

Manuscript Number: JMB-D-15-00738R2

Title: Defining the Intrinsically Disordered C-terminal Domain of SSB
Reveals DNA-mediated Compaction

Article Type: Full Length Article

Section/Category: DNA replication, recombination and repair

Keywords: Intrinsic disorder, SASSIE, SANS, SAXS, Bacillus subtilis

Corresponding Author: Prof. panos sultanas, PhD

Corresponding Author's Institution: University of Nottingham

First Author: Matthew Green

Order of Authors: Matthew Green; louise hatter; Brookes Emre; David
Scott; panos sultanas, PhD

Abstract: The bacterial single stranded DNA binding protein SSB is a strictly conserved and essential protein involved in diverse functions of DNA metabolism, including replication and repair. SSB comprises a well-characterised tetrameric core of N-terminal oligonucleotide binding (OB) folds that bind single-stranded DNA (ssDNA) and four intrinsically disordered C-terminal domains of unknown structure that interact with partner proteins. The generally accepted, albeit speculative, mechanistic model in the field postulates that binding of ssDNA to the OB core induces the flexible, undefined C-terminal arms to expand outwards encouraging functional interactions with partner proteins. In this structural study, we show that the opposite is true. Combined small angle scattering with X-rays and neutrons coupled to coarse-grained modelling reveal that the intrinsically disordered C-terminal arms are relatively collapsed around the tetrameric OB core and collapse further upon ssDNA binding. This implies a mechanism of action, in which the disordered C-terminal domain collapse traps the ssDNA and pulls functional partners onto the ssDNA.

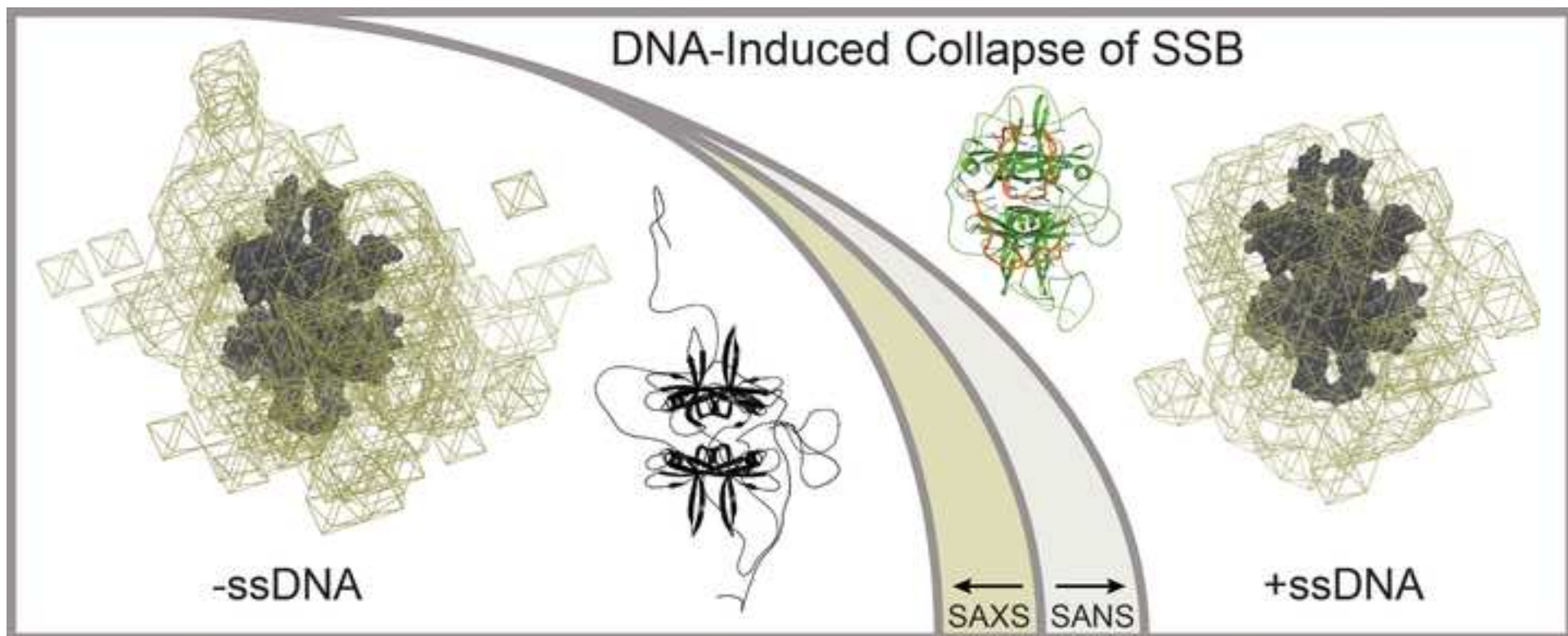
Dear Jane,

Many thanks for your prompt decision.

I have made the requested changes in the last two paragraphs exactly as you suggested.

Yours sincerely,

Panos Soultanas



HIGHLIGHTS

- 1.** The C-terminal domains of tetrameric SSB are intrinsically disordered and undefined
- 2.** In the absence of ssDNA they are relatively compact around the core SSB tetramer
- 3.** ssDNA binding induces maximal compaction around the core SSB tetramer
- 4.** Compaction upon ssDNA binding may localize protein binding partners onto the ssDNA

Defining the Intrinsically Disordered C-terminal Domain of SSB Reveals DNA-mediated Compaction

Matthew Green¹, Louise Hatter², Emre Brookes³, Panos Soultanas^{1,*} and David J. Scott^{2,4,5,*}

¹**Centre for Biomolecular Sciences**, School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.

²**ISIS Spallation Neutron and Muon Source**, Rutherford Appleton Laboratory, Oxfordshire, OX11 0FA, UK.

³**Department of Biochemistry**, MSC 7760, The University of Texas Health Center at San Antonio, 7703 Floyd Curl Drive, San Antonio TX 78229-3900, USA.

⁴**School of Biosciences**, University of Nottingham, Sutton Bonington Campus, Leicestershire, LE12 5RD, UK.

⁵**Research Complex at Harwell**, Rutherford Appleton Laboratory, Oxfordshire, OX11 0FA, UK.

* Corresponding Authors

Panos Soultanas: panos.soultanas@nottingham.ac.uk

David J. Scott: david.scott@nottingham.ac.uk

ABSTRACT

1 The bacterial single stranded DNA binding protein SSB is a strictly conserved and essential
2 protein involved in diverse functions of DNA metabolism, including replication and repair.
3 SSB comprises a well-characterised tetrameric core of N-terminal oligonucleotide binding
4 (OB) folds that bind single-stranded DNA (ssDNA) and four intrinsically disordered C-
5 terminal domains of unknown structure that interact with partner proteins. The generally
6 accepted, albeit speculative, mechanistic model in the field postulates that binding of
7 ssDNA to the OB core induces the flexible, undefined C-terminal arms to expand outwards
8 encouraging functional interactions with partner proteins. In this structural study, we show
9 that the opposite is true. Combined small angle scattering with X-rays and neutrons
10 coupled to coarse-grained modelling reveal that the intrinsically disordered C-terminal arms
11 are relatively collapsed around the tetrameric OB core and collapse further upon ssDNA
12 binding. This implies a mechanism of action, in which the disordered C-terminal domain
13 collapse traps the ssDNA and pulls functional partners onto the ssDNA.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

KEY WORDS

40 Intrinsic disorder, SASSIE, SANS, SAXS, *Bacillus subtilis*
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

INTRODUCTION

1 All organisms protect single stranded DNA (ssDNA) intermediates of DNA metabolism with single
2 stranded DNA binding proteins (SSBs). Bacterial SSBs are ubiquitous well-conserved tetramers
3 comprising a core of four ssDNA binding N-terminal domains (NTD) and four intrinsically
4 disordered C-terminal domains (CTD) that recruit a diverse repertoire of proteins involved in DNA
5 repair and replication [1,2 and 3]. The NTD consists of an oligosaccharide binding (OB) fold that
6 binds to ssDNA in a sequence independent manner and forms tight inter-domain interactions that
7 stabilise the tetramer in solution. SSB tetramers exhibit a degree of cooperativity upon binding to
8 ssDNA forming extended bead-like structures with the ssDNA wrapped around the beads [2,4]
9 *Bacillus subtilis* SSBs have been well-characterised and the crystal structure of the NTD of SSB 2
10 has been solved [5]. Generally, crystallography of bacterial SSB has required either complete or
11 partial removal of the CTD, leading to the assumption that the CTD is intrinsically disordered. In
12 addition, various intrinsically disordered protein (IDP) prediction algorithms predict that the CTD is
13 disordered [6].
14
15
16
17
18
19
20
21
22
23

24 Given its ubiquitous and essential functions across the bacterial kingdom, it is important to fully
25 understand the structure/function relationships that underpin its molecular mechanism of action.
26 This can only be achieved if we define the function of the intrinsically disordered CTD. It is only
27 relatively recently that studies on SSB have begun to tackle the mechanistic coupling of the SSB
28 NTD and CTD and their distinct functions of ssDNA binding and protein binding, respectively [7,8].
29 However, the mechanism by which the CTD carries out its role and the structure-function
30 relationship of the two domains are still not understood. This is mainly due to the limitations of
31 biophysical tools available for studying IDPs [9,10]. Structural studies of IDPs are a challenge and
32 require the appreciation of an ensemble of structures or a mean structure, rather than typical rigid
33 definitions. Nonetheless, it is possible in principle to understand how intrinsic disorder functions
34 mechanistically when multiple techniques are used synergistically. Here, we have combined small
35 angle scattering (SAS) with X-rays (SAXS) and neutrons (SANS) coupled to coarse-grained
36 modelling to uncover the structure-function relationship of the intrinsically disordered CTDs relative
37 to the core tetrameric NTD and how two well characterised single stranded substrates, dT35 and
38 dT70, modulate this relationship.
39
40
41
42
43
44
45
46
47
48
49

50 The recently postulated maintenance hub theory ascribes a speculative role for SSBs as a
51 scaffold hub that recruits proteins involved in DNA metabolism and localises them to ssDNA [1].
52 This theory acknowledges the utility of intrinsic disorder, which is typically associated with
53 promiscuous but reasonably tight binding. The majority of bacterial SSBs have very similar domain
54 organisation and share close homology. The CTD of *B. subtilis* SSB consists of a 60 amino acid
55 glycine and proline-rich region, which is typical of flexible protein regions [11]. This is followed by
56 the protein binding region (PBR), a well-conserved 9 amino acid acidic region (DISDDDLPF) at the
57
58
59
60
61
62
63
64
65

C-terminal end of the arm, which binds to SSB interaction partners. Although this acidic region is well conserved in different species, their SSB interactomes are species dependent [1].

Nuclear magnetic resonance (NMR) studies of *Escherichia coli* SSB have shown that the PBR has a weak affinity for the ssDNA-binding channel on the tetrameric NTD core [12]. Deletion of the PBR enhances NTD ssDNA affinity, suggesting that ssDNA may displace the PBR in order to bind to the tetrameric OB core [7]. Consequently, this has led to a generally accepted but speculative mechanistic theory in the field proposing that ssDNA displaces the PBR releasing the CTD into solution making it more accessible to its binding partners [7]. Our structural study of *apo* and *holo* *B. subtilis* SSB in solution revises this model, showing that ssDNA binding leads to compaction of the CTDs. This implies a molecular mechanism of action, in which the intrinsically disordered CTD collapse traps the ssDNA and pulls functional partners onto the ssDNA.

RESULTS

Small angle scattering of DNA bound and unbound SSB

The effect of ssDNA-binding to the intrinsically disordered CTD of SSB has not been defined. It is, therefore, important to define whether the CTD extend or compact upon binding of the SSB to ssDNA. In order to determine the effect of ssDNA upon the compaction of SSB, the radius of gyration (R_g) and maximum particle dimension (D_{max}) values of *apo* SSB and *holo* SSB complexes were measured by small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS), respectively (**Fig. 1A-B**). Comparison between the two scattering methods is accurate provided the contribution from hydration is accurately taken into account. A detailed explanation of how hydration was considered can be found in the Supplementary Information (**Supplementary Fig. 1**).

In order to study the DNA induced alteration in SSB conformation, contrast matching SANS was used to phase out the scattering from DNA. Hydrogen and deuterium have very different neutron scattering length densities, as do proteins and DNA. As such, at 67% D₂O DNA scattering length density matches that of the solution and hence no excess scattering is seen from the DNA component and only scatter from the protein is observed. As the scattering length densities of the protein and DNA are quite close (corresponding to 40 % D₂O and 67 % D₂O, respectively), the contrast of the protein can be further increased by per-deuteration of the protein, which has a theoretical match point of 120 % D₂O. The scattering power is proportional to the square of the difference between the solvent scattering length density and the protein's match point [13]. Hence, per-deuteration will increase the total scatter from the protein to approximately 7 fold that of the hydrogenated protein, at 67 % D₂O. Per-deuterated SSB was produced in high yields

(**Supplementary Fig. 2**), and was found to be stable and monodisperse to a concentration of up to 20 mg/ml only in the presence of ssDNA. However, it was not possible to produce per-deuterated SSB^{Δ107-171} (armless SSB lacking the disordered CTD) due to insolubility *in vivo*.

SANS experiments were carried out with two SSB:DNA complexes and compared with the SAXS measurements on *apo* SSB (**Fig. 1**). At high concentrations, per-deuterated SSB was prone to some precipitation in the absence of ssDNA. However, the protein was stable at 37°C when complexed to ssDNA. Due to precipitation, it was not possible to collect SANS data at the concentrations necessary for good signal-to-noise for *apo* SSB and armless SSB. Therefore, this study compares SAXS and SANS curves made possible by accurately accounting for hydration and using scale free analysis methods (see also **Supplementary Fig. 1**).

The R_g values, derived by Guinier analysis (**Fig 1B** and **Table 1**) give a parameter related to molecular extension. The R_g for the *apo* SSB, derived from SAXS, is 3.43 nm. The SANS measurements showed that the R_g for equimolar dT35 and dT70-bound SSB both fell to 2.87 and 3.03 nm respectively, indicating that the arms are more compact than in the *apo* complex. Due to scatter length density matching at 67% D₂O, this R_g value does not represent any contribution from the ssDNA. Therefore, this reduction represents a compaction of the CTD around the ssDNA, which tightly wraps around the NTD according to crystal structures [6].

We then measured the approximate occluded site size for *B. subtilis* SSB (**Supplementary Fig. 3**) and the lengths of ssDNA (35mer and 70mer) were chosen to reflect the binding modes of both the *B. subtilis* SSB as well as the highly homologous *E. coli* SSB, which has been extensively studied [2]. Although, any partially free unbound nucleotides will not contribute to the scatter at 67% D₂O. The binding site size of *B. subtilis* SSB was approximated using ssDNA binding induced intrinsic tryptophan quenching (**Supplementary Fig. 3**). In accordance with previous characterisations [15] and under the experimental conditions used, dT70 engages all four monomers forming a fully wrapped complex while dT35 only partially wraps, engaging with 2-3 monomers (**Supplementary Fig. 3**).

Distance distribution functions – P(r) plots – (**Fig 1C**) were generated from each of the scattering curves using GNOM. Each distribution was found to have different maximum dimensions (D_{max}), which are not related to the R_g values (**Table 1**). The origin of this is the ensemble nature of the distribution. In the conformational ensemble, there is a mixture of different conformers, some shorter and some longer. The longer conformers are the ones that contribute most to D_{max} , hence they only need to be present in relatively small amounts to give the same D_{max} value. In contrast, the R_g as the second moment of the distribution gives information about the distribution of mass around the centre of mass of the particle and it is therefore sensitive to conformational change. Thus, the reduction in R_g observed upon DNA binding, is due to a shift in the population to favour more compact conformers. As such, R_g is a more sensitive parameter in

1 the analysis of conformational shift in the populations of conformers in solution. The D_{\max} correlates
2 well with the ssDNA-induced compaction trend supported by the observation that dT35 reduces
3 the D_{\max} from 11.2 to 10.4 nm and dT70 reduces it further to 9.5 nm. This independent SAS
4 analysis corroborates the Guinier analysis. However, this also confirms that a small population of
5 the SSB complex has the ability to adopt extended structures, as expected in a system with high
6 intrinsic flexibility.
7

8
9
10 Kratky [14] curves for *wt* SSB and SSB $^{\Delta 107-171}$ (armless SSB lacking the CTDs) show clear
11 deviation (**Fig. 1D**). In Kratky analysis, well-folded proteins have a distinctive initial parabolic peak,
12 which is exemplified by the curve of the armless mutant. As proteins become more flexible, they
13 deviate from this peak causing broadening of the peak and in the case of fully disordered proteins
14 a plateau. The *wt* SSB deviates from the armless mutant showing that the protein has a degree of
15 flexibility. Since removal of the CTD moved the peak, this confirms that it is exclusively the CTDs
16 (residues 107-171) that contribute to this flexibility in solution. This result is in line with previous
17 observations that the SSB CTD has flexible properties (as predicted through difficulties with protein
18 crystallisation and secondary structure predictions) and the observation that the NTD forms a rigid
19 tetramer with no major flexibility [3,6]. Addition of dT35 or dT70 at one-fold molar excess over
20 tetramer does not change the shape of the curve suggesting that despite compaction, the CTD
21 remains somewhat mobile.
22
23
24
25
26
27
28
29

30
31 As the bound solvent layer in SAXS contributes to the scattering curve, the hydration shell
32 effect was calculated and subtracted. Hydrating the structures gave a maximum increase in R_g of
33 0.2 nm, indicating that the differences seen in the R_g measurements from SANS cannot be due to
34 an incorrect description of hydration effects (for a full description of this normalisation, see
35 **Supplementary Fig. 1**).
36
37
38
39

40 41 42 43 **Analysing the Significance of Compaction**

44
45 In order to estimate the significance of this compaction, an ensemble of 10,000 structures was
46 produced using discrete molecular dynamics, as described in the methods section. A sufficiently
47 broad R_g range was explored in order to give poor fitting to the experimental data at both upper
48 and lower ends (**Fig. 2**). Best and worse curves for each sample are shown in **Supplementary**
49 **Fig. 4**. This modelling reveals that the SSB's theoretical R_g range is between 2.6-6.5 nm, which
50 correlates with the experimental R_g obtained from the armless mutant SSB $^{\Delta 107-171}$ i.e. 2.61 nm.
51 These two independently acquired parameters define the maximum structure compaction and their
52 correspondence validates the Guinier analysis. Based on this minimal expected state, our
53 measurements of ssDNA bound SSB at 2.87 and 3.03 nm suggest a highly compact structure.
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Furthermore, the *apo* SSB measurement of 3.43 also suggests a compact *apo* structure with ssDNA inducing further compaction (**Fig. 2**).

To visualise the degree of compaction the models with the best fit were collated to produce a density plot (**Fig. 3**). These plots show clearly that the CTDs in the SSB tetramer are more collapsed in the presence of dT70 or dT35 than in the absence of ssDNA. Whilst a best single model can be useful to help visualise the degree of compaction (**Fig. 4**), it is critical to recognise that the assay is in bulk phase and due to intrinsic flexibility a single model is an averaged representation of an ensemble. A density plot that contains multiple structures, helps to highlight this issue yet it still represents an averaged representation of the ensemble. Deconvolution of multiple populations is a major limitation in all SAS experiments and is only overcome by ensuring monodispersity and testing multiple conditions that shift the structure of the ensemble. These experiments represent this idea and clearly show that addition of ssDNA shifts the average population to a more compact state.

DISCUSSION

The tetrameric NTD structure of SSB and its interaction with ssDNA are well studied but the inherently disordered structure of the CTD has hindered efforts to define the structure-function relationships of the native SSB, as no reliable structure of the full length SSB currently exists. Previously published models propose that the CTDs in the SSB tetramer are extended to encourage protein capture but such models have not been experimentally verified [7]. Therefore, even with the extended studies of bacterial SSBs in the literature we still do not fully understand how the intrinsically disordered CTDs of the SSB tetramer function. Using a combination of biophysical and molecular modelling approaches, for the first time, we have demonstrated that the CTDs in the SSB tetramer are not extended. Instead, we have shown that they are relatively collapsed around the core NTD tetramer in the absence of ssDNA or other protein-binding partners. This may be related to previously observed interactions between the PBRs and NTDs [12], which may hold the CTDs in close proximity to the core NTD tetramer. Upon ssDNA binding, the CTDs collapse further almost to the maximum possible compaction, suggesting that the CTDs cap the ssDNA binding groove and may interact with the NTD tetramer and/or ssDNA directly. This capping of the ssDNA by the collapsed CTDs will have a functional relevance as it could accommodate the diffusion of the SSB tetrameric beads along ssDNA by ensuring that the ssDNA remains within the binding channel. However, it is likely that these interactions would be transient as a high degree of flexibility is still observed in the presence and absence of ssDNA [16]. Given that modelling suggests the extended CTD arms are highly hydrated (**Supplementary Fig. 1**),

there will also be a large thermodynamic drive to compaction from the release of water molecules upon CTD collapse that may also drive the compaction.

Our data show that in the absence of ssDNA, SSB can bind to multiple protein partners via its PBRs suggesting that the CTDs are loosely associated with the N-terminal tetrameric core but in a relatively collapsed state with the core tetrameric NTD. Upon ssDNA binding, the CTDs collapse further pulling their binding partners towards the ssDNA. This structural/functional model is consistent with the SSB biological role in protecting and/or processing exposed ssDNA and with previous observations that ssDNA binding increases the affinity of SSB for its protein partners [7]. Recent work has shown compaction of SSB nucleoprotein fibers by atomic force microscopy and total internal reflection fluorescence microscopy, in agreement with our observations. The authors suggest that this intramolecular condensation is protein-mediated and highlight the possible significance of the CTD [23]. They go further to suggest that interactions between the CTD and SSB's protein-binding partners may facilitate compaction or expansion of the SSB nucleoprotein fibers thereby regulating access to ssDNA. Our work is completely consistent with these observations. Furthermore, it demonstrates a method by which further work can proceed to fully describe the mechanism of action and effects of SSB's protein-binding partners.

MATERIALS AND METHODS

Small Angle Scattering

Wild type SSB was purified using the method previously described [15,16]. The SSB^{Δ107-171} (CTD removed) construct included an N-terminal histidine tag. This protein was purified using nickel affinity chromatography in place of the ion-exchange column previously described. SAXS experiments were carried out at 37°C on beam line BM29 at the ESRF, Grenoble. All experiments were carried out in 20 mM Tris pH 7.5, 20 mM NaCl and 1 mM DTT at 5 protein concentrations up to a maximum of 10 mg/ml, in order to determine the extent of thermodynamic non-ideality in the system. Samples were dialysed for 16 hours in 2 L of buffer prior to measurement. The dialysis buffer provided a perfect buffer match was used for scattering experiments. 10 measurements were taken of each sample using a flow cell to reduce radiation damage. The first curve was compared to the other 9 using DATcomp (ATSAS package) and damaged curves were rejected accordingly prior to averaging of the other frames. This process was also checked manually to ensure no more than one frame was rejected. All sample concentrations were determined by UV spectrophotometry after dialysis and prior to addition of ssDNA, which was also quantified by the same method.

SANS experiments were carried out at 37°C on the LOQ instrument at ISIS Spallation Neutron and Muon Source, Harwell (UK). All experiments were carried out in 20 mM Tris pH7.5, 20 mM NaCl, 1

1 mM DTT and 67% D₂O with fully deuterated SSB. Samples were dialysed for 16 hours in 2 L of buffer prior to measurement. The dialysis buffer provided a perfect buffer match was used for scattering experiments. Deuterated SSB was prepared from a 1 litre bacterial growth in fully deuterated media (Silantes, UK) and grown to an OD₆₀₀ = 1.5 before induction with 1 mM IPTG. The culture was grown for a further 12 hours at 30°C. Purification of deuterated SSB was the same as for the wild type i.e. the final degree of deuteration was 99.38% ±0.15 as determined by mass spectrometry (**Supplementary Fig. 2**). All ssDNA were purchased from MWG (Germany) and added to SSB in equimolar ratio to tetramer. All SANS experiments were carried out at 5 concentrations up to a maximum of 25 mg/ml, to detect aggregation and thermodynamic non-ideality.

Data were processed with Primus [15] and ScÅtter [16]. CRY SOL and CRYSON [17] were run in command-line mode to generate theoretical scattering curves using the following adjusted parameters, as defined in the programs manual: /lm 50, /fb 18, /ns 2000. Unless otherwise stated, the contrast of the hydration shell parameter (/dro) was set to 0.00 to prevent an automatic hydration correction.

Modelling realistic flexible conformers

A full-length model of *B. subtilis* SSB was created using Swiss Modeller (for an NTD model) [20] and manual CTD building in Coot [21]. This model was subjected to a series of iterative discrete molecular dynamic simulations (using SASSIE) with regressive R_g filtering, to create the most compact structure possible. This minimal structure was used as a starting point for an expansion simulation, without restrictive filtering, to reach sufficiently large models. This simulation was run twice and a total of 10,000 distinct structures were produced. Theoretical SAXS and SANS curves were calculated for these models using CRY SOL and Xtal2sas respectively. In CRY SOL, parameters for Fibonacci grid order and maximum harmonic order were increased to create curves with the maximum resolution. Solvent density and solvation shell contrast were set at default, as suggested by the explicit hydration analysis (**Supplementary Fig. 1**). SASSIE's inbuilt X² filter and density plot generator were used for further analysis, as shown in **Fig. 2** and **Fig. 3** respectively [22].

Dynamic Light Scattering

All dynamic light scattering experiments were carried out on a Malvern Zetasizer Nano at ISIS, Harwell (UK). Measurements were carried out at ambient temperature with SAS samples pre and post measurement to determine monodispersity of all samples.

ACKNOWLEDGEMENTS

We would like to acknowledge Luke Clifton (ISIS) who provided helpful guidance throughout. We are grateful to Dr. Neil Oldham's group for mass spectroscopy, especially Dr. Matthew Jenner who measured per-deuterated SSB. Small angle neutron scattering data was collected through a beam time award to D.J.S.. Small angle X-ray data was collected through the Midlands (UK) BAG allocation award. This work was supported by a Wellcome Trust grant WT091968 and a Biotechnology and Biological Sciences Research Council (UK) grant BB/K021540/1 to P.S.. E.B. is supported by an NIH grant (K25GM090154). This work benefitted from CCP-SAS software developed through a joint Engineering and Physical Sciences Research Council (UK) grant EP/K039121/1 and National Science Foundation (USA) grant CHE-126582. D.J.S. is a Senior Molecular Biology and Neutron Fellow supported by the Science and Technology Facilities Council (UK).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

REFERENCES

1. Costes, A., Lecointe, F., McGovern, S., Quevillon-Cheruel, S. & Polard, P. (2010). The C-terminal domain of the bacterial SSB protein acts as a DNA maintenance hub at active chromosome replication forks. *PLoS Genet.* 6, E1001238.
2. Lohman, T.M. & Ferrari, E.M. (1994). *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.* 63, 527-570.
3. Antony, E., Weiland, E., Yuan, Q., Manhart, C.M., Nguyen, B., Kozlov, A.G., et al. (2013). Multiple C-Terminal Tails within a Single *E. coli* SSB Homotetramer Coordinate DNA Replication and Repair. *J. Mol. Biol.* 425, 4802-4819.
4. Ferrari, E.M., Bujalowski, W. & Lohman, T.M. (1994). Co-operative binding of *Escherichia coli* SSB tetramers to single-stranded DNA in the (SSB)₃₅ binding mode. *J. Mol. Biol.* 236.1, 106-123.
5. Yadav, T., Carrasco, B., Myers, A.R., George, N.P., Keck, J.L. & Alonso, J.C. (2012). Genetic recombination in *Bacillus subtilis*: a division of labor between two single-strand DNA-binding proteins. *Nucleic Acids Res.* 40, 5546-5559.
6. Savvides, S.N., Raghunathan, S., Fütterer, K., Kozlov, A.G., Lohman, T.M. & Waksman, G. (2004). The C-terminal domain of full-length *E. coli* SSB is disordered even when bound to DNA. *Protein Sci.* 13, 1942-1947.
7. Kozlov, A.G., Cox, M.M. & Lohman, T.M. (2010). Regulation of single-stranded DNA binding by the C termini of *Escherichia coli* single-stranded DNA-binding (SSB) protein. *J. Bio. Chem.* 285, 17246-17252.
8. Matsumoto, T., Morimoto, Y., Shibata, N., Kinebuchi, T., Shimamoto, N., Tsukihara, T., et al. (2000). Roles of Functional Loops and the CD-Terminal Segment of a Single-Stranded DNA Binding Protein Elucidated by X-Ray Structure Analysis. *J. Biochem.* 127, 329-335.
9. Ding, F. & Dokholyan, N.V. (2006). Emergence of protein fold families through rational design. *PLoS Comput. Biol.* 2, e85.
10. Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struc. Biol.* 19, 23-30.
11. Radivojac, P., Lakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N. & Dunker, A.K. (2007). Intrinsic disorder and functional proteomics. *Biophys. J.* 92, 1439-1456.
12. Shishmarev, D., Wang, Y., Mason, C.E., Su, X., Oakley, A.J., Graham, B, et al. (2014). Intramolecular binding mode of the C-terminus of *Escherichia coli* single-stranded DNA binding protein determined by nuclear magnetic resonance spectroscopy. *Nucleic Acids Res.* 42, 2750-2757.
13. Sarachan, K.L., Curtis, J.E. & Krueger, S. (2013). Small-angle scattering contrast calculator for protein and nucleic acid complexes in solution. *J. Appl. Crystallogr.* 46, 1889-1893.
14. Rambo, R.P. & Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496, 477-481.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
15. Green, M., Gilhooly, N.S., Abedeen, S., Scott, D.J., Dillingham, M.S. & Soutanas, P. (2014). Engineering a reagentless biosensor for single-stranded DNA to measure real-time helicase activity in *Bacillus*. *Biosens. Bioelectron.* 61, 579-586.
 16. Zhang, W., Dillingham, M.S., Thomas, C.D., Allen, S., Roberts, C.J. & Soutanas, P. (2007). Directional loading and stimulation of PcrA helicase by the replication initiator protein RepD. *J. Mol. Biol.* 371, 336-348.
 17. Petoukhov, M.V., Konarev, P.V., Kikhney, A.G. & Svergun, D.I. (2007). ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* 40, s223-s228.
 18. Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L., Tsutakawa, S.E., et al. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* 6, 606-612.
 19. Svergun, D., Barberato, C. & Koch, M.H.J. (1995). CRY SOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28, 768-773.
 20. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195-201.
 21. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biol. Crystallography* 66, 486-501.
 22. Curtis, J.E., Raghunandan, S., Nanda, H. & Krueger, S. (2012). SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comp. Phys. Com.* 183, 382-389.
 23. Bell, J.C., Bian, L. & Kowalczykowski, S.C. (2015). Imaging and energetic of single SSB-SSDNA molecules reveal intra-molecular condensation and insights into RecOR function. *eLife*, 4, e08646.

FIGURE LEGENDS

Figure 1: SAXS and SANS data

A. SAXS and SANS curves. SAXS data have more resolution than SANS affording a smaller bin size due to higher intensity at source. SANS data are less dense with larger bins due to lower neutron flux. Each curve is arbitrarily separated on the log scale for clarity.

B. Guinier plots of the data showing linearity over the Guinier range to give accurate R_g values ± 0.02 . Again, data is offset by 1 log unit for clarity.

C. Distance distribution functions of SANS and SAXS data. The right hand shoulder is indicative of a population of atoms in a more extended conformation. This shoulder is greatly reduced by removal of the CTD. The maximum distances came from an unconstrained fit to the data and as such reflect the point where the data first crosses the x-axis. Both apo and holo full length SSB data have similar D_{max} values, though a slight reduction can be seen upon ssDNA addition.

D. R_g normalised Kratky [13] plot showing SAXS and SANS curves on the same axis.

Figure 2: Ensemble of Model R_g vs χ^2

The χ^2 value, calculated using internal SASSIE modules CRY SOL and Xtla2SAS, represent the goodness of fit between the theoretical scattering curve of each model in the ensemble versus the experimental data. Dotted vertical lines show the R_g of the best model from each data set (wt in black, SSB-dT35 in red and SSB-dT70 in blue) for clear comparison of shifts between graphs. This clearly shows that the holo SSB curves fit better to models with R_g values between ~ 2.8 - 3.1 nm whereas apo SSB best fits models with R_g values between ~ 3.3 - 3.9 nm, reflecting a major non-overlapping shift in the ensemble. This method of analysis utilises the entire SAS curve unlike Guinier or Porod analysis, which only use a limited q range in the curve. Therefore, this method gives a more dependable measurement.

Figure 3: Density Plots of Best Models

Each density plot represents a combination of structures with the lowest 2000, 1000, 100 or 50 χ^2 values. The compaction trend from apo to holo is less apparent when sampling higher numbers of structures but becomes very clear when looking at the best 100 or 50 models.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

For clarity the spherically averaged model diameter has been drawn around the best 50 models. These were calculated by projection approximation as 17919.93, 10732.47 and 10311.84 Å² for apo SSB, dT70-SSB and dT35-SSB respectively.

Figure 4: Single Best and Worst Models, plus and minus ssDNA.

The best models for apo and holo SSB are the models with the lowest χ^2 and therefore represent the best fit to the experimental curves. The worst fit for apo SSB, the structure with the highest χ^2 , was very similar to the worst fit for the holo SSB sample, therefore only one is shown. As discussed in the text, these models represent a highly flexible ensemble in solution. Therefore, these models are only included to approximate the arm expansion and the average degree of compaction we observe. Defining a single structure is not otherwise useful for highly disordered proteins.

Table 1

	Conc. (c) (mg/ml)	c Range (mg/ml)	$I_{(0)}/c$	Guinier R_g	Dmax	Theoretical Protein MW	MW from $I(0)$	q range
dT35	18.00	1.14-18	0.15 ± 0.08	2.87 (± 0.02)	9.5 (± 0.2)	74915.1	45.24 ± 4.52	0.09-2.9
dT70	18.00	9-18	0.17 ± 0.09	3.03 (± 0.03)	10.4 (± 0.2)	74915.1	49.90 ± 4.99	0.09-2.9
Armless	0.4	0.37-1.5	32.51 ± 1.80	2.47 ± 0.05	7.9 (± 0.2)	48304.3	49.29 ± 5.18	0.04-5.0
wt apo	0.57	0.28-1.14	54.37 ± 0.16	3.37 ± 0.22	11.2 (± 0.2)	74915.1	82.43 ± 0.23	0.04-4.5

Figure 1
[Click here to download high resolution image](#)

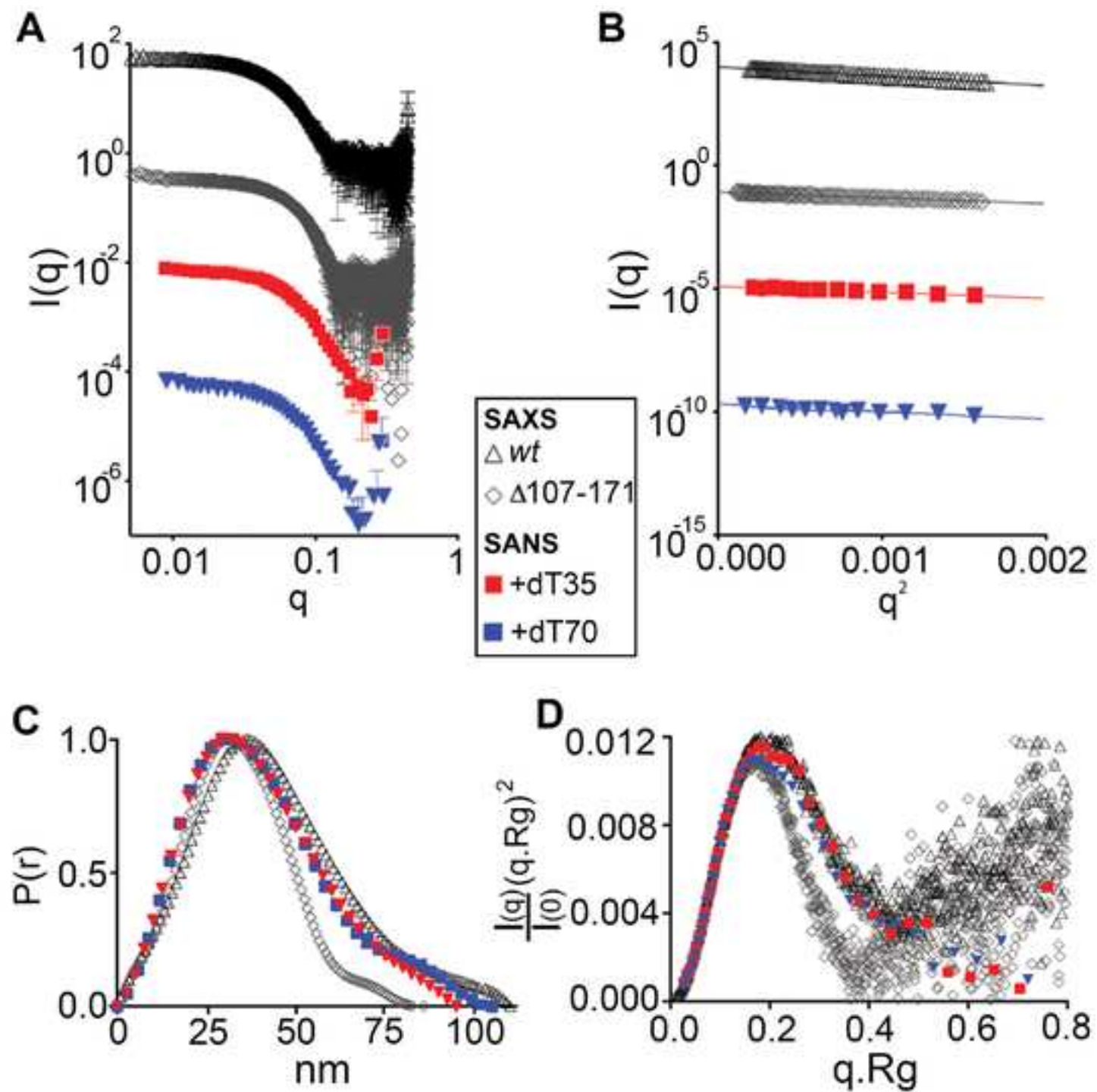


Figure 2

[Click here to download high resolution image](#)

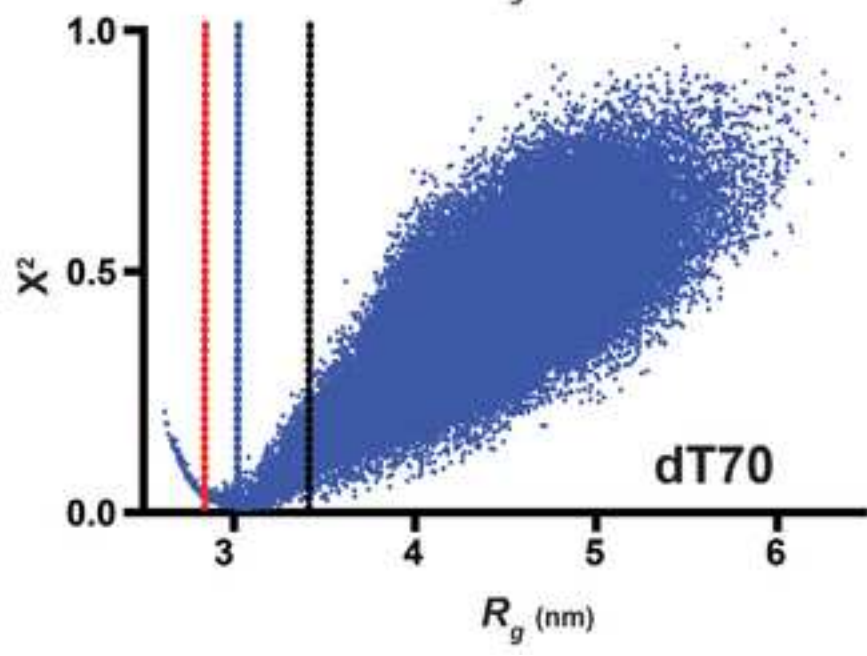
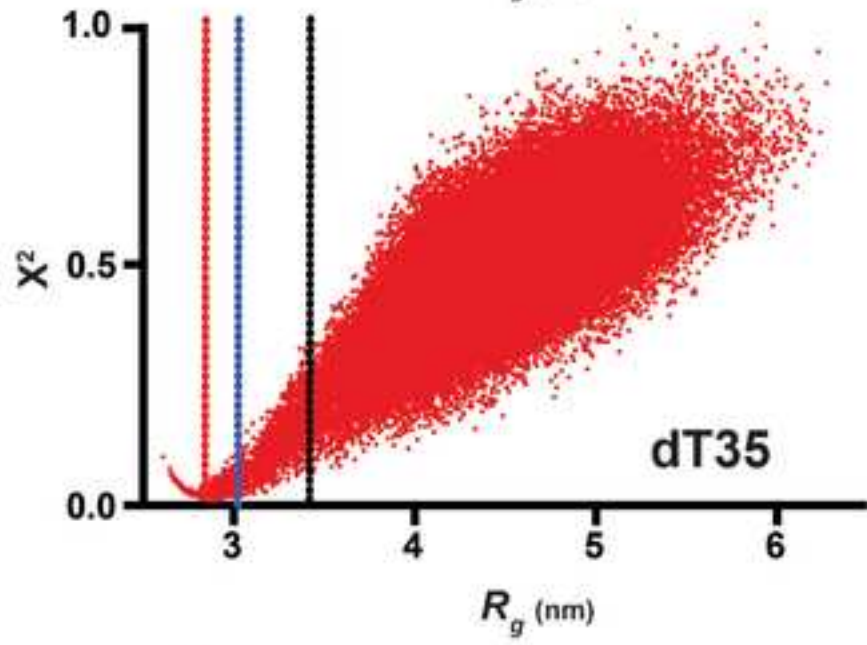
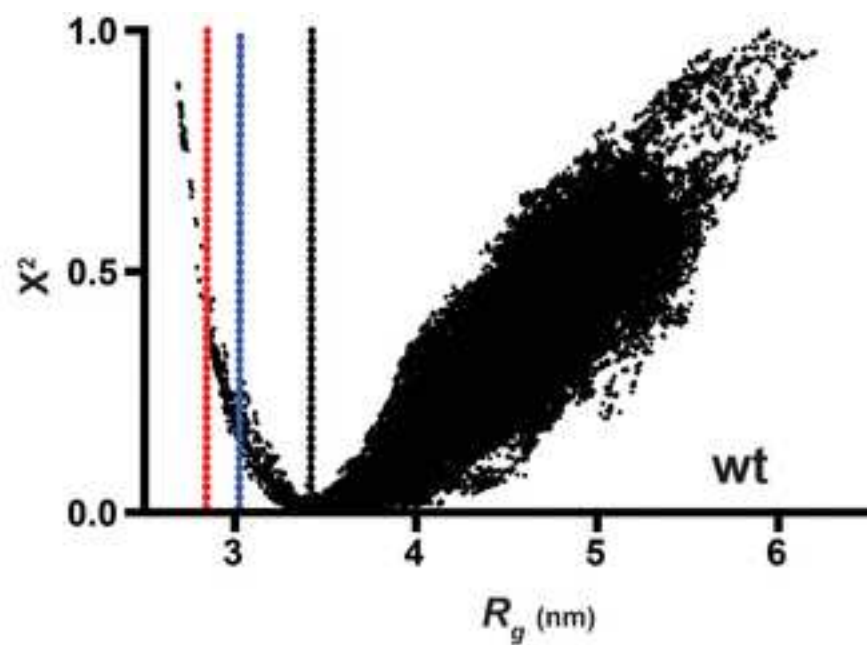


Figure 3
[Click here to download high resolution image](#)

Best Models

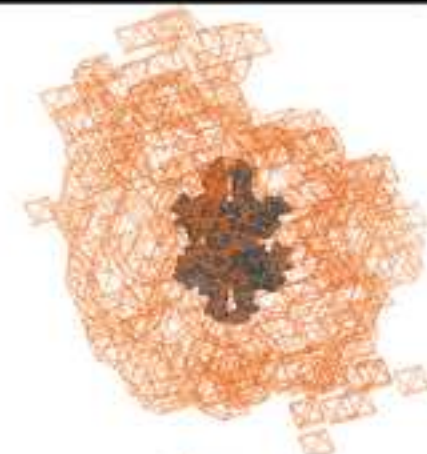
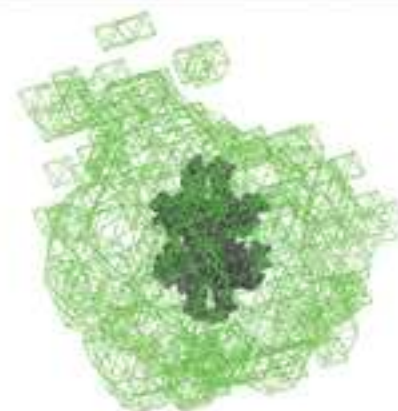
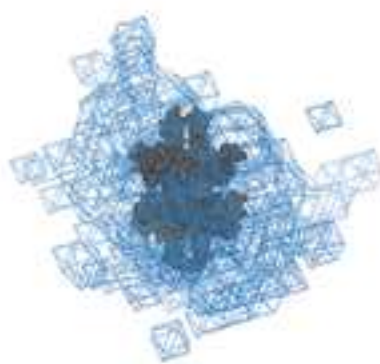
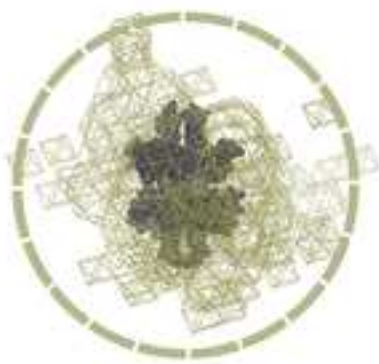
50

100

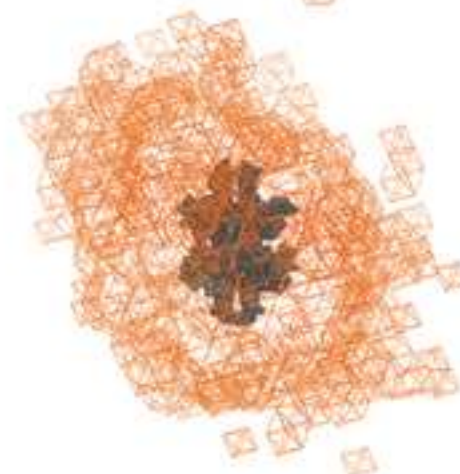
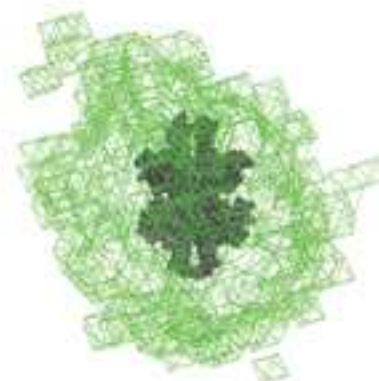
1000

2000

apo



holo
dT70



holo
dT35

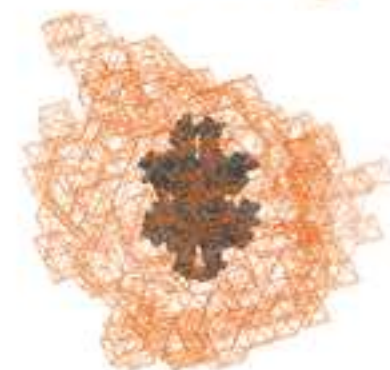
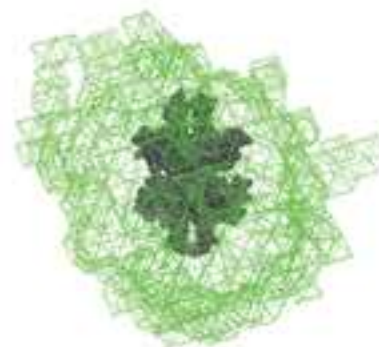
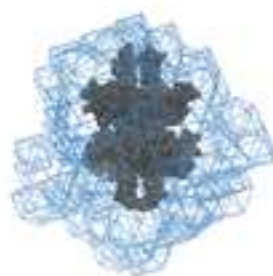


Figure 4
[Click here to download high resolution image](#)

