

# Structure of a low-population intermediate state in the release of an enzyme product

Alfonso De Simone<sup>1</sup>, Francesco A. Aprile<sup>2</sup>, Anne Dhulesia<sup>2</sup>,  
Christopher M. Dobson<sup>2</sup> and Michele Vendruscolo<sup>2</sup>

<sup>1</sup>*Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, UK*

<sup>2</sup>*Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK*

## Abstract

Enzymes can increase the rate of biomolecular reactions by several orders of magnitude. Although the steps of substrate capture and product release are essential in the enzymatic process, complete atomic-level descriptions of these steps are difficult to obtain because of the transient nature of the intermediate conformations, which makes them largely inaccessible to standard structure determination methods. We describe here the determination of the structure of a low-population intermediate in the product release process by human lysozyme through a combination of NMR spectroscopy and molecular dynamics simulations. We validate this structure by rationally designing two mutations, the first engineered to destabilise the intermediate and the second to stabilize it, thus slowing down or speeding up, respectively, product release. These results illustrate how product release by an enzyme can be facilitated by the presence of a metastable intermediate with transient weak interactions between the enzyme and the product.

## Introduction

As it is becoming increasingly clear that proteins populate a variety of ‘intermediate’ states during their function<sup>1,2</sup>, it is essential to determine the structures of such states in addition to defining the native conformations. Protein intermediates are involved in folding, misfolding and aggregation processes, as well as in events associated with molecular recognition, catalysis and allostery<sup>1-3</sup>. These species are transient in nature and as such they have been difficult to characterise. Nuclear magnetic resonance (NMR) spectroscopy has emerged in this context as a powerful technique to define such states as exemplified by the characterisation of the structures of species involved in folding<sup>4</sup>, molecular recognition<sup>5</sup>, and aggregation<sup>6</sup>.

In the present paper we describe a study of the mechanism involved in the process by which an enzyme releases its products. This is one of the three major steps in an enzymatic catalysis process<sup>7</sup>. In the first step the enzyme forms a complex with the substrate. In the second step, the transition state of the reaction is reached within the favourable environment provided by the catalytic site enabling the conversion of the substrate into product. In the third step, which is often rate-limiting, the product is released and the enzyme returns to its original state. Each of these steps is usually rather complex and involves reaction intermediates, which are transient in nature and difficult to characterize.

In order to investigate the third step, we have studied here human lysozyme, the first enzyme to be crystallised<sup>8</sup>, and whose structural properties have been characterised in great detail<sup>8-11</sup>. The native structure of this enzyme is divided into a  $\alpha$  domain (residues 1 to 38, and 86 to 130) and a  $\beta$  domain (residues 39 to 85), containing primarily  $\alpha$ -helical and  $\beta$ -sheet secondary structures, respectively<sup>8-11</sup>. This enzyme degrades bacterial cell walls by catalysing the hydrolysis of the 1,4- $\beta$ -linkages of the cell wall peptidoglycans, with a reaction that has been the object of intense scrutiny<sup>12-15</sup>. According to the mechanism originally proposed by Phillips on the basis of his structure<sup>9</sup>, lysozyme binds to a peptidoglycan molecule in the binding site within the cleft between its two domains thus causing the substrate to adopt a strained conformation similar to that of the transition state of the hydrolysis. Here we study the product release process. To this end, we used a well-characterised oligosaccharide product having an inhibitory effect on the enzyme,

N,N',N''-triacetylchitotriose (triNAG)<sup>16</sup>, which has been frequently used for studying lysozyme-product interactions<sup>14</sup>.

## Results and Discussion

In order to define the structural populations of human lysozyme in the presence and in the absence of an inhibitor, we measured <sup>15</sup>N-<sup>1</sup>H residual dipolar couplings (RDCs)<sup>17,18</sup> in the ligand-free and ligand-bound states (see Methods), and used them as structural restraints in molecular dynamics simulations<sup>19,20</sup>. In this way, we determined two ensembles of structures of the enzyme representing, respectively, the ligand-free and the ligand-bound states of this protein. Our results indicate that large-scale concerted motions between the  $\alpha$  and  $\beta$  domains of the enzyme generate an intermediate state involved in the release of the product.

The use of RDCs as structural restraints assists the conformational sampling in molecular dynamics simulations in order to estimate the free energy landscape of a protein, as recently shown with hen lysozyme for which a large body of experimental data were used for validation purposes<sup>21</sup>. This approach enables the translation of the experimental measurements into structures according to the principle of maximum entropy<sup>22-24</sup>. A number of methods to employ NMR measurements of RDCs for the characterization of the structure and dynamics of proteins have been proposed<sup>25-27</sup>. Since these approaches have generally been used to assess dynamic events of relatively small amplitude, it was necessary to derive a means of extending these methods to enable the description of the large conformational interconversions associated with the function of many protein molecules.

The strategy that we have used for this purpose relies on the ability to extract from time and spatially averaged data the contributions to the experimental observables that come from the low-population states present as a result of conformational fluctuations. Intermediate states determined in this way have been already described using paramagnetic resonance enhancement (PRE)<sup>5</sup> and RDC<sup>21,28</sup> measurements, complementing approaches in which NMR parameters specific for these states are obtained directly, in particular by relaxation-dispersion methods<sup>4,6,29,30</sup>. By applying this

approach using RDCs to human lysozyme we generated a structural ensemble representing the free state of this enzyme that reveals large breathing motions between the  $\alpha$  and  $\beta$  domains (Fig. 1a and Fig. 1 - figure supplements 1 and 2). This motion, which influences the mutual orientation of the two domains thereby altering the structure of the catalytic pocket at their interface, can be accounted for by defining a ‘breathing’ angle  $\theta$  between the  $\alpha$ -domain, the hinge region and the  $\beta$ -domain<sup>21</sup> (Fig. 1 - figure supplement 3).

In order to compare the free and bound states of the enzyme, we measured the RDCs also in the bound state (Fig. 1a - figure supplements 1 and 2). While in the free state the free energy landscape is characterised by a single basin (Fig. 1a), in the bound state a second local minimum appears (Fig. 1b). This change, which reflects the differences in the RDC data of the free and bound states, corresponds to a wider distribution of values of the  $\theta$  angle in the bound state compared to the free state (Fig. 1 and Fig. 1 - figure supplement 4). The bound state ensemble was validated using NOEs, J-couplings, chemical shifts and RDCs (Fig. 1 - figure supplement 5). The relative populations of the ground and excited states are comparable to those that we have studied previously using the method adopted here<sup>20,21,28,31</sup>.

Having in mind the release of the product, we designate the global free energy minimum observed in this study as the ‘locked state’ (i.e. release incompetent), which is centred at  $\theta$  values of about  $58^\circ$  and C $\alpha$ -RMSD values of about 0.9 Å from the X-ray structure of the complex (calculated by considering secondary structure elements only), and the other free energy minimum, which has about a 13% population under the conditions of our experiments, defined as the ‘unlocked state’ (i.e. release competent, Fig. 1b). The unlocked state is a compact conformation that differs from the locked state by a global motion in which the  $\alpha$  and  $\beta$  subunits become closer, with a  $\theta$  value of about  $49^\circ$  in the centre of the basin. This motion generates particularly distorted structures with global RMSD values of about 1.5 Å from the X-ray structure. The angle  $\theta$  provides a simple and effective reaction coordinate to describe the effect of triNAG binding on the energy landscape of human lysozyme (Fig. 2a), which clearly illustrates how the protein is able to explore closed conformations (i.e.  $\theta < 50^\circ$ ) upon ligand binding.

In the structural ensemble representing the complex between human lysozyme and triNAG, unlocked conformations are characterised by less favourable intermolecular Coulomb and van der Waals interactions than those found in the locked state (Fig. 2b-e). A comparison between the locked and unlocked conformations indicates that this difference corresponds to specific interactions between the substrate and the binding pocket (Fig. 2d), which include hydrophobic interactions between a methyl group of triNAG and the side-chain of W131, as well as hydrogen bonds between the ligand and the main chain amide group of N60 and the side chains of W64 and Q104. These interactions are present in essentially all the structures in the ensemble representing the locked state, while they are absent in the structures in the ensemble of the unlocked state. Indeed, because of a partial displacement of the ligand from the binding pocket, the unlocked conformations lose the tight interactions that are stabilise in the locked state and gain new interactions on the external surface of the protein. These interactions, which mainly involve hydrogen bonds between donors and acceptors groups from the ligand and the protein surface, are highly variable and heterogeneous in the unlocked conformations.

Overall, this analysis of the structural ensembles of human lysozyme suggests that, as a consequence of a concerted conformational transition, the enzyme explores conformations in which the specific and tight intermolecular interactions with the substrate in its locked state are largely lost in favour of the formation of weak and non-specific interactions in its unlocked state. This transition is favoured by large-scale conformational motions in which the  $\alpha$  and  $\beta$  domains become closer, thus suggesting that these motions are employed by the enzyme to modulate the affinity with the ligand. The unlocked state therefore represents an intermediate state for product release. In this view (Fig. 3), the enzyme product complex (EP) populates transiently an intermediate state (EP\*) that favours the release of the product (E+P). Thus, the analysis of the structural ensembles that we have determined provides evidence that large-scale conformational transitions are employed by enzymes along their catalytic cycles including key events in the product release step, which often represents the rate-limiting step that governs the turnover of the enzyme. The difficulty for enzymes to release the products can arise from the fact that typically the latter have similar physico-chemical characteristics to the substrates and therefore maintain a significant affinity for the enzyme.

To test the possibility that the structure that we have determined of the unlocked state represents an intermediate state in the release of the product, we compared in detail the energetic contributions that stabilise the locked and unlocked states and identified a stabilising hydrogen bond that can be formed only in the unlocked state (Fig. 4a), which involves side-chain atoms of residue N44 in the  $\beta$ -domain and residue E35 in the  $\alpha$ -domain. These two residues are too far away from each other in the locked state to form a hydrogen bond (Fig. 4b), and therefore this interaction is characteristic only of the unlocked conformation. This finding suggests that this hydrogen bond provides a specific interaction by which one could selectively target the stability of the EP\* intermediate. To this end, we rationally designed the N44A single-amino acid mutation to verify if the ability of the enzyme to release the products is indeed altered by preventing the formation of the hydrogen bond that stabilises the unlocked state.

Comparison of the  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of the wild-type and the N44A variant of human lysozyme shows that the mutation does not affect the structural properties of the mutant in the native state (Fig. 4 - figure supplement 1). This result was expected as the mutation does not affect interactions present in the native state, but was designed explicitly to perturb a hydrogen bond in the intermediate. The ability of the N44A variant to release triNAG from its bound state was assessed by surface plasmon resonance (SPR) experiments (Fig. 4c). By using a double-referenced single chain model (see Methods) we estimated that the destabilisation of the intermediate in the N44A mutant reduces the  $k_{\text{off}}$  by a factor 3, while changing the  $K_d$  by a factor 1/3. The variation of a factor 2 of the  $k_{\text{on}}$  suggests that the pathways of capture and release are not completely distinct and thus perturbing the pathway for release affects in part also that of capture. These results are consistent with our prediction that the N44A mutation affects the stability of the EP\* intermediate, i.e. of the unlocked state. Finally, we tested the catalytic efficiency of the wild type and N44A mutant by using a cellular assay (see Methods), and compared these results with those obtained for an inactive mutational variant lacking the catalytic residue (E35D). The results (Fig. 4d) show that the N44A variant has an intermediate activity between wild type and the totally inactive control E35D variant, which is again consistent with the prediction that the mutation of asparagine to alanine of residue 44 would reduce the efficiency of the product release in such a way to affect the catalytic activity of the enzyme. Finally, to verify that the N44A modified the free energy landscape of lysozyme by reducing the population of the unlocked state, we performed

<sup>15</sup>N-<sup>1</sup>H RDC measurements on the N44A mutant and carried out restrained molecular dynamics simulations to determine its free energy landscape. The results (see Fig. 4 – figure supplement 2a) demonstrate that the unlocked state is not appreciably populated in the N44A mutant.

To further validate the conclusion that the structure that we have determined of the unlocked state represents a release intermediate we designed a second mutational variant to stabilise the unlocked state, rather than destabilising it as the N44A mutation. In the new mutant, N46Q/V110Q, a strong glutamine-glutamine interaction is inserted with the purpose to stabilise the ‘unlocked’ state in its conformation (Fig. 4 – figure supplement 3a). We have verified the folding of the mutant by NMR (Fig. 4 – figure supplement 3b) and measured the binding constants of the ligand for the unlocked state by SPR to show that it corresponds to a weaker binding affinity (Fig. 4 – figure supplement 3c). While the  $K_d$  of the wild type is about 9  $\mu$ M, the  $K_d$  of the N46Q/V110Q mutant is high almost beyond detection, indicating that the mutant essentially does not bind the substrate. These experimentally measured binding constants are consistent with the observation that, considering that the free energy of the free state is the same, the binding free energy of the locked state is larger than that of the unlocked state because the free energy of the former is lower than that of the latter (Fig. 1).

## Conclusions

We have presented the atomic resolution structure of an intermediate associated with the product release in an enzymatic reaction. We have validated this structure by identifying a distinctive structural characteristic of this state, a transient hydrogen bond between the side-chains of residues N44 and E35. As this interaction stabilises the intermediate state but not the ground state, we introduced a mutational variant (N44A) that, by removing specifically the hydrogen bond, reduces the stability of the intermediate state, but not that of the ground state, and thus inhibits the release process. Our results provide an illustration of the manner in which conformational fluctuations can play a central role in enzymatic reactions by creating low-population intermediate states that facilitate the challenging step of release of the products of the catalytic reaction.

## Methods

### *Sample preparation*

Human lysozyme was expressed in *Pichia pastoris* and purified on an ion exchange column, as previously described<sup>32</sup>. <sup>15</sup>N ammonium sulfate and <sup>13</sup>C methanol were used to <sup>15</sup>N and <sup>13</sup>C label the protein, respectively. NMR experiments were carried out using a 700 MHz spectrometer at 37 °C in a buffer at pH 5.0 containing 20 mM potassium phosphate and 10% D<sub>2</sub>O; the pH was re-adjusted after the addition of the protein. Protein concentrations were in the range of 200 to 350 μM. For the measurements of the bound state, N,N',N''-triacylchitotriose (triNAG) sugar was purchased from Sigma and dissolved in water to constitute a concentrated stock solution.

### *Assignment of NMR spectra*

For the assignment of the free state at pH 5.0 and 37 °C we used a previously published assignment<sup>33,34</sup>, which was confirmed using HNCA measurements, which was run with a spectral width of 1561 Hz and 68 points in the <sup>15</sup>N dimension, and a spectral width of 5456 Hz and 64 points in the <sup>13</sup>C dimension<sup>35</sup>. In total, 126 backbone amides were assigned in the <sup>1</sup>H-<sup>15</sup>N spectrum.

For the full assignment of human lysozyme bound to triNAG we performed titrations of <sup>1</sup>H-<sup>15</sup>N HSQC spectra of a 200μM sample of <sup>15</sup>N human lysozyme, which were recorded using progressive concentration of the ligand (0, 0.3, 0.5, 0.8, 1.1, 1.6, 2.4, 3.1, 5.2 and 10 equivalents), allowing us to sample different points along the binding curve. HSQC spectra were recorded with a spectral width of 1621 Hz and 128 points in the <sup>15</sup>N dimension (Figure 1 - figure supplement 2). Additional information was obtained using HNCA and HNCACB experiments of a triNAG-saturated human lysozyme sample<sup>35,36</sup>. The HNCA experiment was carried out with the same settings as for the free state (see above). The HNCACB experiment was carried out with a spectral width of 1561 Hz and 68 points in the <sup>15</sup>N dimension, and with a spectral width of 13210 Hz and 72 points in the <sup>13</sup>C dimension. These complementary data allowed us to obtain the full assignment of the <sup>1</sup>H-<sup>15</sup>N spectra (Figure 1 - figure supplement 1).

### ***Residual dipolar coupling measurements***

Residual dipolar couplings (RDCs) were measured by orienting the free and triNAG-bound states in two different bicelle solutions, neutral and charged<sup>37,38</sup>. The neutral bicelle solution contained 5% w/v of a mixture of DMPC and DHPC (q=2.9), whereas CTAB was used to create a positively charged solution of 10% w/v of the (DMPC:DHPC:CTAB) = (2.9:1:0.2) composition. Splitting of the <sup>2</sup>H signal was recorded before and after the IPAP experiments, to ensure that alignment had remained constant during the course of the NMR experiment. IPAP experiments were recorded on the isotropic sample as well as on the two anisotropic samples (neutral and charged)<sup>39</sup>. These experiments were performed using a spectral width of 2447 Hz with 320 points in the indirect <sup>15</sup>N dimension for the in-phase (IP) or anti-phase (AP) spectra. J-couplings were extracted in each medium and RDCs were derived, discarding overlapping and poorly defined peaks. For the free state, we extracted 109 RDCs in the steric medium and 110 in the charged medium; 109 RDCs were extracted for the bound state, both for steric and charged media.

### ***<sup>3</sup>J scalar coupling measurements***

<sup>3</sup>J HN-Hα couplings were obtained using HNHA experiments<sup>40</sup>, which were performed on the free and bound states using a 700 MHz spectrometer and a spectral width of 1454 Hz with 68 points in <sup>15</sup>N and 9800 Hz with 72-80 points in the indirect <sup>1</sup>H dimension. The <sup>3</sup>J HNHα couplings were extracted using the ratios of intensities of cross- (I<sub>X</sub>) and diagonal (I<sub>D</sub>) peaks<sup>41</sup>

$$I_X/I_D = -\tan^2(2 \pi \xi {}^3J) \quad (1)$$

with  $\xi=13.05$  ms.

Errors in the <sup>3</sup>J <sup>N</sup>HHα coupling values were based either on a 5% uncertainty or on the noise level for cross-peaks with intensities below the RMS noise of the HNHA spectrum, estimated using Sparky (Goddard, T. D., and D. G. Kneller. SPARKY 3. University of California, San Francisco, 2004). Errors on intensities were propagated according to Eq. (1) to yield the error on <sup>3</sup>J HNHα couplings. Residues with overlapping diagonal peaks were discarded, as well as glycine residues.

### ***Molecular dynamics simulations***

As a starting structure for the ligand-free state, we used the crystal structure of human lysozyme at 1.9 Å resolution (PDB code 2ZIJ). For the bound state, we used the crystal structure of the human lysozyme A96L variant bound to triNAG at 1.8 Å resolution (PDB code 1BB5). This structure was modelled by mutating back residue 96 from L to A, as in the wild-type sequence. Molecular dynamics simulations were performed by using AMBER99SB with corrections on backbone<sup>42</sup> and side chains<sup>43</sup> dihedral angles as the force field ( $E^{FF}$ ) for the protein. TriNAG was modeled using the GLYCAM06 force field<sup>44</sup>. The protein and protein/triNAG systems were solvated using the TIP3P water model<sup>45</sup>. A time step of 2 fs was used together with LINCS constraints<sup>46</sup>. Systems were energy minimised and equilibrated with positional restrained simulations of 20 ns, in which the heavy atoms of the protein and triNAG molecules were restrained to their Cartesian coordinates. For the free state, the resulting system box after equilibration was 5.55 x 6.16 x 5.56 nm<sup>3</sup>, with 5,698 water molecules for a total of 19,123 atoms. For the bound state, the resulting system box after equilibration was 6.15 x 5.62 x 5.99 nm<sup>3</sup>, with 6,131 water molecules for a total of 20,509 atoms.

The simulations were performed in the NPT ensemble by weak coupling the pressure and temperature with external baths. Temperature coupling was performed with the v-rescale method<sup>47</sup> with a coupling constant of 0.1 ps. The pressure was kept constant using the Berendsen method<sup>48</sup>, with a coupling constant of 1.0 ps and at a reference pressure of 1 bar. The isotropic compressibility value was set to 4.5 X 10<sup>-5</sup> bar<sup>-1</sup>. Electrostatic interactions were treated by using the particle mesh Ewald method<sup>49</sup>.

### ***Molecular dynamics simulations with RDC restraints***

We used replica-averaged RDC restraints in molecular dynamics simulations<sup>19-21,28</sup>. This method has been tested for its ability to sample interdomain motions in proteins<sup>20,21</sup>, as well as in multiple conformational states in fast exchange in the NMR measurements<sup>28</sup>. A recent study was carried to generate accurate ensembles of hen egg white lysozyme using RDC measured under the same conditions of the present work<sup>21</sup>. This investigation has defined the sampling method that we have used here to characterise the conformational properties of lysozyme using RDC restraints. The accuracy of the resulting ensemble was benchmarked using a large variety of NMR observables, including eight sets of RDCs.

Briefly, in this approach<sup>20,21</sup> the structural information provided by RDC measurements is imposed to restrain the molecular dynamics simulations by adding a term,  $E^{RDC}$ , to a standard molecular mechanics force field,  $E^{Pot}$

$$E^{Tot} = E^{Pot} + E^{RDC} \quad (2)$$

The resulting force field,  $E^{Tot}$ , is employed in the integration of the equations of motion. In this work, the restraint term,  $E^{RDC}$ , is given by<sup>20,21</sup>

$$E^{RDC} = \alpha \sum_i \left( D^{exp} - D^{calc} \right)^2 \quad (3)$$

where  $\alpha$  is the weight of the restraint term, and  $D^{exp}$  and  $D^{calc}$  are the experimental and calculated RDCs, respectively. The RDC of a given bond vector is calculated as<sup>20,21</sup>

$$D^{calc} = \frac{1}{M} \sum_m D_m \quad (4)$$

where  $m$  runs over the  $M$  replicas and  $D_m$  is the RDC of replica  $m$ , which is given by

$$D = D_{\max} \sum_{ij} \langle A_{ij} \rangle \cos j_i \cos j_j \quad (5)$$

where  $j_i$  and  $j_j$  are the angles between the internuclear vector and the molecular reference frame, the indices  $i$  and  $j$  run over the three Cartesian coordinates,  $x$ ,  $y$ , and  $z$ , and  $\langle A_{ij} \rangle$  is the  $(i,j)$  component of the alignment tensor.

The use of replica-averaged molecular dynamics simulations enables one to generate an ensemble of conformations compatible with the experimental data according to the maximum entropy principle<sup>22-24</sup>, at least in the limit of large  $M$  and  $\alpha$ . We have previously shown<sup>22</sup>, however, that it is possible to effectively achieve this limit even if the values of  $M$  and  $\alpha$  remain relatively small and thus obtain conformational ensembles that provide a good agreement between experimental and calculated observables. Following these procedures, we used here  $M = 16$  and for the weight,  $\alpha$ , we first carried out an initial equilibration simulation at 310 K, during which the agreement between the calculated and experimental data was allowed to converge by gradually raising  $\alpha$  to the

largest possible value that did not generate numerical instabilities. Subsequently, we performed a series of 50 cycles of simulated annealing between 310 and 500 K to sample the conformational space. Each cycle was carried out for a total of 250 ps (125,000 molecular dynamics steps). For each cycle we collected 24,000 structures (1 per ps in the final 50 ps of the final 30 cycles of each of the 16 replicas). These were employed for the analyses reported in this study.

The alignment tensor is calculated from the shape and charge of the protein molecule using a procedure recently described<sup>19</sup>. We adopted such an approach here rather than the more commonly used singular value decomposition (SVD) method<sup>26,50</sup> because in the presence of conformational fluctuations of relatively large amplitude, such as those exhibited by hen lysozyme, the SVD method, when used in combination with the replica-averaging procedure of Eqs. 2–5, is less effective in capturing the motions of a protein<sup>21</sup>. The reason is that the SVD method does not necessarily provide the actual alignment tensor of a given structure but rather the alignment tensor that generates the RDC values in closest agreement with the experimental ones and hence is less well suited in describing the specific differences between the structures considered in the averaging procedure in Eq. (3)<sup>19,21</sup>.

This structure based method was here used to calculate the orientations of lysozyme in two alignment media, one steric (DMPC/DHPC) and one electrostatic (DMPC/DHPC/CTAB). The Q factors for the refined ensembles of the free and bound states of human lysozyme were 0.10 in both cases.

In addition to the previous extensive benchmarks performed on the structural ensembles of the hen egg white lysozyme<sup>28</sup>, which were obtained using the same protocol employed in this work, we performed here a set of additional validations based on NMR measurements not used as restraints in the simulations and by comparing the resulting experimental values with those back-calculated from our ensemble of human lysozyme (Figure 1 - figure supplement 5).

### ***Mutagenesis***

N44A mutation and E35D or D53N (control mutations) were introduced into the pPIC9/HuLys wt by using *QuikChange*® XL II mutagenesis kit (Qiagen). The pPIC9 plasmid containing the point mutations of HuLys cDNA was linearised by digestion with

StuI. Transformation into *Pichia pastoris* GS115 was performed by using *Pichia* EasyComp Transformation Kit (Life Technologies), according to manufacturer instructions. Cell colonies were screened for lysozyme expression level by quantifying by NuPAGE analysis the amount of lysozyme produced in 10 ml mini-cultures. Protein expression and purification were performed as previously reported<sup>32</sup>. Protein purity exceeded 95% as estimated by NuPAGE analysis. Protein concentrations were determined by absorbance measurements at 280 nm using theoretical extinction coefficients calculated with Expasy ProtParam.

### ***Surface plasmon resonance***

Surface plasmon resonance (SPR) experiments were performed using a Biacore 3000 system (GE Healthcare). CM5 sensor chip surfaces were activated by using an amine coupling kit (GE Healthcare). WT and N44A lysozyme variants were immobilised to the activated surfaces by amine coupling at a density of 2500–3000 resonance units (RU). Single chain kinetic experiments were performed at 25 °C using a flow rate of 20 µl/min in 50 mM phosphate pH 6.2, 100 mM NaCl. Serial dilutions (200 µM, 100 µM, 50 µM, 25 µM and 12.5 µM) of N,N',N''-Triacetylchitotriose (Tri-NAG, Sigma Aldrich) were sequentially injected every 700 s using a contact time of 250 s for each injection. Data fitting was performed with the single chain kinetic module provided with the Biaevaluation software (Biacore GE lifesciences).

### ***Cellular assay of lysozyme activity***

Hydrolase activity assay was performed using *Micrococcus lysodeikticus* cells (Sigma Aldrich) as the substrate. Cells of *Micrococcus* were suspended at 0.3 mg/ml in 100mM potassium phosphate, pH 6.2, shortly before the assay. The decrease of Absorbance at 450 nm was monitored at 25°C in the presence of 20 nM lysozyme variants.

### **Acknowledgements**

This work was supported by grants from EPSRC, EMBO and EU (ADS), and BBSRC and Wellcome Trust (CMD and MV).

## Competing interests

The authors declare that they have no financial competing interests.

## References

- 1 Dobson CM. Protein folding and misfolding. *Nature* **426**, 884-890 (2003). doi:10.1038/nature02261.
- 2 Sekhar A, Kay LE. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. USA* **110**, 12867-12874 (2013). doi:10.1073/pnas.1305688110.
- 3 Tzeng S-R, Kalodimos CG. Allosteric inhibition through suppression of transient conformational states. *Nature Chem. Biol.* **9**, 462-465 (2013). doi:10.1038/nchembio.1250.
- 4 Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312-1316 (2010). doi:10.1126/science.1191723.
- 5 Tang C, Iwahara J, Clore GM. Visualization of transient encounter complexes in protein-protein association. *Nature* **444**, 383-386 (2006). doi:10.1038/nature05201.
- 6 Neudecker P *et al.* Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362-366 (2012). doi:10.1126/science.1214203.
- 7 Fersht AR. *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding.* (W. H. Freeman, 1999).
- 8 Blake CCF *et al.* Structure of hen egg-white lysozyme - a 3-dimensional fourier synthesis at 2a resolution. *Nature* **206**, 757-& (1965). doi:10.1038/206757a0.
- 9 Phillips DC. Hen egg-white lysozyme molecule. *Proc. Natl. Acad. Sci. USA* **57**, 484-495 (1967)
- 10 Artymiuk PJ, Blake CCF. Refinement of human lysozyme at 1.5 Å resolution analysis of nonbonded and hydrogen-bond interactions. *J. Mol. Biol.* **152**, 737-762 (1981). doi:10.1016/0022-2836(81)90125-x.

- 463 11 Radford SE, Dobson CM, Evans PA. The folding of hen lysozyme involves  
464 partially structured intermediates and multiple pathways. *Nature* **358**, 302-307  
465 (1992). doi:10.1038/358302a0.
- 466 12 Chipman DM, Sharon N. Mechanism of lysozyme action. *Science* **165**, 454-&  
467 (1969). doi:10.1126/science.165.3892.454.
- 468 13 Warshel A, Levitt M. Theoretical studies of enzymic reactions - dielectric,  
469 electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J.*  
470 *Mol. Biol.* **103**, 227-249 (1976). doi:10.1016/0022-2836(76)90311-9.
- 471 14 Post CB *et al.* Molecular-dynamics simulations of native and substrate-bound  
472 lysozyme - a study of the average structures and atomic fluctuations. *J. Mol. Biol.*  
473 **190**, 455-479 (1986). doi:10.1016/0022-2836(86)90015-x.
- 474 15 Vocadlo DJ, Davies GJ, Laine R, Withers SG. Catalysis by hen egg-white  
475 lysozyme proceeds via a covalent intermediate. *Nature* **412**, 835-838 (2001).  
476 doi:10.1038/35090602.
- 477 16 Turner MA, Howell PL. Structures of partridge egg-white lysozyme with and  
478 without tri-n-acetylchitotriose inhibitor at 1.9-angstrom resolution. *Protein Sci.* **4**,  
479 442-449 (1995)
- 480 17 Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules  
481 by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-1114 (1997).  
482 doi:10.1126/science.278.5340.1111.
- 483 18 Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. NMR evidence for slow  
484 collective motions in cyanometmyoglobin. *Nature Struct. Biol.* **4**, 292-297  
485 (1997). doi:10.1038/nsb0497-292.
- 486 19 Montalvao RW, De Simone A, Vendruscolo M. Determination of structural  
487 fluctuations of proteins from structure-based calculations of residual dipolar  
488 couplings *J. Biomol. NMR* **53**, 281-292 (2011). doi: 10.1007/s10858-012-9644-3.
- 489 20 De Simone A, Montalvao RW, Vendruscolo M. Determination of conformational  
490 equilibria in proteins using residual dipolar couplings. *J. Chem. Theor. Comp.* **7**,  
491 4189–4195 (2011).
- 492 21 De Simone A, Montalvao RW, Dobson CM, Vendruscolo M. Characterization of  
493 the interdomain motions in hen lysozyme using residual dipolar couplings as  
494 replica-averaged structural restraints in molecular dynamics simulations.  
495 *Biochemistry* **52**, 6480-6486 (2013)

- 496 22 Cavalli A, Camilloni C, Vendruscolo M. Molecular dynamics simulations with  
497 replica-averaged structural restraints generate structural ensembles according to  
498 the maximum entropy principle. *J. Chem. Phys.* **138**, 094112 (2013)
- 499 23 Pitera JW, Chodera JD. On the use of experimental observations to bias simulated  
500 ensembles. *J. Chem. Theor. Comp.* **8**, 3445-3451 (2012). doi:10.1021/ct300112v.
- 501 24 Roux B, Weare J. On the statistical equivalence of restrained-ensemble  
502 simulations with the maximum entropy method. *J. Chem. Phys.* **138**, 084107  
503 (2013). doi:10.1063/1.4792208.
- 504 25 Bouvignies G, Markwick P, Bruschweiler R, Blackledge M. Simultaneous  
505 determination of protein backbone structure and dynamics from residual dipolar  
506 couplings. *J. Am. Chem. Soc.* **128**, 15100-15101 (2006). doi:10.1021/ja066704b.
- 507 26 Clore GM, Schwieters CD. How much backbone motion in ubiquitin is required  
508 to account for dipolar coupling data measured in multiple alignment media as  
509 assessed by independent cross-validation? *J. Am. Chem. Soc.* **126**, 2923-2938  
510 (2004). doi:10.1021/ja0386804.
- 511 27 Lange OF *et al.* Recognition dynamics up to microseconds revealed from an  
512 RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471-1475 (2008).  
513 doi:10.1126/science.1157092.
- 514 28 De Simone A *et al.* Structures of the excited states of phospholamban and shifts  
515 in their populations upon phosphorylation. *Biochemistry* **52**, 6684-6694 (2013)
- 516 29 Bouvignies G *et al.* Solution structure of a minor and transiently formed state of a  
517 T4 lysozyme mutant. *Nature* **477**, 111-114 (2011). doi:10.1038/nature10349.
- 518 30 Korzhnev DM *et al.* Low-populated folding intermediates of Fyn SH3  
519 characterized by relaxation dispersion NMR. *Nature* **430**, 586-590 (2004)
- 520 31 De Simone A, Richter B, Salvatella X, Vendruscolo M. Toward an accurate  
521 determination of free energy landscapes in solution states of proteins. *J. Am.*  
522 *Chem. Soc.* **131**, 3810-3811 (2009). doi:10.1021/ja8087295.
- 523 32 Johnson RJK *et al.* Rationalising lysozyme amyloidosis: Insights from the  
524 structure and solution dynamics of T70N lysozyme. *J. Mol. Biol.* **352**, 823-836  
525 (2005). doi:10.1016/j.jmb.2005.07.040.
- 526 33 Ohkubo T, Taniyama Y, Kikuchi M. H-1 and N-15 NMR-study of human  
527 lysozyme. *J. Biochem.* **110**, 1022-1029 (1991)
- 528 34 Hagan CL *et al.* A non-natural variant of human lysozyme (i59t) mimics the in  
529 vitro behaviour of the I56T variant that is responsible for a form of familial

530 amyloidosis. *Prot. Eng. Des. Sel.* **23**, 499-506 (2010).  
 531 doi:10.1093/protein/gzq023.

532 35 Grzesiek S, Bax A. Improved 3D triple-resonance NMR techniques applied to a  
 533 31-kDa protein. *J. Mag. Res.* **96**, 432-440 (1992). doi:10.1016/0022-  
 534 2364(92)90099-s.

535 36 Muhandiram DR, Kay LE. Gradient-enhanced triple-resonance 3-dimensional  
 536 NMR experiments with improved sensitivity. *J. Mag. Res. B* **103**, 203-216  
 537 (1994). doi:10.1006/jmrb.1994.1032.

538 37 Schwalbe H *et al.* A refined solution structure of hen lysozyme determined using  
 539 residual dipolar coupling data. *Protein Sci.* **10**, 677-688 (2001).  
 540 doi:10.1110/ps.43301.

541 38 Ottiger M, Bax A. Characterization of magnetically oriented phospholipid  
 542 micelles for measurement of dipolar couplings in macromolecules. *J. Biomol.*  
 543 *NMR* **12**, 361-372 (1998). doi:10.1023/a:1008366116644.

544 39 Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from  
 545 simplified two-dimensional NMR spectra. *J. Mag. Res.* **131**, 373-378 (1998).  
 546 doi:10.1006/jmre.1998.1361.

547 40 Vuister GW, Bax A. Quantitative J correlation - a new approach for measuring  
 548 homonuclear 3-bond J(h(n)h(alpha) coupling-constants in N-15-enriched  
 549 proteins. *J. Am. Chem. Soc.* **115**, 7772-7777 (1993). doi:10.1021/ja00070a024.

550 41 Kuboniwa H, Grzesiek S, Delaglio F, Bax A. Measurement of H-N-H-alpha J-  
 551 couplings in calcium-free calmodulin using new 2D and 3D water-flip-back  
 552 methods. *J. Biomol. NMR* **4**, 871-878 (1994). doi:10.1007/bf00398416.

553 42 Best RB, Hummer G. Optimized molecular dynamics force fields applied to the  
 554 helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004-9015 (2009).  
 555 doi:10.1021/jp901540t.

556 43 Lindorff-Larsen K *et al.* Improved side-chain torsion potentials for the amber  
 557 ff99sb protein force field. *Proteins* **78**, 1950-1958 (2010).  
 558 doi:10.1002/prot.22711.

559 44 Kirschner KN *et al.* Glycam06: A generalizable biomolecular force field.  
 560 Carbohydrates. *J. Comp. Chem.* **29**, 622-655 (2008). doi:10.1002/jcc.20820.

561 45 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison  
 562 of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-  
 563 935 (1983). doi:10.1063/1.445869.

- 46 Hess B. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theor. Comp.* **4**, 116-120 (2008)
- 47 Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126** (2007). doi:10.1063/1.2408420.
- 48 Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690 (1984). doi:10.1063/1.448118.
- 49 Essmann U *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103** (1995)
- 50 Clore GM, Schwieters CD. Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: Correspondence of dipolar coupling and heteronuclear relaxation measurements. *Biochemistry* **43**, 10678-10691 (2004). doi:10.1021/bi049357w.
- 51 Shen Y, Bax A. Sparta+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **48**, 13-22 (2010). doi:10.1007/s10858-010-9433-9.

**Figure 1. Comparison of the free energy landscapes of human lysozyme in the free state (a) and in the bound state with triNAG (b).** The bound state exhibits a ground state (the ‘locked state’) and an additional local minimum with about 13% population (the ‘unlocked state’), which represents an intermediate in the release of the product of

the enzymatic reaction. Free energy landscapes are shown as function of the ‘breathing’ angle  $\theta$  and of the RMSD from the X-ray structure, which was calculated on the C $\alpha$  atoms by including secondary structure regions only, of a human lysozyme variant in complex with triNAG (PDB code 1BB5); free energy landscapes were obtained as  $-k_B T \ln H(\theta, \text{RMSD})$ , where  $H(\theta, \text{RMSD})$  is the number of times conformations with specific  $\theta$  and RMSD values was sampled during the simulations<sup>21</sup>.

**Figure 2. Analysis of the interactions that stabilise the intermediate state in the release of the product (the ‘unlocked state’).** (a) Free energy landscape as a function of the angle  $\theta$ . (b) Potential energy landscape,  $E_{pot}$ , of lysozyme in the free state;  $E_{pot}$  represents the contribution of the force field used in the simulations, i.e. the total force field without the RDC restraint term (see Methods). (c) Potential energy landscape,  $E_{pot}$ , of the lysozyme-triNAG bound state. (d) Structure of the ‘locked state’. (e) Structure of the ‘unlocked state’.

**Figure 3. Schematic illustration of the process of product release.** The product (P) is released by the enzyme (E) in a process that begins in the ground (or ‘locked’) state of the complex (EP), visits a metastable (or ‘unlocked’) intermediate state (EP\*) and reaches the unbound state (E+P). The interactions in the ‘locked state’ (EP) and in the ‘unlocked state’ (EP\*) are shown in light blue.

**Figure 4. Experimental assessment of the role of the intermediate state determined in this work in the product release process.** (a) Structure of the unlocked state illustrating the hydrogen bond between side chains of N44 and E35 that stabilises this intermediate species. (b) The hydrogen bond is not formed in the locked state because N44 and E35 are too far apart. (c) The N44A variant, which lacks the hydrogen bond donor, is unable to form this hydrogen bond, thus destabilising the intermediate state and inhibiting the release of the product. The decrease of the ability of the N44A mutant to release triNAG has been assessed by surface plasmon resonance (SPR) experiments. (d) Cellular assay of lysozyme activity. The N44A variant has an intermediate activity

between wild type and the control E35D variant.

**Figure 1 - figure supplement 1.** Assignments of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the free and triNAG-bound states of human lysozyme. The x-axis represents the  $^1\text{H}$  (in ppm) dimension and the y-axis represents the  $^{15}\text{N}$  (in ppm) dimension

**Figure 1 - figure supplement 2.** Extracts of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra showing the titration of triNAG to human lysozyme for selected residues showing significant chemical shift changes upon binding. Depending on the individual residues, the time scale of exchange is fast or intermediate. Blue: free state (f); Yellow: 0.5 equivalents of sugar; Red: 1.1 equivalents; Green: 2.4 equivalents; Blue: 10 equivalents, corresponding to the bound state (b). The x-axis represents the  $^1\text{H}$  (in ppm) dimension and the y-axis represents the  $^{15}\text{N}$  dimension (in ppm).

**Figure 1 - figure supplement 3.** Illustration of the breathing angle  $\theta$  of lysozyme<sup>21</sup>, which accounts for the large-amplitude motion between the  $\alpha$ -domain and  $\beta$ -domain of lysozyme and is computed from the centres of mass of  $\text{C}\alpha$ -atoms from three protein regions<sup>21</sup>. Region 1 (in the  $\alpha$ -domain) spans residues 28- 31 and 111-114 (in red), region 2 (in the hinge region) spans residues 90-93 (in yellow) and region 3 (in the  $\beta$ -domain) spans residues 44, 45, 51 and 52 (in green).

**Figure 1 - figure supplement 4.** (A,B) Experimentally-measured  $^{15}\text{N}$ - $^1\text{H}$  residual dipolar couplings (RDCs) of human lysozyme in the free state (A) and the triNAG-bound state (B). (C,D) Representation of the regions that are mostly affected by triNAG binding in the steric (C) and electrostatic (D) RDC measurements (red indicates small changes, and blue large changes).

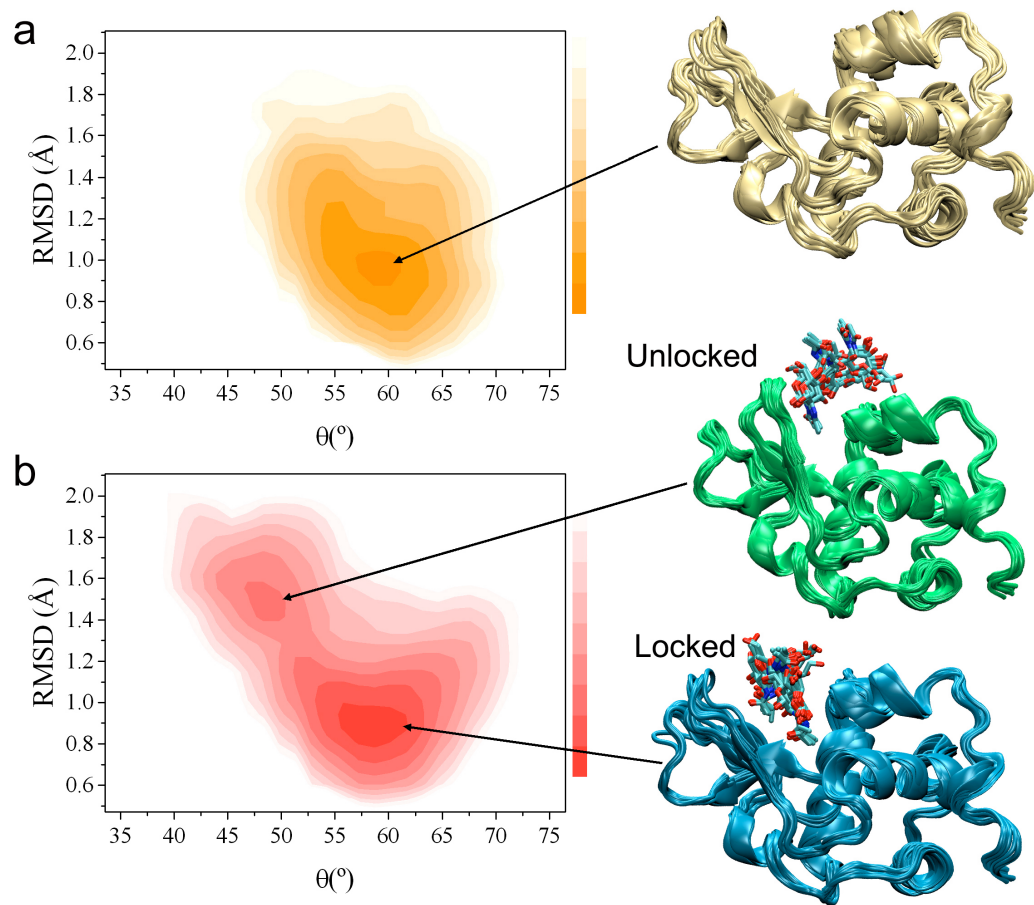
**Figure 1 - figure supplement 5.** Validation of the RDC-refined structural ensembles determined in this work representing the free and triNAG-bound states of human lysozyme. (A) Comparison between calculated and experimental RDCs (black circles for the steric medium and red circles for the electrostatic medium); the Q factors at 0.10 in

both cases. (B) Comparison between experimental (black) and calculated (red)  $^3J$  HN-H $\alpha$  scalar couplings; the RMSD value is 0.49 Hz. (C-G) Comparison between experimental and calculated chemical shifts, which we obtained by using the Sparta+ method<sup>51</sup>: C $\alpha$  (C), C $\beta$  (D), N (E), H $\alpha$  (F), and HN (G); the RMSD values (in ppm) are 0.90, 0.82, 1.90, 0.16 and 0.24 for C $\alpha$ , C $\beta$ , N, H $\alpha$  and HN, respectively. H) Distributions of the satisfied (about 93%, in green) and violated (about 7%, in orange) NOEs in the individual structures of the ensemble.

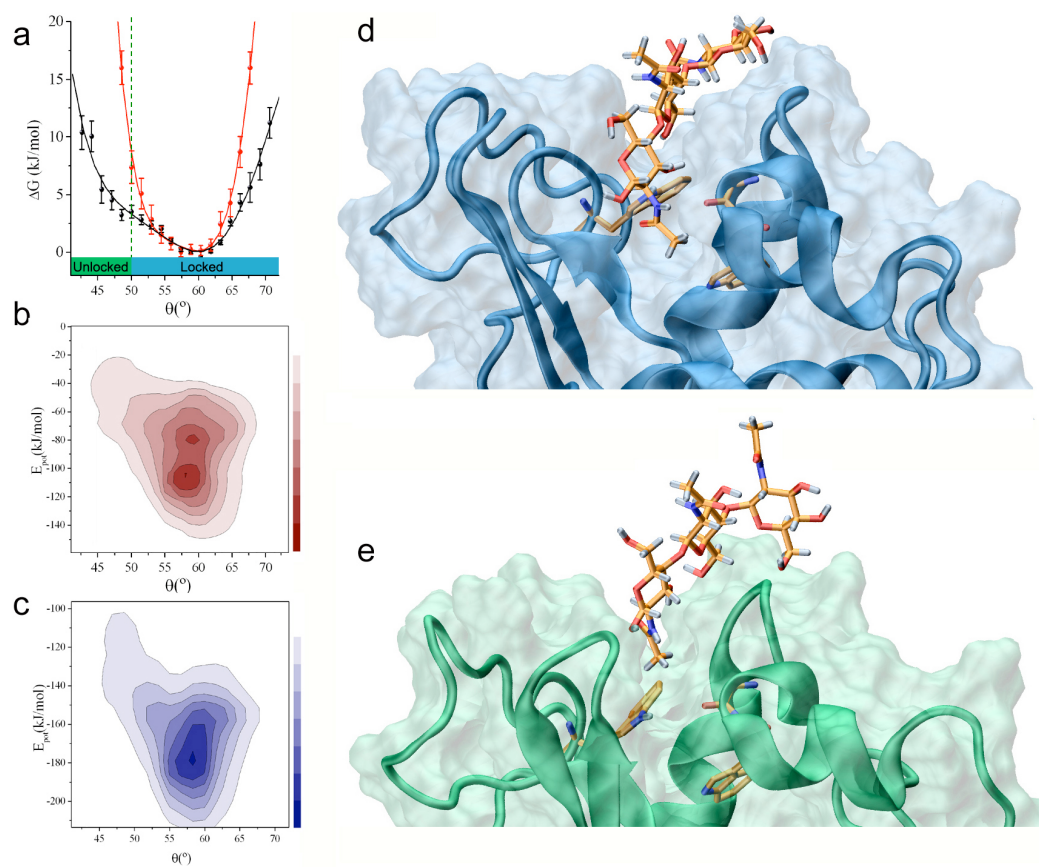
**Figure 4 - figure supplement 1.** Comparison of the  $^1H$ - $^{15}N$  HSQC spectra of WT (black) and N44A mutant (red).

**Figure 4 - figure supplement 2.** Comparison of the free energy landscapes of wild type (red) and N44A mutant (black) lysozyme. In the N44A mutant the unlocked state (see also Fig. 1) is absent. Free energy landscapes were obtained as described in Fig. 1.

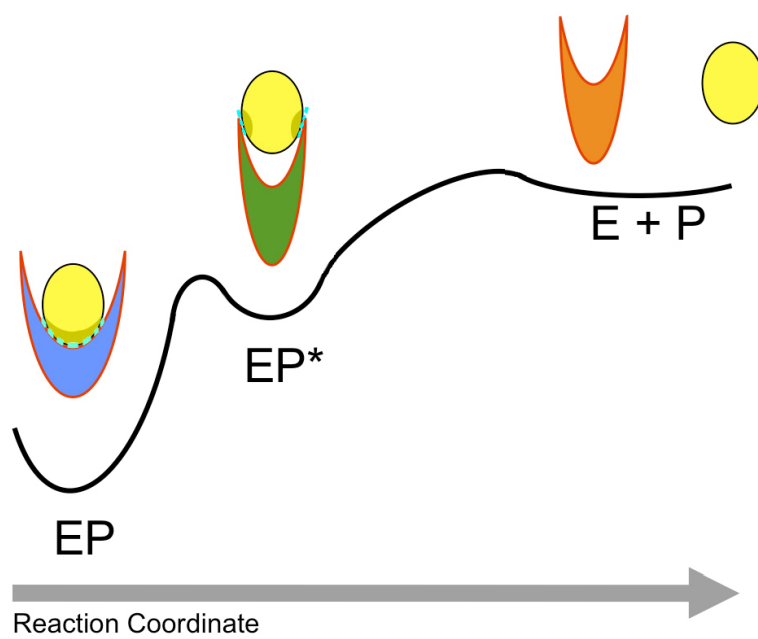
**Figure 4 - figure supplement 3.** Study of the N46Q/V110Q mutant. (a) Illustration of the engineered N46Q/V110Q glutamine-glutamine interactions. (b) The  $^1H$ - $^{15}N$ -HSQC spectrum of the N46Q/V110Q mutant shows that the mutation does not affect the structural properties of the mutant (see also Fig. 4 - figure supplement 1). (c) The increase in the ability of the N46Q/V110Q mutant to release triNAG has been assessed by surface plasmon resonance (SPR) experiments.



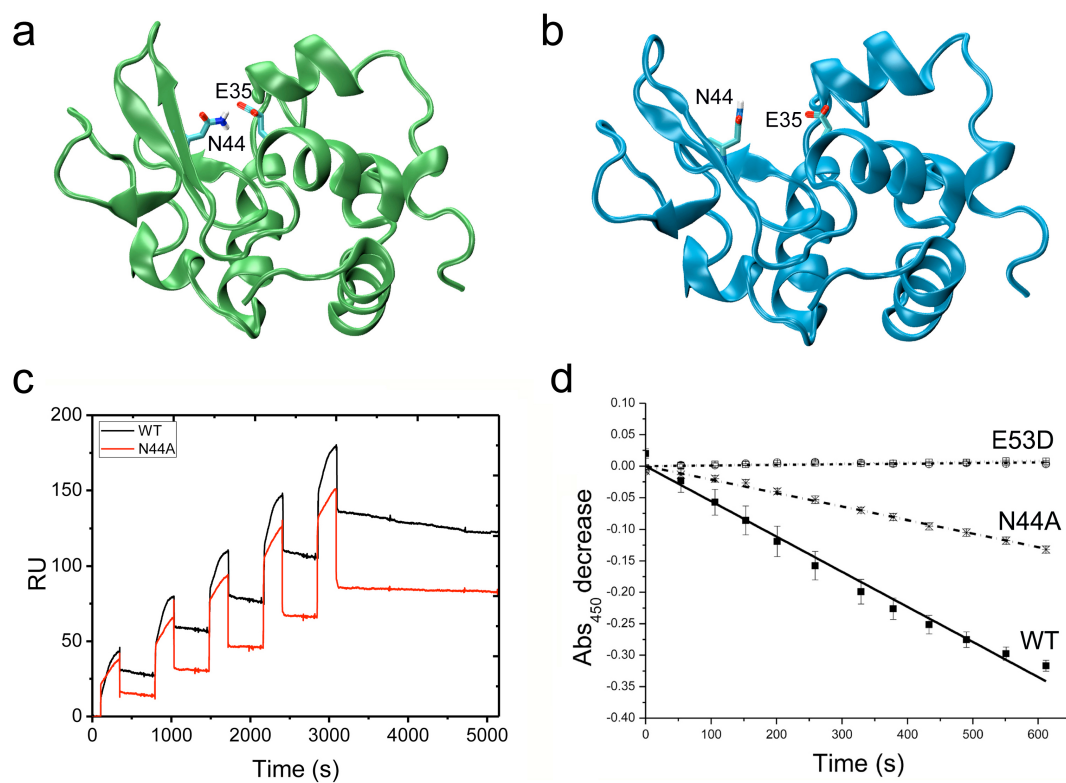
**Figure 1.**



**Figure 2.**



**Figure 3.**



**Figure 4.**