Supplementary Data

S1 CrgA DBD structural comparison

The CrgA DBD is comparable to other members of the bacterial wHTH superfamily including OhrR a member of the MarR family regulators (PDB code 1Z9C RMSD 3.04 Å over 58 equivalent Cα atoms), MecI (PDB code 2D45, RMSD 2.55 Å over 55 equivalent Cα atoms), DtxR (PDB code 1F5T, RMSD 2.37 Å over 52 equivalent Cα atoms), CAP (PDB code 2CGP, RMSD 2.81 Å over 53 equivalent Cα atoms) and BirA (PDB code 2EWN, 2.27 Å over 54 equivalent Cα atoms) (Figure S1). Outside of the wHTH superfamily, NarL (PDB code 1ZG5) a LuxR family two-component response regulator had the highest Z score (Z-score 5.7, RMSD 1.55 Å, length of alignment 50 residues), a measure of the statistical significance of the match in a pairwise structural comparison using SSM with the CrgA DBD (Krissinel & Henrick, 2004).

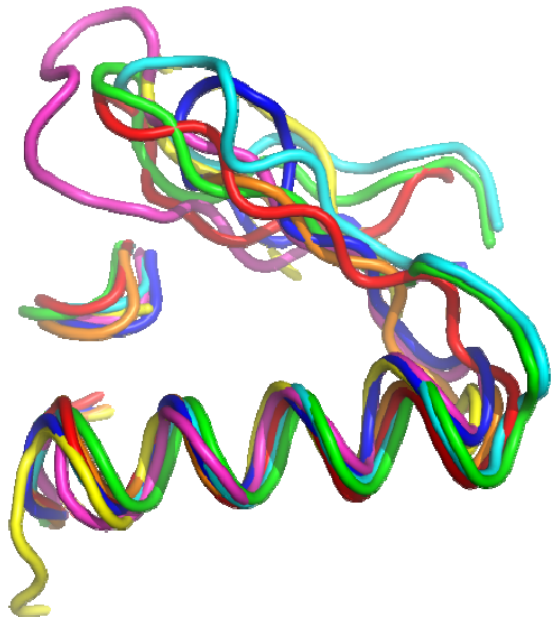


Figure S1 wHTH motif of CrgA

Superimposition of the wHTH motif of CrgA and other members of the winged helix superfamily. CrgA is shown in cyan, CbnR (PDB code 1IZ1) in green, DtxR (PDB code 1F5T) in blue, Ohr (PDB code 1Z9C) in magenta, BirA (PDB code 2EWN) in red, CAP (PDB code 2CGP) in orange and MecI (PDB code 2D45) in yellow. Residues 21-41 (α2-α3) of CrgA and equivalent residues in the other wHTH proteins were used for the superimposition. The first helix of the wHTH motif is not shown.

S2 Cross comparison of the eight CrgA subunits

The eight CrgA subunits are not identical as there is significant variation in the orientation of the individual subdomains (Figure S2A and S2B). Both the DBD-LH (RMSD typically <0.7 Å Cα

residues 2-86 for all chains except chain D) and the regulatory domains (RMSD <0.4 Å Cα residues 94-292) of the eight chains individually superimpose well, therefore implying that, it is a rigid body movement between these two structural modules. The hinge region (H88, E89, I90 P91, Q92 and G93) first identified in CbnR is expected, to a great extent be responsible for the movement. To further understand the nature of the conformational differences between the chains: (i) the dihedral angle between the regulatory domain and the LH was calculated from Cα atoms of T9, E61, E83, and L173 in order to estimate the rotation of the regulatory domain around the LH; (ii) the angle (E61, E83, and L173) and (iii) distance (E61 and L173) between the regulatory domain and LH were measured to quantify the movement of the regulatory domain away from the linker helix. The measured dihedral angles (i) varied by 47 º from 119.8 - 166.7 º in the eight chains, Chains A and H (Figure S2A) which form one DNA binding pair, were most dissimilar. In contrast the (ii) distances (36.1 - 40.2 Å) and (iii) angles (47.1-56.8 º) measured between the LH and regulatory domain, were more comparable in the eight chains (Figure S2B). This indicates that much of the variation between the chains is brought about through a rotation of the regulatory domain around the LH rather than a direct pivot movement of the regulatory domain away from the LH. This is in contrast to CbnR in which the N-terminal of the LH (E65) is 29 Å further away from the regulatory domain (H174) in the extended form compared to the compact form.

Superimposition of the residues of the β3 in CrgA, located immediately after the hinge, shows that the main chain begins to divert away at V94 which is adjacent to a glycine residue (Figure S2C). Proline and glycine residues are frequently present in the hinge regions of LTTRs. The PXG motif of the hinge region of CrgA is conserved in all close homologues as identified from the Uniprot data base (overall sequence identity of the homologues ClustalW scores 44 to 98) whereas the other residues of the hinge which are not highly conserved (Figure S2C) . In contrast to CrgA, the hinge region of CbnR contains two flexible glycine residues (GXXG). The constrained geometry of the proline residue in the hinge of CrgA may cause it to have less structural freedom than CbnR, which can adopt two markedly different conformations. In general the residues of the hinge region of the LTTRs family are variable. This can be observed by looking, at the composition of the hinge region in the 6 LTTRs present in *N. meningitidis* MC58 and the conservation scores measured by Consurf from more distantly related LTTRs. In this case the LTTRs from the SwissProt had lower overall sequence identity to CrgA (ClustalW scores 22 to 60). Residues 88-92 of the hinge were variable with Consurf conservation scores of 1-3, a value of 1 is most variable and 9 most conserved. The only well conserved residue was G93 with a conservation score of 8 and this residue is conserved in all the six LTTRs present in *N. meningitidis* MC58 (Figure S2C).

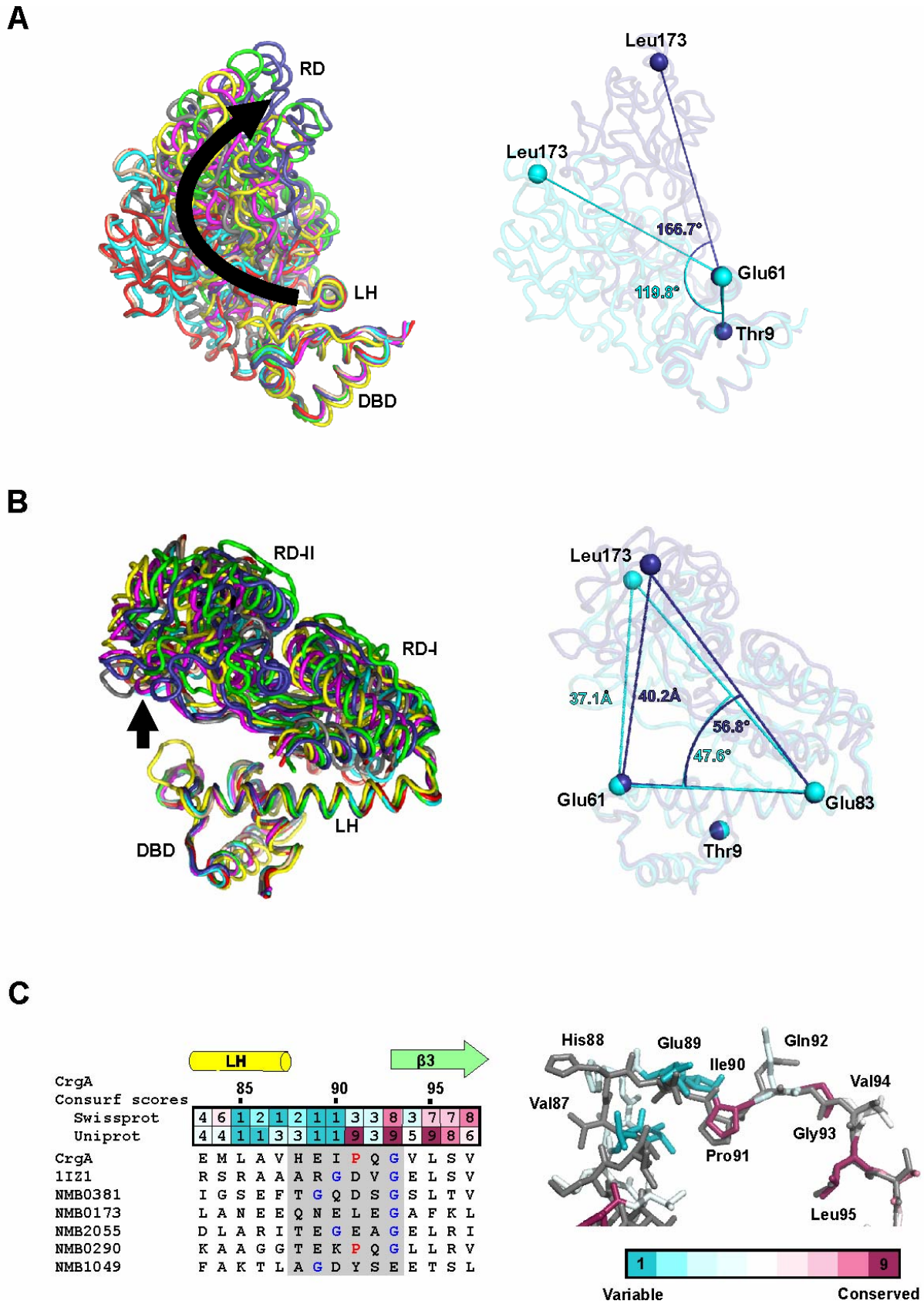


Figure S2 CrgA flexibility and properties of the hinge

(**A**) Overlays of the DBD and LH subunits illustrate the range of rotation of the RD around the LH in the eight CrgA subunits (left panel). The largest difference, in the calculated dihedral angle between the DBD-LH and the regulatory domain from T9, E61, E83 and L173 residues, was observed between chain A (blue) and H (cyan) (right panel) (**B**) Overlays of the DBD and LH subunits show a smaller range of pivot movement of the RD away from the LH in the eight CrgA subunits (left

panel). The distance and angle between the RD and LH is similar in Chain A (blue) and H (cyan) (right panel). (**C**) Multiple sequence alignment of the hinge region of CrgA, CbnR and the 5 other LTTR family members found in *N. Meningitidis* MC58 and structural comparison of the residues of the hinge region of CrgA, chain A (residues coloured by conservation score obtained using Consurf using the Uniprot database) and chain H (grey). The conservation scores for the residues of CrgA using Consurf from 60 related proteins obtained from the Swissprot and Uniprot databases are shown above the alignment. Residues of the hinge region are indicated by a grey background. Proline and Glycine found within the hinge region are highlighted in red or blue.

S3 Cluster analysis

UniProt contained over 20000 proteins with a predicted LTTR-type DNA binding domain. In order to reduce the size and sequence redundancy of the set only proteins contained within the UniRef50 database were used for the cluster analysis, based on pair-wise BLAST similarity scores, using CLANS (Frickey & Lupas, 2004).

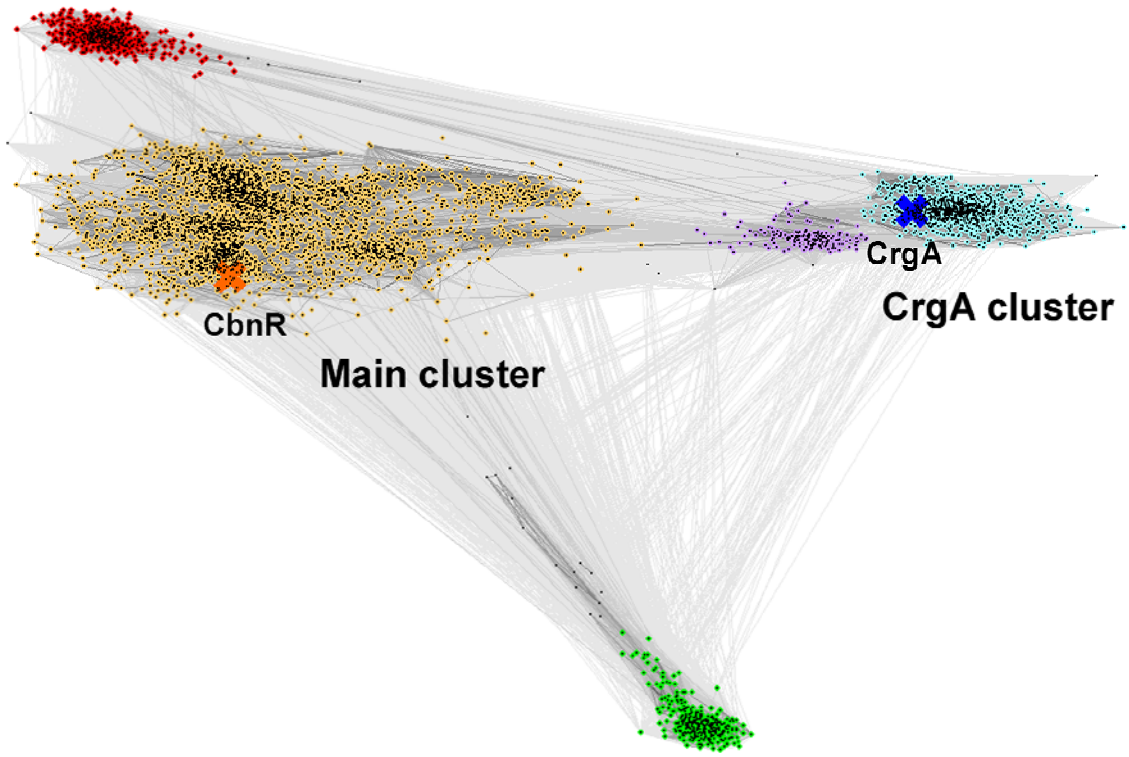


Figure S3 2D cluster analysis of the LTTR family using CLANS. The line colouring represents the BLAST pairwise P-values. The darker lines have lower P-values (i.e higher similarities) and the lighter lines are closer to the cutoff of 10^{-6}. Each dot represents one of the 4843 predicted LTTR sequences obtained from the UniRef50 database.

S4 Ion mobility–mass spectrometry of *apo* and *holo* CrgA

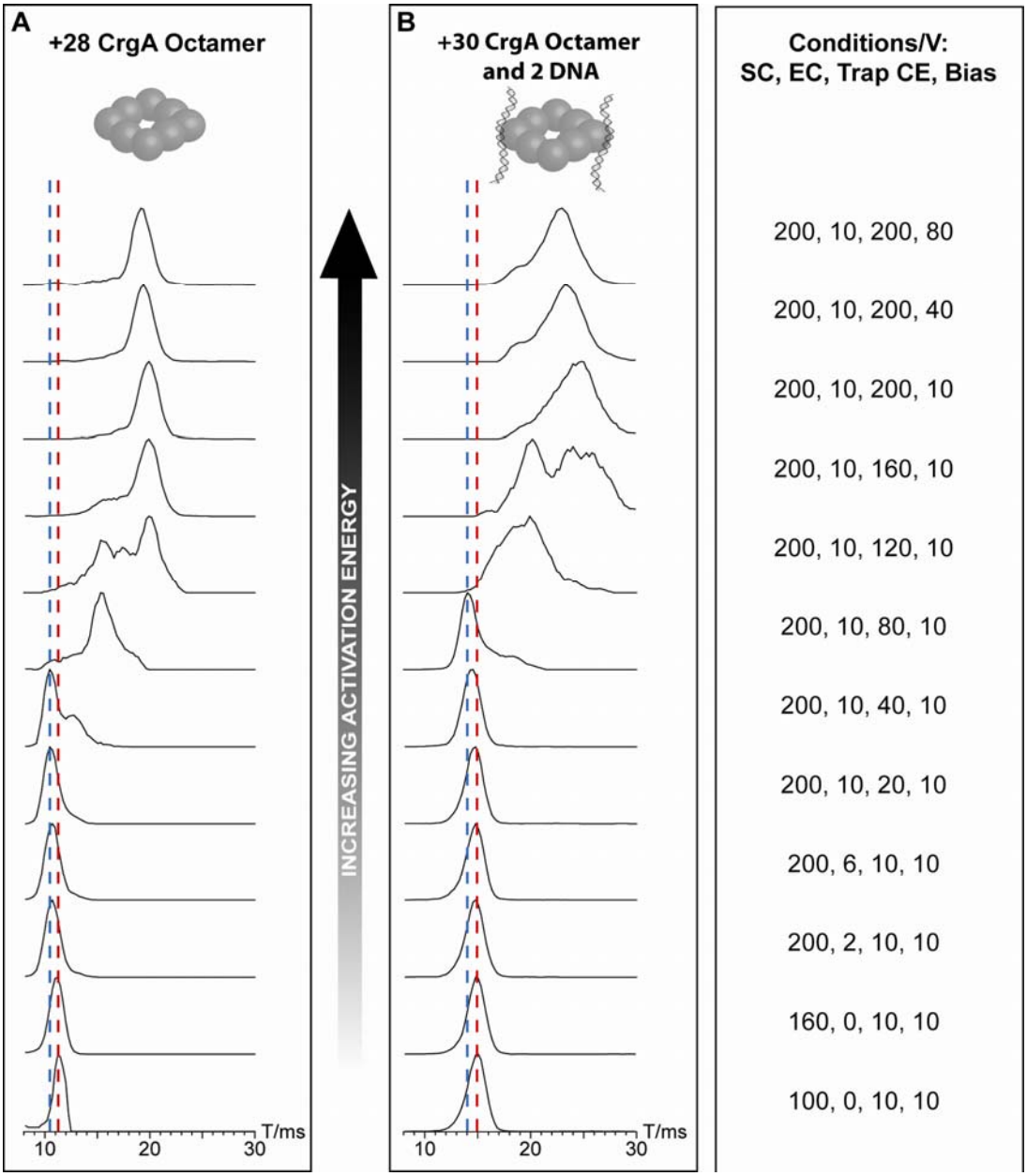
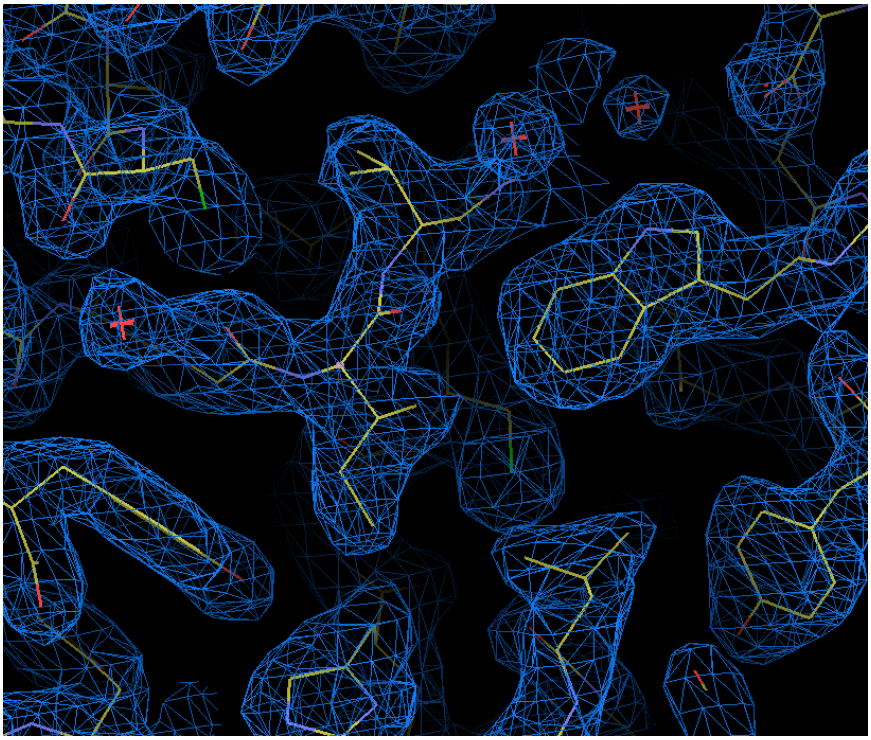


Figure S4: This figure shows the effect of increasing activation energy on the +28 charge state of the *apo* CrgA octamer (A) and the +30 charge state of the *holo* CrgA octamer binding two lengths of DNA (B). The effect of increasing activation energy on drift time, which indicates species mobility, follows a similar pattern for both charge states. Firstly the small decrease in drift time observed with increasing activation energy, from the drift time obtained at the lowest energy conditions, signifies at least some collapse of the *apo* and *holo* CrgA cavity. Following this, a much larger increase in drift time of the charge states is observed as the complexes are activated and unfolded. The red dotted line indicates the drift time obtained at low energy for the lowest charge state of each complex. The blue dotted line represents the minimum drift time achieved by each complex. The conditions under which spectra were acquired are shown to the right hand side of the figure. The first observation is that the energy required to cause maximum collapse of the CrgA octamer ring is greater when the DNA is bound. This suggests that the DNA maybe stabilising the ring. Secondly, the charge normalised maximum drift time reduction (which is achieved by subtracting the blue line from red line and multiplying by the charge of the species) is greater for the *apo* than the *holo* CrgA. This reduction in drift time corresponds to the degree of collapse experienced by each of the assemblies. For the *apo* CrgA the reduction in drift time is 28.84 ms compared to 22.80 ms for the *holo* CrgA. The binding of two lengths of DNA to the CrgA octamer appears to not only require more energy to cause ring collapse but also reduces the extent of the collapse that can be achieved.

S5 Crystallographic supporting information

A



B

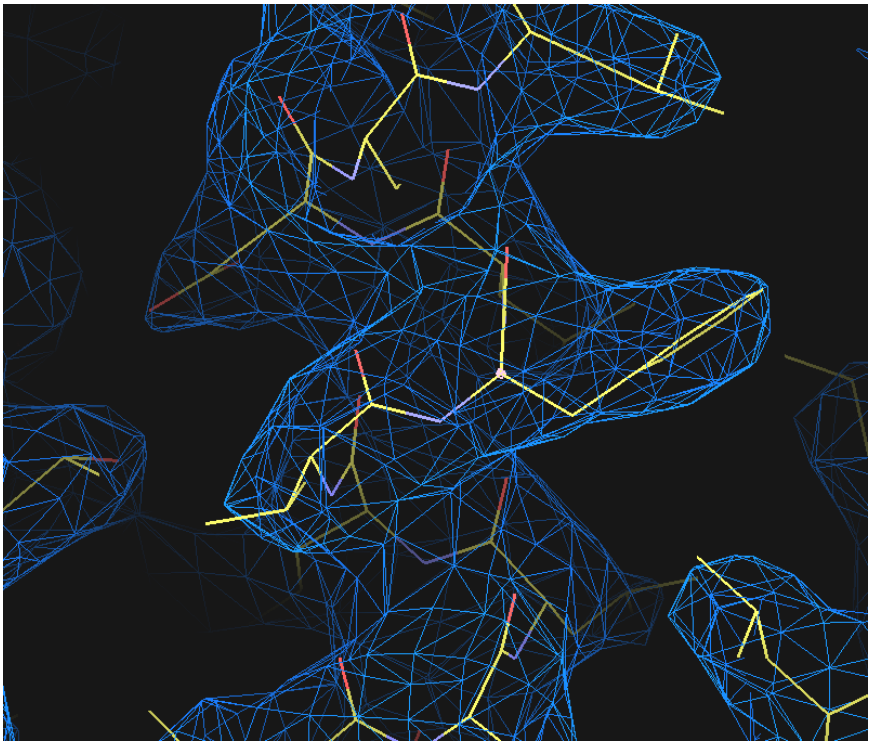
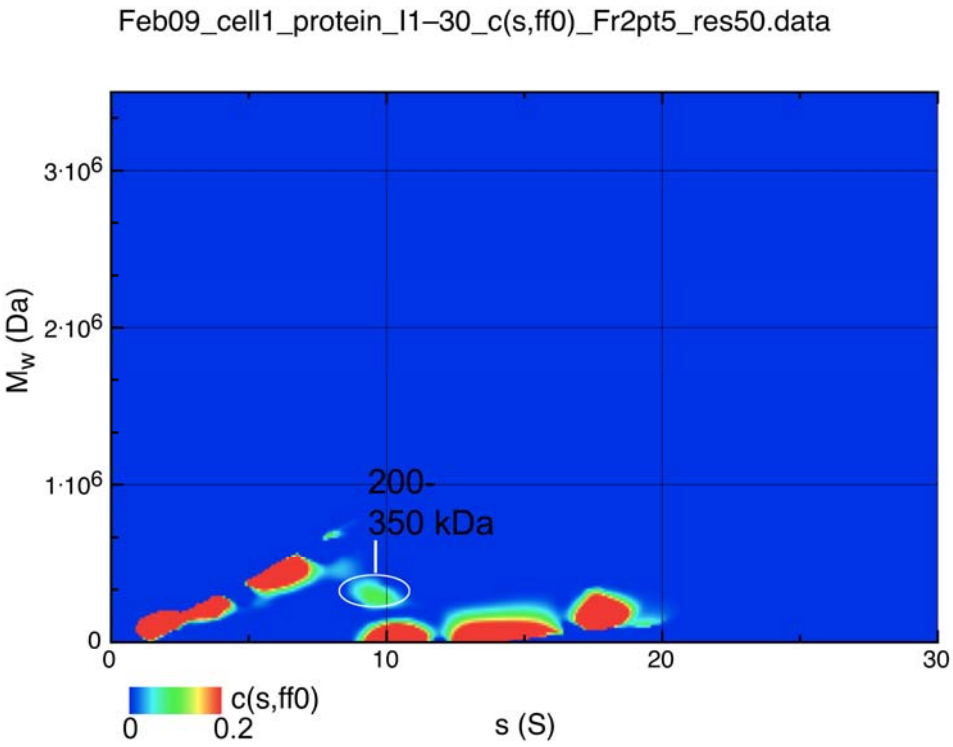


Figure S5 Representative electron density for the RD and full-length crystal forms of CrgA. Electron density maps shown in blue, contoured at 1σ for (**A**) RD monoclinic crystal form at 2.3 Å resolution and (**B**) the full-length orthorhombic crystal form at 3.2 Å resolution.

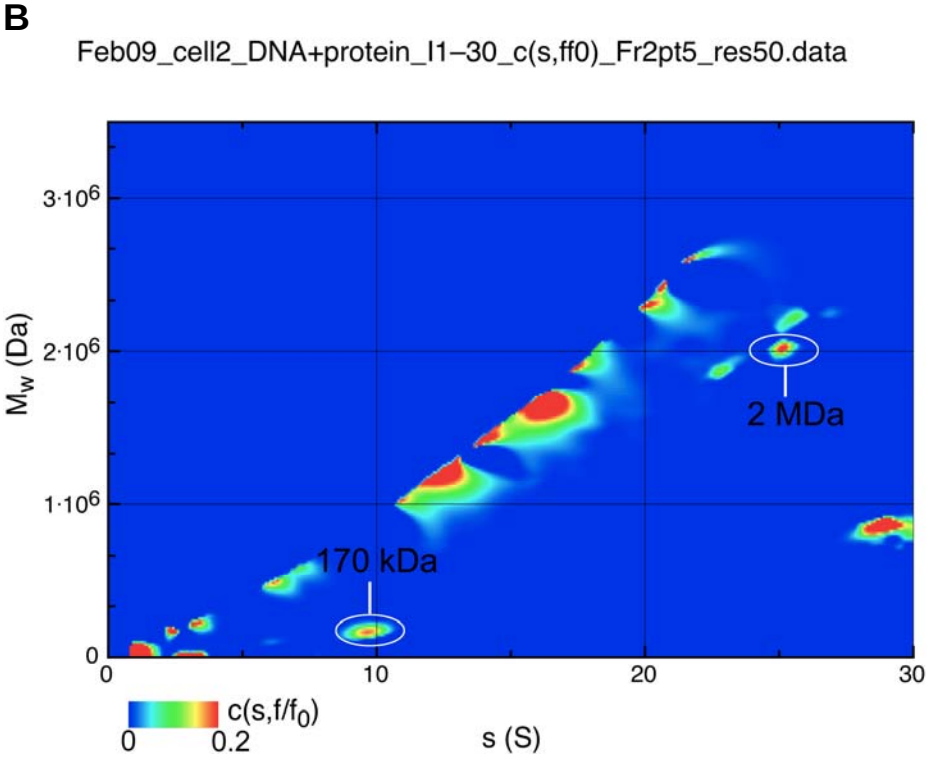
S6 c(s,M) plots

Two-dimensional contour plot representation of the apparent sedimentation coefficient (s) and molecular weight (Mw) distribution in the data shown in the c(s) plots, with the function c(*s,f/f*$_0$) shown as a blue-to-red contour plot (red=high population occupancy). Peaks along the diagonals may be trivial, but those off the diagonals are non-trivial and indicate resolution by these means of the 280 kDa octamer in the protein-only plot (A) and of that species and a 2 MDa species in the protein and DNA plot (B) (ringed in each case).

A



B



References

Frickey T, Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20:** 3702-3704

Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* **60:** 2256-2268