

# STRUCTURE OF THE N-TERMINAL OLIGOMERIZATION DOMAIN OF DnaD REVEALS A UNIQUE TETRAMERIZATION MOTIF AND PROVIDES INSIGHTS INTO SCAFFOLD FORMATION

Schneider, S.<sup>1</sup>, Zhang, W.<sup>1</sup>, Soultanas, P.<sup>2</sup> & Paoli, M.<sup>2</sup>

From the Centre for Biomolecular Sciences, University of Nottingham, University Park,  
Nottingham NG7 2RD, UK

Running title: The oligomerization domain of DnaD

<sup>1</sup>Joint first authors

<sup>2</sup>To whom correspondence should be addressed: Max.Paoli@nottingham.ac.uk, tel. +44-115-8468015 and  
Panos. Soultanas@nottingham.ac.uk, tel. +44-115-9513525

DnaD is a primosomal protein that remodels supercoiled plasmids. It binds to supercoiled forms and converts them to open forms without nicking. During this remodelling process all the writhe is converted to twist and the plasmids are held around the periphery of large scaffolds made up of DnaD molecules. This DNA-remodelling function is the sum of a scaffold-forming activity on the N-terminal domain (Nd) and a DNA-dependent oligomerization activity on the C-terminal domain (Cd). We have determined the crystal structure of the scaffold-forming Nd which reveals a winged-helix (WH) architecture, with additional structural elements extending from both N- and C-termini. Four monomers form dimers that join into a tetramer. The N-terminal extension mediates dimerization and tetramerization, with extensive interactions and distinct interfaces. The wings and helices of the WH domains remain exposed on the surface of the tetramer. Structure-guided mutagenesis and atomic force microscopy (AFM) imaging indicate that these elements together with the C-terminal extension are involved in scaffold formation. Based upon our data we propose a model for the DnaD-mediated scaffold formation.

## INTRODUCTION

Initiation of bacterial DNA replication is orchestrated by the master bacterial replicator DnaA and the staged assembly of a primosomal complex that remodels the DNA at the replication origin *oriC*, to load two replicative hexameric ring DNA helicases on the exposed single strands (1-4). Ring helicases are tethered around the single stranded DNA, translocate the replication forks in opposite directions along the DNA and unwind the double stranded DNA in the process, fuelled by chemical energy released from ATP binding/hydrolysis (5). Occasionally replication forks may be arrested along the DNA, leading to replisome collapse and re-initiation at random sites distant to *oriC*. PriA coordinates the restart primosome and re-assembly of the replisome at such sites (6). The DnaA and PriA proteins are highly conserved in all bacteria but the rest of the primosomal proteins appear to be different, reflecting subtle mechanistic differences in primosome assembly and function in different species. For example, whilst the Gram positive *Bacillus subtilis* and Gram negative *Escherichia coli* both have a helicase loader (DnaI and DnaC, respectively), no homologues of the *E. coli* PriB, PriC and DnaT primosomal proteins are found in *B. subtilis* (2). Instead, the DnaD and DnaB proteins, with no homologues in *E. coli*, have been implicated in primosomal assembly in *B. subtilis* (7,8)

DnaD is an essential protein (9,10) that interacts with the master replication initiator protein DnaA (11) and the primosome-restart helicase PriA (12). It has been implicated in the early steps of the primosomal cascade and is recruited at a membrane attachment point by DnaB (13,14) to complete primosomal assembly and, together with DnaI, the loading of the replicative helicase DnaC (12,15-18). A genetic link between the PriA, DnaD and DnaB proteins reinforces the notion of a coordinated cascade for the assembly of the restart primosome (19). In addition to its specialized role at *oriC*, DnaD has been implicated in DNA repair (10), recombination (8), initiation of plasmid replication (9), and initiation of sporulation in *B. subtilis* (20). It also

exhibits a novel global remodelling activity on supercoiled DNA; it opens up supercoiled plasmids and untwists the duplex thus converting all the superhelical writhe into twist (21,22). DnaD molecules form a large macromolecular ‘scaffold’, around the periphery of which the DNA is forced and held (22,23). This unique DNA-remodelling function is the sum of oligomerization and DNA binding activities of DnaD’s two distinct domains: an N-terminal domain (Nd 16,056 Da) that mediates oligomerization and scaffold formation, and a C-terminal domain (Cd 13,730 Da) that binds DNA and exhibits a second DNA-induced oligomerization activity (24). The two domains must be covalently linked into the same polypeptide chain in order to obtain the full wild-type DNA-remodelling effects. Interestingly, DnaB, which acts in the same primosomal cascade, exhibits a lateral DNA-compaction activity (22) and its molecular architecture resembles that of the *E. coli*  $\gamma$ -complex involved in the loading of the ring-shaped processivity factor onto the DNA (16). The significance of such activities for the initiation of DNA replication is not clear at present, but in addition to its helicase co-loading activity with DnaI (15,16,18), DnaB has been proposed to act as a ‘sensor’ bridging large segments of the nucleoid at its membrane attachment site while at the same time sensing for the *oriC* (22). Once the *oriC* has been located onto the membrane attachment site, DnaD is recruited to remodel locally the nucleoid and assist the interaction of DnaA with the *oriC*. Other architectural proteins are also known to facilitate remodelling of the *oriC* by DnaA *in vitro* (25-28) and proteins such as HU $_{\beta}$  and H-NS suppress (29) or enhance (30) the thermosensitive phenotype of *dnaA46ts*, providing a genetic link between DNA remodelling proteins and initiation of DNA replication.

In an effort to understand the mechanism of the DnaD-mediated DNA-remodelling, we have solved the crystal structure of the scaffold-forming Nd domain and carried out structure-guided mutagenesis and AFM imaging to gain insights into the molecular details that underpin the formation of the DnaD scaffolds. Based upon our data, a structural model for scaffold formation is proposed and discussed.

## RESULTS AND DISCUSSION

### The crystal structure of DnaD Nd reveals an extended winged-helix fold.

The structure of the oligomerization domain of DnaD was refined against data to 2.0 Å spacing (see Supplemental Fig. S1 for an example of representative electron density), with a final R-factor of 18.1% and an R-free of 21.0% (see crystallographic and refinement statistics in Table 1). The overall structure reveals a winged-helix (WH) architecture with polypeptide extensions at both the N- and C-terminal ends of the classical WH fold (Fig. 1). DnaD Nd has no significant sequence homology with any other protein structurally characterised to date: sequence identity values in the range of 6-23% are obtained from structure-based sequence alignments with other WH proteins with which structural similarity was detected using programs such as Dali (31) and TOPS (32) (Supplemental Table SI). Despite their unrelated sequences, the structures of DnaD Nd and WH proteins like the BlaI repressor (33), ADAR1 (34), genesis (33) and the ESCRT II domains (36,37) are superimposable with r.m.s. deviations in the 1.7-3.5 Å range, for selected C $\alpha$  atoms (Supplemental Table SI).

Two molecules of DnaD Nd are present in the asymmetric unit of the lattice (Fig. 1B-C). It is known from biochemical studies (24) that the domain is tetrameric in solution; a tetramer arrangement is observed (Fig. 1D) when crystallographic symmetry is applied. The dimer in Figure 1B shows P2 symmetry about a central axis perpendicular to the plane of the page. Despite the overall symmetry in the arrangement of the monomers, detailed contacts in the dimer interface do not obey perfect symmetry and some equivalent side chains have non-identical conformations. Dimers arrange themselves into the tetramer (Fig. 1D) also by P2 symmetry, about a vertical axis, again perpendicular to the plane of the page but orthogonal to the previously mentioned dimer axis (Fig. 2A).

The most striking feature of the WH fold in DnaD Nd is the presence of extensions at both termini (Fig. 1), which account for more than a third of the whole structure. The N-terminal extension includes a helix-strand-helix (H1'-S1'-H2') structural element which appears to be unique to DnaD; on the other hand, the smaller C-terminal extension, made up of an alpha-helix (H3'), is analogous to a similar extension found in the BlaI repressor (Supplemental Fig. S2). Inspection of the structure shows that both extensions, the N-terminal one in particular, are involved in the dimerization and tetramerization contacts.

### Pairing of monomers and the dimer interface.

The single  $\beta$ -strand present in the 24 amino acid-long H1'-S1'-H2' motif forms a small  $\beta$ -sheet when two monomers pair up with each other with P2 symmetry, as mentioned before (Fig. 1A-B and Fig. 2A). Therefore, it appears that the resulting dimer is stabilised by the hydrogen bonding between the strands, in addition to a wealth of further contacts as outlined below. The dimerization interface is distinguished from the tetramerization interface in terms of surface area and interactions. The dimer contacts, between chains A and B (and chains A' and B', see Fig. 1B-D), cover an area of about 1524 Å<sup>2</sup>, larger than the tetramerization interface which is 1262 Å<sup>2</sup> in total and discontinuous, being made up of B-A', and A'-A' interactions. The dimer contacts include 5 strong (less than 3.0 Å) hydrogen bonds (Table 2) and several buried side chains (more than 90% of their area becomes solvent inaccessible; see Table 2). Amongst these buried groups, Ile17, Leu22, Leu35 and Ile36 are notable because they cluster into an extensive interaction hot spot of interlocking residues. Figure 2B shows an overall view of the interaction surface with the more important side chains highlighted.

While both interfaces are significantly hydrated, water-mediated interactions occur almost exclusively at the periphery of the dimer interface, as opposed to the tetramer in which hydration is more widespread across the contact area (Fig. 3A). Accordingly, only 5 buried (or relatively buried, with less than 5 Å<sup>2</sup> accessible to the solvent) waters are observed in the dimer interface, while the tetramer interface has as many as 12 inaccessible water molecules. Thus, in terms of both the entropic cost of this hydration and the types of interactions observed, the dimer contacts result in a tighter association than the tetramerization interface, consistent with the native DnaD protein existing predominantly as a dimer in solution, with tetramers and higher oligomers forming at higher concentration as the scaffold is assembled (22,23).

### **The tetramer interface and the tetramerization motif.**

Interactions that govern the formation of the tetramer appear generally weaker and less numerous than those involved in the dimer interface (Table 3). Only one strong electrostatic contact is present and the shape of the interacting surfaces is such that there is no extensive hot spot as defined by burial of solvent accessible area. The only remarkable feature is the presence of a highly symmetrical cluster of threonine residues at the centre of the tetrameric assembly. These threonines (Thr14 and Thr16) are provided by the  $\beta$ -strands of the N-terminal extensions which pair up in the dimers; the face of each resulting 2-stranded  $\beta$ -sheet then presents a threonine tetrad to another dimer so that the two sets come together into a tight core of 8 threonine side chains, as shown in Figure 3B. The interlocking is very effective because it is stabilised both by non-polar hydrophobic contacts, contributed by the methyl groups, and by hydrogen-bonding interactions, from the hydroxyl functions of the threonine residues (Fig. 3B). This arrangement appears to be a unique feature of DnaD Nd, as is not observed in any other oligomeric protein with known structure. However, the tetramerization motif does not seem to be conserved across the sequences of homologous DnaD Nd proteins (Supplemental Fig. S2), although the preceding glycine (Gly13) and successive proline (Pro18) clearly show some degree of conservation. Even in the absence of strict conservation it is possible that different residues pack together at the core of the tetramer; in particular, the space occupied by two water molecules, within each monomer and adjacent to the  $\beta$ -strands, mean that some flexibility may allow the strands to rearrange their position and accommodate a range of side chain groups in the tetramer core. In addition to the striking tetramerization motif, another region likely to provide a significant contribution to the interface is the cross-over between helix H1' (residues 3-10) from one dimer and H3' (residues 103-111) from another dimer, where intervening non-polar side chains yield a series of zip-like hydrophobic contacts. Notably, all groups involved in these interactions belong to the N- and C-terminal extensions which appear to be a characteristic of the WH fold of DnaD Nd.

### **The role of the wing-hairpin.**

Since DnaD Nd is involved predominantly in scaffold formation rather than DNA-binding (24), its extended WH domain participates in protein-protein interactions that underpin DnaD-mediated scaffold formation. In addition to the N- and C-terminal extensions, another striking feature of the WH fold in DnaD Nd is the wing-hairpin structure, which is pronounced and relatively long compared to a large set of non-homologous (less than 20% sequence identity) WH proteins (PDB codes: 1Q1H, 1BJA, 1ILG, 1LEA, 1ZAR, 2A61, 2FXA, 1FT9, 1G6N, 1DPR, 1SAX, 1SEJ, 1BIA) (Supplemental Fig. S2), as indicated by structural

superpositions. Given the unique character of this structural feature, we speculate that it might be involved in protein-protein interactions key for scaffold formation. The role of the wing was examined through a protein engineering experiment in which the wing-hairpin was truncated to form a much smaller structure. The EDQNGIK segment was replaced with the GS dipeptide sequence to produce the Nd<sub>GS</sub> and DnaD<sub>GS</sub> proteins (see Supplemental Fig. S3). With this protein engineering experiment we removed the extended structure to generate a new hairpin where two short strands are still connected by a tight, type II  $\beta$ -turn. In our turn design, the sequence GS was chosen considering the conformation of the polypeptide chain and the propensities of amino acids to form and stabilise turn structures (see the section “Plasmid constructs and mutagenesis” in Experimental Procedures). The hairpin truncated mutant proteins were expressed at high levels in the soluble fraction, purified to homogeneity and confirmed to associate into tetramers for Nd, and dimers for DnaD in solution, by size-exclusion chromatography (data not shown). Circular Dichroism (CD) analysis verified that this hairpin truncation did not induce gross structural changes either in the native DnaD or the Nd domain (Supplemental Fig. S4).

AFM was used to compare the scaffolds formed by Nd and Nd<sub>GS</sub>, which revealed striking differences (Fig. 4). At 6 nM, Nd<sub>GS</sub> formed irregular aggregates that were small and elongated. Even at 30 nM abnormal aggregates were apparent, growing elongated or very tall out of the mica surface. The Nd<sub>GS</sub> aggregates are, therefore, distinctly different from the characteristic, scaffolds formed by Nd, indicating that Nd<sub>GS</sub> exhibits a defect in scaffold formation.

### **The role of the H3 helix.**

Given the strategic positioning of the H3 helix, which is prominently exposed at four corners of the tetrameric assembly (Fig. 1D), we envisaged that it might be involved in scaffold formation. Although the H3 helix is used in WH proteins to make contacts to the DNA in the major groove (38) it also appears to be able to engage in protein-protein interactions as seen in the ESCRT-II complex (36,37). Examination of the three-dimensional coordinates together with solvent accessibility calculations indicated that two residues on the H3 helix, Asn70 and Arg73, are very exposed to the solvent and project outwards from the protein surface. It is tempting to postulate that these residues might be mediating scaffold formation. The role of the H3 helix was, therefore, investigated by mutating the N70 and R73 residues to alanines. AFM imaging of the resulting N70A and R73A mutant Nd proteins revealed abnormal aggregates (Fig. 4). Compared to the native Nd, irregular aggregates and smaller abnormal scaffolds were apparent at 6 and 30 nM, indicating that both mutations affect the ability of native Nd to form scaffolds. These data suggest that the wings and the H3 helices in the tetramer are involved in inter-tetramer interactions that are important in scaffold formation.

### **The role of the C-terminal extension.**

Given that the structure shows that the C-terminal (H3') helix is in a prominent, significantly exposed position, we devised two experiments to test its role in scaffold formation. With the first approach, residue Leu112, which points away from the molecule, was mutated to alanines in both the native DnaD and in the Nd domain. In the second approach, three residues (Gln, Ser, His) were added to the N-terminus of the protein which is placed orthogonally, on the surface of the tetrameric assembly and proximal to H3', so that any lengthening of the N-terminal end would project out of the surface of the tetramer and sterically hinder other tetramers interacting with H3'. The GSH tripeptide was added to the native DnaD and used to engineer the GSH-N70A, GSH-R73A and GSH-DnaD<sub>GS</sub> mutant proteins. All GSH-tagged proteins were expressed in the soluble fraction purified to homogeneity and confirmed to associate into tetramers for Nd and dimers for DnaD in solution by size-exclusion chromatography (data not shown). CD analysis also verified that the GSH addition did not induce gross structural changes in the native DnaD or the Nd domain (Supplemental Fig. S4).

The GSH-DnaD proteins were examined by AFM, in the presence of pBSK, and found to form aggregates that were unable to open up supercoiled plasmids, indicating no regular scaffold formation (data not shown). The inability of the GSH-DnaD proteins to form scaffolds was verified further by their inability to shift intact supercoiled plasmids in agarose gels (Fig. 5). Such shifts are dependent on the formation of Nd-mediated scaffolds (22,23,24). By comparison the DnaD Cd is also unable to form scaffolds and fails to shift supercoiled plasmids, despite the fact that it binds to DNA (21,24). The GSH-DnaD proteins appeared to shift only the open

circular R-form of the plasmid, and upon treatment of the samples with proteinase K the R-plasmid was released (compare lanes 2-5 and 7-10 in Fig. 5). Therefore as expected, the GSH-DnaD proteins bind to DNA but are unable to form scaffolds.

In order to verify further that the GSH insertion at the N-terminus affects scaffold formation, we constructed the GSH-Nd protein and examined scaffold formation by AFM. Small aggregates were observed at 6 nM that became larger at 30 nM (Fig. 6A). These aggregates were very different than the characteristic scaffolds observed with the native Nd (compare Fig. 5 and Fig. 6A). By comparison, Nd-L112A even at 30 nM also formed abnormal scaffolds compared to native Nd (compare Fig. 4 and Fig. 6B), while DnaD-L112A retained its ability to bind to DNA (Fig. 7). However, the scaffolds formed by Nd-L112A are more regular compared to other mutant scaffolds and appear to grow high up from the surface of the mica (see enlarged images in Fig. 6B). In that respect they are also different than the wtDnaD scaffolds. This difference could be the result of subtle changes in tetramer-tetramer association.

The combined data show that the GSH insertion and the L112A mutation cause defects in scaffold formation without affecting binding to DNA. This is consistent with previous studies showing that the major DNA-binding activity resides on Cd and binding to DNA results in Cd-oligomerization and duplex untwisting (21,24). The Nd domain mediates scaffold formation that appears to enhance duplex untwisting by Cd (21). It forms tetramers in solution (24) and the crystal structure suggests that the H1'-S1'-H2' insertion at the N-terminus is intimately involved in dimerization and tetramerization. Indeed, deletion of the H1'-S1'-H2' insertion results in insoluble protein which is likely to be the result of dimer and tetramer destabilization and exposure of the surface hydrophobic patches involved in dimerization and tetramerization (Supplemental Fig. S5). In conclusion, these experiments suggest that the H3' helices in the tetramer are involved in inter-tetramer interactions that are important in scaffold formation.

### ***The effects of mutations in the H3, H3' helices and the wing on the DNA-remodelling function of native DnaD.***

While scaffolds formed by the Nd domain appear similar to those formed by the native DnaD (23,24), the DNA-binding activity resides on the Cd domain (24), and the two domains are likely to co-operate for DNA-remodelling. In fact the formation of scaffolds opens up supercoiled plasmids and is manifested by shifts of the plasmids in agarose gels (24). Also binding of Cd to DNA is sufficient to induce helix untwisting but the native DnaD appears to be more efficient at untwisting the DNA helix than Cd (21). Therefore, the presence of the scaffold appears to enhance the DNA untwisting activity of Cd. We argued that defective scaffold formation may be manifested indirectly by effects on the shifts of supercoiled plasmids in agarose gels which will be different than that caused by native DnaD. The N70A, R73A (H3 helix), the L112A (H3' helix) and the DnaD<sub>GS</sub> (wing truncation) mutants in the native DnaD were used to study the effects on gel shift assays.

Agarose shift assays revealed that all mutants formed nucleoprotein complexes that shifted higher up the gel but subtle differences were apparent (Fig. 7). Higher concentrations of the N70A and DnaD<sub>GS</sub> proteins were required to produce complete shifts compared to native DnaD (Compare lanes 1-3, 4-6 and 13-15 in Fig. 7). Also the DnaD<sub>GS</sub> shifts appeared more diffuse relative to the rest. The R73A mutant did not produce intermediate shifts exhibiting very big nucleoprotein complexes that stayed in the well (lanes 7-9), while the L112A mutant produced slightly bigger nucleoprotein complexes at 105  $\mu$ M compared to native DnaD (compare lanes 3 and 12). These data are consistent with irregular scaffold formation and intact DNA-binding activities, hence resulting in subtle differences in the positions of the shifted nucleoprotein species.

### **A model for scaffold formation.**

Canonical WH proteins are usually versatile DNA-binding proteins utilizing contacts between the so-called recognition helix H3, the wing and the DNA. The H3 and the wing interact either with the major or minor grooves of DNA, depending on whether these elements are located on the most basic electrostatic surface of a protein (38). However, some WH proteins do not bind DNA and their H3 and wing elements are involved in protein-protein interactions instead (37,38). Although DnaD is a non-specific DNA-binding protein, the main function of Nd is to mediate scaffold formation rather than to bind DNA. Instead a major DNA-binding activity resides on a separate Cd (24). The Nd forms a tetramer, with three key secondary structural elements exposed on

the surface: H3 recognition helices, wings and C-terminal H3' extensions. The defects observed in scaffold formation by the DnaD<sub>GS</sub>, N70A and R73A mutants suggest that the exposed H3 helices and wings participate in the formation of scaffolds. In addition, both the L112A mutation and the GSH addition at the N-terminus affected scaffold formation. The N-terminal H1'-S1'-H2' insertion participates in dimerization and tetramerization at the core of the structure and is unlikely to be involved in scaffold formation directly. However, the GSH insertion at the N-terminus is expected to project out of the structure in the vicinity of the orthogonal C-terminal H3' helix, thus sterically preventing the H3' helix from participating in inter-tetramer interactions.

Given that Nd forms concentration-dependent scaffolds in a manner similar to the native DnaD protein, we propose a possible structural model for scaffold formation (Fig. 8A) based on the arrangement of the molecules as seen in the crystal lattice. It is reasonable to assume that some of the interactions which govern scaffold formation are also present in the lattice. Indeed, contacts between molecules across the crystal are made almost exclusively by groups from the H3 helix and the wings (Fig. 8B). The H3' terminal extension is not involved in crystal contacts. Presumably this can be explained because in the lattice a particular set of scaffold interactions dominate, thus giving the crystal its order and directionality of molecules which is lacking from the seemingly irregular scaffolds observed by AFM. Also, since the wing is associated with high flexibility (Supplemental Fig. S6), it is likely that a particular or more common conformation of the  $\beta$ -hairpin is selected by the crystallization conditions, whilst different wing conformations may be present in molecules across the scaffold affecting the direction and orientation of the molecular packing. Regardless of possible differences in the exact orientation of molecules in the lattice versus the scaffold, it is striking to see that both Asn70 and Arg73 of H3 are involved in a number of intermolecular crystal contacts. Interestingly, in the Nd lattice there is sufficient space between Nd tetramers to accommodate four Cds, which would be required in the scaffold of the native DnaD. It is, therefore, attractive to propose that the lattice packing of Nd molecules provides a valid snapshot of some of the key interactions that are likely to govern the association of molecules into the unique structural scaffolds which underlie the DNA-remodelling function of DnaD.

## EXPERIMENTAL PROCEDURES

### Plasmid constructs and mutagenesis.

The native DnaD and Nd proteins were produced from the pET22b-DnaD (23) and pET22b-Nd (24) constructs, respectively. The plasmid construct for the GSH-DnaD and GSH-Nd proteins were engineered by cloning the NdeI-HindIII or the NdeI-XhoI fragments, containing the *dnaD* and Nd genes, respectively, into the same sites of pET28a to produce the pET28a-DnaD and pET28a-Nd constructs, respectively. The pET28a-DnaD construct produced a protein with a removable N-terminal His<sub>6</sub> tag while the pET28a-Nd construct produced a protein with removable N-terminal and non-removable C-terminal His<sub>6</sub> tags. The removable tag was cleaved by thrombin (Novagen) cleavage, according to the manufacturer's instructions. After thrombin treatment a GSH tag remained attached at the N-terminus producing the GSH-DnaD or GSH-Nd proteins.

The N70A and R73A mutants were constructed in the native DnaD, GSH-DnaD and Nd proteins, while the L112A mutant in the native DnaD and Nd proteins. All single point mutants were engineered using the pET22b-DnaD (23), pET22b-Nd (24) or pET28a-DnaD constructs, as appropriate and the QuickChange II-E-Site directed mutagenesis kit (Stratagene), according to the manufacturer's instructions.

The wing structure typical of the WH fold has a type II  $\beta$ -hairpin as assigned using the program PROMOTIF (39). In order to shorten the wing, residues 87-93 were replaced with a Gly-Ser dipeptide sequence (Supplemental Fig. S3) commonly used in type II  $\beta$ -hairpin structures (40) and which are also generally accommodated amino acids substitutions in type II beta-hairpins (41). The new, shorter designed  $\beta$ -hairpin was modeled in COOT (42) and analyzed with PROMOTIF (39). Supplemental Figure S1 shows the predicted model of the shortened hairpin in comparison with the wild-type. The wing truncations were constructed in the native DnaD (DnaD<sub>GS</sub>), GSH-DnaD (GSH-DnaD<sub>GS</sub>) and Nd (GSH-Nd<sub>GS</sub>) proteins using a two-step PCR splice by overlap mutagenesis procedure (43).

The constructs coding for the N-terminal H1'-S1'-H2' truncated DnaD and Nd proteins were engineered using suitable primers (Supplemental Table S2) and PCR with the pET22b-DnaD and pET22b-Nd substrates, respectively, followed by cloning into the NdeI-HindIII and NdeI-XhoI sites of pET22b, respectively.

All the oligonucleotides and proteins used in this study are shown in the Supplemental Tables S2 and S3, respectively.

### **Protein purifications.**

Native DnaD and its mutants (N70A, R73A, L112A and DnaD<sub>GS</sub>) were purified as described before (23,24). The only difference is that DnaD<sub>GS</sub> was over-expressed at 20°C.

For the purifications of the GSH-DnaD and its mutants (N70A, R73A and DnaD<sub>GS</sub>) over-expressions were carried out at 37°C (N70A, R73A) and 20°C (DnaD<sub>GS</sub>). Cells were harvested in buffer A (50 mM sodium phosphate, pH 7.5, 500 mM NaCl, 20 mM imidazole) and 10% w/v sucrose, sonicated and then incubated for 30 min at 37°C with benzonase (Novagen). The cell extract was clarified by centrifugation at 17,500 rpm for 30 min. Total protein was precipitated with ammonium sulfate (2.9 g per 10 ml) and centrifuged at 17,500 rpm, 30 min. The protein pellet was re-suspended in buffer A and loaded onto His-trap-Ni<sup>2+</sup> chelating column pre-equilibrated with buffer A. The protein was eluted with a 20–500 mM imidazole gradient in buffer A. Relevant fractions were pooled and N-terminal His-tag was removed by thrombin (Novagen) treatment at 20°C, overnight. The protein solution was then re-loaded onto His-trap-Ni<sup>2+</sup> chelating column pre-equilibrated with buffer A. The flow through of the His-trap column containing the protein was collected, precipitated with ammonium sulfate and centrifuged at 17,500 rpm, 30 min. The protein pellet was re-suspended in TED (50 mM Tris pH7.5, 2 mM EDTA, 1 mM DTT) with 600 mM NaCl and loaded onto a superdex S-75 gel filtration column pre-equilibrated with TED, 350 mM NaCl. Final samples were made up to 10% v/v glycerol before snap-freezing in liquid nitrogen for storage at -80°C.

The Nd and GSH-Nd proteins and their mutants were purified as described elsewhere (24), the only difference being that the over-expressions of Nd<sub>GS</sub> and GSH-Nd<sub>GS</sub> were carried out at 20°C and for the GSH-Nd proteins treatment with thrombin to remove the N-terminal His-tag preceded the final gel filtration step.

The purity of all protein preparations was checked by SDS-PAGE analysis and their correct oligomerisation states were verified by comparative analytical gel filtration (data not shown).

### **Crystallization and preparation of protein for phase determination.**

Crystallization of the native DnaD Nd was carried out as described before (44). For obtaining phase information, samples of the DnaD Nd were prepared in the presence of selenomethionine (SeMet) by transforming the expression plasmid in methionine auxotroph *E. coli* B834 (DE3) cells. An over-night culture (2 mL) of B834 (DE3) cells, grown at 310 K in Luria-Bertani medium supplemented with 50 µg mL<sup>-1</sup> ampicillin, was transferred to 1 L of LeMaster medium (45) with 50 mg L<sup>-1</sup> SeMet (Calbiochem) and 50 µg mL<sup>-1</sup> carbenicillin (Melford) in a baffled 2 L flask. Cells were grown at 310 K and shaken at 200 rev min<sup>-1</sup> until an OD<sub>A<sub>600nm</sub></sub> of 0.7 was reached. Protein expression was then induced by adding isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 mM and cultures were further grown over-night. Purification and crystallization of Se-Met Nd were carried out as described for the native protein but at the lower protein concentration of 10 mg mL<sup>-1</sup> in contrast to 20 mg mL<sup>-1</sup> for the unlabelled protein (44).

### **Data collection and phasing.**

Phase information was obtained by means of the multi-wavelength anomalous dispersion (MAD) method using a highly redundant two-wavelength SeMet-MAD dataset collected from a single SeMet derivative crystal at the ID23-1 beamline of the European Synchrotron Radiation Source (ESRF), Grenoble, France. The selenium absorption edge was determined by an X-ray fluorescence scan and the peak (0.9798 Å) and the remote (0.976 Å) data sets were collected with 1° oscillation, an exposure of 1 second with a 98 % attenuated beam. The SeMet derivative crystals diffracted to 2.2 Å (peak) and 2.1 Å (remote) spacing. Data on the native DnaD Nd crystals were collected in house as previously described (44). All data sets were processed with the programs MOSFLM (46) and SCALA (47) of the CCP4 suite (48).

Heavy atom location and phase calculation on the SeMet-MAD data set was carried out with SHELXD (49) and SHARP/autoSHARP (50,51), using all 10 possible SeMet sites of the two molecules in the asymmetric unit cell and showing excellent phasing statistics and a high quality initial map.

## Model building and refinement of the atomic coordinates.

Manual building of the model was carried out using COOT (42) and refined against the native data using TLS refinement (52) in REFMAC5 (53). The optimal number of TLS groups was determined using the TLSMD server (54). The model consists of chains A and B, with 1-115 and 1-117 residues respectively. The residues 116-128/118-128 (chains A/B) are not included in the model, because only weak, broken peaks of electron density were obtained for this regions. The sequence of C-terminal tail is mainly made up of almost exclusively small, polar and charged amino acids, consistent with this region forming a random coil, disordered structure connecting the N- and C-terminal domains. Also, relatively weak density was obtained for the loop region 88-92 of chain A, which is consistent with the high flexibility of this region (see B-factor plot in Supplemental Fig. S6). In chain B this loop region is engaged in crystal contacts and therefore less flexible. Data processing and model refinement statistics are summarized in Table 1. The model was evaluated using the MOLPROBITY server (55,56) and shows excellent geometry and no Ramachandran outliers. Surface area calculations were carried out with the program AREAIMOL (48) and interactions between the molecules were analysed with LIGPLOT (57) and visually examined in COOT. All structural figures were prepared using PyMol (58).

## Electrophoretic Mobility Shift Assays (EMSA).

Agarose EMSA was carried out as described before (22), using 4.2 nM supercoiled pBSK and various concentrations (15, 45 and 105  $\mu$ M) of proteins, as appropriate.

## Atomic Force Microscopy (AFM).

Nd-DnaD and its mutants (N70A, R73A, L112A, Nd<sub>GS</sub> and GSH-Nd) were diluted to 6-30 nM in TED, 100 mM NaCl. For imaging in air, 10  $\mu$ l samples were incubated with freshly cleaved mica (Agar Scientific) for 1 minute, rinsed gently with 300  $\mu$ l of dH<sub>2</sub>O and allowed to dry under a gentle flow of nitrogen gas. All AFM imaging was carried out with a Digital Instruments Nanoscope IIIa Mutlimode AFM with a type E scanner (Veeco, Santa Barbara, CA) in tapping mode. The cantilevers used were silicon-tapping probes with a spring constant of 34.4–74.2 N/m (Olympus, OMCL-AC160TS). The tapping set point was adjusted to minimise probe-sample interactions. Images were recorded in both topography and phase modes with a pixel size of 512×512, flattened and analysed with the Nanoscope software.

## ACKNOWLEDGEMENTS

We are very grateful to A McCarthy at the ID23-1 beamline (ESRF, Grenoble) for the help with the MAD experiment, to C Roberts and S Allen for use of the AFM, to G Balkwill for carrying out the CD analysis and to G Chaconas for critical reading and comments on the manuscript. This work was supported by a BBSRC grant to PS and MP (BB/E006450/1) and by the University of Nottingham. The atomic coordinates (PDB code XXXX) have been deposited in the Protein Data Bank, European Bioinformatics Institute (EBI) - Macromolecular Structure Database, Cambridge, United Kingdom (<http://www.ebi.ac.uk/msd/>).

## REFERENCES

1. Kaguni, J. M. (2006) DnaA: controlling the initiation of bacterial DNA replication and more. *Annu Rev Microbiol* **60**, 351-375
2. Lemon, K. P., Moriya, S., Ogasawara, N., and Grossman, A. D. (2002) *Bacillus subtilis and Its Closest Relatives*; Chromosome replication and segregation. 73-86. Editors Sonenshein, A.L., Hoch, J.A. and Losick, R, ASM Press, Washington D.C.
3. Marians, K. J. (1992) Prokaryotic DNA replication. *Annu Rev Biochem* **61**, 673-719
4. Paulsson, J., and Chatteraj, D. K. Origin inactivation in bacterial DNA replication control. (2006) *Mol Microbiol* **61**, 9-15



5. Patel, S. S., and Picha, K. M. Structure and function of hexameric helicases. (2000) *Annu Rev Biochem* **69**, 651-697
6. Sandler, S. J., and Marians, K. J. (2000) Role of PriA in replication fork reactivation in *Escherichia coli*. *J Bacteriol* **182**, 9-13
7. Bruand, C., Ehrlich, S. D., and Janniere, L. (1995) Primosome assembly site in *Bacillus subtilis*. *EMBO J* **14**, 2642-2650
8. Bruand, C., Farache, M., McGovern, S., Ehrlich, S. D., and Polard, P. (2001) DnaB, DnaD and DnaI proteins are components of the *Bacillus* restart primosome. *Mol Microbiol* **42**, 245-255
9. Bruand, C., Sorokin, A., Serror, P., and Ehrlich, S. D. (1995) Nucleotide sequence of the *Bacillus subtilis* *dnaD* gene. *Microbiology* **141**, 321-322
10. Li, Y., Kurokawa, K., Matsuo, M., Fukuhara, N., Murakami, K., and Sekimizu, K. (2004) Identification of temperature sensitive *dnaD* mutants in *Staphylococcus aureus* that are defective in chromosomal DNA replication. *Mol Genet Genomics* **271**, 447-457
11. Ishigo-Oka, D., Ogasawara, N., and Moriya, S. (2001) DnaD protein of *Bacillus subtilis* interacts with DnaA, the initiator protein of replication. *J Bacteriol* **183**, 2148-2150
12. Marsin, S., McGovern, S., Ehrlich, S. D., Bruand, C., and Polard, P. (2001) Early steps of *Bacillus subtilis* primosome assembly. *J Biol Chem* **276**, 45818-45825
13. Hoshino, T., McKenzie, T., Schmidt, S., Tanaka, T., and Sueoka, N. (1987) Nucleotide sequence of *Bacillus subtilis* *dnaB*: a gene essential for DNA replication initiation and membrane attachment. *Proc Natl Acad Sci U S A* **84**, 653-657
14. Rokop, M. E., Auchtung, J. M., and Grossman, A. D. (2004) Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Mol Microbiol* **52**, 1757-1767
15. Ioannou, C., Schaeffer, P. M., Dixon, N. E., and Soutanas, P. (2006) Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading *Bacillus subtilis*. *Nucleic Acids Res* **34**, 5247-5258
16. Nunez-Ramirez, R., Velten, M., Rivas, G., Polard, P., Carazo, J. M., and Donate, L. E. (2007) Loading a ring: Structure of the *Bacillus subtilis* DnaB protein, a co-loader of the replicative helicase. *J Mol Biol* **367**, 764-769
17. Soutanas, P. (2002) A functional interaction between the putative primosomal protein DnaI and the main replicative DNA helicase DnaB in *Bacillus*. *Nucleic Acids Res* **30**, 966-974
18. Velten, M., McGovern, S., Marsin, S., Ehrlich, S. D., Noirot, P., and Polard, P. (2003) A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Mol Cell* **11**, 1009-1020
19. Bruand, C., Velten, M., McGovern, S., Marsin, S., Serena, C., Ehrlich, S. D., and Polard, P. (2005) Functional interplay between the *Bacillus subtilis* DnaD and DnaB proteins essential for initiation and re-initiation of DNA replication. *Mol Microbiol* **55**, 1138-1150
20. Lemon, K. P., Kurtser, I., Wu, J., and Grossman, A. D. (2000) Control of initiation of sporulation by replication initiation genes in *Bacillus subtilis*. *J Bacteriol* **182**, 2989-2991
21. Zhang, W., Allen, S., Roberts, C. J., and Soutanas, P. (2006) The *Bacillus subtilis* primosomal protein DnaD untwists supercoiled DNA. *J Bacteriol* **188**, 5487-5493
22. Zhang, W., Carneiro, M. J., Turner, I. J., Allen, S., Roberts, C. J., and Soutanas, P. (2005) The *Bacillus subtilis* DnaD and DnaB proteins exhibit different DNA remodelling activities. *J Mol Biol* **351**, 66-75
23. Turner, I. J., Scott, D. J., Allen, S., Roberts, C. J., and Soutanas, P. (2004) The *Bacillus subtilis* DnaD protein: a putative link between DNA remodelling and initiation of DNA replication. *FEBS Lett* **577**, 460-464
24. Carneiro, M. J., Zhang, W., Ioannou, C., Scott, D. J., Allen, S., Roberts, C. J., and Soutanas, P. (2006) The DNA-remodelling activity of DnaD is the sum of oligomerization and DNA-binding activities on separate domains. *Mol Microbiol* **60**, 917-924
25. Dixon, N. E., and Kornberg, A. (1984) Protein HU in the enzymatic replication of the chromosomal origin of *E. coli*. *Proc Natl Acad Sci U S A* **81**, 424-428

26. Hwang, D. S., and Kornberg, A. (1992) Opening of the replication origin of *E. coli* by DnaA protein with HU or IHF. *J Biol Chem* **267**, 23083-23086
27. Krause, M., Ruckert, B., Lurz, R., and Messer, W. (1997) Complexes at the replication of *B. subtilis* with homologous and heterologous DnaA protein. *J Mol Biol* **274**, 365-380
28. Roth, A., Urmoneit, B., and Messer, W. (1999) Functions of histone-like proteins in the initiation of DNA replication at *oriC* of *E. coli*. *Biochimie* **76**, 917-923
29. Bahloul, A., Boubrik, F., and Rouviere-Yaniv, J. (2001) Roles of *E. coli* histone-like protein HU in DNA replication: HU<sub>β</sub> suppresses the thermosensitivity of *dnaA46ts*. *Biochimie* **83**, 219-229
30. Katayama, T., Takata, M., and Sekimizu, K. (1996) The nucleoid protein H-NS facilitates chromosome DNA replication in *E. coli dnaA* mutants. *J Bacteriol* **178**, 5790-5792
31. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, 123-138
32. Michalopoulos, I., Torrance, G. M., Gilbert, D. R., and Westhead, D. R. (2004) TOPS: an enhanced database of protein structural topology. *Nucleic Acids Res* **32**, D251-254
33. Safo, M. K., Zhao, Q., Ko, T. P., Musayev, F. N., Robinson, H., Scarsdale, N., Wang, A. H., and Archer, G. L. (2005) Crystal structures of the BlaI repressor from *Staphylococcus aureus* and its complex with DNA: insights into transcriptional regulation of the *bla* and *mec* operons. *J Bacteriol* **187**, 1833-1844
34. Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A., and Rich, A. (1999) Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841-1845
35. Jin, C., Marsden, I., Chen, X., and Liao, X. (1999) Dynamic DNA contacts observed in the NMR structure of winged helix protein-DNA complex. *J Mol Biol* **289**, 683-690
36. Hierro, A., Sun, J., Rusnak, A. S., Kim, J., Prag, G., Emr, S. D., and Hurley, J. H. (2004) Structure of the ESCRT-II endosomal trafficking complex. *Nature* **431**, 221-225
37. Teo, H., Perisic, O., Gonzalez, B., and Williams, R. L. (2004) ESCRT-II, an endosome-associated complex required for protein sorting: crystal structure and interactions with ESCRT-III and membranes. *Dev Cell* **7**, 559-569
38. Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B., and Burley, S. K. (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916-921
39. Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212-220
40. Sibanda, B. L., and Thornton, J. M. (1993) Accommodating sequence changes in beta-hairpins in proteins. *J Mol Biol* **229**, 428-447
41. Blanco, F., Ramirez-Alvarado, M., and Serrano, L. (1998) Formation and stability of beta-hairpin structures in polypeptides. *Curr Opin Struct Biol* **8**, 107-111
42. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
43. Clackson, T., Gussow, D., and Jones, P. T. (1991) *PCR-A practical approach.*, Editors McPherson, M.J., Quirke, P. & Taylor, G.R., IRL Press, Oxford, UK
44. Schneider, S., Carneiro, M. J. V. M., Ioannou, C., Soultanas, P., and Paoli, M. (2007) Crystallization and X-ray diffraction analysis of the DNA-remodelling protein DnaD from *Bacillus subtilis*. *Acta Crystallographica Section F* **F63**, 110-113
45. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three dimensional structure. *EMBO J* **9**, 1665-1672
46. Leslie, A. G. W. (1992) *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, 26
47. Evans, P. (1997) *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* **33**, 22-24
48. CCP4. (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 760-763
49. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779
50. de La Fortelle, E., and Bricogne, G. (1997) *Methods in Enzymology* **276**, 472-494
51. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2006) *Methods Mol Biol* **364**, 215-230

52. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* **374**, 300-321
53. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255
54. Painter, J., and Merritt, E. A. (2006) *Journal of Applied Crystallography* **39**, 109-111
55. Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, W615-619
56. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by C $\alpha$  geometry and C $\beta$  deviation. *Proteins* **50**, 437-450
57. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**, 127-134
58. DeLano, W. L. (2002) The PyMol molecular graphics system. DeLano Scientific, San Carlos, CA.
59. Ha, S. C., Lokanath, N. K., Van Quyen, D., Wu, C. A., Lowenhaupt, K., Rich, A., Kim, Y. G., and Kim, K. K. (2004) A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Z $\alpha$  bound to DNA. *Proc Natl Acad Sci U S A* **101**, 14367-14372
60. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268
61. Stevens, T., and Paoli, M. (2007) *Proteins: Structure, Function, and Bioinformatics* (**in press**)

	<b>DnaD N-Domain</b>		
	native	Se-Met peak	Se-Met remote
<b>Data collection</b>			
Spacegroup	P 3 <sub>1</sub> 21	P 3 <sub>1</sub> 21	P 3 <sub>1</sub> 21
Unit cell parameters (Å)			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.7, 78.7, 124.6	79.1, 79.1, 124.4	
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	
Wavelength (Å)	1.5418	0.9798	0.976
Resolution range (Å)	28.0-2.0 (2.1-2.0)	30-2.2 (2.3-2.2)	30.0-2.1 (2.2-2.1)
R <sub>sym</sub> (%)	4.8 (37.6)	8.9 (41.8)	6.4 (36.6)
Mean <i>I</i> / $\sigma$ <i>I</i>	24.4 (4.0)	19.0 (5.1)	24.4 (6.0)
No. of observations	177843 (25869)	310519 (50593)	360835 (58215)
No. of unique reflections	30662 (4429)	23412 (3354)	26920 (3875)
Completeness (%)	99.8 (100.0)	100.0 (100.0)	99.9 (100.0)
Anomalous completeness (%)	-	97.4 (100.0)	98.1 (100.0)
Multiplicity	5.8 (5.8)	13.3 (15.1)	13.4 (15.0)
Anomalous multiplicity (%)	-	7.0 (7.8)	7.1 (7.8)
<b>Phasing statistics</b>			
Number of Se sites		10	
Figure-of-merit (centric/acentric)		0.84 / 0.93	
Phasing power		1.53	
<b>Refinement</b>			
Resolution (Å)	28.0-2.0 (2.05-2.0)		
No. of reflections	29103 (2119)		
R <sub>work</sub> /R <sub>free</sub> * (%)	18.1-21.0 (25.3-30.6)		
No. atoms			
Protein (chain A / B)	1944 (958 / 986)		
Waters / ions	231 / 14		
B-factors			
Protein (chain A / B)	37.8 / 34.4		
Waters / Ions	43.1 / 39.1		
R.m.s deviations			
Bond length (Å)	0.012		
Bond angles (°)	1.126		

\* 5 % randomly selected reflections were excluded from the refinement for R<sub>free</sub> calculation.

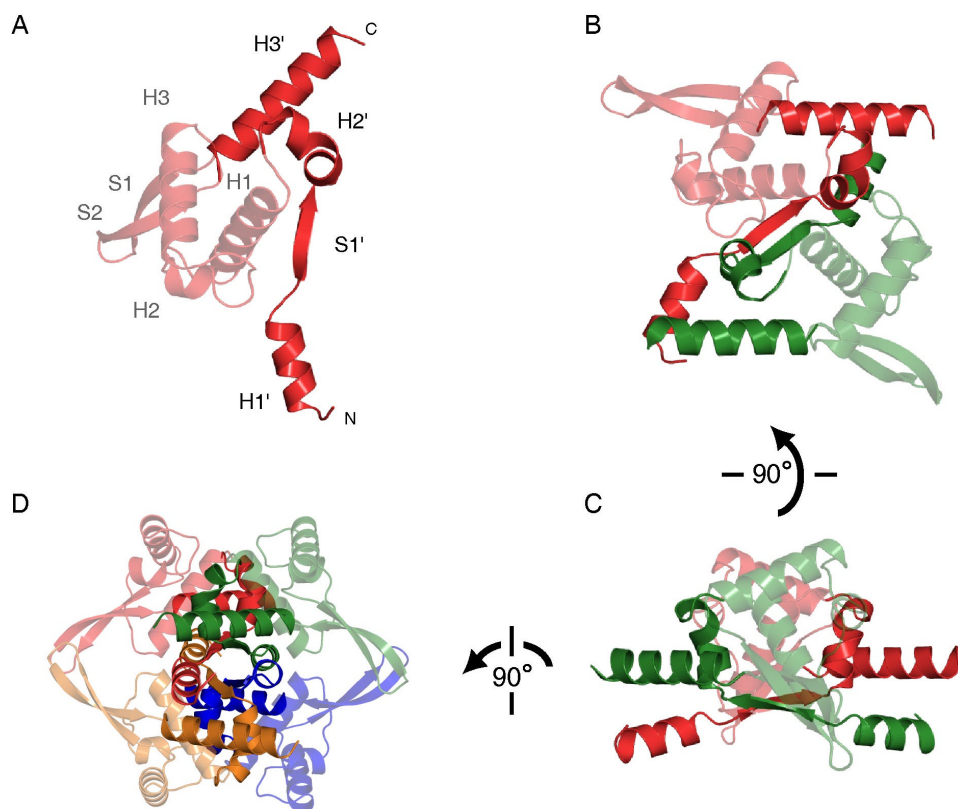
**TABLE 1; Data collection, processing, phasing and refinement statistics.** Numbers in parenthesis are for the high resolution shell.

	Dimer contacts			Tetramer contacts					
Interface area	1524 Å <sup>2</sup>			1026 Å <sup>2</sup>			236 Å <sup>2</sup>		
chain ID	chain A	chain B	dist. (Å)	chain B	chain A'	dist. (Å)	chain A	chain A'	dist. (Å)
polar /electrostatic	M1 (mc)	Q111 (sc)	2.8	K3 (sc)	E108 (sc)	2.9	M1 (mc)*	M1 (mc)*	2.9
	Q12 (mc)*	N19 (sc)*	2.9	Y107 (sc)*	M1 (mc)*	3.1			
	G13 (mc)	N19 (mc)	3.1						
	S15 (mc)*	I17 (mc)*	2.8						
	S15 (mc)*	N19 (sc)*	2.5						
	E32 (sc)*	S62(sc/mc)*	2.6						
	M61 (mc)	K26 (sc)	3.1						
van der Waals	M1 (sc)*	I110 (sc)*	3.7	K3 (sc)*	G104 (mc)*	3.6	M1 (sc)*	M1 (sc)*	3.8
	M9 (sc)*	L20 (sc)*	3.7	K3 (sc)	E108 (sc)	3.8	F6 (sc)*	M9 (sc)*	3.8
	T16 (sc)	T14 (sc)	3.6	F6 (sc)*	Y107 (sc)*	3.5	M9 (sc)*	Q10 (sc)*	3.9
	N19 (sc)*	M43 (sc)*	3.8	F6 (sc)*	I110 (sc)*	3.7			
	L22 (sc)	M43 (sc)	3.7	I7 (sc)*	Q100 (sc)*	3.8			
	T23 (sc)*	M43 (sc)*	3.7	I7 (sc)*	G104 (mc)*	3.7			
	K26 (sc)*	G60 (mc)*	3.7	Q10 (mc/sc)*	W103 (sc)*	3.6			
	E32 (sc)	I36 (sc)	3.8	E11 (sc)*	Q100 (sc)*	3.7			
	E32 (sc)	T33 (mc)	3.6	T16 (sc)	T16 (sc)	3.8			
	T33 (mc)	E32 (sc)	3.8						
	L35 (sc)*	I36 (sc)*	3.7						
	I36 (sc)	I36 (sc)	3.4						

**TABLE 2; Interactions identified in the dimer and tetramer interfaces.** Interface area refers to the solvent accessible surface area buried by each chain upon association. A' is the symmetry-related molecule to A (refer to the colour scheme in Figure 1: chain A=red, chain B = green, chain A' = blue, chain B' = orange). The symbol \* refers to contacts mirrored by symmetry, although within the dimer symmetry is not strictly obeyed at the level of individual side chains. The abbreviations sc and mc refer to side chain and main chain, respectively. Interactions were examined and identified using the programs AREAIMOL and DISTANG, of the CCP4 suite (46), and LIGPLOT (55).

Residue	SIS (%) chain A/B
<b>Ser 15</b>	<b>95 / 93</b>
<b>Ile 17</b>	<b>95 / 100</b>
<b>Asn 19</b>	<b>92 / 92</b>
Leu 20	73 / 77
<b>Leu 22</b>	<b>99 / 100</b>
Thr 23	61 / 71
<b>Glu 32</b>	<b>98 / 95</b>
<b>Leu 35</b>	<b>95 / 100</b>
<b>Ile 36</b>	<b>100 / 100</b>
Met 43	78 / 85
<b>Met 61</b>	<b>92 / 100</b>
Ile 110	68 / 75
Glu 114	61 / 42

**TABLE 3; List of amino acid residues the surface of which is predominantly buried, in the dimer interface.** Those side chains that are completely buried, or virtually so, are highlighted in bold and are likely to make the strongest contribution to the stability of the interface.



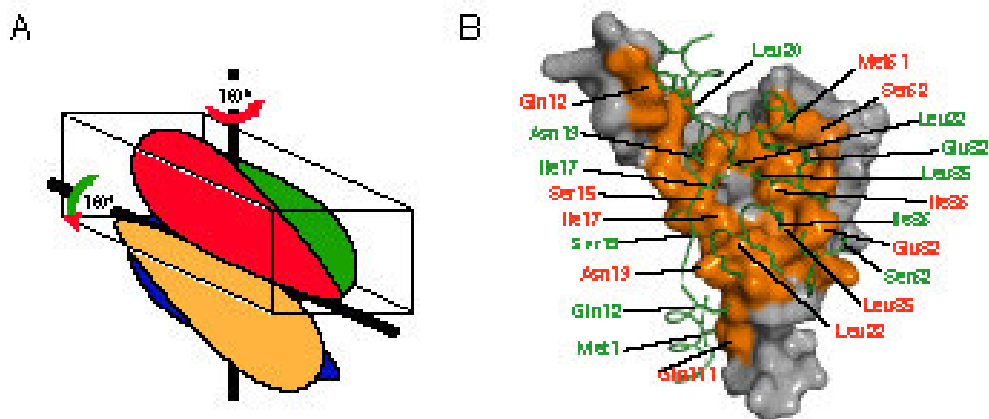
**FIGURE 1; Ribbon representations depicting the structure of the DnaD Nd.**

**A.** Schematic representation of the monomer with its wing-helix (WH) folding topology. The typical structural elements of the WH architecture are shown in a semi-transparent mode. The N- and C-terminal extensions unique to the WH fold of DnaD, termed H1', S1', H2' and then H3' respectively, are highlighted.

**B.** Two monomers of DnaD Nd form a symmetrical dimer through the pairing of their S1' strands (for a diagram of the dimer and tetramer symmetry see Figure 2A).

**C.** Side view of the dimer rotated 90° relative to the previous image.

**D.** Figure showing the tetrameric assembly, which is the predominant oligomeric state of the DnaD Nd in solution.

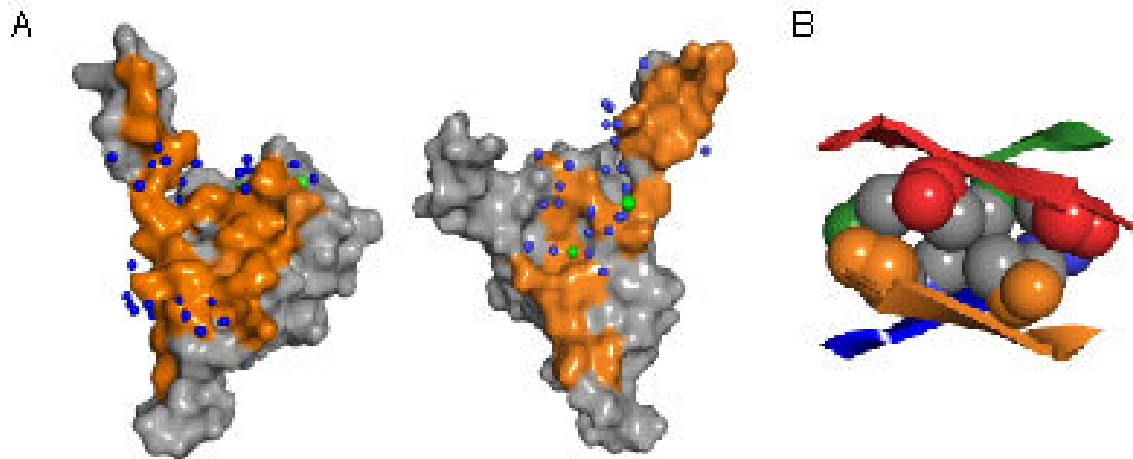


**FIGURE 2; Association of DnaD Nd into dimers and tetramer.**

**A.** Schematic diagram indicating the symmetrical relationship between the monomers and dimers in the DnaD Nd tetramer: monomers (eg A and B, shown in red and green) come together to form a dimer with P2 symmetry, as shown by the vertical axis. Two dimers then associate into a tetramer through another P2 symmetry axis, as highlighted by the rotation of the boxed AB dimer.

**B.** Interactions at the dimer interface: the surface representation of one monomer is colour-coded (orange) according to the areas engaged in contacts with the other monomer, which is shown in a green wire model. Each monomer buries over 1500 Å<sup>2</sup> upon association into the dimer. All amino acid side chains involved in either strong polar interactions (see Table 2) or extensive hydrophobic contacts (see Table 3) are annotated. The residues Ile17, Leu22, Leu35 and Ile36 are buried into the interface and interlock into an interaction hot spot.

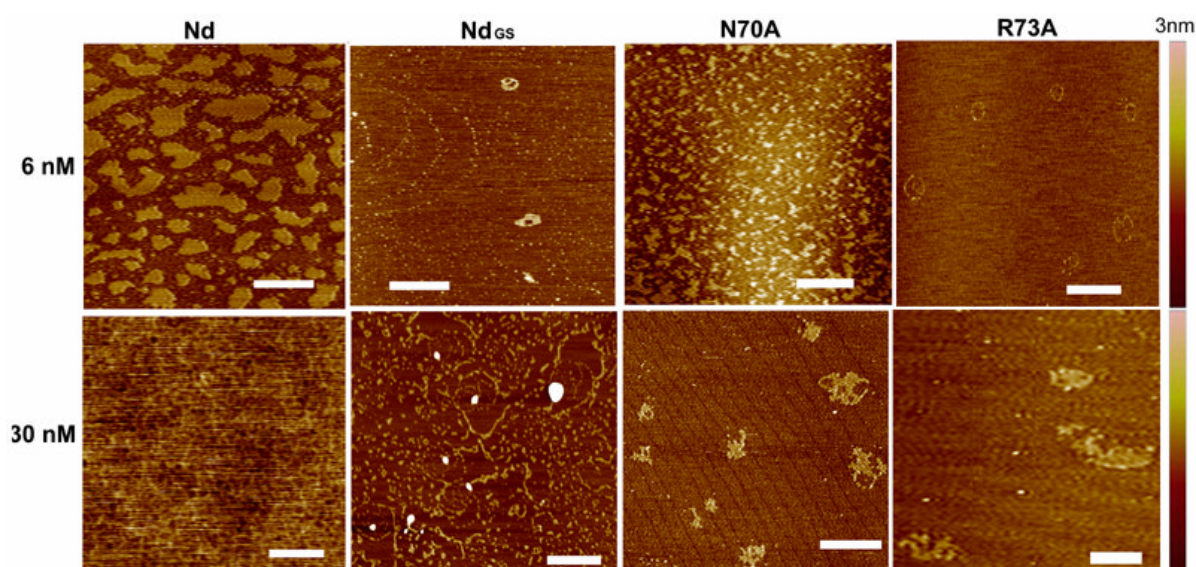




**FIGURE 3; Key features at the interfaces.**

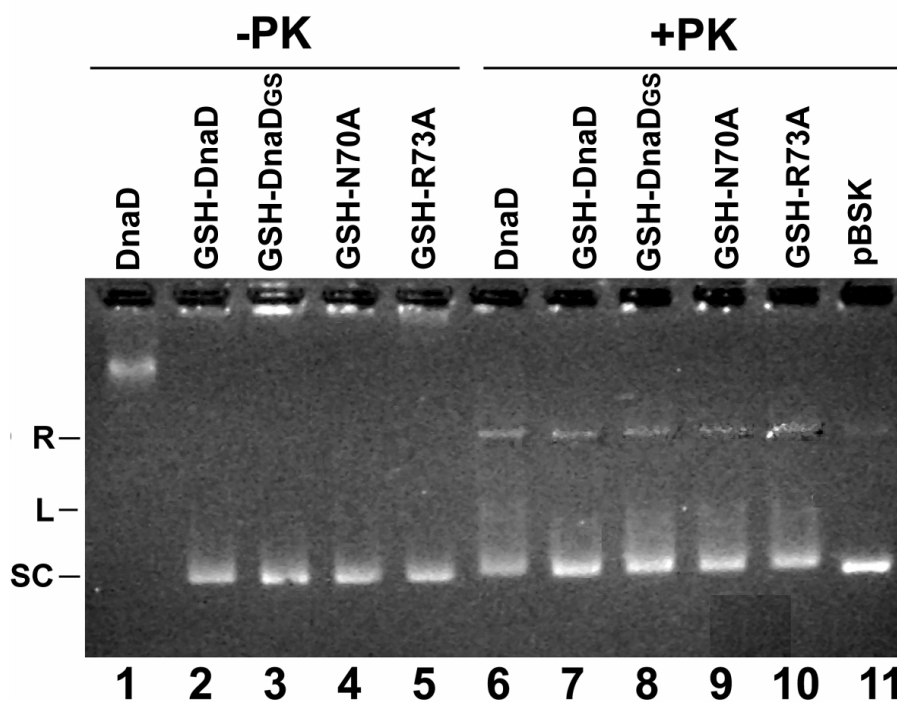
**A.** Hydration at the dimer and tetramer interfaces: surface representations of the dimer interface, on the left hand side, and the tetramer interface, on the right hand side. The diagrams show how water molecules (shown in blue) and ions (in green) mediate interactions exclusively at the periphery of the contact area in the case of the dimer interface. On the other hand, several waters are found interspersed across the tetramer contacts, resulting in over twice as many water molecules becoming buried in the latter case compared to the dimer interface.

**B.** The tetramerization motif: the tetramer appears to be stabilized by a striking structural feature: one threonine tetrad, with two Thr14/Thr16 pairs (shown in a space-filling model) provided by the N-terminal extensions  $\beta$ -strands hydrogen-bonded in the dimers, associates with another one to pack onto a tight core of 8 threonine side chains. The interlocking involves both non-polar hydrophobic contacts, contributed by the methyl groups (highlighted in grey), and hydrogen-bonding interactions, from the hydroxyl functions of the threonine residues.



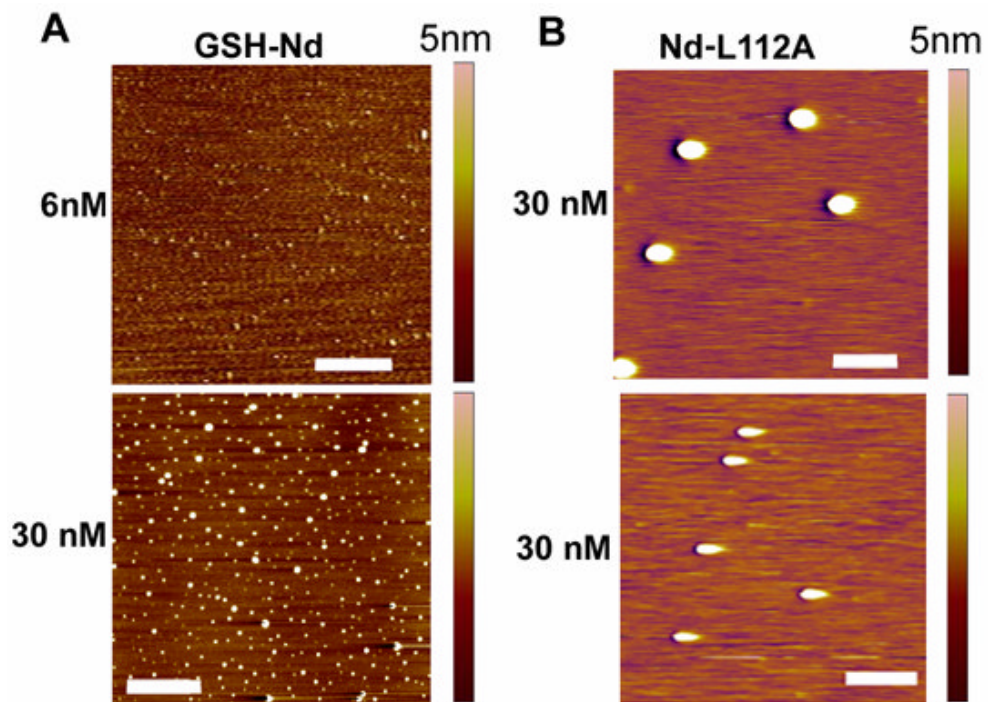
**FIGURE 4; AFM images of scaffolds formed by Nd and its mutants.**

AFM images of scaffolds formed by Nd and its mutants. Scaffolds formed by native Nd and the Nd<sub>GS</sub>, N70A and R73A mutants are shown at 6 and 30 nM, as indicated. At 30 nM native Nd forms a scaffold that covers the entire surface (bottom left), while mutants can only form irregular aggregates and smaller abnormal scaffolds. Scale bars represent 400 nm with Z-range 3 nm.



**FIGURE 5; The interaction of GSH-DnaD and its mutants with DNA.**

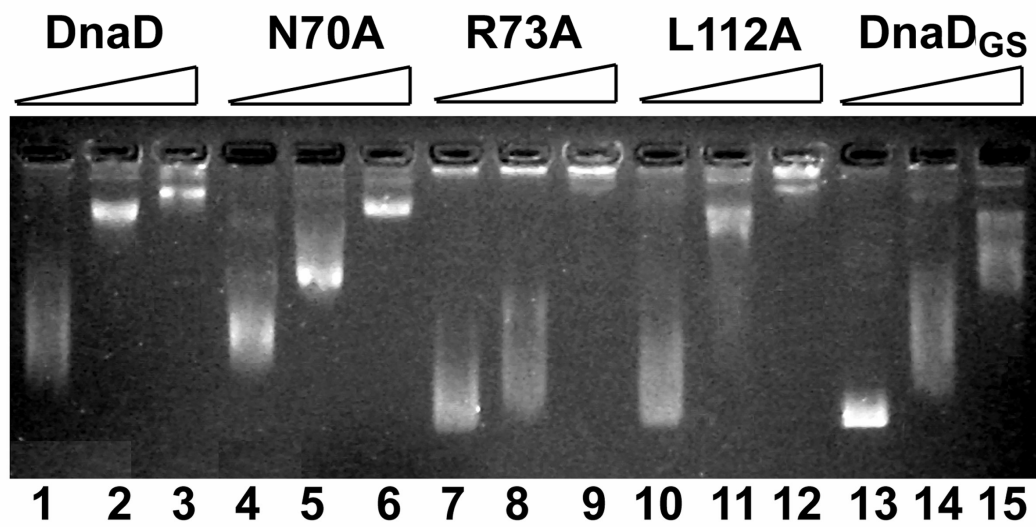
Agarose gel shifts of supercoiled pBSK plasmid (4.2 nM) treated with GSH-DnaD and its mutant proteins (105  $\mu$ M), as indicated. Control reactions with native DnaD and just pBSK plasmid are shown in lanes 1, 6 and 11, respectively. Samples from binding reactions were treated or not with proteinase K (PK), as indicated. Only the relaxed R-form of pBSK was shifted by the GSH-DnaD and its mutants compared to the all of the plasmid with the native DnaD protein.



**FIGURE 6; The GSH-Nd and Nd-L112A proteins form defective scaffolds.**

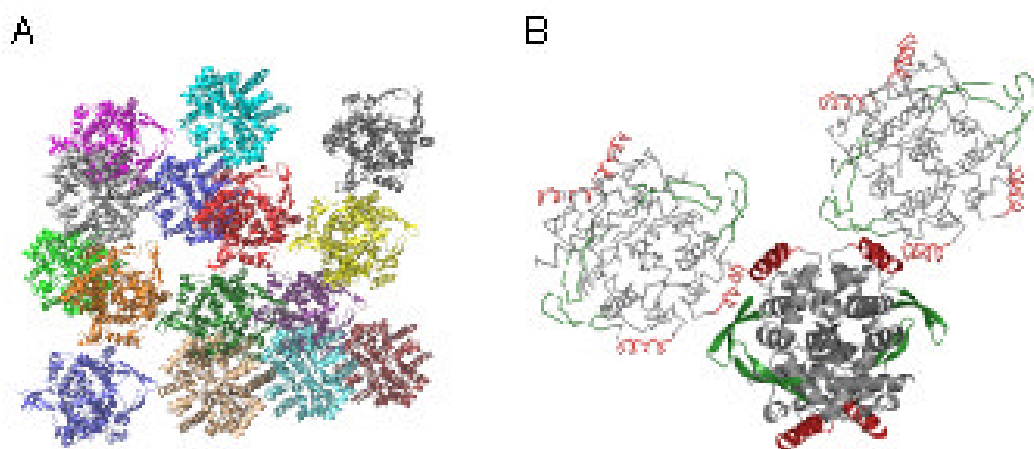
**A.** AFM images of GSH-DnaD scaffolds at 6 and 30 nM, as indicated. Scale bars represent 200 nm (top) and 1  $\mu\text{m}$  (bottom).

**B.** AFM images of L112 scaffolds at 30 nM. Zoomed (top) and distant (bottom) images are shown. The scale bars indicate 1 (top) and 2  $\mu\text{m}$  (bottom), for the zoomed and distant imaged, respectively with Z-range 5 nm.



**FIGURE 7; The interaction of native DnaD and its mutants with DNA.**

Comparative agarose gel shift assays of supercoiled pBSK (4.2 nM) in the presence of increasing concentrations (15, 45 and 105  $\mu$ M) of native DnaD and its mutants, as indicated.



**FIGURE 8; A lattice-based model for some key interactions governing scaffold formation.**

**A.** Cartoon representation of DnaD Nd molecules packing in the crystal lattice. Molecules along a plane or layer are shown and overlapping molecules have been removed from the figure to make the association between molecules easier to view.

**B.** A selection of three molecules from the lattice is displayed, highlighting the inter-tetramer interactions involving the H3 helix and the wing hairpin.

## SUPPLEMENTARY INFORMATION

pdb	chain	name	species	remark	seq. ID	RMSD	No. aligned
1XSD	A	BlaI repressor	<i>Staphylococcus aureus</i>	DNA	10.1	3.5	69/125
1W7P/1U5T	A	ESCRT II / Vps22 WH1	<i>Saccheromyces cerevisiae</i>		11.8	2.6	81/81
1W7P/1U5T	A	ESCRT II / Vps22 WH2			9.8	2.6	61/106
1W7P/1U5T	B/C (C/D)	ESCRT II / Vps25 WH1			7.7	3.1	52/86
1W7P/1U5T	B/C (C/D)	ESCRT II / Vps25 WH2			8.5	2.5	47/96
1W7P/1U5T	B (D)	ESCRT II / Vps36 WH1			6.5	2.9	46/96
1W7P/1U5T	B (D)	ESCRT II / Vps36 WH1			14.3	3.0	57/75
1QBJ	A	ADAR1 Z-alpha domain	<i>Homo sapiens</i>	DNA	17.6	2.3	51/65
	B				17.6	1.6	61/63
	C				17.3	1.7	52/66
2HDC	A	Genesis	<i>Rattus norvegicus</i>	DNA	6.4	1.7	47/97
2A07	J	FOXP2	<i>Homo sapiens</i>	DNA	4.7	2.8	64/83

**Supplemental Table SI** Comparison of the WH fold of DnaD Nd with other selected WH domains.

The tables reports the PDB codes, values of sequence identity, and r.m.s. deviations and the associated numbers of aligned residues obtained from secondary structure matching (60) carried out with COOT (54).

**Oligonucleotides used in this study.**

**N70A:**

N70AF: 5'-GAA GAA TGT ACA GCG AGA TTG CGG ATG TTT A-3'

N70AR: 5'-TAA ACA TCC GCA ATC TCG CTG TAC ATT CTT C-3'

**R73A:**

R73AF: 5'-GAA TGT ACA AAC AGA TTG GCG ATG TTT ATT CAA-3'

R73AR: 5'-TTG AAT AAA CAT CGC CAA TCT GTT TGT ACA TTC-3'

**L112A:**

L112AF: 5'-GAG TAT ATT CAG GCG GCA CAG AAT CAA ACA-3'

L112AR: 5'-TGT TTG ATT CTG TGC CGC CTG AAT ATA CTC-3'

**Wing truncation:**

GS2 FOR: 5'-ATT GAA GAA TGC GGC AGC TTT GAG AAA TAT TCT CTT CAG CCT TTA-3'

GS2 REV: 5'-ATA TTT CTC AAA GCT GCC GCA TTC TTC AAT AAA CAG AAA GCC TTT-3'

**Truncation of the N-terminal H1'-S1'-H2' element**

T29F: 5'-TAT AAA CAG CTT CAT ATG GGG CTT AAT GAA AC-3'

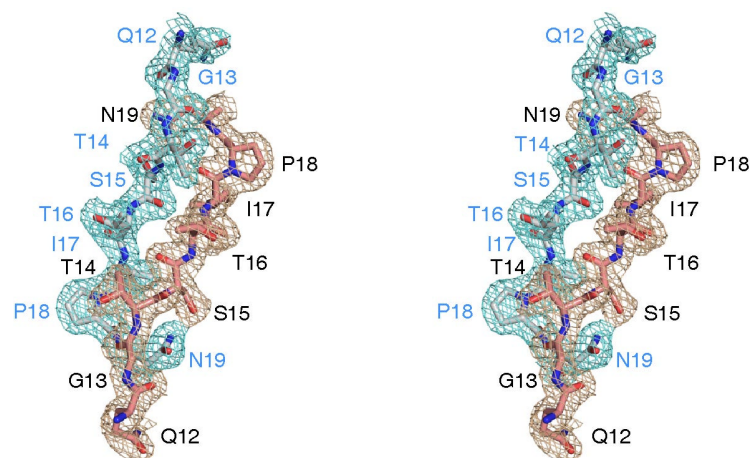
T7rev: 5'-GCT AGT TAT TGC TCA GCG G-3'

**Supplemental Table SII Oligonucleotides used in this study.**



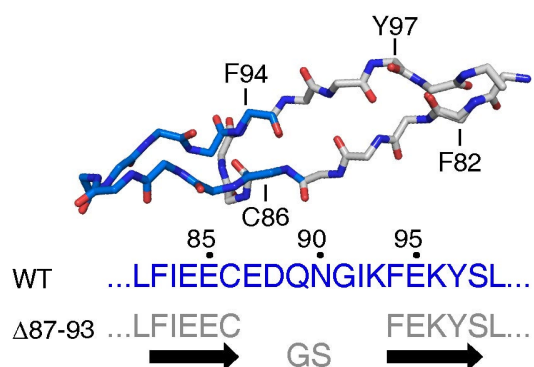
<b>Protein</b>	<b>GSH tag (N-terminus)</b>	<b>His<sub>6</sub> tag (C-terminus)</b>	<b>H1'-S1'-H2' truncation</b>	<b>Wing truncation</b>
DnaD	-	-	-	-
DnaD N70A	-	-	-	-
DnaD R73A	-	-	-	-
DnaD L112A	-	-	-	-
DnaD <sub>GS</sub>	-	-	-	+
GSH-DnaD	+	-	-	-
GSH-DnaD N70A	+	-	-	-
GSH-DnaD R73A	+	-	-	-
GSH-DnaD <sub>GS</sub>	+	-	-	+
Nd	-	+	-	-
Nd N70A	-	+	-	-
Nd R73A	-	+	-	-
Nd L112A	-	+	-	-
Nd <sub>GS</sub>	-	+	-	+
GSH-Nd	+	+	-	-
DnaD	-	-	+	-
Nd	-	-	+	-

**Supplemental Table SIII** Proteins and constructs used in this study.

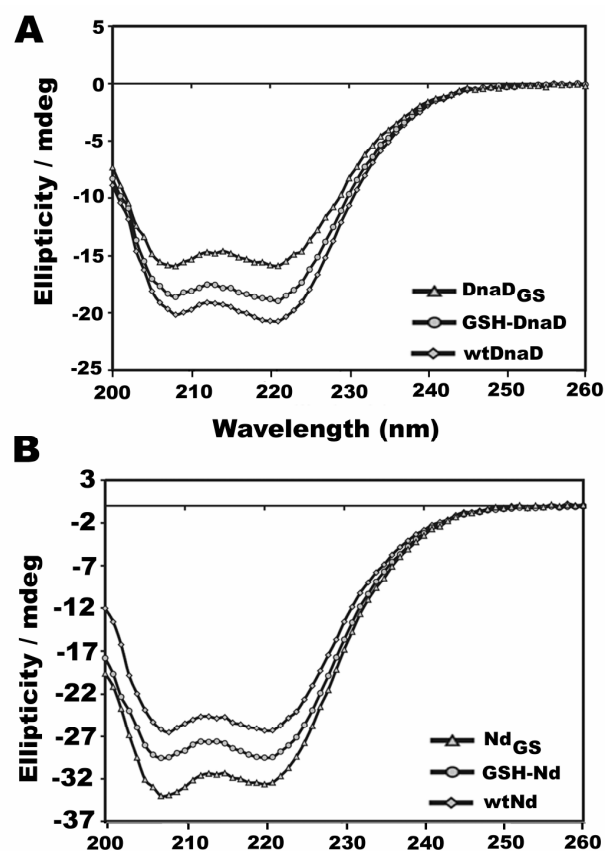


**Supplemental Figure S1** Stereo view of the electron density of representative area of the DnaD Nd molecule. Strands (residues 12 to 18) from the dimer are shown, with the electron density in different colours for each strand, contoured at  $1\sigma$  level.

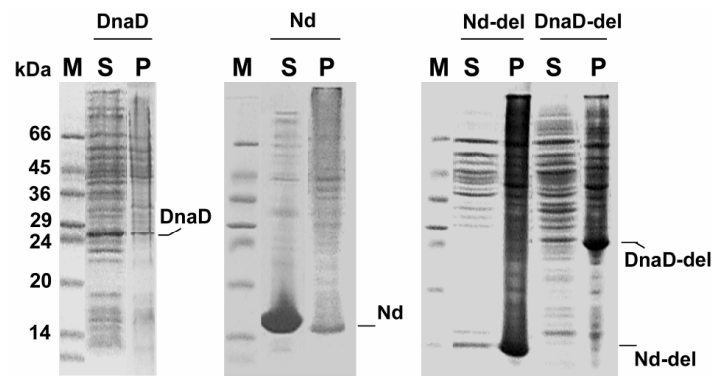




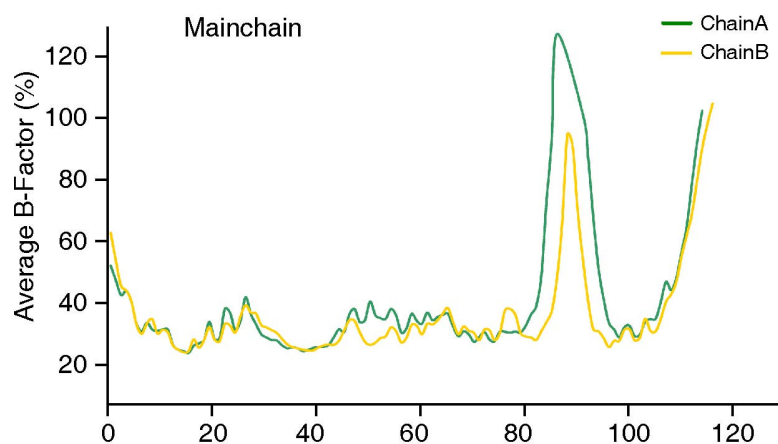
**Supplementary Figure S3** The predicted structure of the shortened wing-hairpin. A shorter, type-II  $\beta$ -hairpin (with carbon atoms shown in grey) was designed by replacing residues 87-93 (carbon atoms shown in light blue) with a Gly-Ser dipeptide sequence commonly found in type-II  $\beta$ -hairpin structures (52,53).



**Supplementary Figure S4** Comparative Circular Dichroism (CD) analysis of the loop truncated (DnaDGS) and GSH N-terminal insertions in the wtDnaD (panel **A**) and Nd proteins (panel **B**). The CD spectra were recorded on an Applied Photophysics Pi-Star-180 Spectrophotometer at 25 °C. The temperature was regulated using a Neslab RTE-300 circulating programmable water bath and a thermoelectric temperature controller (Melcor). Spectra were obtained from 300  $\mu$ l of sample in a 1 mm path length cuvette. A blank spectrum was also recorded and subtracted from the protein spectrum. Secondary structure content was investigated by recording spectra from 260 nm to 200 nm, using a 1 nm step and 3 nm entrance and exit slit widths.



**Supplementary Figure S5** Truncation of the N-terminal H1'-S1'-H2' element results in insoluble protein. The N-terminal H1'-S1'-H2' element was truncated in the native DnaD and the Nd, producing the DnaD-del and Nd-del proteins. Both proteins were over-expressed in *Escherichia coli* (BL21 DE3) but were insoluble. Various attempts to improve solubility, including different concentrations of IPTG, low temperature expressions and the use of solubilising agents such as triton X-100 were tried but no improvement in solubility was achieved (data not shown). The soluble (S) and insoluble (P) fractions of cell extracts containing the DnaD-del and Nd-del proteins are shown. For comparison the equivalent fractions from the over-expressions of native DnaD and Nd are also shown. The relative positions of the various proteins in the gels are indicated for clarity.



**Supplementary Figure S6** Average B-factors of main chain atoms in the DnaD Nd crystal structure. The plot indicates that the region corresponding to the wing-hairpin (residues 80-98) of the WH fold has the highest B-factors in both monomers and it therefore expected to have significant flexibility.