Supporting Information

Wall et al. 10.1073/pnas.1416744111

SI Text

Rerefined Crystal Structure of Staphylococcal Nuclease. The coordinates for the liquid-like motions model in Wall et al. (1) were originally obtained by rigid-body refinement of RCSB Protein Data Bank (PDB) entry 2SNS (2) against the Bragg diffraction data; however, these coordinates were never deposited in the PDB. During revision of the present manuscript, the crystal structure of staphylococcal nuclease in complex with Ca^{2+} and thymidine-3′-5′-bisphosphate (pdTp) was refined more thoroughly using the Bragg diffraction data collected by Wall et al. (1). Initial model phases for this refinement were obtained by molecular replacement using PDB entry 1SNC (3). The structure and reflections data along with refinement statistics were deposited at the Worldwide Protein Data Bank (4) (www.wwpdb.org) and were released as entry 4WOR (1) (doi:10.2210/pdb4wor/pdb). The 4WOR structure was used for structure comparisons.

Simulations Using a Protein Complex. To assess the robustness of our results to changes in the MD model, we also performed a simulation using protein and ligand atomic coordinates obtained from PDB entry 2SNS (2), which was used for MD simulations in ref. 5. As entry 2SNS has poor geometry (see validation report at www.rcsb.org/pdb/d/file?b=validation&n=2sns_full_validation.pdf), we first refined it in *Phenix* (6) using the Bragg data from ref. 1. The structure of the protein-thymidine 3′,5′-bisphosphate-calcium ion complex refined to a final R-work, R-free = 18.8%, 20.8%. The RMSD of the coordinates between the 2SNS structure before and after refinement and alignment was 0.86 Å (0.58 Å) for heavy atoms (α-carbons) [calculated using PyMOL (7)]. The RMSD between the coordinates of 4WOR and the 2SNS structure before and after refinement were 1.0 Å (0.61 Å) and 0.64 Å (0.18 Å), respectively. The RMSD of protein atomic coordinates between PDB entry 1STN (8) (used for the simulation in the main text) and 2SNS after refinement is 1.4 Å (0.70 Å). Entry 1STN does not include an inhibitor, whereas the 2SNS structure includes thymidine-3′,5′-bisphosphate and a Ca^{2+} ion, which were present in the crystal from which the experimental data were collected (1).

The MD model based on refined entry 2SNS was prepared using University of California San Francisco Chimera (9) to generate the symmetry related copies of the protein in the unit cell. Ligand parameters were obtained using the SwissParam server (10). The initial restrained NVT equilibration simulation was prepared as for the 1STN simulation, except two iterations of adding first 200, and then 40 waters using genbox were performed before the final 100-ps equilibration, to achieve near-atmospheric pressure. This yielded a model with 9,196 protein atoms, 6,684 TIP3P water atoms, 148 ligand atoms, 40 Cl^- ions, and 4 Ca^{2+} ions. The CHARMM27 force field was selected for this simulation (11).

The 2SNS-derived model simulation was performed in GROMACS 4.6.5 single precision (12) using a NVT (constant volume instead of pressure) ensemble at 300 K using a 2-fs time step and a modified Berendsen thermostat with a 0.1-ps time constant for temperature equilibration. The diffuse scattering calculations using either the 1STN or the refined model 100-ns trajectory were very similar (correlation coefficient of 0.98). The overall correlation of the MD model using the 100-ns refined model trajectory with the data are 0.90 (0.34) isotropic (anisotropic), compared with 0.93 (0.35) for the 100-ns 1STN trajectory. Calculations using the 1STN trajectory therefore agreed slightly better with the experimental data. The small difference might reflect variations that would occur using different random number seeds in different MD runs, differences in using a NPT vs. NVT ensemble, or CHARMM27 vs. OPLS-AA force field differences.

MD Model Diffuse Scattering Calculations and Reproducibility. To extract ensembles of atomic coordinates from the full trajectory in 5-ns chunks, first the trajectory was adjusted to remove discontinuous jumps of atomic positions to symmetry-related positions during the course of the simulation (trjconv –pbc nojump). Next, the coordinates were adjusted to remove translational drift of the protein (trjconv –fit translation). The protein and solvent coordinates were then extracted and processed further. To process the protein, the coordinates were adjusted to preserve the covalent bonding structure (trjconv –pbc mol) and were then extracted as a series of multiple model .pdb files. To process the solvent, the coordinates were first adjusted so that all atoms were in the simulation box (trjconv –pbc atom); then, they were adjusted so that the covalent bonding structure of each molecule was preserved (trjconv –pbc mol). They were then extracted as a series of multiple model .pdb files.

Diffuse Scattering Data. We used the 64,335 experimental diffuse scattering observations $D_o(hkl)$ from Wall et al. (1) for validation. As described in refs. 13 and 14, these data were obtained by an integration procedure that yielded a single diffuse intensity measurement per lattice point hkl (figure 2.13 in ref. 13): each pixel in each diffraction image was mapped to a scattering vector $h'k'l'$; the nearest lattice point hkl was identified; if $h'k'l'$ was within a 1/2 × 1/2 × 1/2 cube about hkl, the measurement was rejected as being too close to a Bragg peak; the mean value of all retained diffuse intensity measurements (each appropriately scaled) was used as the measurement of diffuse intensity in the neighborhood of hkl. This procedure yielded views of measured diffuse intensity like those in figure 3 of ref. 1 and figures 5.14–5.17 of ref. 13.

As for the calculated $D_{md}(hkl)$, we enforced the P4/m Patterson symmetry; the symmetry-expanded map was very similar to the original map [Pearson correlation coefficient = 0.999, which is higher than for $D_{md}(hkl)$] and consisted of 120,845 points. The isotropic and anisotropic components $D_o(s_n)$ and $D'_o(hkl)$ were calculated from $D_o(hkl)$, like for $D_{md}(s_n)$ and $D'_{md}(hkl)$ (*Methods*). We used the correlation coefficient r_{oc} to compare the total calculated diffuse scattering, $D_{md}(hkl)$, to the experimental data, $D_o(hkl)$, and used the correlation coefficient r'_{oc} to compare the anisotropic component, $D'_{md}(hkl)$ to the experimental data, $D'_o(hkl)$. Before performing the comparison, we reindexed the calculated map by rotating it 180° about the h axis. The rotation was required to place the model and data in an equivalent orientation.

We also double-checked the data integration procedure by using LUNUS software (14) to reprocess the raw frames, and using a custom python script created for integration using indexing information from LABELIT (15). As in ref. 1, the integration produced one measurement per Miller index, which is suitable for modeling the large-scale diffuse features under the assumption of independent unit cell electron density variations (a finer sampling was used in ref. 16, which focused on the small-scale, streaked features). The resulting 3D set of intensity measurements $D_o(hkl)$ were very similar to the original dataset obtained from the same frames (Pearson correlation before enforcing symmetry = 0.997; after enforcing symmetry = 0.999).

1. Wall ME, Ealick SE, Gruner SM (1997) Three-dimensional diffuse x-ray scattering from crystals of *Staphylococcal* nuclease. *Proc Natl Acad Sci USA* 94(12):6180–6184.
2. Cotton FA, Hazen EE, Jr, Legg MJ (1979) Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme-thymidine 3′,5′-bisphosphate-calcium ion complex at 1.5-Å resolution. *Proc Natl Acad Sci USA* 76(6):2551–2555.
3. Loll PJ, Lattman EE (1989) The crystal structure of the ternary complex of staphylococcal nuclease, Ca^{2+}, and the inhibitor pdTp, refined at 1.65 Å. *Proteins* 5(3):183–201.
4. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 10(12):980.
5. Meinhold L, Smith JC (2005) Fluctuations and correlations in crystalline protein dynamics: A simulation analysis of staphylococcal nuclease. *Biophys J* 88(4):2554–2563.
6. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
7. DeLano WL, The PyMOL Molecular Graphics System (Schrödinger, LLC, New York), Version 1.7.1.1.
8. Hynes TR, Fox RO (1991) The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins* 10(2):92–105.
9. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
10. Zoete V, Cuendet MA, Grosdidier A, Michielin O (2011) SwissParam: A fast force field generation tool for small organic molecules. *J Comput Chem* 32(11):2359–2368.
11. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30(10):1545–1614.
12. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel modecular dynamics implementation. *Comput Phys Commun* 91(1-3):43–56.
13. Wall ME (1996) Diffuse features in X-ray diffraction from protein crystals. PhD thesis (Princeton University, Princeton).
14. Wall ME (2009) Methods and software for diffuse X-ray scattering from protein crystals. *Methods Mol Biol* 544:269–279.
15. Sauter NK, Grosse-Kunstleve RW, Adams PD (2004) Robust indexing for automatic data collection. *J Appl Cryst* 37(Pt 3):399–409.
16. Wall ME, Clarage JB, Phillips GN (1997) Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Structure* 5(12):1599–1612.

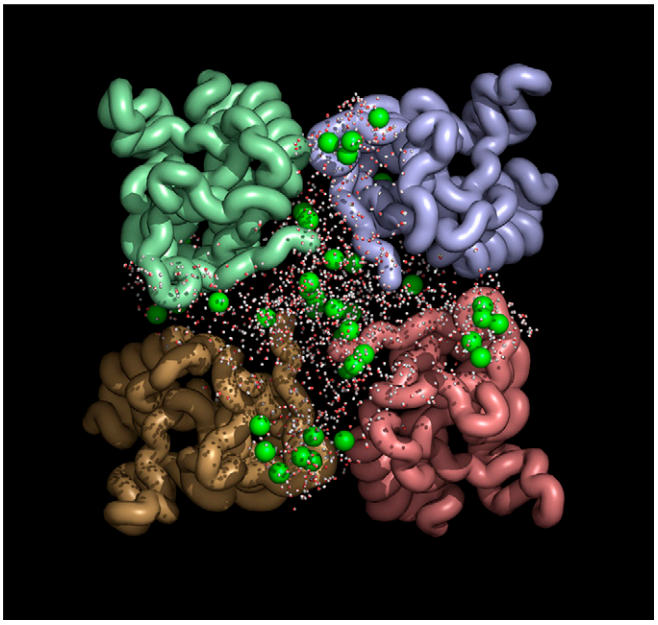


Fig. S1. Starting model for the MD simulation derived from PDB entry 1STN (8). The proteins (green, purple, pink, and yellow tubes) are arranged with $P4_1$ symmetry in a 48.5 Å × 48.5 Å × 63.5-Å unit cell. Although the starting protein model has $P4_1$ symmetry, this symmetry is not enforced in the MD simulation; the system therefore quickly relaxes to P1 symmetry. Also in the model are 6,477 TIP3P waters (small CPK molecules) and 40 Cl^- counterions (green spheres). View is down the *c* axis. The waters and Cl^- ions are displayed entirely within the simulation volume. The protein is displayed as covalently linked whole molecules, with some parts shifted to a symmetry equivalent position outside of the unit cell. The image was created using PyMOL (7).

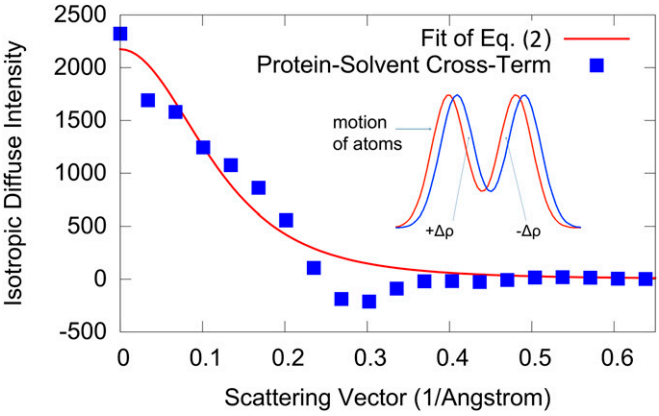


Fig. S2. Fit of Eq. **2** to the protein–solvent cross term, $D_{md,x}$, calculated from decomposition of the simulated diffuse scattering. The best fit was obtained using $\gamma = 0.89 \pm 0.08$ Å (with a scale factor of 119 ± 27). Positive values of $D_{md,x}$ correspond to negative correlations in the density fluctuations; the *Inset* illustrates how a negative correlation can arise from atoms moving together (*Discussion*).

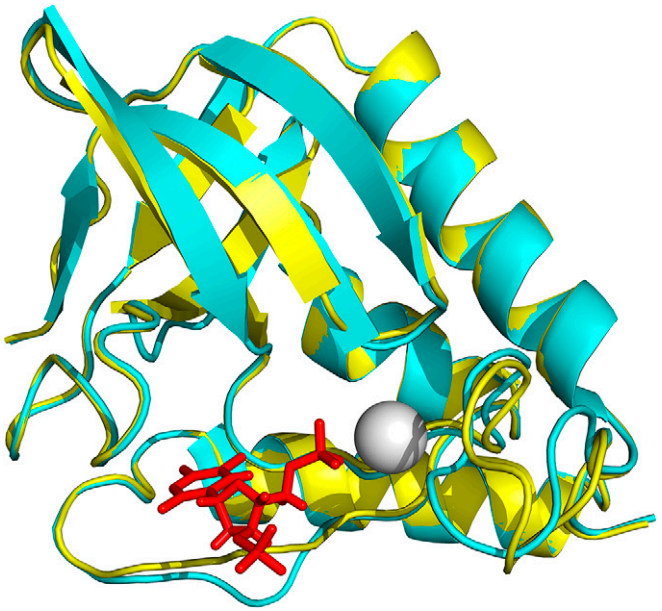
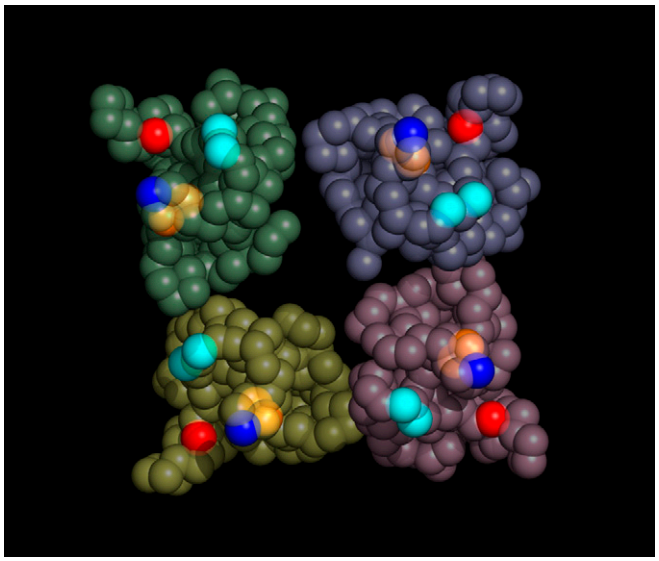


Fig. S3. Superposition of PDB entries 1STN (cyan cartoon, used for MD model) and 4WOR (yellow cartoon, refined against Bragg data). The pdTp (red sticks) and Ca^{2+} (white sphere) coordinates were taken from the 4WOR structure. Backbone deviations are concentrated in loops near the Ca^{2+}-pdTp ligand: residues 43–51, to the right of the Ca^{2+}; and residues 112–114, below the pdTp.

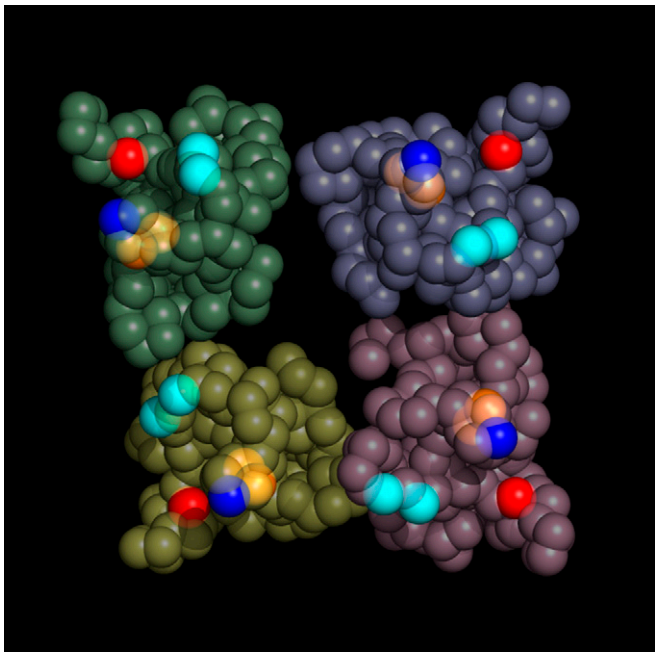
Table S1. Pearson correlation coefficients of total intensity r_{oc} (and anisotropic intensity r'_{oc}, in parentheses) of the experimental data with diffuse scattering calculated for MD models using either all atoms or just the heavy atoms

	Correlation	
Model	All atoms	Heavy atoms
0- to 10-ns MD	0.90 (0.43)	0.85 (0.43)
0- to 100-ns MD	0.93 (0.35)	0.91 (0.34)
0- to 1,100-ns MD	0.94 (0.40)	0.92 (0.40)
100- to 1,100-ns MD	0.94 (0.40)	0.92 (0.40)



Movie S1. Motions of α-carbons (colored spheres) in the largest nonzero principal component of the MD trajectory. The active site is indicated using brightly colored spheres, using the same color code as in Fig. 2. The chain colors are the same as the *Inset* of Fig. 2. The movie was created using PyMOL (7).

Movie S1



Movie S2. Motions of α-carbons (colored spheres) in the second-largest nonzero principal component of the MD trajectory. Details are the same as for Movie S1.

Movie S2