LETTER

doi:10.1038/nature10349

Solution structure of a minor and transiently formed state of a T4 lysozyme mutant

Guillaume Bouvignies[1,2,3]*, Pramodh Vallurupalli[1,2,3]*, D. Flemming Hansen[1,2,3], Bruno E. Correia[4,5], Oliver Lange[4], Alaji Bah[6], Robert M. Vernon[4,6], Frederick W. Dahlquist[7], David Baker[4] & Lewis E. Kay[1,2,3,6]

Proteins are inherently plastic molecules, whose function often critically depends on excursions between different molecular conformations (conformers)[1–3]. However, a rigorous understanding of the relation between a protein's structure, dynamics and function remains elusive. This is because many of the conformers on its energy landscape are only transiently formed and marginally populated (less than a few per cent of the total number of molecules), so that they cannot be individually characterized by most biophysical tools. Here we study a lysozyme mutant from phage T4 that binds hydrophobic molecules[4] and populates an excited state transiently (about 1 ms) to about 3% at 25 °C (ref. 5). We show that such binding occurs only via the ground state, and present the atomic-level model of the 'invisible', excited state obtained using a combined strategy of relaxation-dispersion NMR (ref. 6) and CS-Rosetta[7] model building that rationalizes this observation. The model was tested using structure-based design calculations identifying point mutants predicted to stabilize the excited state relative to the ground state. In this way a pair of mutations were introduced, inverting the relative populations of the ground and excited states and altering function. Our results suggest a mechanism for the evolution of a protein's function by changing the delicate balance between the states on its energy landscape. More generally, they show that our approach can generate and validate models of excited protein states.

A detailed characterization of the conformers along a protein's energy landscape is important for understanding the structure–function relationship and also because such an analysis provides insight into fundamental aspects of protein structure and dynamics. In this vein, numerous detailed studies of mutant lysozymes and lysozyme complexes from phage T4 have greatly increased our understanding of the inter-relation between structure, stability, folding and motion in proteins[8]. Among the approximately 700 mutant lysozymes and lysozyme complexes that have been characterized is a family where each member contains an engineered cavity in its hydrophobic core, generated by replacing larger amino acids with alanine (ref. 9). The point mutant causing the most pronounced stability change involved the replacement of a leucine at position 99 (referred to in what follows as L99A T4L), creating a cavity of ~150 Å^3 in the carboxy terminus of the enzyme that is able to bind hydrophobic ligands[4]. Interestingly, X-ray studies showed that the L99A mutant undergoes the least rearrangement at the site of mutation, with the structure essentially unchanged[9].

Despite the fact that the wild-type and L99A T4L structures are virtually identical in the crystalline state, solution NMR studies of the L99A mutant indicated that many of the peaks were significantly broadened relative to the corresponding resonances in data sets recorded of the wild-type protein[5]. Spectral broadening is indicative of dynamics on the microsecond–millisecond timescale[6], and in this case provides a clear indication that cavity creation introduces one or more dynamic processes that are not observed in the wild-type enzyme. Such dynamics can be studied by NMR transverse spin-relaxation experiments, in which the relaxation rates of probe nuclei are measured as a function of the strength of applied radio-frequency fields[6,10]. These experiments provide a powerful approach to quantify structural transitions in proteins because they are sensitive to microsecond–millisecond exchange processes in which a highly populated ground state (G) interconverts with conformers that can have much lower populations (>0.5%), referred to in what follows as excited states (E).

Initial ^{15}N Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion NMR experiments indicated that L99A T4L undergoes a dynamic process involving residues that are proximal to the cavity. The relaxation data were well fitted to a model in which a highly populated ground state (97%, 25 °C) interconverts with a second state that because of its low population (3%) and short lifetime (~1 ms) is 'invisible' in NMR spectra[5]. Using recently developed CPMG dispersion experiments[11], we have obtained nearly all of the backbone ^{1}H, ^{15}N and ^{13}C chemical shifts—as well as side-chain methyl ^{13}C chemical shifts—of the invisible excited state with a high level of accuracy (Supplementary Fig. 1, Supplementary Tables 1–5). Such chemical shifts are powerful constraints in structure calculations; when combined with computational protocols[7,12] they can be used to calculate accurate folds of small proteins, even in the absence of additional information, such as internuclear distances[13,14].

A comparison of the chemical shifts of the ground and excited states shows that conformational rearrangements occur in the vicinity of the cavity involving the C-terminal region of helix E and helices F, G, H and I (Fig. 1a, b). These regions do not become disordered in the excited state, as calculated squared order parameters reporting on the amplitudes of backbone motion from chemical shifts[15], S^2_{RCI}, change little between states (Fig. 1c). However, a decrease in helix propensity is noted for the C-terminal region of helix E, with a very significant concomitant increase in the helix content for the loop connecting helices F and G (Fig. 1d).

The ^{15}N, $^1H^N$, $^1H^\alpha$, $^{13}C^\alpha$ and $^{13}C'$ chemical shifts of the excited state were used to guide Rosetta 'loop' building and refinement[16] to generate structural models of the excited state (described in Supplementary Information, Supplementary Fig. 2). Only regions with significant chemical shift changes (residues 100–120, 132–146; Fig. 1a) were allowed to deviate from the X-ray structure of the L99A cavity mutant. The CS-Rosetta-based excited state conformers so produced are well converged, with pair-wise backbone root-mean-squared deviations (r.m.s.d.) for ten representative, low energy structures of 0.7 ± 0.2 Å over the region that was allowed to vary in the calculations (Supplementary Table 6). As a control, an identical protocol was used to generate the structure of the ground state in the vicinity of the cavity mutant, based on the same number of chemical shifts as for the excited

[1]Department of Molecular Genetics, The University of Toronto, Toronto, Ontario M5S 1A8, Canada. [2]Department of Biochemistry, The University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]Department of Chemistry, The University of Toronto, Toronto, Ontario M5S 1A8, Canada. [4]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [5]Program in Computational Biology, Instituto Gulbenkian de Ciência, P-2780-156 Oeiras, Portugal. [6]Hospital for Sick Children, Program in Molecular Structure and Function, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [7]Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, California 93106, USA.
*These authors contributed equally to this work.

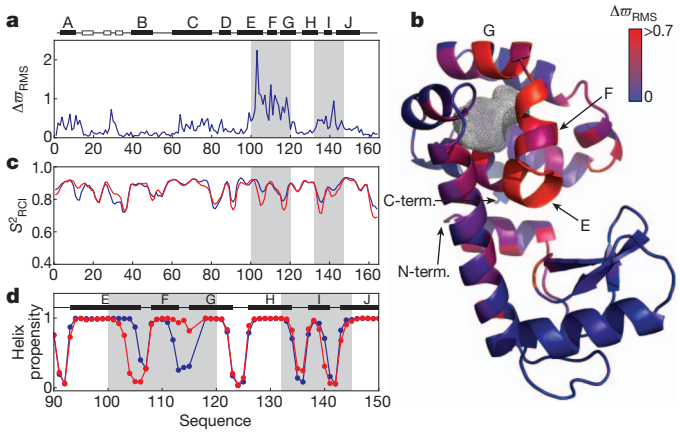

Figure 1 | L99A T4L exchanges between ground (visible) and excited (invisible) states, each with distinct conformations. **a**, Plot of $\Delta\varpi_{RMS} = \sqrt{\frac{1}{N}\sum_i \left(\frac{\Delta\varpi_i}{\Delta\varpi_{i,STD}}\right)^2}$ as a function of residue, where $\Delta\varpi_i$ is the shift difference in p.p.m. between states, $\Delta\varpi_{i,STD}$ is a nucleus specific value that corresponds to the range of shift values (1 s.d.) that are observed in a database of protein chemical shifts (http://www.bmrb.wisc.edu) for the nucleus in question ($^1H^N$, ^{15}N, $^{13}C^\alpha$, $^1H^\alpha$ and $^{13}C'$) and N is the number of nuclei ≤5 that are included in the average. Significant $\Delta\varpi_{RMS}$ differences are localized to a pair of regions (100–120, 132–146) that are highlighted in grey. The secondary structure of the ground state of L99A T4L is illustrated. **b**, Values of $\Delta\varpi_{RMS}$ colour-coded onto the X-ray structure of L99A T4L (PDB: 3DMV[28]), ranging from blue ($\Delta\varpi_{RMS} = 0$) to red ($\Delta\varpi_{RMS} > 0.7$). The mesh surface indicates the position of the cavity formed by the Leu to Ala substitution at position 99. **c**, S^2 values for the backbone amide groups in the ground (blue) and excited (red) states of L99A T4L as predicted by the RCI approach[15]. **d**, Helix propensity values, predicted using TALOS+ (ref. 29), highlighting important changes in secondary structure between ground (blue) and excited (red) L99A T4L conformers.

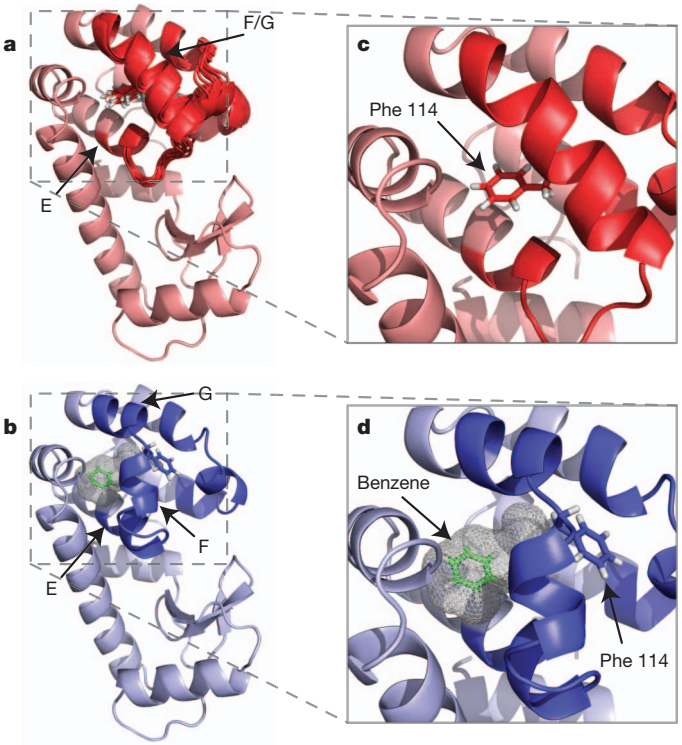

Figure 2 | The structure of the invisible, excited state of L99A T4L. **a**, Superposition of the 10 lowest energy structures of the L99A T4L excited state. The locations of helices E, F and G are indicated, along with the side chain of Phe 114 (see **c**). Only residues 100–120 and 132–146 were allowed to deviate from the L99A T4L ground state X-ray structure in calculations of the excited conformer (Methods). **b**, Ground state structure of L99A T4L, showing helices E, F and G and the position of Phe 114 (PDB: 3DMV[28]) or of benzene (green; PDB: 3DMX[28]) when it is bound inside the cavity (see **d**). **c**, **d**, Expanded regions of the excited state (**c**) and ground state (**d**) structures, focusing on the differences between helices F and G and the position of Phe 114.

state conformer. The lowest energy structures so obtained are in excellent agreement with the L99A crystal structure (r.m.s.d. of 0.6 ± 0.2 Å, Supplementary Table 6). Figure 2 shows an overlay of ten low energy representative excited state structures (Fig. 2a), along with the X-ray structure of the ground state (Fig. 2b) for comparison. As predicted on the basis of the input chemical shift data (Fig. 1a), there are clear structural differences between ground and excited states. These occur in a region immediately surrounding the cavity, involving rearrangement of the pair of short helices F and G that are orthogonal in the wild-type structure and that form a single, continuous and nearly straight helix in the excited state. This conformational rearrangement also includes a significant change in the backbone dihedral angle (Ψ) of Phe 114 to a helical value in the excited state (+49° to −36°) and a reorientation of its side chain caused by a change in the torsion angle (χ_1) from a *gauche*− to a *trans* conformation (see below). The change in χ_1 projects the Phe 114 benzyl moiety into the cavity of L99A T4L, significantly decreasing its volume (Fig. 2c).

To cross-validate the excited state structures, we used Rosetta structure based design calculations to identify substitutions predicted to stabilize the excited state relative to the ground state (Supplementary Table 7). One such substitution is G113A, which replaces one of the most helix destabilizing residues (Gly) with the most favourable (Ala)[17] in a region of the structure that is predicted to become more helical in the excited state. ^1H-^{15}N and ^1H-^{13}C spectra of L99A,G113A T4L (recorded at low temperature (1 °C) to slow down the exchange and hence improve spectral quality) show two sets of cross-peaks that can be connected by magnetization exchange[18,19] (Fig. 3a, Supplementary Fig. 3). The first set corresponds to those observed for the L99A ground state, with a second set occurring at the positions predicted for the excited state on the basis of the chemical shifts obtained from CPMG relaxation dispersion experiments recorded on L99A T4L (Fig. 3b). Intensities of peaks from magnetization exchange experiments can be fitted to extract the population of the excited state, p_E, and an exchange rate, $k_{ex} = k_{GE} + k_{EG}$ (k_{ij} is the exchange rate from state i to j); values of $p_E = 34 \pm 2\%$ and $k_{ex} = 48 \pm 1\ s^{-1}$, at 1 °C, are obtained (Supplementary Fig. 4, Supplementary Table 8). Thus, the G113A mutation shifts the G $\rightleftharpoons$ E equilibrium, as expected from the excited state structure, from $p_E < 0.5\%$ to 34% at 1 °C.

In a previous set of studies based on analysis of ^{15}N and methyl ^{13}C CPMG relaxation dispersion profiles, we speculated that the excited state of L99A T4L is an open conformation where ligands can access the cavity[5]. The solution structure of the low populated L99A T4L conformer, however, predicts that hydrophobic ligands would not bind the excited state because the cavity is occupied by the side chain of Phe 114 (Fig. 2c). As a second cross-validation of the structure, we measured the binding of benzene to the ground and excited conformers independently using a sample of L99A,G113A T4L, where separate peaks can be observed for each state (Fig. 3a). A previous study has established that benzene binds to L99A T4L with a millimolar K_D and a dissociation rate of close to 1,000 s^{-1} at 20 °C (ref. 20). Lowering the temperature to 1 °C decreases both the rate of benzene binding and the rate of exchange between ground and excited states; these rates are reduced to the point where separate peaks are observed for the methyl group of Met 102 in ^1H-^{13}C HSQC spectra of the ground, excited and benzene-bound states of L99A, G113A T4L to which one molar equivalent of benzene was added (Fig. 3c). Rates of exchange between the three states can be quantified by analysis of magnetization exchange experiments[18,19,21]. From fits of the time-dependencies of the auto-peaks (labelled 'G', 'E', 'B' in Fig. 3c, diagonal panels of Fig. 3d) and cross-peaks (cross panels) to a model of three-site exchange (Fig. 3e and Supplementary

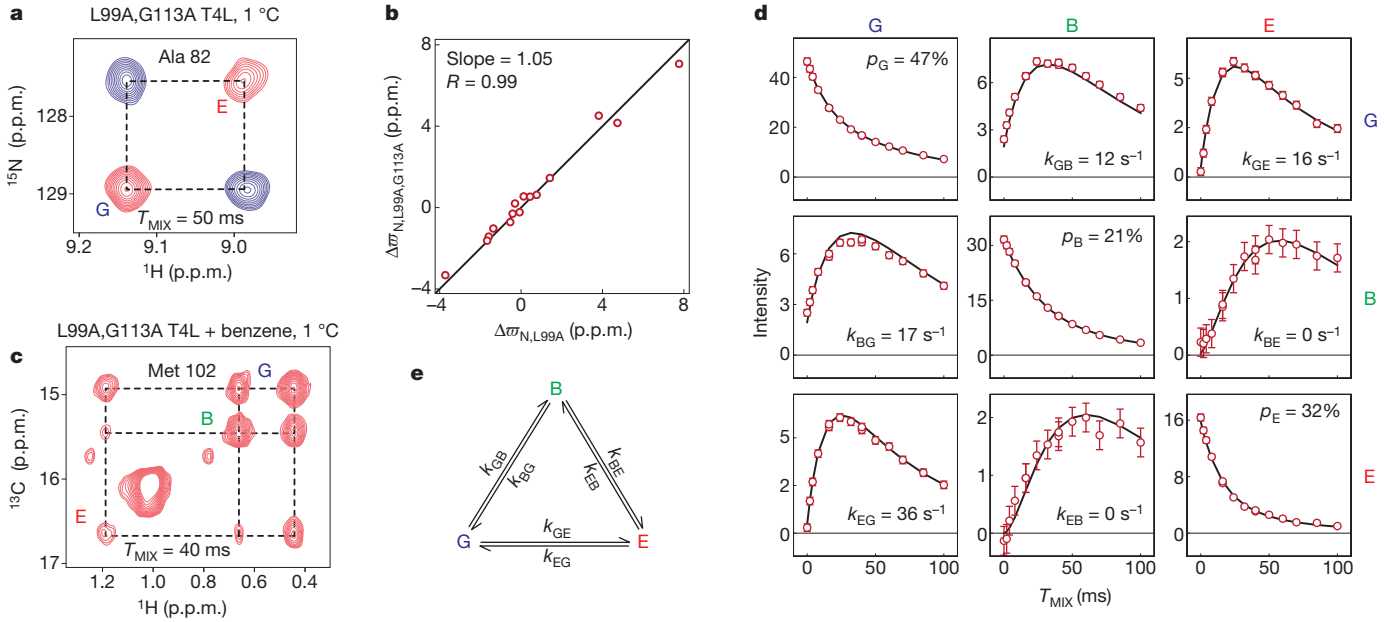

Figure 3 | Hydrophobic ligands do not bind the excited state of L99A T4L. **a**, Selected region of the ^{1}H–^{15}N correlation map from a magnetization exchange experiment recorded on L99A,G113A T4L at 1 °C, showing separate peaks for the ground (G) and excited (E) states. A pair of data sets are obtained, with the mixing time (T_{MIX} = 50 ms) recorded before or after the ^{15}N chemical shift evolution period, and the data sets subtracted so that diagonal- (cross-) peaks are positive (negative)[30]. **b**, Correlation between $\Delta\varpi_N$ values measured directly from the spectrum in **a** (*y* axis) and corresponding values from CPMG relaxation dispersion measurements of L99A T4L (25 °C; *x* axis). **c**, Magnetization exchange spectrum (T_{MIX} = 40 ms) recorded on an L99A,G113A T4L sample with a 1:1 molar equivalent of benzene, focusing on Met 102 that shows well resolved correlations from ground, excited and benzene-bound (B) states. **d**, Intensity of auto- and cross-peaks for residue Met 102 from magnetization exchange experiments recorded as a function of T_{MIX} (red circles), along with the best fit of the data (solid lines) to the exchange model of **e**. Values of rates of exchange, k_{ij}, are indicated.

Table 8), the six relevant rates, k_{ij}, are extracted. Best fit values for k_{EB}, k_{BE} are $<0.1\ s^{-1}$ (Supplementary Fig. 5), and F-test analyses establish that there is no difference in the quality of the fits when these rates are set to zero, indicating that binding to the excited state does not occur. By contrast, $k_{GB} = 11.6 \pm 0.3\ s^{-1}$ and $k_{BG} = 17.4 \pm 0.4\ s^{-1}$, so that ligand binding proceeds via the ground state. The mechanism by which this occurs is not at present known, but it must involve excursions of the ground state to additional conformations, presumably on a timescale faster than those that are accessible to the dispersion experiments described here.

Whereas the G113A mutation shifts the fractional population of the excited state from approximately 3% to 34%, we were interested in perturbing the equilibrium still further. The R119P substitution is predicted by Rosetta to further favour the excited state because the X-ray structure of the L99A ground state[9] is incompatible with a Pro at position 119 due to steric clashes involving an H^{δ} proton and the C′ of Thr 115. Consistent with this prediction, the ^{1}H–^{15}N HSQC spectrum of G113A,R119P,L99A T4L contains a single set of peaks (Supplementary Fig. 6) at resonance positions identical to those of the invisible, excited state of L99A that were determined by relaxation dispersion measurements (Supplementary Fig. 7). Further relaxation dispersion experiments recorded on the triple mutant established that the dominant conformer in solution, corresponding to the ligand inaccessible state (Supplementary Table 6), interconverts with a minor state conformer whose structure is that of the L99A ground state (Supplementary Fig. 8) (population of 3.8 ± 0.1%, $k_{ex} = 806 \pm 28\ s^{-1}$, 1 °C). Thus, the pair of mutations G113A,R119P inverts the populations of ground and excited states relative to L99A T4L.

This population inversion, rendering the ligand-inaccessible state the major conformer, allows an additional test of the structure of the invisible L99A T4L excited state. Quantitative *J*-based scalar coupling experiments[22] recorded on L99A,G113A,R119P T4L confirm that the χ_1 rotamer state for Phe 114 is *trans*, as observed in the CS-Rosetta based structure of the excited state (Supplementary Fig. 9, Supplementary Table 9, see above). By contrast, similar experiments recorded on the ground state of L99A T4L are consistent only with a *gauche*− conformation, as expected from the X-ray structure[9]. The trans χ_1 rotameric state for Phe 114 is a novel feature of the L99A T4L excited state and L99A,G113A,R119P T4L. In a G113A variant T4L, as examined here but where Leu is retained at position 99, a *gauche*− conformation is observed[23] for Phe 114; a *trans* χ_1 angle would lead to steric clashes with Leu 99 (Supplementary Fig. 10).

The L99A mutation in T4L creates an energy landscape in which a low-lying excited state is transiently populated. We have shown that this invisible, excited state has different functional properties from the

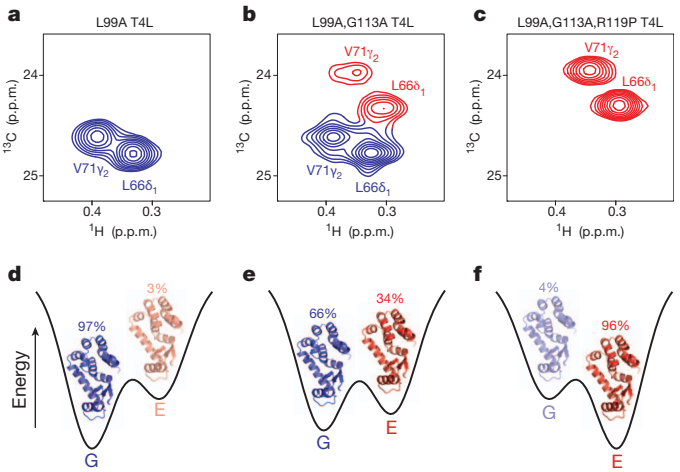

Figure 4 | The delicate balance between states on the energy landscape can be readily manipulated through mutation, providing a path for protein evolvability. **a**–**c**, Selected regions from ^{1}H–^{13}C HSQC spectra (recorded at 1 °C) of L99A T4L (**a**), L99A,G113A T4L (**b**) and L99A,G113A,R119P T4L (**c**), with the peaks from the ground and excited states coloured in blue and red, respectively. **d**–**f**, Corresponding energy landscapes, showing the structures of the ground and excited states and their fractional populations.

ground state—it does not bind hydrophobic ligands. This divergence in function can be controlled through mutation, with the G113A single mutation and the G113A,R119P double mutation changing the ratio of binding-competent to binding-incompetent states from 97% to 66% (G113A) and to less than 5% (G113A,R119P) (Fig. 4). The picture of evolving protein function suggested by our studies of L99A T4L is consistent with an emerging view of protein plasticity[24,25], with each molecule sampling a range of structures. Each unique conformer can, in turn, potentially carry out a different function[24,26,27]. A small number of mutations can then shift the relative populations of the conformers, thereby changing the activity of the protein, as has been observed in directed evolution experiments involving the introduction of mutations into flexible loop regions of enzymes[24]. Insight into the relation between protein dynamics, structure and evolvability is greatly facilitated through the powerful combination of relaxation dispersion NMR and Rosetta enabling individual protein states that populate the energy landscape to be investigated, even in cases where these conformers are invisible to other biophysical techniques. As applications of this methodology continue to grow, so too will our understanding of how protein dynamics control function, increasing the scope for the rational design of proteins with specific properties.

METHODS SUMMARY

Protein expression and purification. All NMR samples were prepared following previously published protocols, as described in detail in Methods.

NMR experiments and structure calculations. NMR experiments were recorded and analysed as described in Methods. Structure calculations of the L99A T4L excited state were based on experimental backbone ^{15}N, ^{13}C and ^{1}H chemical shifts obtained from CPMG dispersion experiments. The structures of regions in the excited state with significant variations in chemical shift values relative to those in the ground state were computed using the loop modelling application of Rosetta[16].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 March; accepted 30 June 2011.
Published online 21 August 2011.

1. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl Acad. Sci. USA* **102,** 6679–6685 (2005).
2. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313,** 1638–1642 (2006).
3. Henzler-Wildman, K. A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450,** 838–844 (2007).
4. Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **355,** 371–373 (1992).
5. Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of protein by NMR spectroscopy. *Nature Struct. Biol.* **8,** 932–935 (2001).
6. Palmer, A. G., Kroenke, C. D. & Loria, J. P. NMR methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339,** 204–238 (2001).
7. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105,** 4685–4690 (2008).
8. Baase, W. A., Liu, L., Tronrud, D. E. & Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci.* **19,** 631–641 (2010).
9. Eriksson, A. E. *et al.* Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255,** 178–183 (1992).
10. Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **41,** 442–451 (2008).
11. Hansen, D. F., Vallurupalli, P. & Kay, L. E. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J. Biomol. NMR* **41,** 113–120 (2008).
12. Cavalli, A., Salvatella, X., Dobson, C. M. & Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl Acad. Sci. USA* **104,** 9615–9620 (2007).
13. Vallurupalli, P., Hansen, D. F. & Kay, L. E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **105,** 11766–11771 (2008).
14. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329,** 1312–1316 (2010).
15. Berjanskii, M. V. & Wishart, D. S. A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127,** 14970–14971 (2005).
16. Wang, C., Bradley, P. & Baker, D. Protein-protein docking with backbone flexibility. *J. Mol. Biol.* **373,** 503–519 (2007).
17. Fersht, A. *Structure and Mechanism in Protein Science* (Freeman and Company, 1999).
18. Montelione, G. T. & Wagner, G. 2D chemical-exchange NMR-spectroscopy by proton detected heteronuclear correlation. *J. Am. Chem. Soc.* **111,** 3096–3098 (1989).
19. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4,** 727–734 (1994).
20. Feher, V. A., Baldwin, E. P. & Dahlquist, F. W. Access of ligands to cavities within the core of a protein is rapid. *Nature Struct. Biol.* **3,** 516–521 (1996).
21. Religa, T. L., Sprangers, R. & Kay, L. E. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* **328,** 98–102 (2010).
22. Hu, J. S., Grzesiek, S. & Bax, A. Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings. *J. Am. Chem. Soc.* **119,** 1803–1804 (1997).
23. Nicholson, H., Tronrud, D. E., Becktel, W. J. & Matthews, B. W. Analysis of the effectiveness of proline substitutions and glycine replacements in increasing the stability of phage T4 lysozyme. *Biopolymers* **32,** 1431–1441 (1992).
24. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324,** 203–207 (2009).
25. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30,** 409–425 (1976).
26. Walden, W. E. *et al.* Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science* **314,** 1903–1908 (2006).
27. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449,** 1078–1082 (2007).
28. Liu, L. J., Baase, W. A. & Matthews, B. W. Halogenated benzenes bound within a non-polar cavity in T4 lysozyme provide examples of I···S and I···Se halogen-bonding. *J. Mol. Biol.* **385,** 595–605 (2009).
29. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223 (2009).
30. Rodríguez, J. C., Jennings, P. A. & Melacini, G. Using chemical exchange to assign non-covalent protein complexes in slow exchange with the free state: enhanced resolution and efficient signal editing. *J. Biomol. NMR* **30,** 155–161 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Forman-Kay for providing laboratory space, and J. Forman-Kay and T. Alber for discussions. G.B. acknowledges The European Molecular Biology Organization and The Canadian Institutes of Health Research (CIHR) for postdoctoral fellowships. L.E.K. holds a Canada Research Chair in Biochemistry. This work was supported by the CIHR and the Natural Sciences and Engineering Research Council of Canada, with computations performed on the GPC supercomputer at the SciNet HPC Consortium.

Author Contributions G.B., P.V. and D.F.H. made samples. G.B., P.V., D.F.H. and L.E.K. designed and performed all NMR experiments. G.B., P.V., B.E.C., O.L. and R.M.V. performed structure calculations. G.B., P.V. and A.B. carried out initial crystallization trials. F.W.D, D.B., G.B., P.V. and L.E.K. designed experiments and wrote the paper.

Author Information The structural ensembles of the L99A excited state and L99A,G113A,R119P T4L have been deposited with PDB (accession codes 2LCB, 2LC9). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.E.K. (kay@pound.med.utoronto.ca).

METHODS

Protein expression and purification. The gene for expression of L99A T4L was optimized for protein production (Genscript) and placed in a pET-29b plasmid. Additional plasmids for the expression of the mutants used in this study (L99A,G113A T4L; L99A,G113A,R119P T4L) were constructed from the L99A T4L plasmid. T4L proteins were expressed in *Escherichia coli* BL21(DE3) cells grown in M9 minimal media with glucose (3–4 $g\,l^{-1}$, typically) and ammonium chloride (NH_4Cl, 1 $g\,l^{-1}$) as the sole carbon and nitrogen sources, respectively. Proteins with specific labelling patterns (see below) were obtained by expression in M9 media containing the appropriately labelled glucose, $^{15}NH_4Cl$ and solvent (H_2O or D_2O)[31]. The labelling patterns (carbon source/solvent) used in this study are: uniform (U) ^{15}N (glucose/H_2O), U-$^{13}C/^{15}N$ ([$^{13}C_6$]-glucose/H_2O), U-$^{2}H/^{13}C/^{15}N$ ([$^{13}C_6$,$^{2}H_7$]-glucose/100% D_2O), U-$^{2}H/^{15}N$ ([$^{2}H_7$]-glucose/100% D_2O), $^{13}C^{\alpha}/^{15}N$ ([2-^{13}C]-glucose/100% H_2O), U-^{13}C/U-^{15}N/50%^{2}H ([$^{13}C_6$,$^{2}H_7$]-glucose/50% D_2O), $^{13}CH_3$-Met/^{15}N (glucose/H_2O, supplemented with 100 $mg\,l^{-1}$ of $^{13}CH_3$-Met added to the media 30 min before induction of protein overexpression). *E. coli* BL21(DE3) cells transformed with the appropriate plasmid were grown in one or two litres of media at 37 °C until an OD_{600} of ~1. The temperature was then reduced to 16 °C and protein expression was induced with 1 mM IPTG for 14–18 h. The cells were harvested by centrifugation and frozen. The protein was purified from the cells as described[32].

Samples. NMR samples (~1.5 mM in protein) were prepared in a buffer consisting of 50 mM sodium phosphate, 25 mM NaCl, 2 mM EDTA, 2 mM NaN_3, pH 5.5 in either 10% or 100% D_2O. The experiments were performed on Varian Inova spectrometers operating at frequencies (^1H) of 500, 600 and 800 MHz, at a temperature of 25 °C, unless stated otherwise.

Assignments. Complete assignments for L99A T4L have been reported previously[5]. The major (ground) state peaks in the L99A,G113A T4L mutant were assigned by comparison with assigned spectra of L99A T4L. Minor (excited-) state ^1H–^{15}N resonance assignments were obtained using an ^{15}N magnetization exchange experiment[19] recorded at 800 MHz, 1 °C, with a mixing time (T_{MIX}) of 50 ms. At 1 °C the excited state is populated to ~34%, $k_{ex} \approx 50\ s^{-1}$, so that very clear exchange peaks correlate ground and excited state resonances. Backbone $^1H/^{13}C/^{15}N$ and $^{13}C^{\beta}$ assignments for the L99A,G113A,R119P T4L mutant were obtained using standard triple resonance experiments[33] recorded at 34 °C either at 500 or 800 MHz. Assignments were very close to complete. The elevated temperature (34 °C) and lower field (500 MHz) were used to minimize signal loss due to chemical exchange.

CPMG experiments. $^1H/^{15}N/^{13}C$ constant-time (CT) CPMG relaxation dispersion experiments were (usually) performed at two static magnetic fields. Typically, dispersion curves were composed of a large number (~15) of ν_{CPMG} values with errors estimated based on two or three repeat values[34]. Here $\nu_{CPMG} = 1/(4\tau_{CPMG})$, where $2\tau_{CPMG}$ is the spacing between the refocusing π pulses applied during the CT delay of length T_{Relax}. Details of the experiments[6,35,36] used to characterize the excited state of L99A T4L are summarized in Supplementary Table 1.

Sign experiments. Single-quantum CPMG relaxation dispersion experiments provide only the magnitude of the change in chemical shift $|\Delta\varpi_{GE}| = |\varpi_E - \varpi_G|$ between the two exchanging states. Signs were obtained by comparing the positions of the ground state peaks in HSQC spectra recorded at different static magnetic fields and/or between peak positions in HSQC/HMQC spectra recorded at the same field[37,38]. Once the signs of amide nitrogen $\Delta\varpi$ values were obtained, the corresponding signs of the amide protons were generated from zero-quantum (ZQ) and double-quantum (DQ) CPMG experiments[39]. Experiments used to obtain this information are listed in Supplementary Table 2.

Quantitative *J*-modulated experiments. χ_1 angles of aromatic residues in both L99A T4L (ground state) and L99A,G113A,R119P T4L (ground state that is a mimic of the L99A T4L excited state structure, see Supplementary Fig. 7 and text) were determined by measurement of three-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ scalar couplings using quantitative *J*-based experiments[22,40] recorded with dephasing delays of 100 ms ($J_{C'C\gamma}$) and 120 ms ($J_{NC\gamma}$), respectively. All experiments were obtained at 35 °C on a 600 MHz spectrometer, using uniformly ^{15}N, ^{13}C enriched samples. Difference spectra for both experiments and both proteins are shown in Supplementary Fig. 9. Measured scalar coupling values are summarized in Supplementary Table 9 along with the χ_1 angles for the aromatic residues of the ground and excited state structures.

Magnetization-exchange experiments. ^{15}N magnetization exchange experiments, recorded at 1 °C, 800 MHz, were used for assignment of excited state correlations of L99A,G113A T4L, as described above. A pair of experiments were recorded in which the exchange mixing period, T_{MIX} (50 ms), was placed (1) before and (2) after indirect detection of ^{15}N magnetization. Subtraction of the two data sets so obtained generates a two-dimensional spectrum where correlations from ground and excited states (positive) are connected by cross-peaks (negative), forming a 'rectangular' structure[30] (Fig. 3a). Quantitative methyl ^{13}C magnetization exchange experiments to quantify exchange, $G \underset{k_{EG}}{\overset{k_{GE}}{\rightleftharpoons}} E$ (see below or text), in L99A,G113A T4L were performed at 600 MHz, 1 °C, using a sample in which only Met-C^{ε} was ^{13}C enriched. A second similarly labelled sample to which a small amount of benzene was added (approximately 1:1 molar equivalents of benzene and protein) was used to study exchange between the ground, excited and benzene-bound states of L99A,G113A T4L. Experiments were recorded with T_{MIX} values ranging between 0 and 85 ms (100 ms in the presence of benzene) with errors estimated based on repeat measurements.

Data processing. The NMRpipe software package[41] was used to processes all of the NMR data. Subsequent visualization and peak picking was achieved using the program Sparky[42]. The intensities of peaks (I) were obtained using the program FuDA (http://pound.med.utoronto.ca/software.html), while the CcpNmr set of programs[43] was used to analyse some of the triple resonance assignment experiments.

Analysis of CPMG data. Relaxation dispersion (RD) profiles, $R_{2,eff}(\nu_{CPMG})$, were generated from peak intensities, $I(\nu_{CPMG})$, measured in a series of 2D correlation maps recorded at various CPMG frequencies, ν_{CPMG}. The effective relaxation rates, $R_{2,eff}(\nu_{CPMG})$, were computed via the relation:

$$R_{2,\mathrm{eff}}(\nu_{\mathrm{CPMG}}) = -\frac{1}{T_{\mathrm{Relax}}}\ln\left(\frac{I(\nu_{\mathrm{CPMG}})}{I_{\mathrm{o}}}\right)$$

where I_o is the peak intensity extracted from a reference spectrum recorded without the CPMG block. RD profiles were analysed assuming a two-site exchanging system, $G \rightleftharpoons E$, where the major state, G, interconverts with the minor state, E, as described previously in the context of the L99A system[5,44]. The model parameters defining the chemical exchange process, that is the exchange rate, k_{ex}, the population of the minor state, p_E, and the absolute difference in chemical shifts between the two states, $|\Delta\omega_{GE}| = |\omega_E - \omega_G|$, were determined by minimizing the target function:

$$\chi^2(\zeta) = \sum_{i=1}^{N}\left(\frac{R_{2,\mathrm{eff}}^{\mathrm{Exp}} - R_{2,\mathrm{eff}}^{\mathrm{Calc}}(\zeta)}{\Delta R_{2,\mathrm{eff}}^{\mathrm{Exp}}}\right)^2$$

where $R_{2,eff}^{Exp}$ and $\Delta R_{2,eff}^{Exp}$ are the experimental effective transverse relaxation rates and their associated uncertainties, $R_{2,eff}^{Calc}(\zeta)$ are back-calculated relaxation rates obtained by numerical integration of the Bloch-McConnell equations[45] using the program CATIA (http://pound.med.utoronto.ca/software.html), ζ represents the set of adjustable model parameters and the sum is over all the experimental data points.

Analysis of quantitative magnetization exchange data. As described above and in the text, methyl ^{13}C magnetization exchange experiments were recorded on L99A,G113A T4L (1 °C) without and with added benzene using a pulse scheme described previously[21,46]. Data from Met 102 were analysed because separate, well-resolved correlations are obtained for the ground, excited and benzene-bound states that could be accurately quantified. At the low temperature used (1 °C), the interconversion between ground and excited states as well as benzene binding are in the slow exchange regime, a requirement for the magnetization exchange experiment. The intensity I of auto- (cross-) peaks corresponding to magnetization that is not (is) transferred between states during a mixing time, T_{MIX}, has been analysed by numerically solving the Bloch-McConnell equations for the evolution of magnetization in the presence of chemical exchange[45]. The time dependence of magnetization during the entire pulse sequence was simulated (details are available from the authors on request). The resonance frequencies of the peaks and their transverse relaxation rates were obtained from the positions and linewidths of the peaks in spectra, respectively. The fitted parameters include the total magnetization, longitudinal relaxation rates (R_1) for each state, the fractional populations of each state, p_i (subject to the constraint that $\sum_i p_i = 1$), and the rates of exchange between states i and j, $k_{exij} = k_{ij} + k_{ji}$. The fitting parameters were optimized using a simplex procedure to minimize the function:

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{I_i^{\mathrm{Exp}} - I_i^{\mathrm{Calc}}}{\sigma_i^{\mathrm{Exp}}}\right)^2$$

Here the summation is over all the experimental data points, I^{Exp} is the experimental intensity, I^{Calc} is the calculated intensity and σ^{Exp} is the error in the intensity. In the case of two-state exchange (no benzene), 48 data points from four peaks associated with Met 102 (two auto- and two cross- peaks × 12 T_{MIX} values) were fitted using five fitting parameters. For the three-state exchange (approximately 1:1 molar equivalent protein:benzene), 117 data points from nine Met 102 peaks (three auto- and six cross-peaks × 13 T_{MIX}; Fig. 3d) were fitted using nine fitting parameters. The minimum error in the intensities was assumed to be 3%. Errors in the fitted parameters were estimated using a Monte Carlo

procedure[47]. Here 50 synthetic data sets were generated using the best-fit parameters in which random error was added to magnetization intensities and $^{13}C/^{1}H$ R_2 values (based on the experimental errors) and each of the data sets fitted as per the experimental data. Errors are calculated as 1 s.d. in the extracted values.

These experiments clearly indicate that only the ground state binds benzene. As shown in Fig. 3d, fits of the time dependencies of diagonal- and cross-peaks from magnetization exchange spectra establish that k_{BE}, $k_{EB} < 0.1\ s^{-1}$. To further support the result that benzene does not bind the excited state, Supplementary Fig. 5 plots the reduced χ^2 obtained from the fit of the magnetization exchange data as a function of $k_{exBE} = k_{BE} + k_{EB}$. A clear minimum occurs for $k_{exBE} \approx 0$ ($\chi^2_{red} = 1.1$). Of note, when k_{exBE} is fixed to the relatively small value of $0.5\ s^{-1}$, χ^2_{red} increases by fivefold (to 5.4), clearly indicating that k_{exBE} is very small. From the principle of detailed balance for the equilibrium denoted in Fig. 3e, it is predicted that the ratios $\frac{p_E}{p_E + p_G}, \frac{k_{GE}}{k_{EG}}$ will be independent of ligand (benzene) concentration, as observed to within experimental error (Supplementary Table 8). As expected, binding of benzene shifts the equilibrium from the excited (binding incompetent) to the ground/bound states. These results are in complete agreement with the structure of the excited state of L99A T4L, showing clearly that Phe 114 is inserted into the cavity, hence obstructing the binding of hydrophobic ligands.

L99A excited state chemical shifts. Excited state amide nitrogen, amide proton and carbonyl chemical shifts are available for nearly all the (164) L99A T4L residues (Supplementary Tables 3–5). In cases where the sign of $\Delta\varpi_{GE}$ is obtained, the chemical shift of the excited state is readily calculated, $\varpi_E = \varpi_G + \Delta\varpi_{GE}$. In cases where the sign is not available, the excited state chemical shift is ambiguous but restricted to two values, $\varpi_E = \varpi_G + \Delta\varpi_{GE}$ or $\varpi_E = \varpi_G - \Delta\varpi_{GE}$. The magnitude of the change in chemical shift, $|\Delta\varpi_{GE}|$, also provides useful information (see section on Rosetta calculations below).

Calculation of L99A T4L excited state structures. The structure of the excited state of L99A T4L was obtained by using the CS-Rosetta approach developed for the determination of ground state protein structures[7], with a number of important differences relative to the standard protocol. Based on the $\Delta\varpi_{RMS}$ values (see text), we assumed that only the regions encompassing residues 100–120 and 132–146 adopt different conformations in the ground and excited states. The structure of the rest of the molecule was fixed to the ground state crystal structure of L99A T4L (3DMV[28]). The structures of these two regions in the excited state were computed using the loop modelling application of Rosetta[16]. As a control, identical CS-Rosetta computations of the ground state structure were also performed using the same limited set of ground state shifts (that is, that are available for the excited state). To avoid any bias, T4 lysozyme structures were removed from the fragment databases in all the computations described here. Two hundred starting 3mer and 9mer fragments were selected for each position using the CS-Rossetta approach[7], modified to include ambiguous excited state shifts in cases where the sign of $\Delta\varpi_{GE}$ could not be determined. Fragments were scored against ambiguous shifts by selecting the shift which agreed best with the one predicted for the fragment. Ambiguous chemical shifts were similarly taken into account during the scoring of the final structures. The selected fragments were then used in a standard Rosetta loop modelling protocol[16] to generate 9,600 structures (the target secondary structure propensities input into Rosetta are those predicted by TALOS+[29]). Supplementary Fig. 2a and b plots the energies of the resultant ground state structures, generated with ground state chemical shifts (Supplementary Fig. 2a, CS-Rosetta energy; Supplementary Fig. 2b, chemical shift component of the CS-Rosetta energy term) versus r.m.s.d. to the L99A ground state X-ray structure. The characteristic funnel shape energy profile so obtained is an excellent indicator of convergence and indeed the lowest energy 10 (50) structures have pair-wise backbone r.m.s.d. values of 0.6 ± 0.15 (0.55 ± 0.15) Å relative to the ground state crystal structure, including only those residues that were allowed to move in the calculations. The corresponding plots for the excited state structures are shown in Supplementary Fig. 2c and d.

Some of the low energy CS-Rosetta structures produced with the excited state chemical shifts are very similar to the ground state structure (Supplementary Fig. 2c). This is not surprising, given the experimental finding that the ground state is more stable than the excited state by $\sim 2\ kcal\ mol^{-1}$. However, the chemical shift score by itself clearly indicates that the ground state structure is not a 'good' solution (Supplementary Fig. 2d). Hence a two-step selection procedure was used. Out of the 9,600 structures that were generated initially, 960 structures with the best CS-Rosetta score were selected for further analysis. As a second selection step, 96 of these structures with the best chemical shift score were retained. In this way, structures with both low energy CS-Rosetta and low energy chemical shift scores are selected. It is noteworthy that the structures so generated were essentially identical irrespective of whether selection was first performed on the basis of the Rosetta score followed by selection according to chemical shifts (as described here) or whether the opposite order of scoring was used. We find that there is a major cluster of structures with an r.m.s.d. of ~1.4 Å to the ground state crystal structure and a much smaller cluster with an r.m.s.d. of ~1 Å to the ground state conformer; the final 96 structures selected are indicated in green in Supplementary Fig. 2c and d. Only one out of the forty lowest energy structures after the second round of filtering is part of the second cluster (10% of the 96 structures). A major difference between the two sets of structures is the χ_1 angle of Phe 114, with this dihedral angle assuming a *trans* (*gauche*−) conformation in the major (minor) cluster. Quantitative-*J* experiments[22,40] recorded on the L99A,G113A,R119P T4L mutant that is an excellent mimic of the excited state (Supplementary Fig. 7, Supplementary Table 6, see text) clearly show that the *trans* conformation is the only one populated (Supplementary Fig. 9, Supplementary Table 9). Notably, CS-Rosetta computations performed with the full set of chemical shifts for this mutant (L99A,G113A,R119P T4L) produced structures that have a conformation that is essentially identical to that obtained for the excited state of L99A T4L (Supplementary Table 6). All structure calculations were performed on the University of Toronto SciNeT super-computer cluster[48]. Each set of 9,600 structures required 10 h of computational time using 256 processor cores. Pymol[49] and Chimera[50] were used to visualize and analyse the resultant structures.

Mutations shifting ground and excited state populations. Predictions of free energy differences ($\Delta\Delta G$) were performed as described[51]. Briefly, single point mutants of the 19 amino acids (except cysteine) were made *in silico* in the region corresponding to residues 105–120 and $\Delta\Delta G$ values were computed for representative structures of both ground and excited states. The crystal structure of L99A T4L (3DMV[28]) was used as the ground state, with a representative low energy structure obtained from CS-Rosetta simulations performed with excited state chemical shifts used for the excited state. The free energy difference between corresponding point mutants in the excited and ground states was computed ($\Delta\Delta G_E - \Delta\Delta G_G$; negative values indicate relative stabilization of the excited state). We screened for single point mutants that energetically favoured the excited state conformation and simultaneously disfavoured the ground state conformation (Supplementary Table 7). Mutations for experimental characterization were selected according to two criteria: secondary structure propensity of the excited state and Rosetta $\Delta\Delta G$ predictions.

31. Lundström, P., Vallurupalli, P., Hansen, D. F. & Kay, L. E. Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy. *Nature Protocols* **4,** 1641–1648 (2009).
32. Vallurupalli, P., Hansen, D. F., Lundstrom, P. & Kay, L. E. CPMG relaxation dispersion NMR experiments measuring glycine $^{1}H\alpha$ and $^{13}C\alpha$ chemical shifts in the 'invisible' excited states of proteins. *J. Biomol. NMR* **45,** 45–55 (2009).
33. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34,** 93–158 (1999).
34. Korzhnev, D. M. *et al.* Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430,** 586–590 (2004).
35. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E. & Kay, L. E. Measurement of bond vector orientations in invisible excited states of proteins. *Proc. Natl Acad. Sci. USA* **104,** 18473–18477 (2007).
36. Hansen, D. F., Vallurupalli, P., Lundstrom, P., Neudecker, P. & Kay, L. E. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: how well can we do? *J. Am. Chem. Soc.* **130,** 2667–2675 (2008).
37. Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. Reconstructing NMR spectra of ''invisible'' excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* **124,** 12352–12360 (2002).
38. Bouvignies, G. *et al.* A simple method for measuring signs of $^{1}H^{N}$ chemical shift differences between ground and excited protein states. *J. Biomol. NMR* **47,** 135–141 (2010).
39. Orekhov, V. Y., Korzhnev, D. M. & Kay, L. E. Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. *J. Am. Chem. Soc.* **126,** 1886–1891 (2004).
40. Hu, J. S. & Bax, A. Determination of φ and χ_1 angles in proteins from ^{13}C-^{13}C three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J. Am. Chem. Soc.* **119,** 6360–6368 (1997).
41. Delaglio, F. *et al.* NMRpipe – a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
42. Goddard, T. D. & Kneller, D. G. SPARKY 3 (University of California, San Francisco, 2006).
43. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct. Funct. Bioinform.* **59,** 687–696 (2005).
44. Mulder, F. A., Hon, B., Muhandiram, D. R., Dahlquist, F. W. & Kay, L. E. Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. *Biochemistry* **39,** 12614–12622 (2000).
45. McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28,** 430–431 (1958).
46. Sprangers, R., Gribun, A., Hwang, P. M., Houry, W. A. & Kay, L. E. Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. *Proc. Natl Acad. Sci. USA* **102,** 16678–16683 (2005).
47. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C. The Art of Scientific Computing* 2nd edn (Cambridge Univ. Press, 1992).
48. Loken, C. *et al.* SciNet: Lessons learned from building a power-efficient top-20 system and data centre. *J. Phys. Conf. Ser.* **256,** 012026 (2010).

49. The PyMOL Molecular Graphics System, Version 1.3 (Schrödinger, LLC, 2010).
50. Pettersen, E. F. *et al.* UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
51. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79,** 830–838 (2011).