
**Volume 74 (2018)**

**Supporting information for article:**

# Effects of protein crystal hydration and temperature on side-chain conformational heterogeneity in monoclinic lysozyme crystals

**Hakan Atakisi, David W. Moreau and Robert E. Thorne**

**Table S1** Protein volume, solvent volume, and ordered solvent volume per protein molecule for monoclinic lysozyme crystals as determined by the rolling probe method using the program 3VEE with a probe size of 1.4 Å (radius of water) and a grid size of 0.07 Å.

The calculated volumes for lysozyme become nearly constant for a grid size below ~0.1. For comparison, the solvent content and volume derived from the Matthews coefficient are also given, where the solvent content (volume solvent/volume unit cell) is calculated as $1-1.23/V_m$. This latter value assumes a protein density of 1.35 g/cm$^3$. The 3VEE-determined protein volume at 298 K is 16908 Å$^3$, corresponding to a protein density of ~1.41 g/cm$^3$, for the native structure, and these values do not vary appreciably with dehydration. and without a systematic dependence on hydration level. The solvent excluded volume of lysozyme calculated by 3VEE is consistent with results we calculated with other widely used programs based on rolling probe method such as MSMS (Sanner *et al.*, 1996) (16637 Å$^3$ or protein density of ~1.43 g/cm$^3$, using a probe radius of 1.4 Å and the highest density of pixel points of 10) and EDTSurf (Xu & Zhang, 2009) (16441 Å$^3$ or protein density of ~1.45 g/cm$^3$, using a probe radius of 1.4 Å, a scale factor of 20, and MC triangulation.) Previous analysis using the Connolly algorithm and properly accounting for the surface atoms obtained an average protein density near 1.43 g/cm$^3$ (Quillin & Matthews, 2000), and found that protein density increases with decreasing molecular weight, especially for proteins with MW below 20 kDa (Fischer *et al.*, 2004). Based on the reported exponential fit, the predicted density for lysozyme with molecular weight of 14.3 kDa is ~1.46 g/cm$^3$ (volume of ~16270 Å$^3$), consistent with our results.

| | | Matthews coefficient method | | Rolling probe method (3VEE) | | | |
|---|---|---|---|---|---|---|---|
| | Cell volume (Å$^3$) | Solvent content (%) | Solvent volume (Å$^3$) | Protein volume (Å$^3$) | Solvent volume (Å$^3$) | Solvent content (%) | Ordered solvent volume (Å$^3$) |
| **h99** | 27055 | 34.9 | 9442 | 16908 | 10148 | 37.5 | 1880 |
| **h93** | 25583 | 31.2 | 7982 | 16722 | 8861 | 34.6 | 2374 |
| **h85** | 25215 | 30.2 | 7615 | 16581 | 8634 | 34.2 | 2789 |
| **h75** | 22550 | 22.0 | 4961 | 16552 | 5998 | 26.6 | 2451 |
| **h58** | 21750 | 19.1 | 4154 | 16575 | 5175 | 23.8 | 1404 |
| **h33** | 20840 | 15.6 | 3251 | 16689 | 4151 | 19.9 | 994 |
| **h11** | 20475 | 14.1 | 2887 | 16747 | 3728 | 18.2 | 413 |
| **h99-c** | 25213 | 30.2 | 7614 | 16496 | 8717 | 34.6 | 4230 |
| **h85-c** | 24160 | 27.2 | 6572 | 16421 | 7740 | 32.0 | 5025 |
| **h75-c** | 21681 | 20.9 | 4531 | 16472 | 5209 | 24.0 | 3697 |
| **h58-c** | 21428 | 16.7 | 3578 | 16441 | 4987 | 23.3 | 2634 |
| **h11-c** | 20215 | 13.0 | 2628 | 16465 | 3750 | 18.6 | 912 |

**Table S2** Data collection and refinement statistics for room temperature and T=100 K data collected using tetragonal thaumatin crystals.

| | **r.h. 99%** | **r.h. 93%** | **r.h. 89%** | **r.h. 75%** | **r.h. 99% cryo** |
|---|---|---|---|---|---|
| Wavelength (Å) | 0.6299 | 0.6309 | 0.6309 | 0.6309 | 0.9770 |
| Space group | P $4_12_12$ | P $4_12_12$ | P $4_12_12$ | P $4_12_12$ | P $4_12_12$ |
| a | 58.80 | 58.32 | 57.79 | 53.15 | 57.91 |
| b | 58.80 | 58.32 | 57.79 | 53.15 | 57.91 |
| c | 152.37 | 152.22 | 151.65 | 152.48 | 150.19 |
| Solvent content (%) | 58.5 | 57.8 | 56.8 | 49.3 | 56.6 |
| Total reflections | 57979 | 63573 | 39139 | 39416 | 46060 |
| Redundancy | 3.3 (1.9) | 4.9 (2.7) | 4.4 (1.8) | 8.5 (4.0) | 6.8 (4.5) |
| Completeness (%) | 99.4 (96.1) | 97.7 (77.4) | 92.6 (58.2) | 99.2 (90.6) | 99.6 (93.4) |
| Mosaicity (°) | 0.04-0.08 | 0.16-0.47 | 0.04-0.58 | 0.07-0.50 | 0.28-0.40 |
| $CC_{1/2}$ | 0.996 (0.503) | 0.988 (0.154) | 0.996 (0.197) | 0.992 (0.323) | 1.000 (0.756) |
| Mean I/σ | 9.0 (0.8) | 8.6 (0.5) | 6.9 (0.5) | 7.7 (0.6) | 10.0 (1.2) |
| Resolution (Å) (at I/σ=2) | 1.52 | 1.55 | 1.77 | 1.67 | 1.51 |
| R-merge/R-meas (%) | 0.068/0.074 | 0.110/0.119 | 0.135/0.137 | 0.201/0.211 | 0.065/0.075 |
| Wilson B-factor (Å) | 16.8 | 20.2 | 20.1 | 17.6 | 15.6 |
| Refinement resolution range (Å) | 27.06-1.36 | 27.23-1.30 | 28.38-1.49 | 23.76-1.45 | 45.86-1.45 |
| R-work/R-free (%) | 12.0/14.2 | 12.9/16.1 | 13.6/18.1 | 14.0/17.8 | 13.2/15.5 |
| Test set for R-free (%) | 5 | 5 | 5 | 5 | 5 |
| RMS bonds (Å)/angles (°) | 0.007/0.938 | 0.008/0.934 | 0.008/0.951 | 0.009/0.982 | 0.007/0.941 |
| Ramachandran favored/outliers (%) | 96.5/0.0 | 98.0/0.0 | 97.8/0.0 | 98.6/0.0 | 98.2/0.0 |
| Rotamer outliers (%) | 1.1 | 1.5 | 2.8 | 2.1 | 0.0 |
| Clashscore | 2.4 | 3.3 | 4.3 | 6.1 | 2.8 |
| B-factor ($Å^2$) | 22.1 | 27.5 | 25.6 | 23.8 | 18.6 |
| Macromolecule B-factor ($Å^2$) | 21.2 | 26.5 | 24.9 | 23.2 | 16.6 |
| Solvent B-factor ($Å^2$) | 35.0 | 41.9 | 38.0 | 34.7 | 31.2 |
| Observed waters | 157 | 158 | 133 | 120 | 325 |

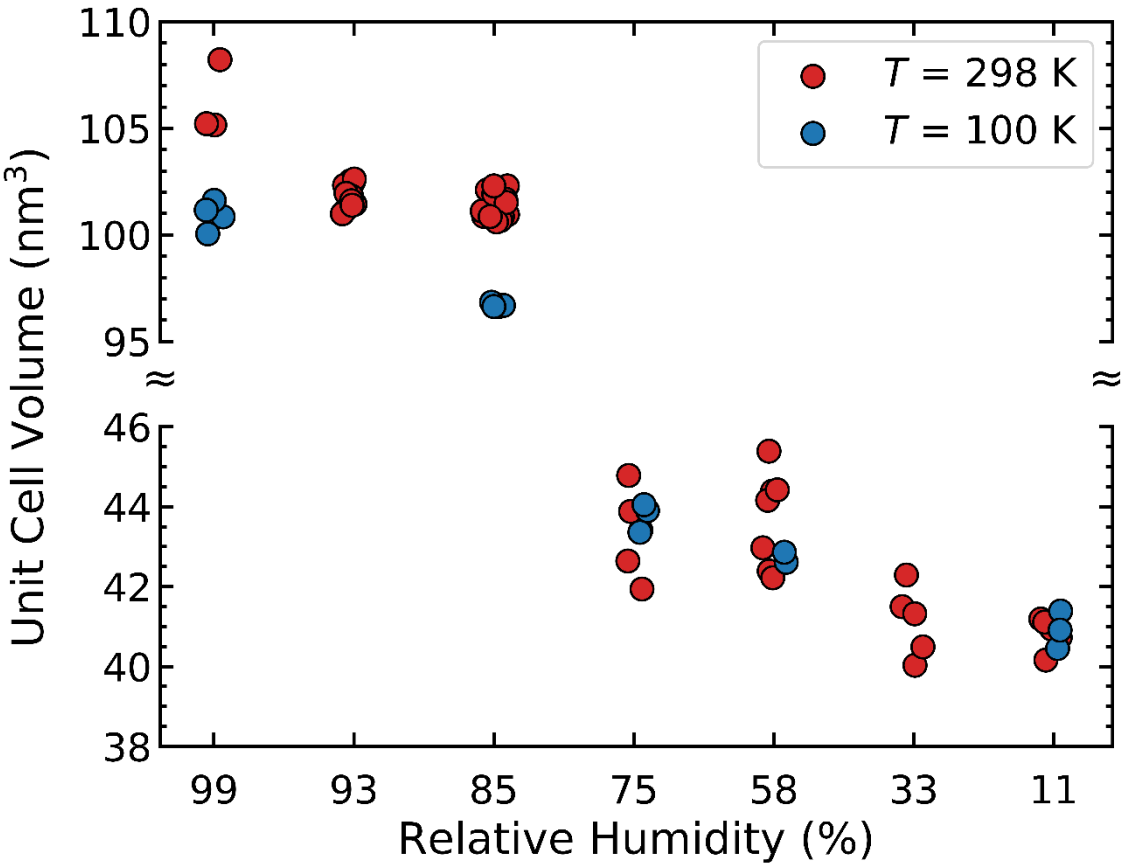



**Figure S1** Measured unit cell volumes for all crystals examined in this study at room temperature (red symbols) and at T=100 K (blue symbols). Cell volumes obtained at relative humidities below that of the structural transition that occurs between 86% and 75% r.h. show more scatter, perhaps reflecting differences in unit cell evolution through the transition and in the lattice disorder created by the transition.

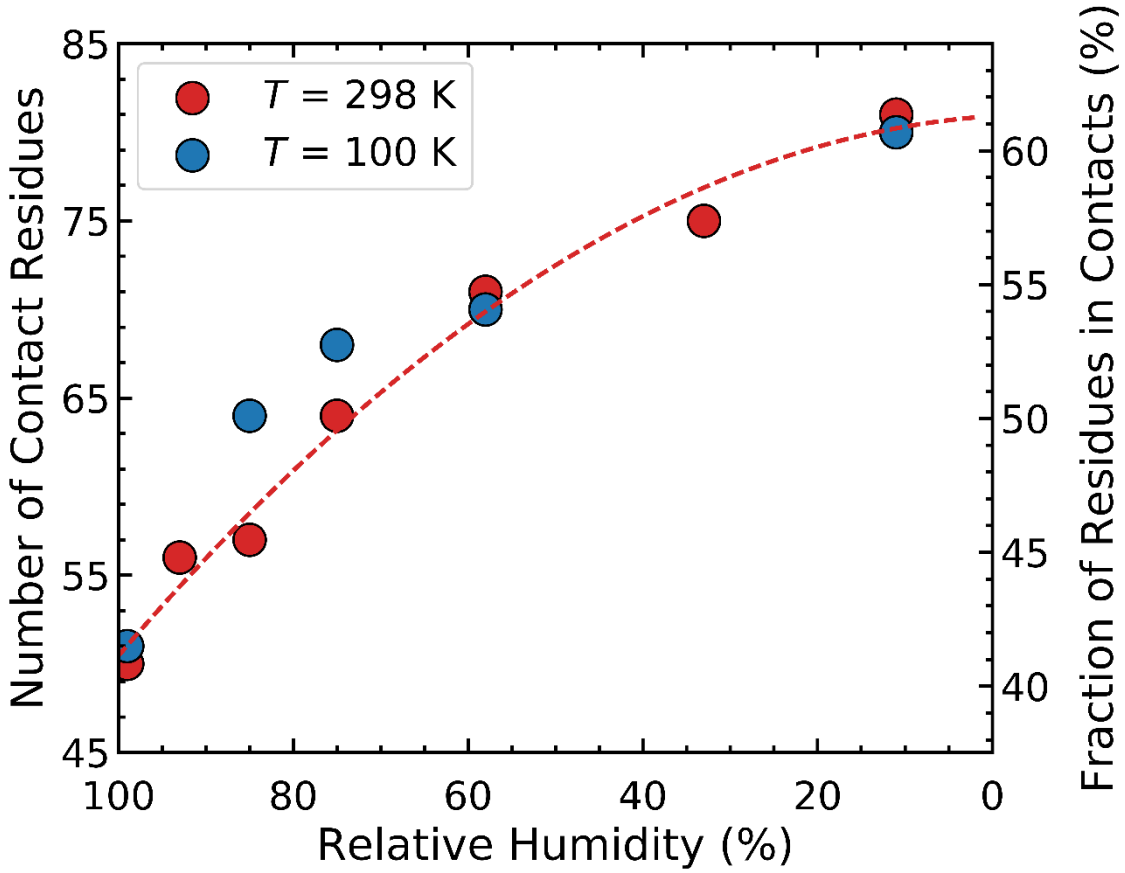



**Figure S2** Evolution of number of residues involved in crystal contacts as a function of relative humidity. Two atoms are considered to be in contact if their center to center distance is less than 4 Å.

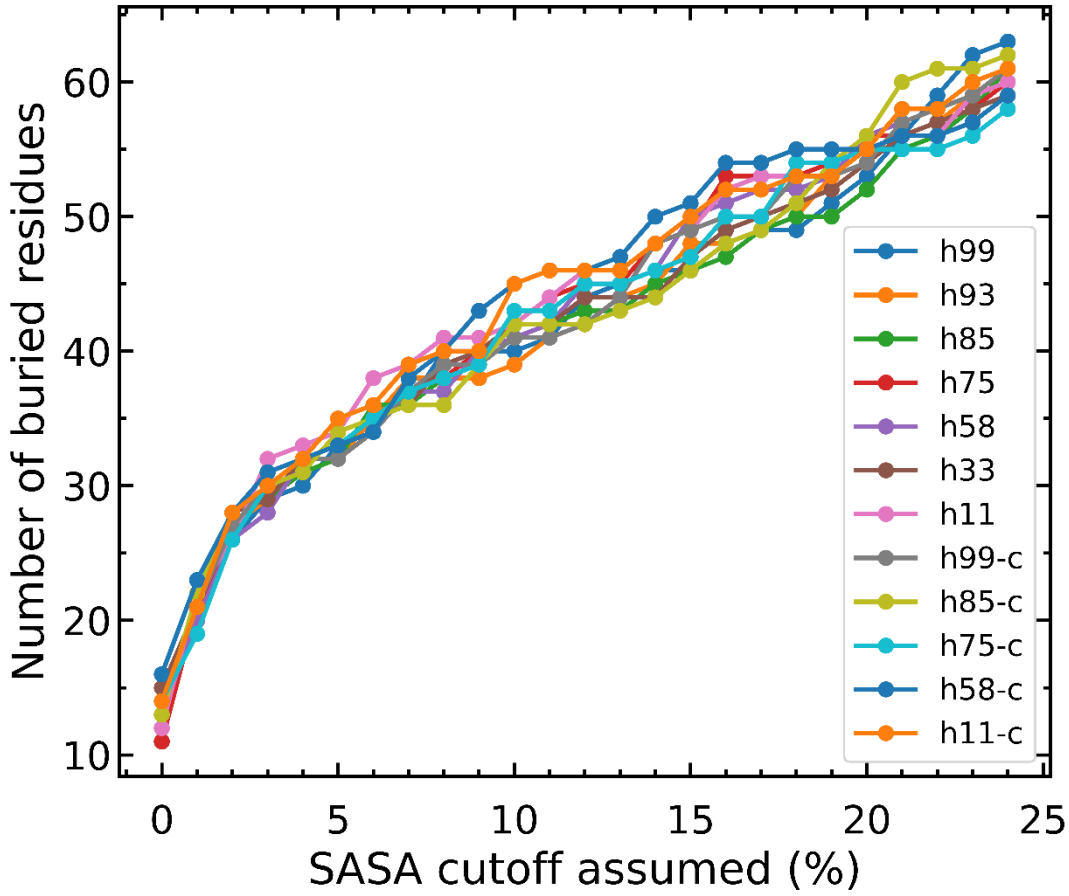



**Figure S3** Number of buried residues as a function of the cut-off percent assumed for the fraction of a given residue's "free" solvent accessible surface area surface area (SASA) that is solvent accessible in the crystal.

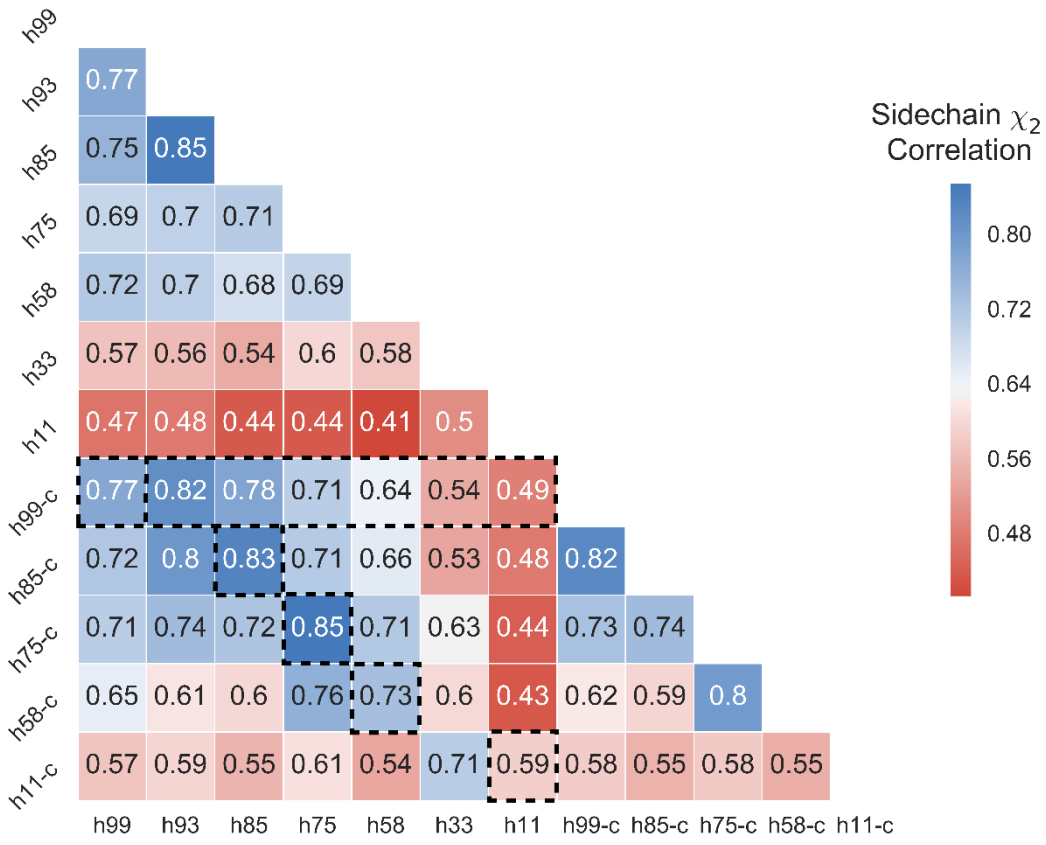



**Figure S4** Pairwise Pearson correlation coefficients for Ringer plots at $\chi_2$, averaged over all 72 residues with $\chi_2$ rotamers in lysozyme, for all possible pairs of dehydrated and cryocooled structures. Data sets collected at 100 K are labelled with "-c". ". The leftmost column indicates correlations with the native structure at room temperature; the diagonal boxed entries indicate correlations between room temperature and 100 K structures for crystals dehydrated to the same r.h., and the horizontal rectangle highlights correlation between the 100 K native structure and dehydrated structures at room temperature.

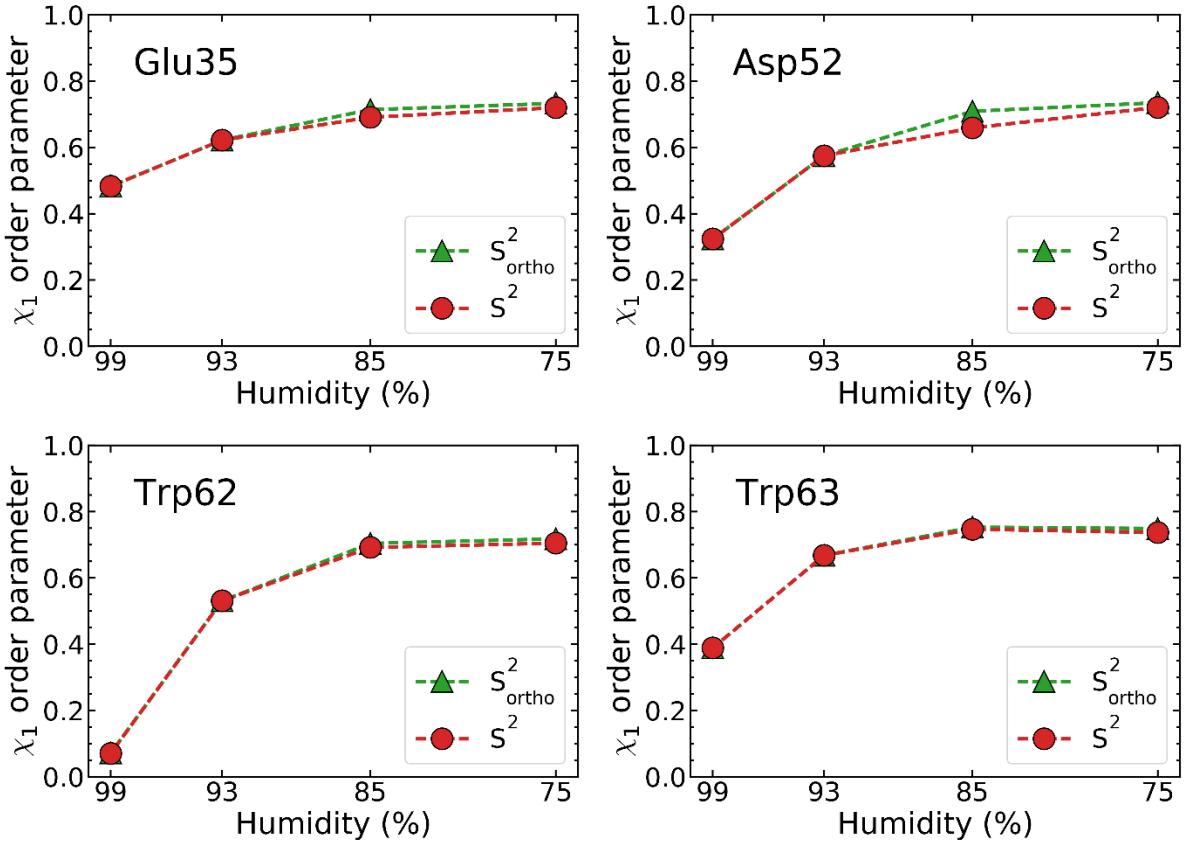



**Figure S5** Order parameter $S^2$ and its component $S^2_{ortho}$ versus relative humidity, for a selection of lysozyme residues at $\chi_1$, as deduced using multiconformer refinement from room temperature data sets. The same data plotted as $1-S^2$ versus unit cell volume is shown in Fig. 8.

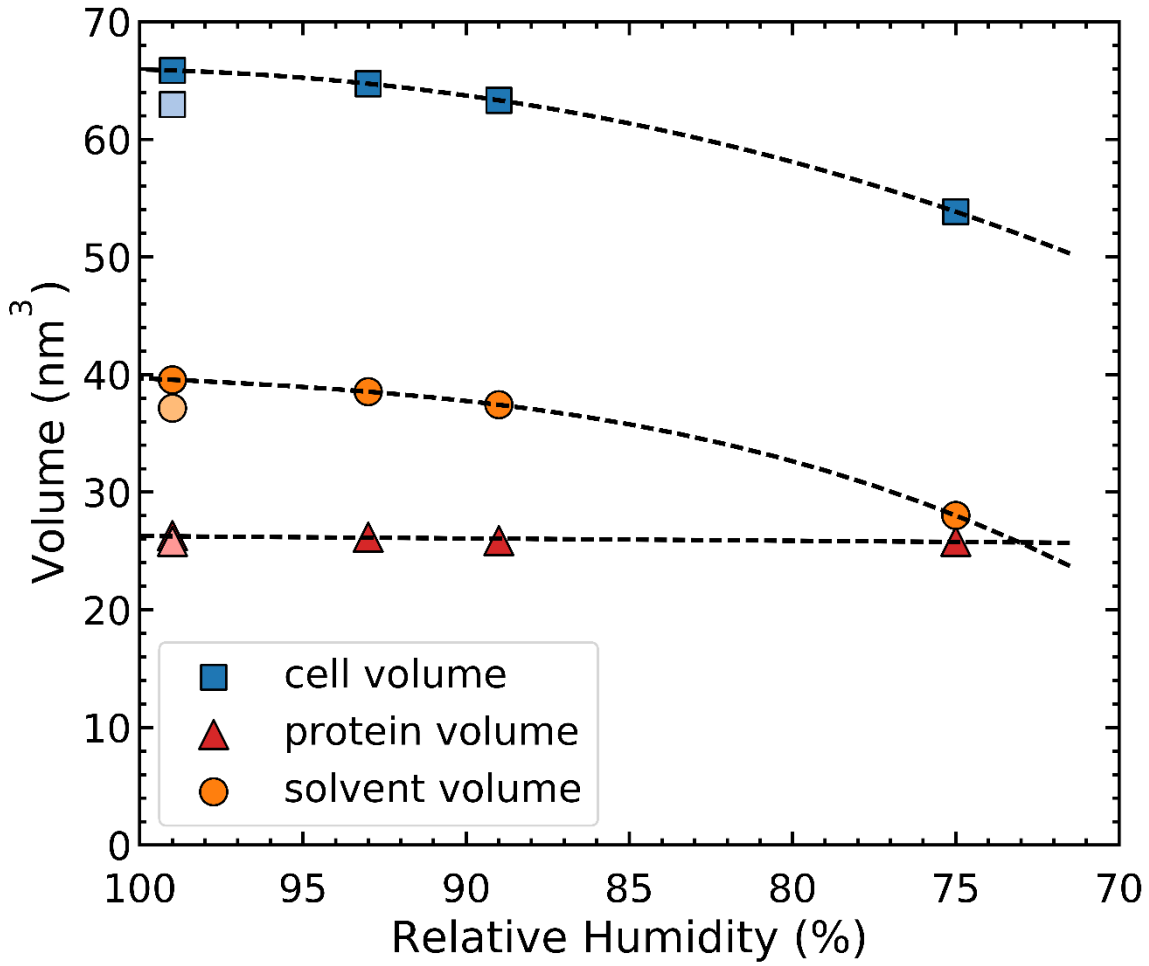



**Figure S6** Crystallographic cell volume ( ■ ), protein volume ( ▲ ) (calculated as the volume enclosed by solvent excluded surface (SES) area of the protein molecule), and solvent volume ( ● ) (given by the difference between cell and protein volume) per protein molecule for tetragonal lysozyme crystals versus relative humidity. Dark and light symbols are for room temperature and T=100 K values, respectively. The dashed lines are guides to the eye through the room temperature data points.

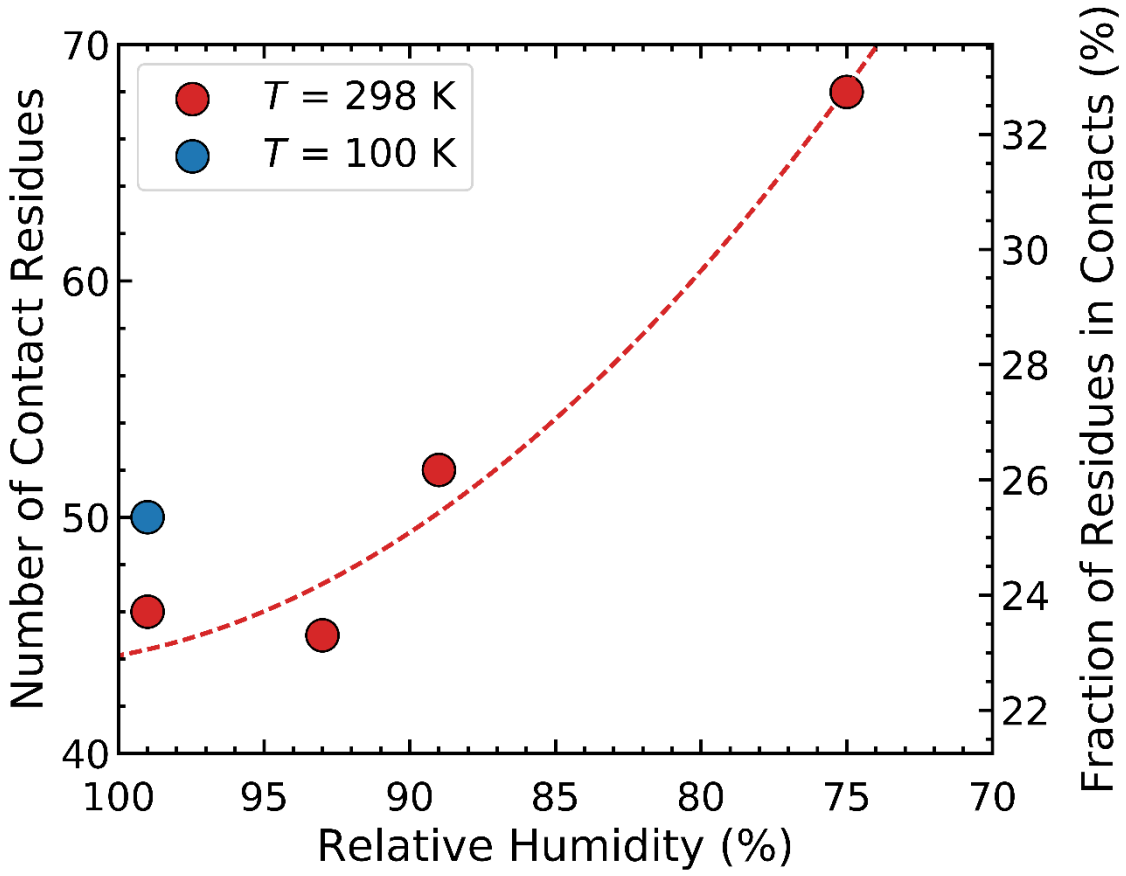
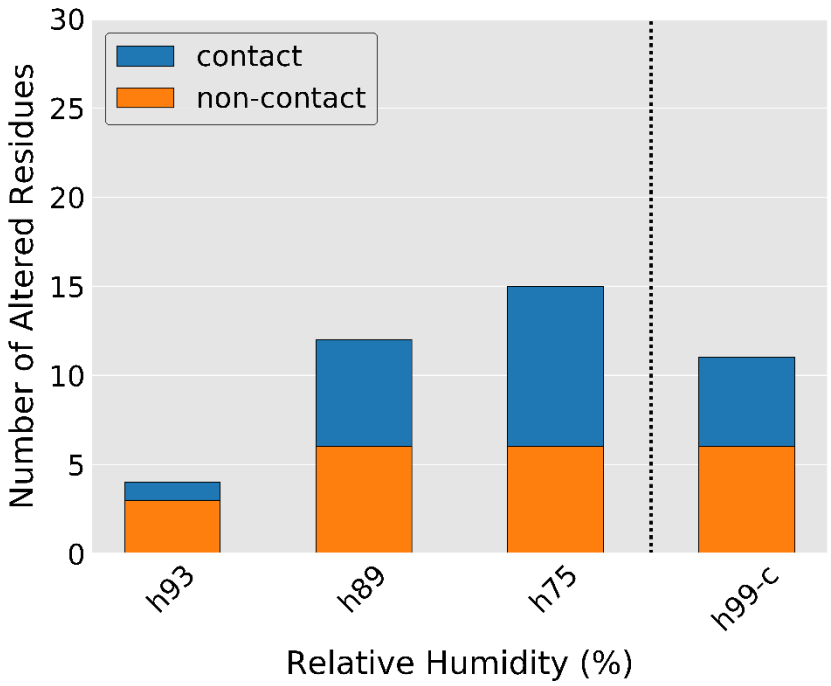
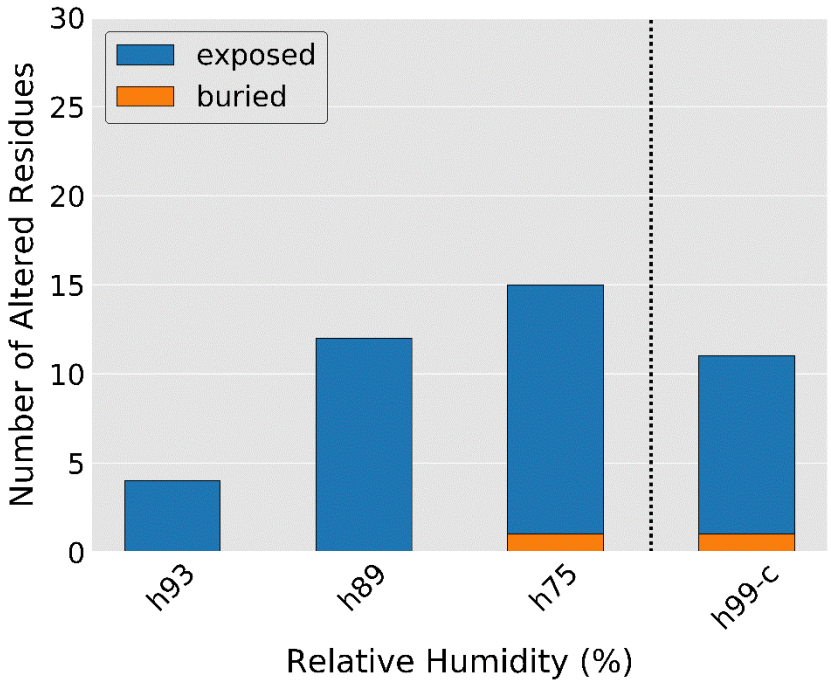



**Figure S7** Number of residues involved in crystal contacts versus relative humidity at room temperature (red) and T=100 K (blue) for tetragonal thaumatin crystals. The dashed red line is a guide to the eye for the room temperature data.




(a)




(b)

**Figure S8** Number of altered residues, of the 156 thaumatin residues having $\chi_1$ rotamers, due to dehydration and cryocooling, relative to the native (99% r.h.) structure, categorized according to (a) whether or the not the residues were involved in crystal contacts in the final, non-native structure; and (b) whether the residues were solvent exposed or buried. Here, side chains were deemed to be altered if the Pearson correlation coefficient between Ringer curves for the native and non-native structures was less than 0.85.

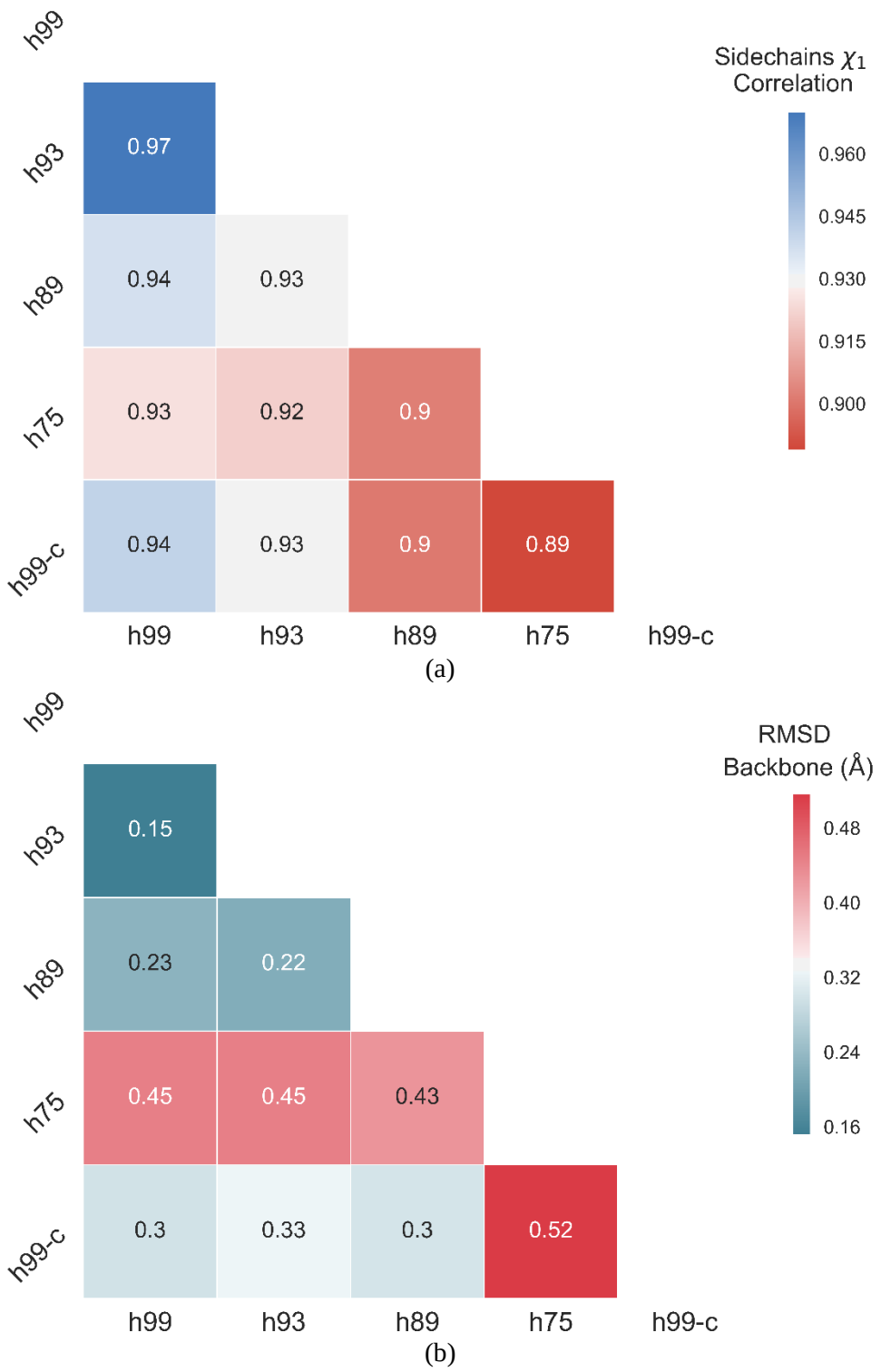



**Figure S9** a) Pairwise Pearson correlation coefficients for Ringer plots at $\chi_1$ as in Fig. 4, averaged over all residues in thaumatin with $\chi_1$ rotamers, for all possible pairs of dehydrated and cryo-cooled structures. Data sets collected at 100 K are labelled with "-c". (b) Backbone RMSD between all possible pairs of dehydrated and cryo-cooled structures. Only backbone atoms were used for alignment of the two structures and calculation of RMSD.