).

3.5 Conformational equilibrium probed by PREs

The very strong distance dependence of PREs allows detection of little populated conformational states if they involve short distances between the nuclear spins and the paramagnetic centre.⁴⁹ In a protein undergoing fast conformational exchange, the experimentally measured PREs are averages of the PREs of all conformational species in solution, but the average is heavily biased towards the states with short distances to the paramagnetic centre.^{43,51-53}

The PRE data measured for the Gd³⁺ complex of T4-L S44C-4MMDPA showed a good fit to calculated values for the structure 1SSY, in particular in the vicinity of residue 108, which is sensitive to the overall conformational state of the enzyme (Fig. 6a). While the PREs do not allow distinction between the very open conformation 1L97 or the less

open conformation 1SSY, the closed state is predicted to produce large PREs for the polypeptide segment containing residue 108 (Fig. 6b). As these were not observed, an equilibrium between approximately equal populations of very open conformation (represented by 1L97) and the closed state (as represented by 1L51) appears unlikely. We note that the different structures also predict large PREs in the vicinity of residue 70. The magnitude of these predicted PREs, however, does not report on the openness of the conformation, as a smaller PRE is predicted for, e.g., residue 70 in the intermediate conformation 1SSY than for the very open and closed conformations 1L97 and 1L51 (Fig. 6). Regardless, the PRE results do not indicate a significant population of the closed state.

4 Discussion

4.1 Conformational space in crystals and in solution

Hundreds of conformations have been determined of T4-L and its mutants in the crystalline state.⁸ These data indicate that the protein readily undergoes a hinge-bending motion in which the N- and C-domain behave as structurally conserved entities. The hinge-bending angles between the N- and C-domains span a 50° range between the different mutants.^{8,10} Despite the wealth of atomic-resolution information, however, it is difficult to disentangle the effects of the mutations, the crystal environment, and the crystallisation conditions on the hinge bending of the enzyme. The best approximation to the solution conditions was obtained by fusion of T4-L to the crystallization module 2TEL, which left the N-domain without crystal contacts.¹⁴ The model 2QAR built to fit the low-resolution electron density observed in the crystal, however, displays very large *B*-factors particularly for the N-domain, which appears distorted compared to other T4-L crystal structures. Nonetheless, the structure 2QAR is in remarkably good agreement with the open conformation 1SSY, which we identified as the best-fitting single structure with regard to PCS and RDC data (Fig. 3b).

Earlier EPR measurements also indicated a preferentially open conformation but associated with conformational heterogeneity. The structures of the different states, however, could not be elucidated, and the structural impact of the nitroxide tags at different sites with engineered cysteine residues was unclear.¹⁵ In general, flexibility of

conventional nitroxide tags allows alternative interpretations of the EPR data arising from changes in nitroxide tag conformation.^{54,55} FRET measurements are subject to similar problems.

4.2 Analysis of conformational space by paramagnetic NMR spectroscopy

The present study illustrates the structural information that can be gleaned from a single paramagnetic lanthanide binding tag, even when the protein may populate different conformations that exchange so rapidly with each other that only average NMR observables can be measured. The PCSs and RDCs show that WT* assumes on average a predominantly open conformation, which is slightly more open than the conformation 150L found previously to match RDCs measured in four different alignment media,²³ and in excellent agreement with the conformation found in the crystal structure 2QAR, in which the N-domain is free of crystal contacts (Fig. 3b).

In contrast to FRET studies, where a large range of uniformly populated tag conformations facilitates the prediction of measurable distances, PCS data are easier to interpret when the tag is immobilized. In the present study, the paramagnetic centre was constructed from a lanthanide ion that was held in place by a small lanthanide-binding tag and by vicinity to a carboxyl side chain that occurs naturally in the wild-type protein. Based on previous experience with DPA tags,^{24,56,57} the magnitudes of the $\Delta\chi$ tensors suggested the absence of significant tensor averaging due to lanthanide mobility. Notably, even if the metal were mobile, local reorientation of the metal complex would preserve the relation between alignment and $\Delta\chi$ tensor (Eq. 8). To break the relationship of Eq. 8 requires motions causing a significant change in metal coordinates.⁴⁶

In summary, the combined PCSs and RDCs strongly suggest that, on average, WT* occupies an open conformation in solution that is more open than the crystal structure of the wild-type protein. The PREs (Fig. 6) are in agreement with this model. In contrast to PCSs, however, which do not report on intermolecular interactions unless the protein molecules assume a preferential orientation relative to each other as in, *e.g.*, a dimer, PREs are more difficult to assess. Intermolecular PREs can be pronounced, excess Gd^{3+} ions (for example due to incomplete yields in tag ligation) can bind non-specifically and thus contribute to the PRE, and a sub-stoichiometric amount of Gd^{3+} ions would lead to

cross-peaks that contain both paramagnetic and diamagnetic components, making the PRE measurement more difficult. As PREs generated by Gd^{3+} ions are very large, a larger domain than the N-domain of T4-L would be required for calibration against known metal-proton distances to eliminate the unknown electron relaxation time τ_s as a variable. In contrast to PREs, PCSs generate resolved peaks for the paramagnetic species, which can readily be measured even with incomplete tag ligation or incomplete titration with metal ion, and the magnitude of the associated PREs can be tuned by the choice of lanthanide.

4.3 Population of the closed conformation

Our PCS and RDC data could be fitted to either a single structure (1SSY), which is in an open conformation, or to a population-weighted 2-state model comprising the very open conformation 1L97 and the closed conformation 1L51, but the improvement achieved by the 2-state model was marginal. Fitting to three crystal structures did not yield any significant improvement. The result is reminiscent of the attempt by Kay and co-workers attempted to improve the fit of RDCs, which had been generated by alignment media, using a linear combination of a closed conformation (PDB code 3LZM) and a more open structure (PDB code 172L). While the fit allowed 50% population of the closed state, the overall improvement of the fit over a fit with a single structure was insignificant.²³ In our case, the PREs showed no signature of the closed state, which is in agreement with the average conformation found in a crystal structure, in which the N-domain made no crystal contacts.¹⁴ Population of a narrow range of open conformations is also in agreement with the results from single molecule studies performed with single-walled carbon nanotube field-effect transistor devices, which revealed a much narrower range of structural fluctuations in the absence than in the presence of substrate.^{17,18}

5 Conclusion

The present study shows that PCSs generated by paramagnetic lanthanides provide a powerful tool for probing the average structure and accessible conformational space of a protein in solution. In contrast to RDC data generated by steric or electrostatic alignment in dilute liquid crystals, the approach is not potentially affected by interactions with the

liquid crystalline media. With the advent of different technologies for rigid site-specific labelling of proteins with lanthanide ions, paramagnetic NMR spectroscopy is set to become a prime tool for the analysis of conformational changes accompanying enzyme function.

Acknowledgements

Financial support by the 973 program (grant 2013CB910200), the National Science Foundation of China (grants 21073101 and 21273121), and the Australian Research Council is greatly acknowledged.

References

- 1 K. Henzler-Wildman and D. Kern, *Nature*, 2007, **150**, 964–972.
- 2 R. K. Allemann, R. M. Evans and E. J. Loveridge, *Biochem. Soc. Trans.*, 2009, **37**, 349–353.
- 3 I. Bahar, T. R. Lezon, L. W. Yang and E. Eyal, *Annu. Rev. Biophys.*, 2010, **39**, 23–42.
- 4 S. C. Kamerlin and A. Warshel, *Proteins*, 2010, **78**, 1339–1375.
- 5 S. R. Tzeng and C. G. Kalodimos, *Curr. Opin. Struct. Biol.*, 2011, **21**, 62–67.
- 6 V. C. Nashine, S. Hammes-Schiffer and S. J. Benkovic, *Curr. Opin. Chem. Biol.*, 2010, **14**, 655–651.
- 7 R. Nussinov, C. J. Tsai and B. Ma, *Annu. Rev. Biophys.*, 2013, **42**, 169–189.
- 8 W. A. Baase, L. Liu, D. E. Tronrud and B. W. Matthews, *Protein Sci.*, 2010, **19**, 631–641.
- 9 B. W. Matthews and S. J. Remington, *Proc. Natl. Acad. Sci. USA*, 1974, **71**, 4178–4182.
- 10 X. J. Zhang, J. A. Wozniak and B. W. Matthews, *J. Mol. Biol.*, 1995, **250**, 527–552.
- 11 M. M. Dixon, H. Nicholson, L. Shewchuk, W. A. Baase and B. W. Matthews, *J. Mol. Biol.*, 1992, **227**, 917–933.
- 12 L. H. Weaver and B. W. Matthews, *J. Mol. Biol.*, 1987, **193**, 189–199.
- 13 R. Kuroki, L. H. Weaver and B. W. Matthews, *Science*, 1993, **262**, 2030–2033.
- 14 S. Nauli, S. Farr, Y.-J. Lee, H.-Y. Kim, S. Faham and J. U. Bowie, *Protein Sci.*, 2007,

- 16, 2542–2551.
- 15 H. S. Mchaourab, K. J. Oh, C. J. Fang and W. L. Hubbell, *Biochemistry*, 1997, **36**, 307–316.
- 16 Y. Chen, D. Hu, E. R. Vorpapel and H. P. Lu, *J. Phys. Chem. B*, 2003, **107**, 7947–7956.
- 17 Y. Choi, I. S. Moody, P. C. Sims, S. R. Hunt, B. L. Corso, I. Perez, G. A. Weiss and P. G. Collins, *Science*, 2012, **335**, 319–324.
- 18 Y. Choi, G. A. Weiss and P. G. Collins, *Phys. Chem. Chem. Phys.*, 2013, **15**, 14879–14895.
- 19 R. B. Yirdaw and H. S. Mchaourab, *Biophys. J.*, 2012, **103**, 1525–1536.
- 20 G. E. Arnold, J. I. Manchester, B. D. Townsend and R. L. Ornstein, *J. Biomol. Struct. Dyn.*, 1994, *12*, 457–474.
- 21 G. E. Arnold and R. L. Ornstein, *Biopolymers*, 1997, **41**, 533–544.
- 22 B. L. de Groot, S. Hayward, D. M. F. van Aalten, A. Amadei and H. J. C. Berendsen, *Proteins*, 1998, **31**, 116–127.
- 23 N. K. Goto, N. R. Skrynnikov, F. W. Dahlquist and L. E. Kay, *J. Mol. Biol.*, 2001, **308**, 745–764.
- 24 X.-C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle and G. Otting, *J. Am. Chem. Soc.*, 2008, **130**, 10486–10487.
- 25 M. Matsumura and B. W. Matthews, *Science*, 1989, **243**, 792–794.
- 26 I. Bertini, C. Luchinat and G. Parigi, *Prog. NMR Spectrosc.*, 2002, **40**, 249–273.
- 27 J. R. Tolman, J. M. Flanagan, M. A. Kennedy and J. H. Prestegard, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 9279–9283.
- 28 M. Zweckstetter and A. Bax, *J. Biomol. NMR*, 2002, **23**, 127–137.
- 29 L. Russo, M. Maestre-Martinez, S. Wolff, S. Becker and C. Griesinger, *J. Am. Chem. Soc.*, 2013, **135**, 17111–17120.
- 30 O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger and F. L. de Groot, *Science*, 2008, **320**, 1471–1475.
- 31 I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani and M. A. Zoroddu, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 6841–6846.

- 32 I. Bertini, A. Giachetti, C. Luchinat, G. Parigi, M. V. Petoukhov, R. Pierattelli, E. Ravera and D. I. Svergun, *J. Am. Chem. Soc.*, 2010, **132**, 13553–13558.
- 33 S. Dasgupta, X. Hu, P. H. J. Keizers, W. M. Liu, C. Luchinat, M. Nagulapalli, M. Overhand, G. Parigi, L. Sgheri and M. Ubbink, *J. Biomol. NMR*, 2011, **51**, 253–263.
- 34 M. Nagulapalli, G. Parigi, J. Yuan, J. Gsponer, G. Deraos, V. V. Bamm, G. Harauz, J. Matsoukas, M. R. R. de Planque, I. P. Gerothanassis, M. M. Babu, C. Luchinat and A. G. Tzakos, *Structure*, 2012, **20**, 522–533.
- 35 W. Andrałojć, C. Luchinat G. Parigi and E. Ravera, *J. Phys. Chem. B*, 2014, **118**, 10576–10587.
- 36 X.-C. Su, T. Huber, N. E. Dixon and G. Otting, *ChemBioChem*, 2006, **7**, 1599–1604.
- 37 M. Ottiger, F. Delaglio and A. Bax, *J. Magn. Reson.*, 1998, **131**, 373–378.
- 38 M. Stanton-Cook, X.-C. Su, G. Otting and T. Huber, <http://compbio.anu.edu.au/mscook/PPT>
- 39 J. L. Battiste and G. Wagner, *Biochemistry*, 2000, **39**, 5355–5365.
- 40 M. Benmelouka, A. Borel, L. Morriggi, L. Helm and A. E. Merbach, *J. Phys. Chem. B*, 2007, **111**, 832–840.
- 41 I. Solomon, *Phys. Rev.*, 1955, **99**, 559–565.
- 42 N. R. Skrynnikov, F. A. Mulder, B. Hon, F. W. Dahlquist and L. E. Kay, *J. Am. Chem. Soc.*, 2001, **123**, 4556–4566.
- 43 D. H. Powell, O. M. Nidhubghaill, D. Pubanz, L. Helm, Y. S. Lebedev, W. Schlaepfer and A. E. Merbach, *J. Am. Chem. Soc.*, 1996, **118**, 9333–9346.
- 44 I. Bertini, C. Luchinat, M. Nagulapalli, G. Parigi and E. Ravera, *Phys. Chem. Chem. Phys.*, 2012, **14**, 9149–9156.
- 45 C. Schmitz, M. J. Stanton-Cook, X.-C. Su, G. Otting and T. Huber, *J. Biomol. NMR*, 2008, **41**, 179–189.
- 46 D. Shishmarev and G. Otting, *J. Biomol. NMR*, 2013, **56**, 203–216.
- 47 M. M. He, Z. A. Wood, W. A. Baase, H. Xiao and B. W. Matthews, *Protein Sci.*, 2004, **13**, 2716–2724.
- 48 S. Daopin, T. Alber, W. A. Baase, J. A. Wozniak and B. W. Matthews, *J. Mol. Biol.*, 1991, **221**, 647–667.
- 49 X. J. Zhang and B. W. Matthews, *Protein Sci.*, 1994, **3**, 1031–1039.

- 50 G. M. Clore, *Mol. BioSyst.*, 2008, **4**, 1058–1069.
- 51 C. Tang, C. D. Schwieters and G. M. Clore, *Nature*, 2007, **449**, 1078–1082.
- 52 Q. Bashir, A. N. Volkov, G. M. Ullmann and M. Ubbink, *J. Am. Chem. Soc.*, 2010, **132**, 241–247.
- 53 N. J. Anthis, M. Doucleff and G. M. Clore, *J. Am. Chem. Soc.*, 2011, **133**, 18966–18974.
- 54 R. Langen, K. J. Oh, D. Cascio and W. L. Hubbell, *Biochemistry*, 2000, **39**, 8396–8405.
- 55 Z. Zhang, M. R. Fleissner, D. S. Tipikin, Z. Liang, J. K. Moscicki, K. A. Earle, W. L. Hubbell and J. H. Freed, *J. Phys. Chem. B*, 2010, **114**, 5503–5521.
- 56 B. Man, X.-C. Su, H. Liang, S. Simonsen, T. Huber, B. A. Messerle and G. Otting, *Chem. Eur. J.*, 2010, **16**, 3827–3832.
- 57 X. Jia, A. Maleckis, T. Huber and G. Otting, *Chem. Eur. J.*, 2011, **17**, 6830–6836.

Table 1. $\Delta\chi$ -tensor parameters of T4-L S44C-4MMDPA in complex with Tb^{3+} , Tm^{3+} , or Yb^{3+} ^a

	Ln^{3+}	$\Delta\chi_{\text{ax}}$	$\Delta\chi_{\text{rh}}$	α	β	γ	Q [%]
N-domain ^b	Tb^{3+}	11.1 ± 0.1	3.9 ± 0.1	17	74	170	3.4
	Tm^{3+}	-9.4 ± 0.1	-3.7 ± 0.1	25	81	1	2.4
	Yb^{3+}	3.6 ± 0.1	2.4 ± 0.1	117	80	169	3.2
C-domain ^b	Tb^{3+}	10.8 ± 0.2	6.0 ± 0.2	26	91	171	12.8
	Tm^{3+}	-9.4 ± 0.2	-4.9 ± 0.3	31	98	175	18.3
	Yb^{3+}	-4.3 ± 0.3	-2.7 ± 0.2	39	103	6	30.8
N-domain ^c	Tb^{3+}	11.2 ± 0.4	3.9 ± 0.2	17	73	170	3.4
	Tm^{3+}	-9.4 ± 0.2	-3.7 ± 0.1	25	81	1	2.4
	Yb^{3+}	3.6 ± 0.1	2.4 ± 0.1	117	80	169	3.2

^a The $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ parameters are in units of 10^{-32} m^3 . The Euler angles α , β and γ are in degrees. The fits were performed to the crystal structure of the closed conformation (PDB code: 2LZM) using the program PyParaTools.³⁸ The fits to the N-domain used only the data of residues 14-65. The fits to the C-domain used the data of all other residues except for the helix connecting the N- and C-domain (residues 66-81). For comparison, the axial components of the $\Delta\chi$ tensors determined previously for the corresponding lanthanide complexes of ArgN-4MMDPA²⁴ were reported to be 12.9 ± 0.9 (Tb^{3+}), 12.2 ± 0.6 (Tm^{3+}), and 5.7 ± 0.4 (Yb^{3+}), respectively, in units of 10^{-32} m^3 . The error ranges of the axial and rhombic components of the $\Delta\chi$ tensors were derived from Monte-Carlo simulations, repeating the fits 100 times while randomly omitting 10% of the PCSs. The tensors are reported in the unique tensor representation.⁴⁵ The coordinates of the paramagnetic ion determined by the best fit to the N-domain in the structure 2LZM were $x = 37.736$, $y = 7.815$, $z = 25.425 \text{ \AA}$. Q -factors are reported for PCS data only.

^b Fits using PCSs only. The Q -factors of the fits to the C-domain are large because the fits used the structure 2LZM with the metal coordinates determined for the N-domain.

^c Fits using PCSs and RDCs simultaneously. The RDCs presenting the two largest outliers in the Tm^{3+} data (residues 25 and 53, Fig. 5a and c) were excluded from the fit; both residues are in polypeptide segments of irregular secondary structure. The coordinates of the paramagnetic ion used were those determined by the best fit to the N-domain in the structure 2LZM (see footnote a).

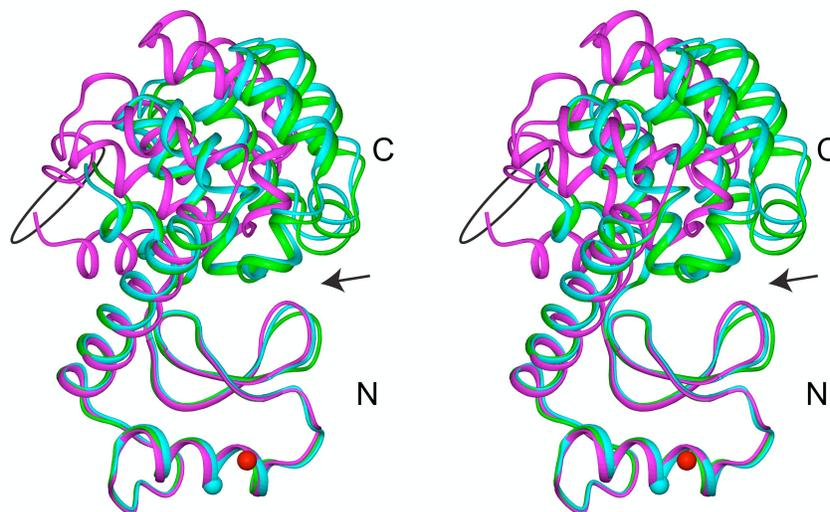


Fig. 1 Crystal structures of wild-type T4-L with and without bound substrate are very similar, whereas large-amplitude hinge-bending has been observed for mutants. The structures are displayed as stereo views of ribbon drawings of the backbones after superimposition of the N-domains. Superimposition excluded the N-terminal helix (N-termini marked by a circle) as it moves with the C-domain. The N- and C-terminal domains are marked and an arrow identifies the active-site cleft. Cyan: structure of the wild-type protein (PDB ID: 2LZM).¹² The cyan ball identifies the C^α atom of Ser44, which was mutated to cysteine in the present study to attach a lanthanide tag. The red sphere marks the position of the lanthanide. Green: structure of the active-site mutant T28E with bound substrate (PDB ID: 148L;¹³ substrate not shown). We refer to structures similar to 2LZM and 148L as closed conformation. Magenta: I3P mutant (PDB ID: 1L97, conformer B),¹¹ which is one of the most open conformations of T4-L crystallized.

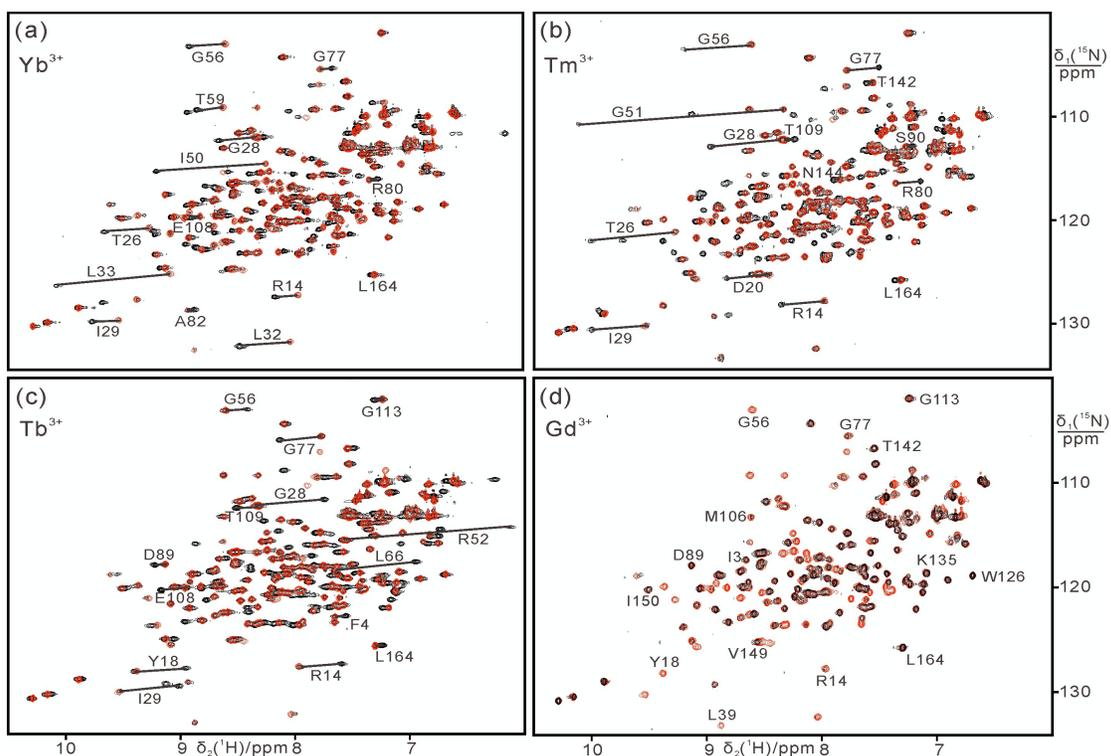


Fig. 2 Superimposition of ^{15}N -HSQC spectra of T4-L S44C-4MMDPA in complex with one equivalent of diamagnetic Y^{3+} (red) or paramagnetic lanthanide (black). The spectra were recorded of 0.1 mM protein solutions in 20 mM MES buffer, pH 6.5, at 298 K, using a 600 MHz NMR spectrometer. The resonance assignment of a number of resolved cross-peaks is indicated, and their PCSs are indicated by lines connecting the cross-peaks observed for the paramagnetic and diamagnetic samples. The paramagnetic ions were (a) Yb^{3+} , (b) Tm^{3+} , (c) Tb^{3+} , and (d) Gd^{3+} .

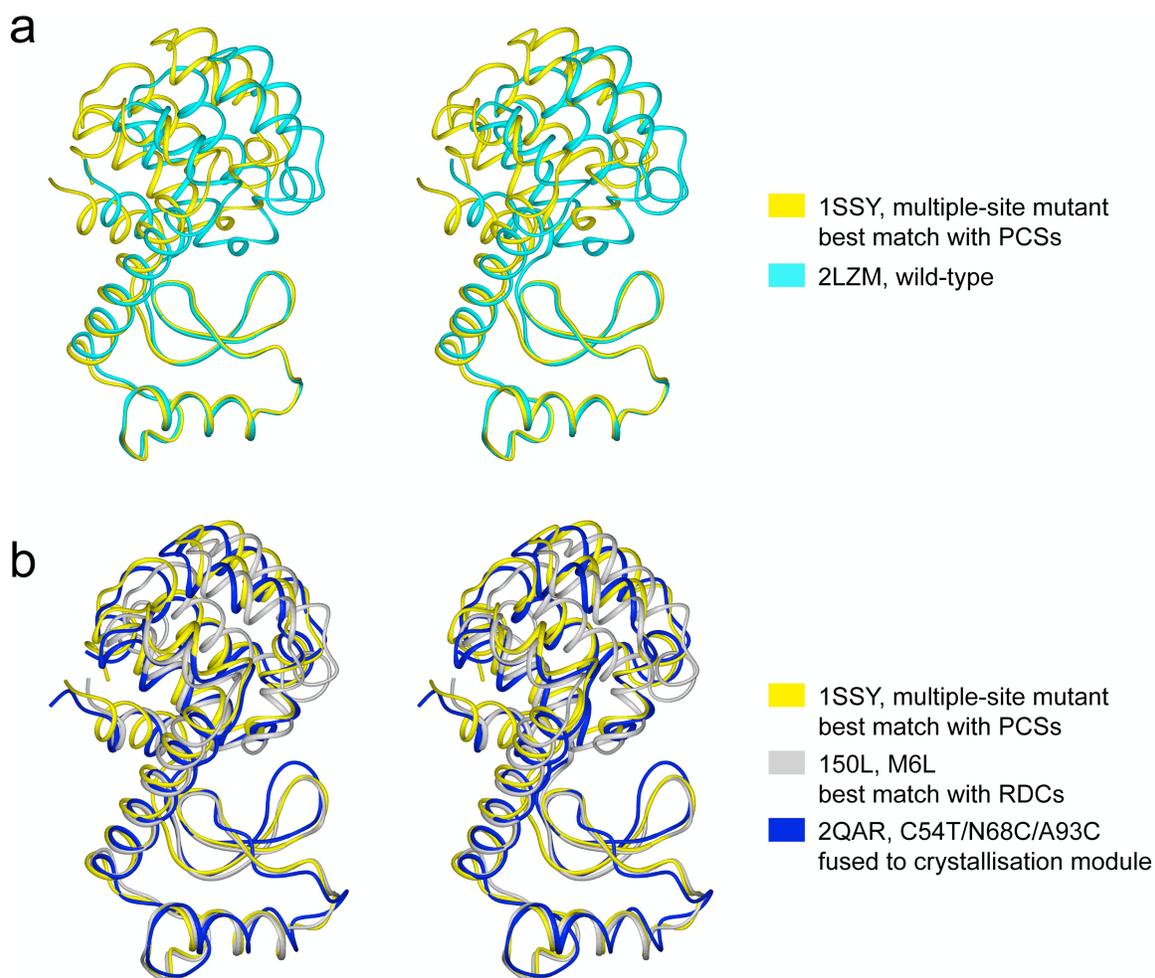


Fig. 3 Superimpositions of crystal structures of different T4-L wild-type and mutant constructs. The structures are shown as stereo views in the same orientation as in Fig. 1, following superimposition of the N-domains. (a) Yellow: multi-site mutant G28A/I29A/G30A/C54T/C97A (1SSY),⁴⁶ which best fulfils the combined PCS, RDC and PRE data. Cyan: closed state observed for wild-type T4-L without substrate (2LZM).¹² The multi-site mutant A98V/V149I/T152S (PDB ID 1L51)⁴⁷ is structurally very similar to 2LZM (RMSD of 0.2 Å). (b) Yellow: multi-site mutant G28A/I29A/G30A/C54T/C97A (1SSY). Grey: M6L mutant (150L)⁴⁸ that was earlier identified as the crystal structure best-fitting the RDCs measured for the wild-type protein by NMR in several alignment media.²³ This structure is more open than 2LZM. Blue: multiple-site mutant C54T/N68C/A93C in a fusion with the 2TEL crystallisation module (2QAR).¹⁴

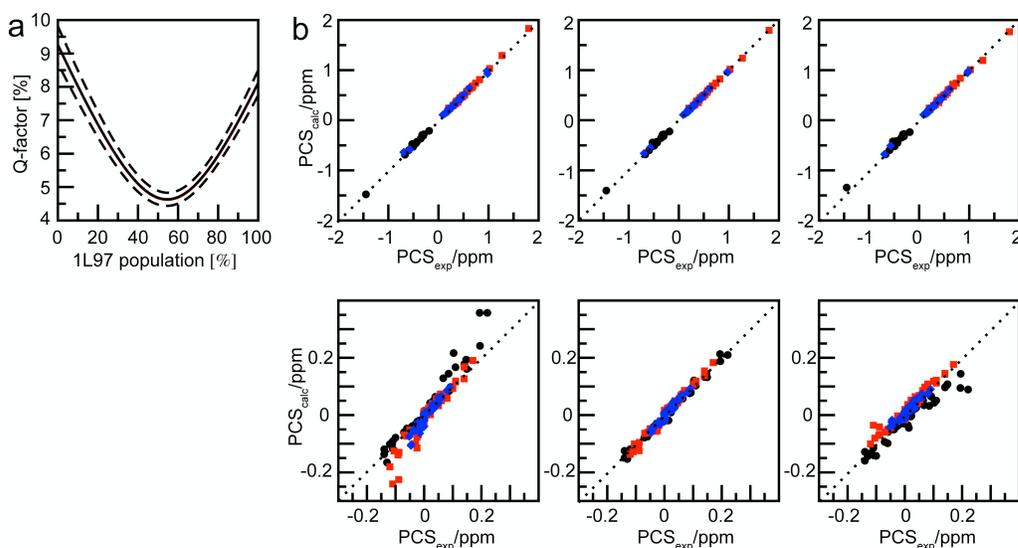


Fig. 4 Correlation of back-calculated and experimental PCSs obtained in a PCS+RDC fit using a pair of crystal structures of T4-L. The best-fitting pair comprised the I3P mutant 1L97 (conformer B), which is a wide-open conformation and the multi-site mutant 1L51, which is closely similar to the closed conformation 2LZM of wild-type T4-L (Fig. 1). (a) Q -factors calculated for the PCS data are plotted versus the percentage of more open conformation (1L97). The Q -factor was calculated for the combined N- and C-domains, excluding residues 66-81 of the interconnecting helix. The uncertainty band delineated by the dashed lines indicates the range of Q -factors obtained after randomly omitting 10% of the experimental data. (b) Correlation plots of the back-calculated versus experimental PCS values for population percentages of 0, 55 and 100% of 1L97 conformation in the binary mixture of 1L97 and 1L51 conformers (left to right). The plots in the first row display the correlations using PCSs of the N-domain (residues 14-65) only. These were used together with the RDCs of the N-domain to fit the $\Delta\chi$ tensors, which were subsequently used to back-calculate the PCSs of the first helix and C-domain. The correlations obtained in this way for the first helix (residues 1-13) and C-domain (residues 82-164) are displayed in the second row. PCS data of Tb³⁺, Tm³⁺, and Yb³⁺ are displayed by black, red and blue symbols, respectively. The estimated uncertainty in PCS measurements was ± 0.015 ppm.

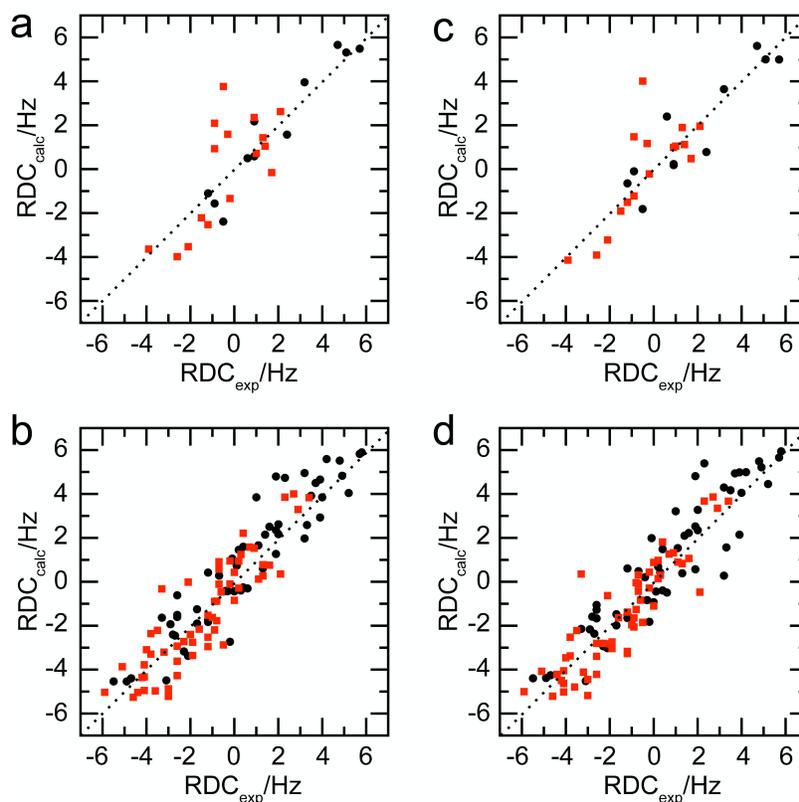


Fig. 5 Correlation plots of calculated RDCs versus the experimentally measured ^1H - ^{15}N RDCs (Table S2). Black and red data points represent data obtained with Tb^{3+} and Tm^{3+} ions, respectively. The RDCs were predicted using the tensor axes and axial and rhombic components of the $\Delta\chi$ -tensors determined from the PCS+RDC fit (Table 1), except that the alignment tensor was scaled by 0.9 to account for structural uncertainties. (a) RDCs of the N-domain (residues 14-65) of 1SSY (conformer B). (b) Same as (a), except for RDCs of the C-domain and N-terminal α -helix of 1SSY. (c) RDCs of the N-domain using a weighted average (55 and 45%, respectively) of the crystal structures 1L97 (conformer B) and 1L51 for back-calculation of the RDCs. (d) Same as (c), except for the RDCs of the C-domain and N-terminal α -helix of 1SSY.

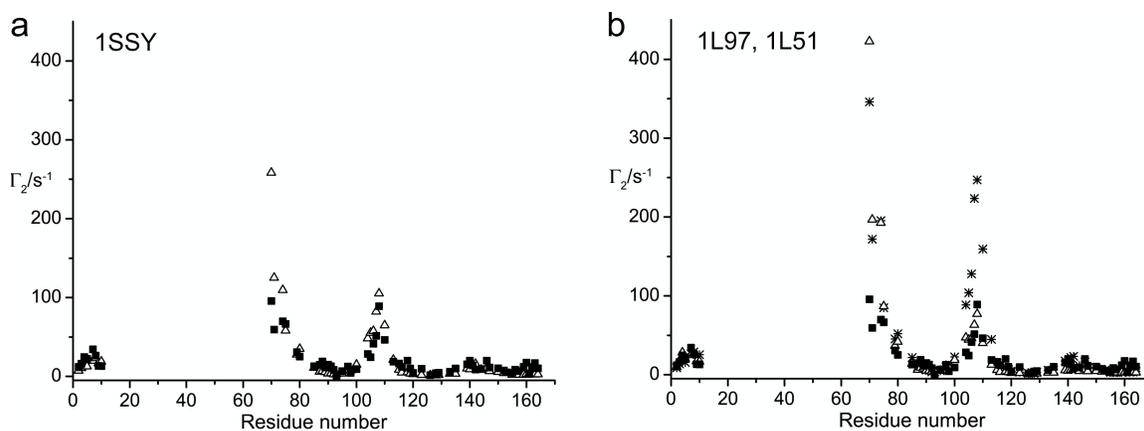


Fig. 6 Comparison of predicted and experimental PRE rates for different conformations of T4-L. (a) Filled squares: experimentally observed values; open triangles: values predicted based on the structure 1SSY, using equation 7 with $K = 11.1 \times 10^9 \text{ \AA}^6/s$. The metal coordinates were those determined by the $\Delta\chi$ -tensor fits to the crystal structure 2LZM, after superimposition of the N-domains of 2LZM and 1SSY. Due to their excessive size, no PREs could be measured for the N-domain. (b) Filled squares: experimentally observed values; open triangles: values predicted based on the wide-open structure 1L97. Stars: values predicted based on the closed conformation 1L51. The simultaneous fit of PCSs and RDCs suggested populations of 55% 1L97 and 45% 1L51, but this is not supported by the experimental PREs of residues near position 108.