Supporting Information

Bartesaghi et al. 10.1073/pnas.1402809111

SI Materials and Methods

Motion Correction. To compensate for drift and beam-induced motion during the electron exposure, individual frames of each movie were aligned by cross-correlation by using the cumulative average of previously aligned frames as a reference to align the remaining frames. Adopting principles similar to those used in iterative image alignment in single-particle cryo-EM (1), we start by aligning two consecutive frames to each other (typically the middle frames of each movie) and then use their aligned average as a reference to align the next frame. The average at each step is updated to include all of the previously aligned frames, and this process is repeated until all frames have been aligned. The entire process is then iterated until incremental changes in the estimated shifts of each frame are negligible; usually 5–8 iterations are sufficient to achieve convergence. All cross-correlation operations were implemented by using IMOD's tiltxcorr command using the cumulative option (2). For the comparison of our frame alignment strategy with the results produced by motioncorr (3) (Fig. S4), unbinned superresolution images were used as input to the program by using default parameter settings (increasing the offset for the number of frames used for comparison with values up to 10, using larger values for the B-factor parameter or binning of the data, did not improve the results). The corrected sum produced by motioncorr was then used as input to ctffind3 (4) with the range of frequencies used to fit the contrast transfer function (CTF) set from 50 to 3.5 Å. Identical settings were used to estimate the CTF of the uncorrected averages and the corrected averages by using our cumulative cross-correlation strategy.

CTF Estimation. Parameters of the contrast transfer function for each micrograph were obtained by using radially averaged power spectra obtained by periodogram averaging with tiles extracted from individual frames of each movie. The power spectra from individual movie frames were obtained by using tomops (5) and then radially averaged to estimate a single defocus value by using tomoctffind (5). Although the lower signal-to-noise ratio of individual frames is known to deteriorate the signal in the power spectra (6), the subsequent radial averaging operation results in a significant increase in the cross-correlation values against the 1D theoretical CTF profile (Fig. S5). This improvement is attributed to the combination of the radial averaging operation (assuming astigmatism is negligible), with the fact that the power spectrum obtained from the individual frames has no influence of either local or global drift during the exposure (due to the shift invariant property of the power spectrum) and does not suffer from dampening of signal at high resolution due to interpolation and averaging operations needed to produce drift-corrected images.

Model Refinement. Side-chain densities were visualized clearly throughout the map, allowing us to build and refine the entire atomic model in the environment of the program COOT (7). More specifically, rigid body fitting of an X-ray structure [Protein Data Bank (PDB) ID code 1DP0] to the map was first carried out by using Chimera (8). The transformed structure was used as a starting point to the refinement in COOT (version 0.7.1), which was possible after translating the map to a new origin (0,0,0, using EMAN). The structure was refined into the density map in a semimanual way by using planar peptide and Ramachandran restraints, no secondary structure restraints, with a weight matrix of 60% during multiple rounds of processing for the entire 1,023-aa peptide chain. Side-chain rotamers were tested for their consistency with the density. The N-terminal domain, and multiple segments of poor agreement with the initial structure, were constructed de novo. Noncrystallographic D2 symmetry was applied at the end.

1. Frank J, Oxford University Press. (2006) Three-dimensional electron microscopy of macromolecular assemblies visualization of biological molecules in their native state. (Oxford Univ Press, Oxford).
2. Kremer JR, Mastronarde DN, McIntosh JR (1996) Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* 116(1):71–76.
3. Li X, et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10(6):584–590.
4. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142(3):334–347.
5. Fernández JJ, Li S, Crowther RA (2006) CTF determination and correction in electron cryotomography. *Ultramicroscopy* 106(7):587–596.
6. Glaeser RM, McMullan G, Faruqi AR, Henderson R (2011) Images of paraffin monolayer crystals with perfect contrast: minimization of beam-induced specimen motion. *Ultramicroscopy* 111(2):90–100.
7. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
8. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.

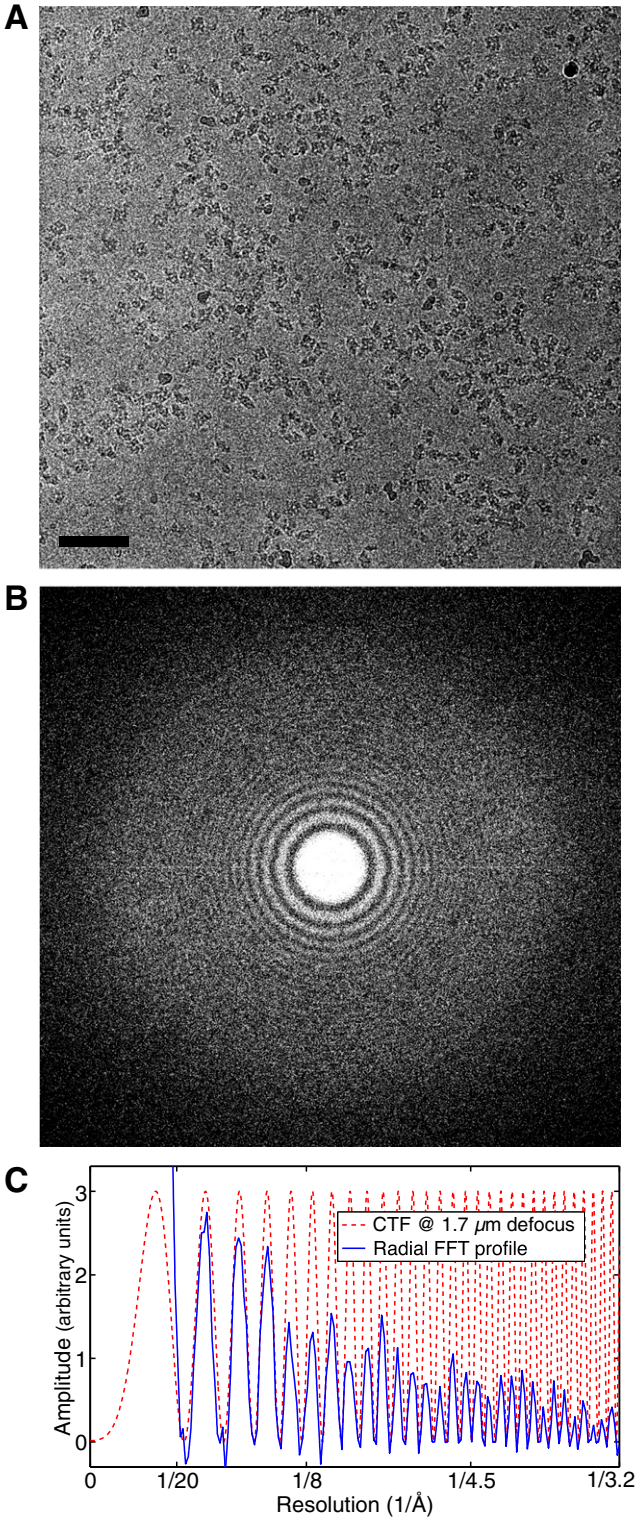

Fig. S1. Cryo-EM of *Escherichia coli* β-galactosidase. (*A*) Average of 38 aligned movie frames acquired over a 15.2-s exposure window (45 e^{-}/Å^2 total accumulated dose) with images recorded at 300 kV and −1.7 μm defocus in superresolution mode by using a physical pixel size of 1.275 Å. (Scale bar: 50 nm.) (*B*) Fast Fourier transform (FFT) of image in *A* showing the extent of Thon rings present in the data extending to ∼2/3 of the edge of the transform, which corresponds to 2.55 Å. (*C*) One-dimensional power spectrum profile obtained by radially averaging the Fourier transform of the image in *A*; peaks are clearly visible out to 3.2-Å resolution.

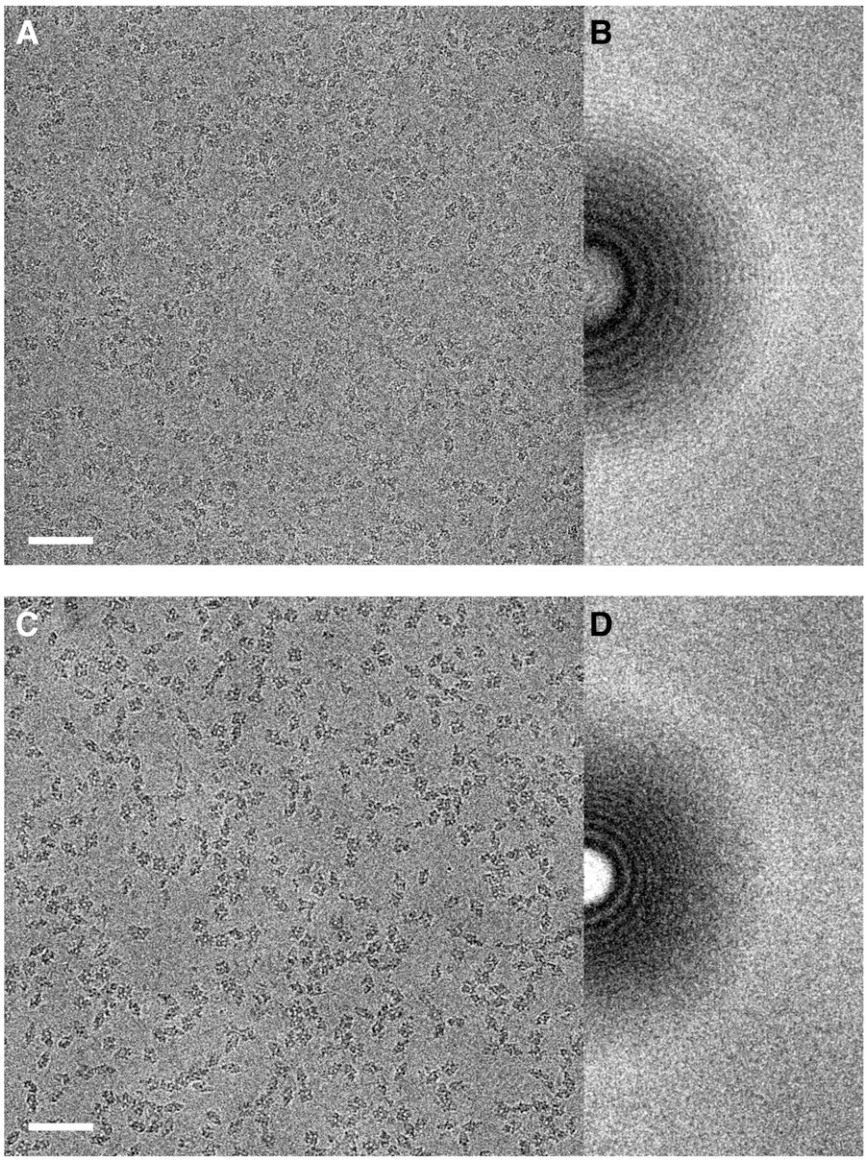

Fig. S2. Cryo-EM imaging of *E. coli* β-galactosidase at 300 kV in superresolution mode by using a physical pixel size of 1.275 Å. Averages of 38 aligned movie frames acquired over a 15.2-s exposure window (45 e^{-}/Å^2 total accumulated dose) using −1 μm (*A*) and −2 μm (*C*) defocus and corresponding power spectra (obtained by periodogram averaging using tiles from all frames) showing the resolution to which Thon rings are visible (*B* and *D*). The edge of the transform corresponds to 2.55 Å, and only the nonredundant half of the FFT is shown. (Scale bars: 50 nm.)

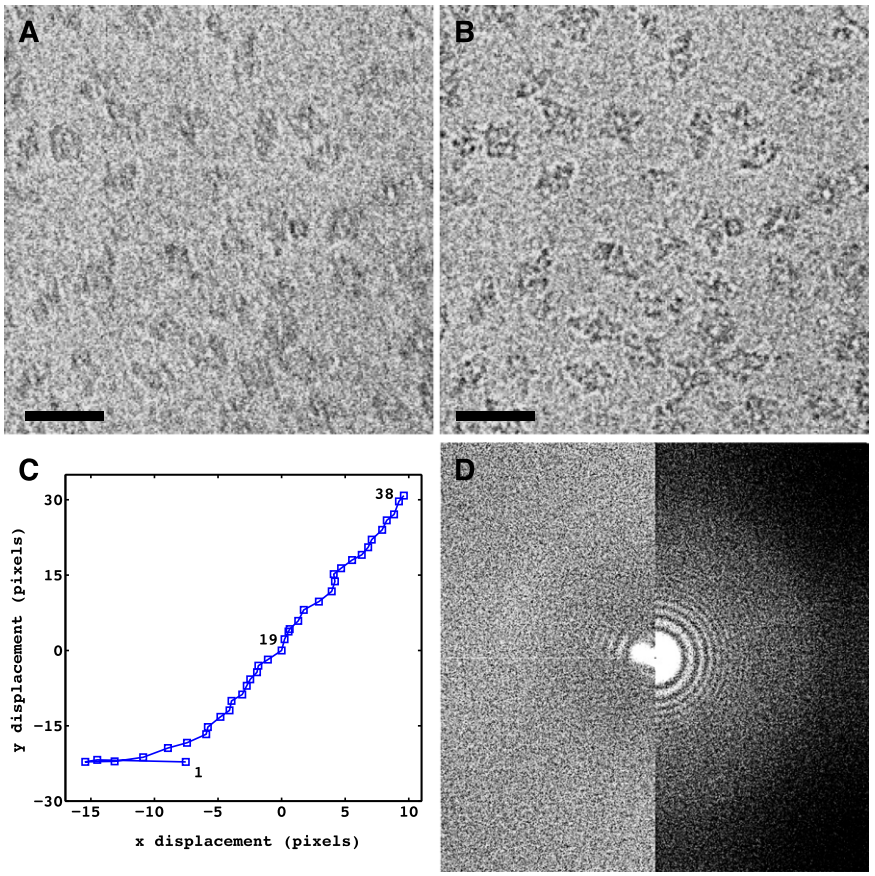

Fig. S3. Improvement in image quality obtained by aligning individual movie frames. (*A* and *B*) Comparison between averages of unaligned (*A*) and aligned (*B*) movie frames showing correction of the blurring effect caused by beam-induced motion (extreme case). (Scale bars: 25 nm.) (*C*) Trajectory of movement for each of the 38 frames during the 15.2-s exposure used in *A*. Displacements are measured in multiples of the superresolution pixel (1 pixel = 0.6375 Å). (*D*) Comparison between the FFTs of the averages of unaligned (*Left*) and aligned (*Right*) movie frames showing recovery of uniform Thon rings after correction for image movement during the exposure.

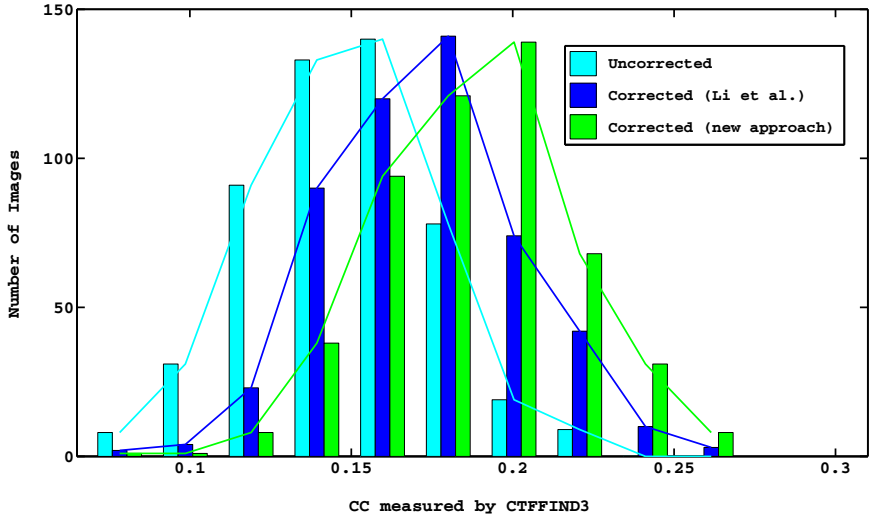

Fig. S4. Improved accuracy of frame alignment by using a cumulative cross-correlation approach. Distribution of cross-correlation values for the dataset of 509 images reported by ctffind3 (1) between theoretical CTF curves and power spectra obtained from uncorrected frame averages (cyan), corrected-averages using motioncorr (2) (blue), and corrected averages using our cumulative cross-correlation strategy (green). The plots show that the use of cumulative cross-correlation increases the accuracy of frame alignment beyond that achieved by motioncorr (2).

1. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142(3):334–347.
2. Li X, et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10(6):584–590.

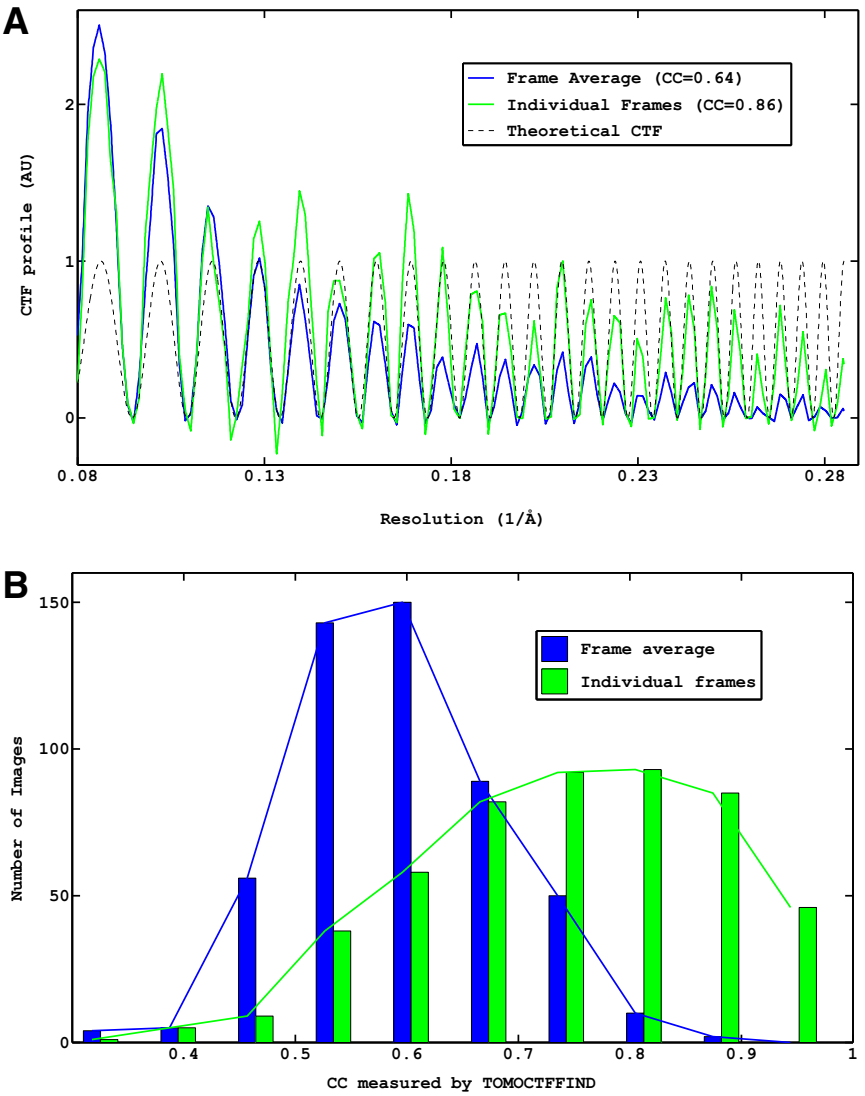

Fig. S5. Use of intermediate frames to improve accuracy of CTF determination. (*A*) Radially averaged power spectra used for CTF determination obtained from the aligned average of frames (blue) and from the raw individual frames without alignment (green). Cross-correlation values reported by tomoctffind (1) against the theoretical CTF curve at −1.7 μm defocus improve from 0.64 to 0.86 when using the individual frames. (*B*) Distribution of cross-correlation values reported by tomoctffind for the set of 509 images using frame averages corrected by cumulative cross-correlation (blue), and for power spectra obtained from individual movie frames (green), showing a substantial improvement in the quality of the CTF fits and increase in accuracy of defocus estimation when using the individual movie frames.

1. Fernández JJ, Li S, Crowther RA (2006) CTF determination and correction in electron cryotomography. *Ultramicroscopy* 106(7):587–596.

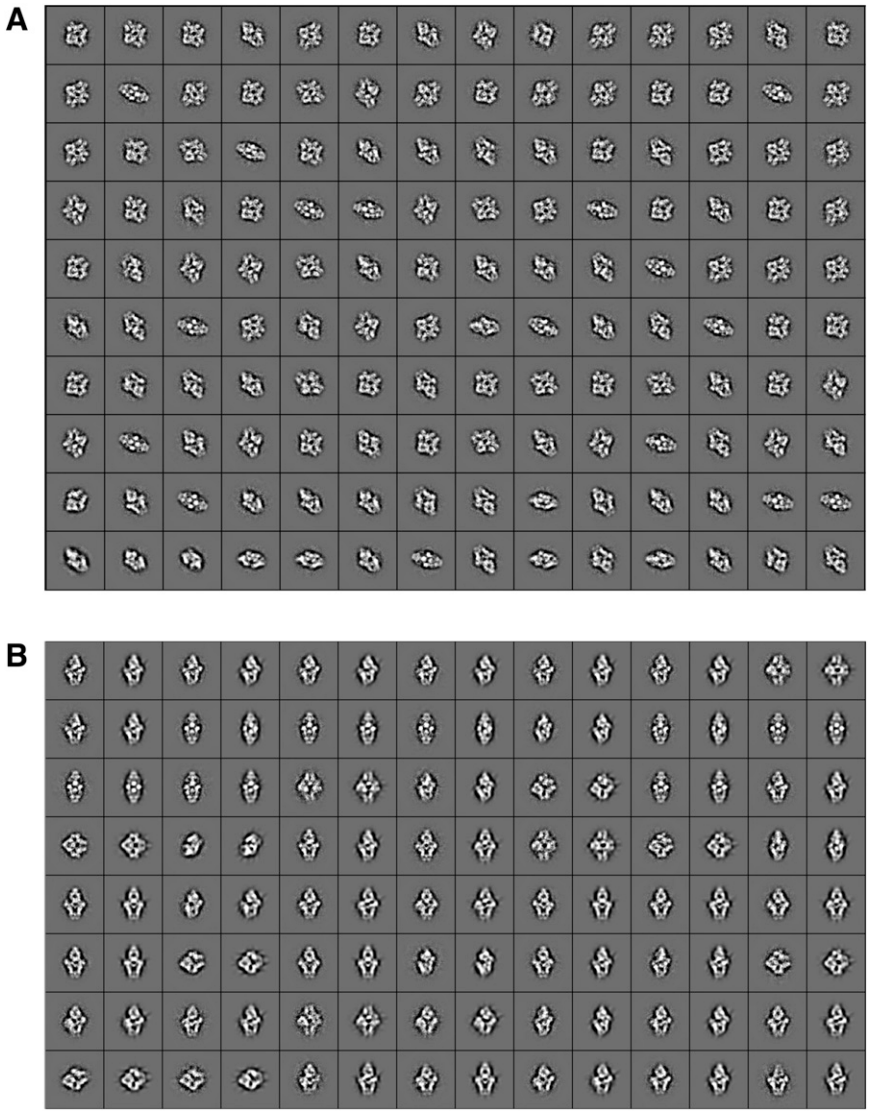

Fig. S6. Two-dimensional reference-free classification and ab initio generation of an initial model. (*A*) Representative 2D class averages of cryo-EM β-galactosidase images obtained with the program e2refine2d.py in the EMAN2 suite shows characteristic views of β-galactosidase. (*B*) Alternating class averages (odd-numbered columns) and corresponding reprojections of the ab initio 3D model obtained with e2initialmodel.py (even-numbered columns) show a good match between computed model projections and reference-free class averages.

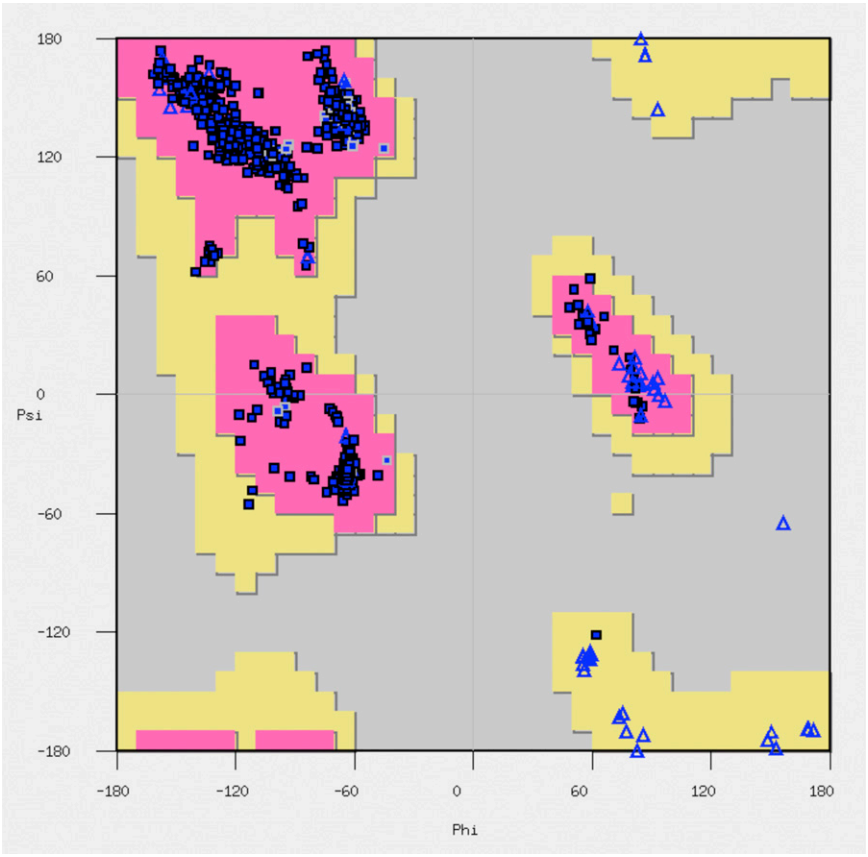

Fig. S7. Ramachrandran plot of the atomic model of the *E. coli* β-galactosidase structure derived from cryo-EM data, indicating 4,004 residues (>98%) in preferred and 72 residues (<2%) in allowed regions.

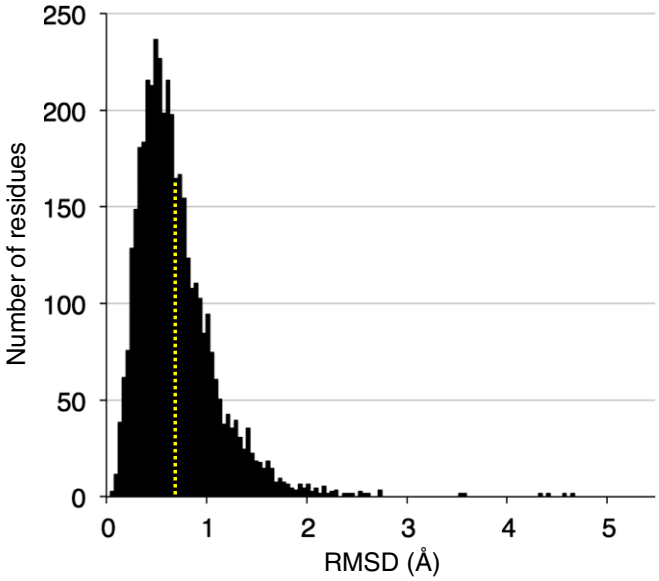

Fig. S8. Histogram showing the distribution of rmsd values of individual Cα atoms between corresponding residues in the cryo-EM and X-ray (PDB ID code 1DP0) structures. Values range between 0.04 and 4.64 Å, with an average of 0.66 Å (yellow dashed line).

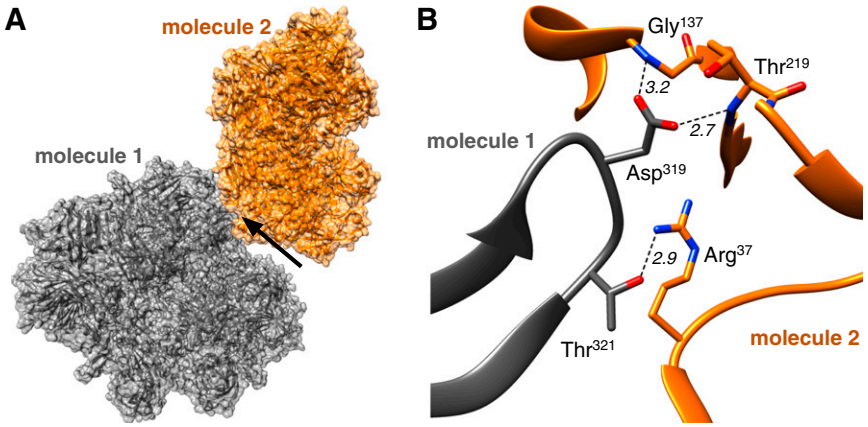

Fig. S9. Residues involved in crystal contacts. (*A*) Visualization of one of the crystal contacts (region ** in Fig. 3) highlighting the intermolecular interaction (indicated by the arrow) of two β-gal tetramers (shown in gray and orange) in the 3D crystal lattice (PDB ID code 1DP0). The presence of this crystal contact correlates with the difference in local backbone conformation between the crystal and cryo-EM structures. (*B*) Detailed view of the interaction between a peripheral loop from one tetramer (shown in gray) with three distinct regions on the neighboring tetramer in the crystal lattice.

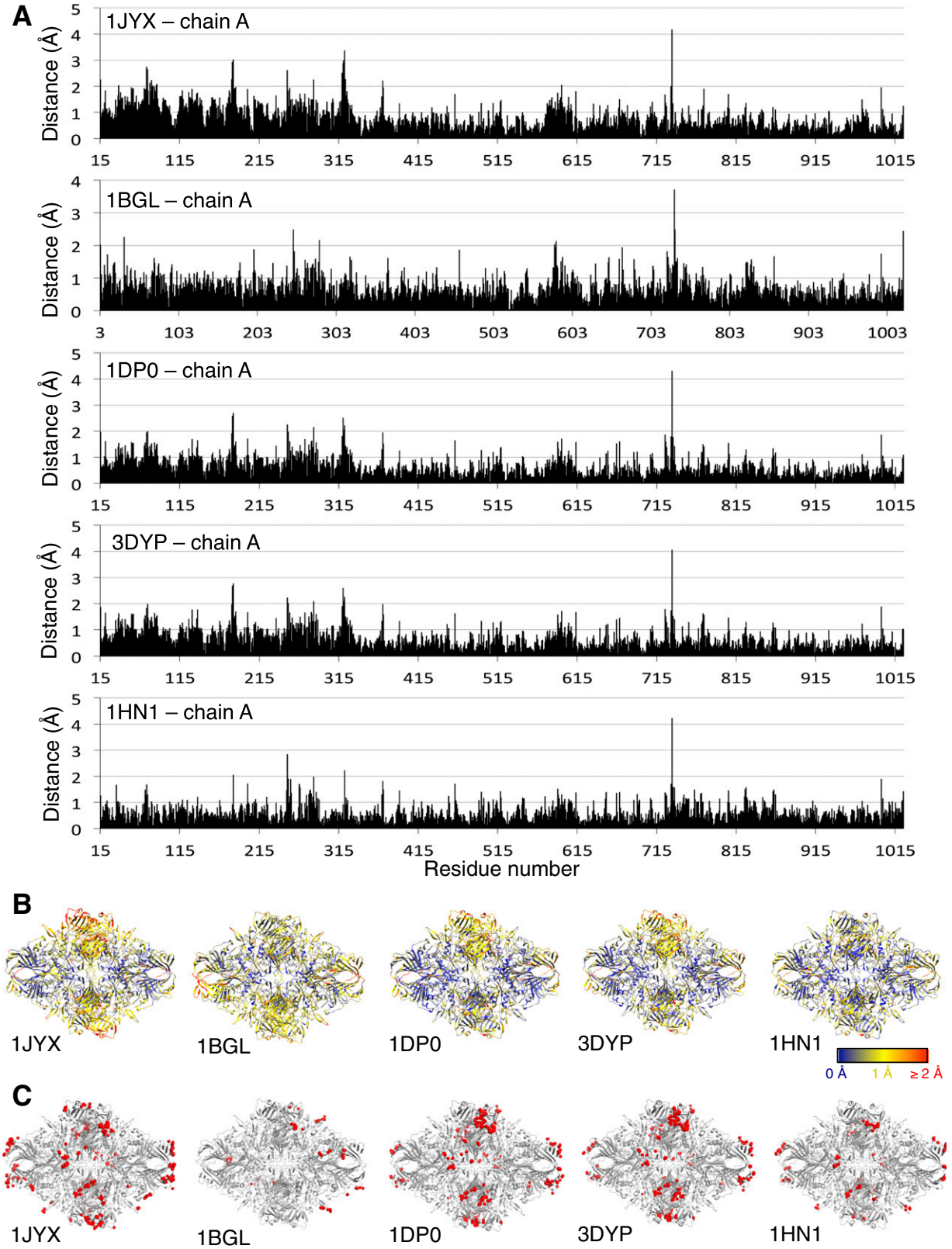

Fig. S10. Comparison between cryo-EM and X-ray structures of *E. coli* β-galactosidase. (*A*) rmsd values of all Cα atoms between corresponding residues in five representative X-ray models [PDB ID codes 1JYX, 1BGL, 1DP0, 3DYP, and 1HN1] relative to the cryo-EM–derived atomic model are plotted according to the amino acid sequence. These five structures were chosen based on uniqueness of space group (1BGL is monoclinic and the rest are orthorhombic), presence of a ligand (1JYX is complexed with IPTG), best and worst resolutions of the determined structures (1DPO at 1.70 Å and 1HN1 at 3.0 Å), and an active-site mutant structure (3DYP). (*B*) X-ray structure color-coded for variation in rmsd values. (*C*) X-ray structure color-coded to display regions of the protein in close contact with neighboring molecules in the crystals used to obtain the five X-ray structures. Atoms that are within 4 Å of neighboring tetramers in the crystal lattice are marked as red spheres, indicating areas of protein-protein contact in the crystal lattice.

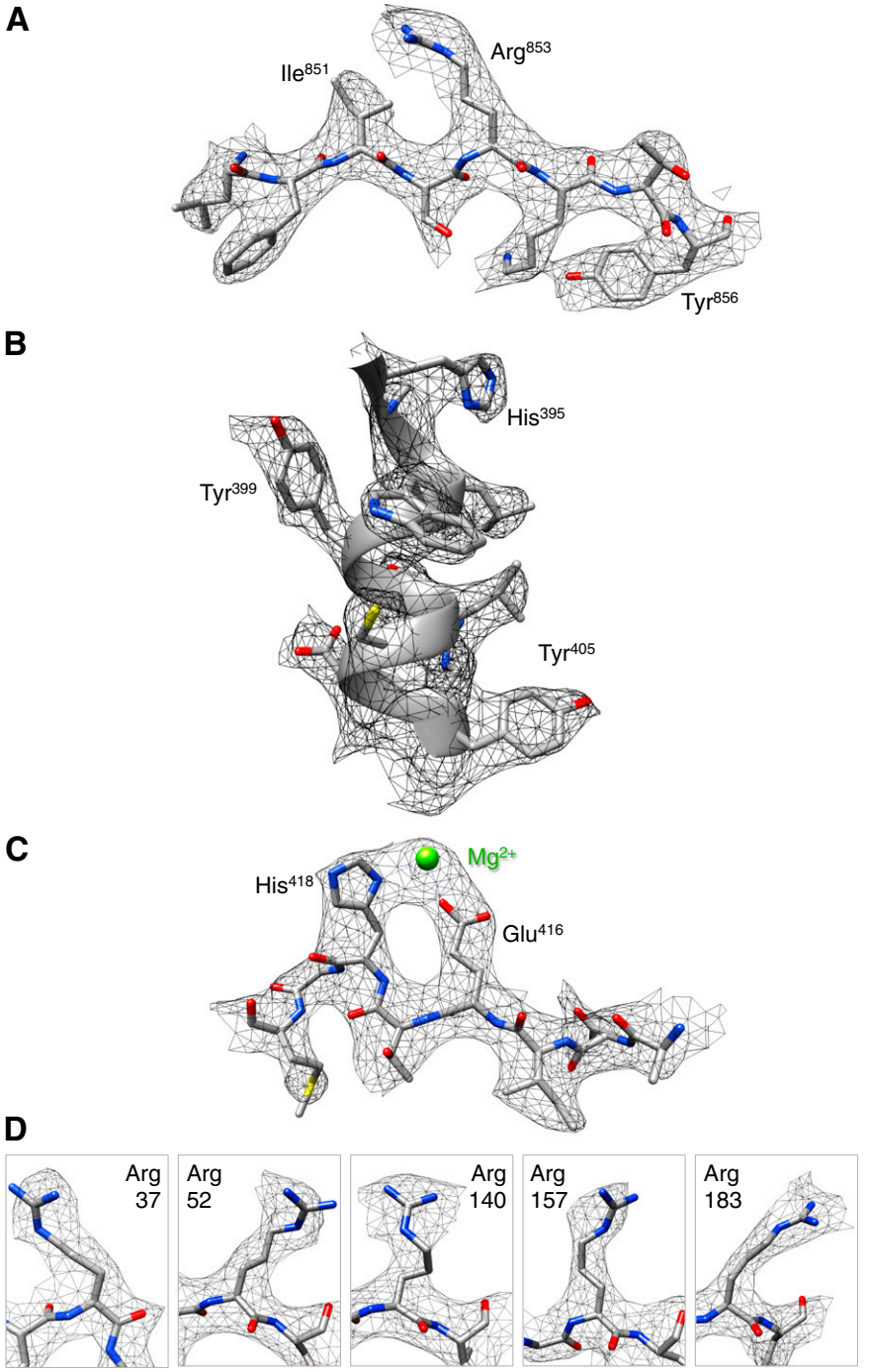

Fig. S11. Electron density map of β-galactosidase derived by X-ray crystallography at 3.0-Å resolution. To compare the cryo-EM density map at 3.2-Å resolution with a map obtained at comparable resolution using X-ray crystallography, $2F_o$-F_c maps obtained from the X-ray structure at 3.0-Å resolution (PDB ID code 1HN1) of the same regions of the protein shown in Figs. 4 and 5 are presented. Residues 849–856 in a β-strand (*A*), residues 395–405 in an α-helix (*B*), and residues 413–420 in the active site along with density for a Mg^{2+} ion partly coordinated by His-418 and Glu-416 (*C*) (compare *A*–*C* with corresponding images shown in Fig. 4 *A*–*C* for the cryo-EM density map). (*D*) Densities for the same set of Arg residues as shown in Fig. 5 for the cryo-EM density map are shown for the $2F_o$-F_c map corresponding to the 1HN1 X-ray structure.

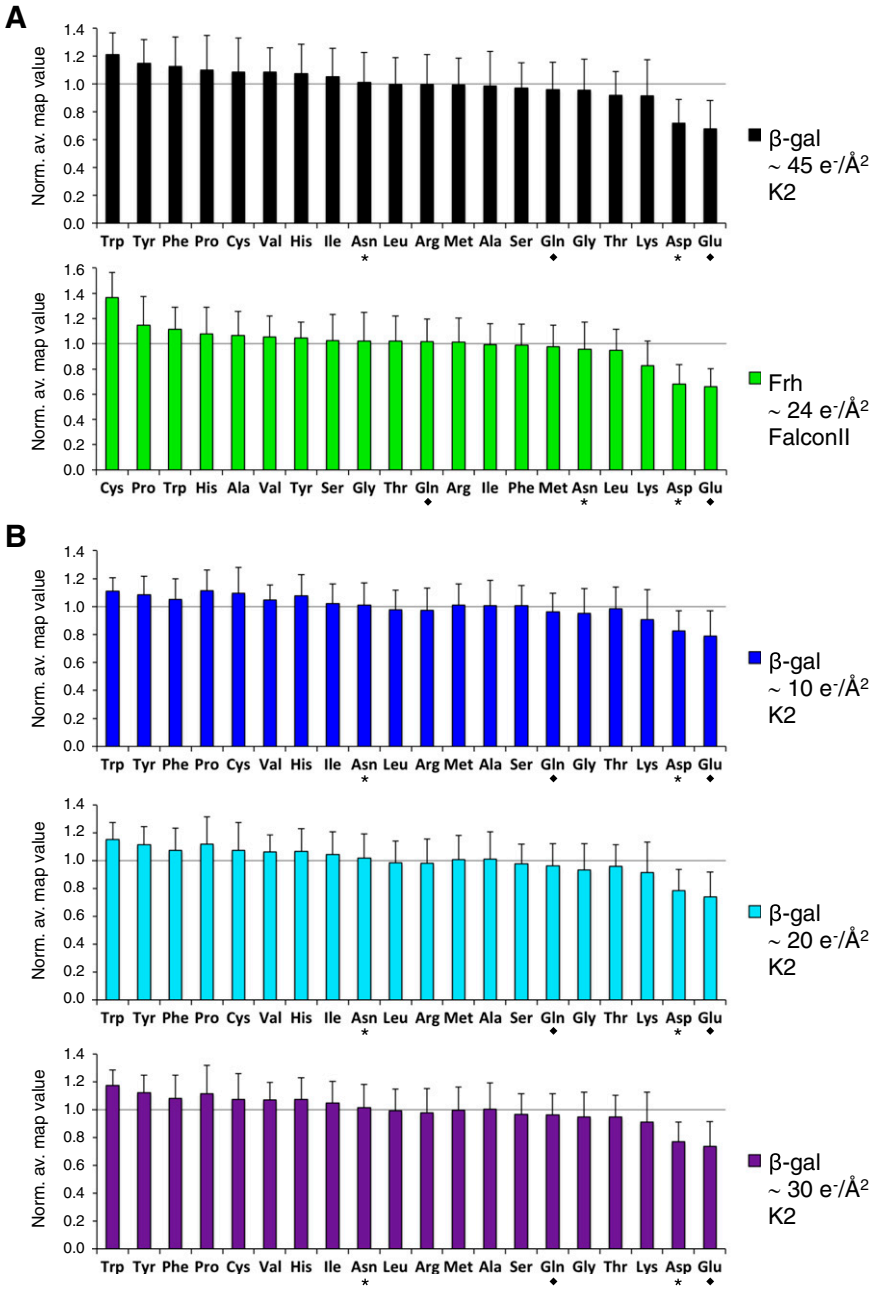

Fig. S12. Proportional variation in side-chain density measured for each residue type on different cryo-EM maps. (*A*) Plots of normalized average map values for each type of residue for the ~3.2-Å map of β-galactosidase (β-gal, *Upper*, black) and the 3.36-Å map of F_{420}-reducing [NiFe] hydrogenase (1) (Frh, *Lower*, green). Density for each type of residue is measured by averaging density map intensities sampled at the position of each nonhydrogen atom for all residues of that type. Values are normalized with respect to the mean density over all residues, resulting in numbers greater than 1 for values above average and numbers smaller than 1 for values below average. SD values are shown as intervals on all plots. Loss of density as a result of radiation damage is seen dominantly in negatively charged residues, Asp and Glu, independent of their exposure to solvent. All Asp (marked by asterisk) and Glu (marked by diamond) residues show on average 30% and 29% less density than the corresponding Asn (asterisk) and Gln (diamond) residues in the case of β-gal, and 35% and 29% in the case of Frh. Most of the cysteine residues in Frh are bonded to FeS clusters, suggesting a possible explanation for the increased averaged map values for this type of residue in Frh compared with β-gal. (*B*) Plots of normalized average map values per residue on three β-gal maps using a total dose of 10 (blue), 20 (cyan), or 30 $e^{-}/Å^2$ (purple). Proportional loss of density at Glu and Asp with respect to Gln or Asn increases with higher total dose, with 18% less density measured for Glu and Asp compared with Gln or Asn in the map calculated at ~10 $e^{-}/Å^2$, 23% at ~20 $e^{-}/Å^2$, and 24% at ~30 $e^{-}/Å^2$. This tendency is also shown in Fig. 6 on selected side chains with map density visualized as iso-surfaces.

1. Allegretti M, Mills DJ, McMullan G, Kühlbrandt W, Vonck J (2014) Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *eLife* 3:e01963.

Table S1. Comparison of the amino acid sequences of *E. coli* β-galactosidase encoded in expression vectors used for protein expression and the subset of the residues in the N-terminal region that are resolved and reported in the PDB entries

PDB ID code	Resolution, Å	N-terminal residues of construct used for protein expression	N-terminal residues in PDB
4DUV	2.3	MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPMIDPVVLQRRDWENPGV	MIDPVVLQRRDWENPGV
3VD3	2.8	MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPMIDPVVLQRRDWENPGV	RRDWENPGV
3SEP	2.1	MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPMIDPVVLQRRDWENPGV	VVLQRRDWENPGV
3MVO	2.2	MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPMIDPVVLQRRDWENPGV	RRDWENPGV
3MUY	2.5	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
3IAP	2.0	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
3I3E	2.0	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
3DYP	1.8	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
3EIF	3.0	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
1PX3	1.6	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
1JYX	1.5	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
1HN1	3.0	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
1F4H	2.8	ITDSLAVVLQRRDWENPGV	ITDSLAVVLQRRDWENPGV
1DP0	1.7	GSHMLEDPVVLQRRDWENPGV	RRDWENPGV
1BGL	2.5	TMITDSLAVVLQRRDWENPGV	ITDSLAVVLQRRDWENPGV

Conserved residues are underlined.