Structural basis for the complex DNA binding behavior of the plant stem cell regulator WUSCHEL

Jeremy Sloan[1,+], Jana P. Hakenjos[2,#], Michael Gebert[2], Olga Ermakova[2], Andrea Gumiero[1,§], Gunter Stier[1], Klemens Wild[1], Irmgard Sinning[1] and Jan U. Lohmann[2*]*

[1] Biochemistry Center, Heidelberg University, Im Neuenheimer Feld 328, 69120 Heidelberg, Germany

[2] Department of Stem Cell Biology, Centre for Organismal Studies, Heidelberg University, Im Neuenheimer Feld 230, 69120 Heidelberg, Germany

[+] Present address: BASF SE, Carl-Bosch-Strasse 38, 67056 Ludwigshafen, Germany

[#] Present address: Celonic AG, Eulerstrasse 55, 4051 Basel, Switzerland

[§] Present address: Instituto Poligrafico e Zecca dello Stato S.p.A., Via Salaria, I-712-00138 Roma, Italy

[*] To whom correspondence should be addressed:

Jan U. Lohmann (email: jan.lohmann@cos.uni-heidelberg.de)
Department of Stem Cell Biology
Centre for Organismal Studies, Heidelberg University,
Im Neuenheimer Feld 230, 69120 Heidelberg, Germany.

Irmgard Sinning (email: irmi.sinning@bzh.uni-heidelberg.de)
Heidelberg University Biochemistry Center (BZH)
Im Neuenheimer Feld 328, 69120 Heidelberg, Germany.

Abstract

Stem cells are one of the foundational evolutionary novelties that allowed the independent emergence of multicellularity in the plant and animal lineages. In plants, the homeodomain (HD) transcription factor WUSCHEL (WUS) is essential for the maintenance of stem cells in the shoot apical meristem. WUS has been reported to bind to diverse DNA motifs and to act as transcriptional activator and repressor. However, the mechanisms underlying this remarkable behavior have remained unclear. Here, we quantitatively delineate WUS binding to three divergent DNA motifs and resolve the relevant structural underpinnings. We show that WUS exhibits a strong binding preference for TGAA repeat sequences, while retaining the ability to weakly bind to TAAT elements. This behavior is attributable the formation of dimers through interactions of specific residues in the HD that stabilize WUS DNA interaction. Our results provide a mechanistic basis for dissecting WUS dependent regulatory networks in plant stem cell control.

Introduction

Plant stem cells are embedded into specialized tissues that promote their life-long maintenance, which are called meristems. These meristems are located at the growth points of all plants, namely the shoot and root tips, as well as the vascular cylinder, to support apical-basal and lateral growth respectively[1]. Similar to animal stem cell systems, signals controlling stem cell identity and activity within the meristem emanate from niche cells located adjacently to stem cells[2, 3]. However, the cellular and molecular mechanisms of this communication are highly divergent between the two kingdoms. While cell-cell contact and secreted ligands play a central role in animals, direct cytoplasmic connections between neighboring cells, called plasmodesmata, take center stage for the maintenance of plant stem cells[4]. Interestingly, the related homeodomain (HD) transcription factors (TFs) that define the niche cells in shoot and root, namely WUSCHEL (WUS) and WUSCHEL HOMEOBOX 5 (WOX5), move to stem cells and execute their function primarily in these cells[4, 5, 6]. Consequently, there is no need for downstream niche to stem cell signaling cascades, since the DNA binding specificities of these TFs will directly dictate the repertoire of genes expressed in stem cells.

Prokaryotes and eukaryotes use different strategies to target TFs to distinct genomic locations. In bacteria, it seems to be sufficient that TFs recognize an extended DNA sequence, whereas TFs in eukaryotes typically bind to shorter DNA recognition motifs and therefore require clustering of sites to achieve specificity[7]. Numerous gene regulatory proteins of higher eukaryotes, such as leucin zipper and zinc finger TFs, bind as symmetric dimers to DNA in a sequence-specific manner, which allows each monomer to bind in a similar fashion and greatly increases the DNA binding affinity[8, 9]. Consistently, the DNA recognition sequences that are bound by these TFs often are

arranged as inverted or everted repeat elements. Many TFs, however, can also associate with nonidentical proteins to form heterodimers composed of two different subunits. As heterodimers are typically composed of two distinct proteins with different DNA-binding specificities, the combination of multiple TFs immensely expands the repertoire of recognized DNA sequences and greatly improves the binding specificity[10]. The eukaryotic superfamily of homeobox TFs is characterized by the presence of a HD, a short stretch of amino acids (60-66 residues) that forms a helix-loop-helix-turn-helix DNA-binding domain consisting of three alpha helices[11]. HD-TFs play a wide variety of roles in developmental and growth processes such as embryonic patterning, stem cell maintenance and organ formation in all kingdoms of life[11, 12, 13]. In animals, the HOX transcription factors are the best studied family of HD proteins and specify segment identity during embryo development along the head-tail axis[14, 15].

Several studies have addressed WUS DNA binding in some detail and at least three divergent sequence motifs bound by WUS, specifically sequences with a TAAT core, a G-Box like and a TGAA repeat element, have been identified[6, 16, 17, 18, 19, 20, 21]. The TAAT sequence was originally identified since it represents the canonical binding element for HD proteins and was subsequently experimentally confirmed to be bound by WUS by electrophoretic mobility shift assays (EMSAs) and reporter genes for multiple independent targets[6, 16, 17, 20]. More recently, dimerization of WUS on TAAT repeats was suggested to control expression of the stem cell specific signaling factor *CLAVATA3* (*CLV3*) based on EMSA and reporter gene assays[21]. The G-Box like (TCACGTGA) motif was found in a combination of systematic evolution of ligands by exponential enrichment (SELEX) and *in vivo* WUS chromatin binding data derived from chromatin immunoprecipitation followed by detection by microarrays (ChIP-chip)[18]. SELEX using a recombinant WUS-HD fragment resulted in the enrichment of TCA

containing sequences and the G-Box like element, which represented an inverted repeat of TCA bases, was found to be the most overrepresented DNA sequence in chromatin regions bound by WUS[18]. Binding of WUS to this motif was confirmed by EMSA and reporter gene analysis. Lastly, the TGAA repeat motif was identified in a large-scale approach using recombinantly expressed TFs and genomic DNA in a highly parallel protein-DNA interaction screen, but has not been verified independently so far[19]. In addition to binding to multiple DNA motifs, WUS also exhibits further functional complexity by acting as transcriptional activator and repressor[17, 18, 21, 22] and the mechanistic basis for both unusual behaviors has remained largely elusive so far. Here we have combined molecular, biochemical and structural approaches to address how the WUS-HD recognizes specific DNA target sites. We find that the DNA-binding preferences of WUS-HD depend on appropriately arranged sequence motifs in a direct tandem repeat and that homodimerization is one of the key determinants to achieve high sequence specificity. We also show that disrupting the dimer interface, either on the protein level or the DNA level, severely reduces DNA-binding affinity.

Results

WUS has a canonical HD fold with unique structural features

As an entry point to elucidate the mechanisms by which the WUS-HD carries out its functions, we recombinantly produced and purified a fragment containing residues 34-103 of WUS (**Fig. 1a** and **Supplementary Fig. 1**) and determined its crystal structure to a resolution of 1.4 Å (**Fig. 1b**, **Table 1**). The overall WUS-HD fold reflected that of a canonical HD structure, consisting of a three-helix bundle with an N-terminal arm. Interestingly, superposition with the structure of Engrailed (En) (PDB code 3HDD[23]) and comparative sequence analyses identified unique structural features in WUS-HD.

First, the loop regions connecting the three α-helices are expanded (**Fig. 1b**,**c**). Loop region I is slightly longer compared to other HDs and is characterized by a distortion at the end of helix α1. This so called π-helix or π-bulge is typically characterized by a single amino acid insertion into an existing α-helix (Y54 in WUS) and usually correlates with a particular functional role[24, 25]. Loop region II has an even longer insertion, which also extends helix α2 by an additional turn compared to En. Secondly, the N-terminal arm is anchored by docking of a tryptophan residue (W39) into a groove formed by helices α2 and α3, whilst a F, Y or I residue typically performs this function in canonical HDs (**Fig. 1c,d**). Furthermore, this docking residue is shifted by one residue relative to the conserved DNA-contacting arginine (R38) in contrast to three residues in canonical HDs (**Fig. 1 c**). Electrostatic surface calculations showed a large positively charged surface formed by the C-terminal recognition helix (α3) and the N-terminal arm (**Fig. 1e**), which represent the most conserved part of the HD (**Fig. 1f**). Finally, comparison with the sequence and structure of En suggested that the readout of DNA bases is mediated by the conserved residues R38, N90 and R94 of WUS-HD (**Fig. 1b,c,f**).

WUS prefers tandemly arranged DNA recognition sequences

WUS has been reported to bind to at least three divergent DNA sequences and these motifs have been proposed to be important to determine the transcriptional output of WUS[16, 18, 19]. Despite the obvious importance of this issue for resolving the mechanisms of WUS activity *in vivo*, quantitative data comparing the binding affinities to these motifs were still lacking. In order to elucidate the DNA binding preferences of WUS-HD, we therefore analyzed its interaction with three of the best studied sequences (**Fig. 2a,b**), namely a TAAT element from the *AG* enhancer[16], a G-Box from the *CLV1* promoter[18] and a TGAA repeat element identified in a large *in vitro* screen[19].

Both TGAA and G-Box harbor two atypical HD recognition motifs[26, 27] that are arranged as a direct repeat and inverted repeat, respectively (**Supplementary Fig. 2a,b**), while the TAAT DNA only contains one typical HD recognition motif (**Fig. 2b**).

We employed microscale thermophoresis (MST) with a N-terminal YFP fusion of WUS-HD and 16-bp double stranded DNA probes corresponding to naturally occurring regulatory sequences containing either the TAAT, G-Box, or TGAA repeat motif. In line with earlier results[18] and in agreement with the low affinity generally reported for other HDs[11], the WUS-HD bound the TAAT probe with lower affinity compared to the G-Box containing probe with dissociation constants (K_d) of 10.60 ± 1.67 μM and 3.78 ± 0.42 μM, respectively (**Fig. 2a,b**).

Intriguingly, the TGAA repeat probe was bound by WUS-HD with much higher affinity (K_d = 0.27 ± 0.03 μM) than the other two sequences (**Fig. 2a,b**). As the 4-bp recognition motifs of the TGAA sequence are not significantly different from the G-Box sequence (**Supplementary Fig. 2b**), we hypothesized that the relative position of recognition motifs may be a major determinant of binding specificity. To rule out any contribution of the YFP-tag to DNA-binding specificity a control measurement with YFP alone was performed, which showed no binding to DNA (**Fig. 2a**).

In order to compare the observed DNA-binding of WUS-HD to that of the full length (FL) protein, we expressed a fusion of WUS-FL to maltose-binding protein (MBP), and performed EMSA experiments (**Supplementary Fig. 3**). In accordance with our MST results, WUS-FL also exhibited divergent DNA-binding behavior, comparable to WUS-HD, when probed with TAAT, G-Box, and TGAA fluorescently labeled oligonucleotides. Consistent with the observations from the MST analysis, the TGAA repeat sequence had the highest binding affinity (K_d = 0.36 ± 0.06 μM) of the three probes (**Supplementary Fig. 3a**). In addition, WUS-FL bound the G-Box probe with higher

affinity compared to the TAAT containing probe (K_d = 1.68 ± 0.30 µM and K_d = 3.15 ± 0.26 µM, respectively) (**Supplementary Fig. 3b,c**).

Overall, the K_d values from MST and EMSA experiments were in good agreement and deviations were mainly within measurement errors. Earlier studies had shown that WUS has the ability to homodimerize via protein domains outside the HD, which was suggested to be critical for WUS function[18, 21]. However, our results showed that DNA-binding preference of WUS is dictated by the WUS-HD alone. To test whether these results reflect WUS chromatin binding behavior in living plant cells, we analyzed WUS ChIP-seq data[28] using read counts associated with the three DNA binding motifs as a proxy for affinity (**Fig. 2c**). Specifically, we analyzed the probability of TAAT, G-Box, or 2xTGAA repeat motifs to be present in chromatin regions strongly bound by WUS and therefore being covered by a large number of ChIP-seq reads. Since the motifs occur in the genome at vastly divergent numbers, we converted read counts into relative binding probabilities. To this end, we plotted the relative occurrence of motifs against the number of Chip-seq reads per motif. In such an analysis, very steep curves in the left part of the coordinate system indicate motifs that occur most frequently in regions with low ChIP-seq coverage, whereas curves that are shifted to the right indicate an association of the motif with regions of higher ChIP-seq reads and hence are suggestive of higher affinity. Our analyses showed that native WUS was indeed associated with 2xTGAA repeat sequences more often than with G-Box containing genomic regions, which was followed by TTAATGG sites. Taken together, our results demonstrated that WUS strongly prefers the TGAA repeat sequence over the G-Box motif and the canonical TAAT element, both *in vitro* and *in vivo*.

WUS-HD uses a general binding mode for different DNA sequences

To elucidate the structural basis for these differential interactions, we solved crystal structures of WUS-HD bound to TAAT, G-Box and TGAA repeat probes to resolutions of 2.8, 2.7 and 1.6 Å, respectively (**Fig. 2d**). In all crystal structures of WUS-HD/DNA complexes, the unit cell contained two DNA molecules which were occupied by at least two WUS-HDs (**Supplementary Fig. 4**). The structure of the HD fold of each WUS molecule was not modified by the formation of the ternary complex, with an overall root mean square deviation (rmsd) before and after DNA-binding of 0.9 Å over 62 residues, although the length of the N- or C-termini vary in a context dependent manner. Interestingly, in the case of G-Box and TAAT, one of the two protein-DNA complexes in the asymmetric unit contained an additional bound HD, whereas both TGAA structures only included two HDs per DNA (**Fig. 2d** and **Supplementary Fig. 4**). In both cases, this additional molecule inserted its C-terminal recognition helix into a major groove of the DNA on the complementary strand of one of the prevalent recognition motifs, but did not make contact to the other two protein molecules. In the G-Box structure, this extra HD (**Fig. 2d** cyan) was stabilized by crystal contacts from the neighboring complex and thus likely represents a crystallization artefact (**Supplementary Fig. 4c**). Furthermore, initial low resolution (>3 Å) crystal structures of the G-Box complex only ever included two HDs per DNA molecule, similar to the structure seen in **Supplementary Fig. 5a**.

The binding behavior in the TAAT structure was more complex. Whilst the additional HD in the TAAT complex (**Fig. 2d,e** light blue) bound on the opposite side of the TAAT recognition motif, one of the other HDs (**Fig. 2d,e** teal) contacted an unexpected DNA sequence, with less clear DNA-base interactions and an overall higher flexibility, as indicated by elevated B-factors (**Supplementary Fig. 6**). In order to understand the

significance of these protein-DNA contacts and to delineate the critical binding regions of the TAAT sequence we performed EMSA experiments, which clearly demonstrated a change in the DNA-binding behavior of WUS-HD to the TAAT sequence probes (**Supplementary Fig. 7**). DNA binding to the T4C probe was largely impaired and no distinct band shifts were visible, indicating that this DNA position is important to form a stable protein-DNA complex (**Supplementary Fig. 7a**). In contrast, the T12C probe gave a similar band shift as the wild-type (wt) TAAT sequence, suggesting no or little interference with WUS-HD binding. However, when probed with the double mutant T12C, T15C the band shift intensities decreased compared to the TAAT wt sequence, which would imply a relevant protein-DNA contact site (**Supplementary Fig. 7b**).

Collectively, the structural analysis and the EMSA results suggested that two HDs, which bind on opposite sides of the TAAT recognition motif (light and dark blue in **Fig. 2d,e**), are crucial for an efficient interaction with the TAAT DNA. An additional HD observed in the TAAT crystal structure (teal in **Fig. 2d,e**) seems to be less important for DNA-binding in solution, consistent with a less defined structure and more ambiguous contact sites.

All WUS-HDs were bound to the expected 4-bp recognition motifs, except in the TAAT structure, where one molecule occupied a sequence distinct from the TAAT motif (**Fig. 2e**). Despite this, comparison of the ternary complex structures revealed a very similar mode of DNA-binding for each HD; the N-terminal arm spanned the DNA minor groove, whereas the C-terminal recognition helix inserted into the major groove (**Supplementary Fig. 8**). Helix α3 made extensive backbone contacts, whilst both N- and C-terminal regions were engaged in establishing base-specific contacts. In almost all WUS-HD molecules, the N-terminal arm inserted R38 into the minor groove to hydrogen-bond with base pairs and typically specified a thymine at the -2 position[29]

(**Fig. 2e** and **Supplementary Fig. 8**). The hydrogen-bond donor-acceptor pattern was neither specific to the sense- or antisense-strand, however the readout by R38 was mediated by base pair recognition and may have also be dependent on DNA shape[30]. The majority of DNA contacts were established by major-groove interactions (**Fig. 2e** and **Supplementary Fig. 5b**), involving extensive backbone contacts as well as base pair recognition.

The readout of bases in the major groove was mediated by the conserved residues Q89, N90 and R94 (**Fig. 1f, Fig. 2e** and **Supplementary Fig. 5b**). N90 that specified adenine (position 0), crucial for HD binding, appeared as most relevant[23, 31]. R94 favored a guanine at position -1 from the adenine (**Supplementary Fig. 2b**), in agreement with the specificity of atypical HDs[26, 27]. Interestingly, in the TAAT structure position -1 was an adenine, similar to the DNA recognition motif of typical HDs; thus, in this complex R94 was not involved in base recognition and instead contacted the sugar phosphate backbone (**Fig. 2e** and **Supplementary Fig. 8**). The role of Q89 was less clearly defined by the structures; in some cases, it did not interact directly with DNA and in others it contacted a base at position +2 or +3 on either side of the double strand, consistent with the idea that the conserved Q89 promotes the recognition of bases at these positions[32, 33].

Residues K82, N83 and Y86 formed a cluster which bound consecutive phosphate groups of the DNA backbone. K92, R96 and R100 were also involved in backbone contacts, although to a lesser extent as these interactions were not present in all structures and thus presumably were dependent on protein-protein interactions (**Fig. 2e** and **Supplementary Fig. 5b**). Importantly, all side-chains involved in the readout of base pairs were among the most highly conserved residues of the WUS-HD (**Fig. 1c** and **Fig. 2e**). In addition, the observed protein-DNA contacts were consistent with

results obtained with other HDs, where specific interactions are established with a 4 to 7 bp DNA binding site[26, 27]. Most other conserved amino acids without indicated functions appeared to have structural roles in maintaining the overall HD fold (**Fig. 1c**).

WUS-HD prefers the atypical TGAA over the typical TAAT motif

Having identified the DNA recognition preferences of WUS, we compared the binding mode with typical and atypical HDs from metazoans with similar interaction motifs (**Fig. 3**). The 'typical' Antennapedia (Antp) HD binds to a core TAAT motif (PDB code 4XID[34]) as found in our *AG* derived TAAT probe. The residues involved in establishing base-specific contacts are conserved, however there are notable differences in the interactions formed by Antp-HD and WUS-HD (**Fig. 3b**). Commonly, arginine (or lysine) as residue R2 or R3 enables specific read-out of the adenine in the -1 position[29, 35], the hallmark of the typical recognition motif. In contrast, the equivalent residues in WUS-HD (T35 and S36) did not form this base-specific contact. However, the N-terminal arm of WUS-HD still contacted adenine -1 via R38, equivalent to the highly conserved R5 that conventionally reads out the -2 position only (**Fig. 3a,b**).

The binding of the 'atypical' Extradenticle (Exd) HD (PDB code 2R5Y[30]) to a core TGAT motif, although very similar in sequence, also shows some differences in the hydrogen-bond pattern compared to WUS-HD bound to the G-Box probe (**Fig. 3b**). Notably, the conserved R38 of the N-terminal arm in our structure neither contacted the -1 nor the -2 position of the sense-strand. Instead, R38 bound to positions -2 and -3 of the antisense-strand, which highlighted the broader specificity of HDs at position -1 and -2 of the core recognition motif[26, 27]. However, the guanine at position -1 was specified by R94 in the C-terminal recognition helix, following the usual mechanism for HDs contacting atypical recognition motifs (**Fig. 3**). Taken together, our crystal structures indicated that WUS-HD is able to establish the canonical base-specific contact with

guanine in position -1 but not adenine. As the nucleotide base in this position is the main determinant in preferential recognition of typical or atypical motifs, this would suggest that WUS-HD prefers binding to atypical motifs. Regardless, WUS-HD is still able to form specific interactions with the typical TAAT motif, likely reflecting the inherent broad specificity of HDs for DNA sequence recognition[26, 27].

WUS-HD binding specificity depends on DNA shape

We noticed that position +1 of the DNA motif made no hydrogen bonds to the protein in any of our structures and therefore we analyzed whether WUS had any base-preference at this position (**Fig. 4a**). Intriguingly, MST experiments showed a strong preference of WUS-HD for A or T, with binding to A containing probes roughly twofold stronger than to probes with a T at this position (**Supplementary Fig. 9a,b**). In contrast, probes with C or G were bound less tightly, decreasing affinity by ~7-fold and ~26-fold, respectively (**Supplementary Fig. 9c,d**). Interestingly, the recognition motifs in the TGAA repeat sequence each have adenine at position +1, consistent with the observation that this sequence was bound with a higher affinity than the other two crystallized variants (**Fig. 2a**,**b**).

Why does the +1 DNA motif position have such a strong influence on WUS-HD affinity, despite the fact that this base is not involved in hydrogen bond interactions? Since DNA shape can have a substantial effect on specificity and affinity of HD-DNA complexes[36], we computationally investigated potential structural differences in the DNA sequences experimentally tested. To this end, we used the DNAShape tool[37] to predict the intrinsic conformation of unbound DNA probes differing in the +1 position focusing on minor groove width (MGW) (**Fig. 4b**). Consistent with the overall similarity of the sequences, the predicted MGW profiles are similar in all cases with two minor groove width minima occurring around the different nucleotides at the +1 position.

Interestingly, these minima spatially coincide with where the side chain of WUS R38 would insert, establishing contacts to thymine 7 and thymine 11 (**Fig. 4b**,**c**).

Our analysis showed that the local MGW minima for the 3xTGAA and 3xTGAT sequences are much more pronounced than in the 3xTGAC and 3xTGAG sequences. The 3xTGAA sequence, which had had the highest affinity of the four sequences (K_d = 0.06 ± 0.01 μM), showed two strong minima that overlapped best with the binding position of WUS R38 (**Fig. 4b**). In contrast, the 3xTGAT sequence, which had a slightly weaker affinity (K_d = 0.12 ± 0.01 μM), also exhibited two strong minima, however, they are shifted to position +2. In addition, the 3xTGAC and 3xTGAG sequences had even weaker affinities (K_d = 0.39 ± 0.05 μM and K_d = 1.42 ± 0.17 μM, respectively), consistent with the local MWG minima being in a different position and a less narrow minor groove. Consequently, the DNA Shape tool predictions suggested that the 3xTGAC and 3xTGAG sequences are less well pre-organized for WUS DNA-binding and require larger conformational changes compared to the 3xTGAA and 3xTGAT sequences.

Besides the widely recognized hydrogen bond interactions of specific bases, hydrophobic contacts can also be an important determinant for protein-DNA specificity[38]. Analysis of the DNA contacts in the structures of TGAA and G-Box, where the +1 motif position was an adenine or thymine respectively, revealed that hydrophobic residues of WUS-HD made contact to the C5 methyl group of a thymine base (**Fig. 4d**). In the G-Box crystal structure, Y86 formed Van der Waals interactions with thymine of the +1 position of the TGAT motif. In addition, the aliphatic chain of K82 interacted with Y86 and thus contributed to the local hydrophobic environment. In contrast, in the TGAA structure A93 contacted the thymine from the complementary strand, which base paired with the adenine of the +1 position (**Fig. 4d**). Thus, despite

the fact that the +1 DNA motif position was not involved in base-recognition via hydrogen bonds with WUS-HD, the bases at this position had a substantial influence on local DNA conformation and, together with Van der Waals contacts of hydrophobic side chains from WUS-HD, led to a strong preference for A/T over G/C. These findings were also consistent with the experimental observation that the G-Box probe, which contained TGAG and TGAT motifs, was bound with much lower affinity than the TGAA probe, which has two TGAA motifs (**Fig. 2a**,**b**). Furthermore, thymine is the most common base at position +1 in typical recognition motifs and correlates with the presence of an aliphatic residue contacting this position[26, 27].

The WUS-HD undergoes DNA-mediated dimerization

One of the surprising findings of our crystallization experiments was that two WUS-HDs were found to bind every DNA molecule, even though the canonical TAAT motif is usually only bound by a single HD[26, 27]. In addition, we observed that irrespective of the probe sequence, the two WUS-HD molecules are engaged in protein-protein contacts (**Fig. 5a** and **Supplementary Fig. 4**) even though multi angle light scattering (MALS) demonstrated that the WUS-HD was monomeric in solution (**Supplementary Fig. 10a**). Interestingly, these DNA-bound dimers had a unique relative orientation in all structures. Bound to the G-Box probe, the two monomers were positioned Head-to-Head on the same side of the DNA and had almost identical binding features, probably due to the palindromic nature of the DNA recognition sequence and the negligible interaction between them (**Fig. 2d**, **Fig. 5a and Supplementary Fig. 11a**). In contrast, the two HD molecules interacting with the TAAT and the TGAA repeat probes were on opposite sides of the DNA and formed specific protein-protein interactions between each other (**Fig. 2d** and **Fig. 5a**). One of the WUS-HD molecules bound to the TGAA probe made additional DNA contacts through stabilization of the helix α3 C-terminus

by the other WUS molecule (**Fig. 2e** and **Supplementary Fig. 11b**), which might explain the higher affinity for the TGAA sequence (**Fig. 2a**,**b**). In particular, R96 and R100 of the recognition helix established new contacts to the DNA-backbone, not observed in any of the other WUS-DNA complexes. Similarly, extensive protein-protein interactions with the HD bound to the typical core TAAT sequence likely allowed an additional WUS-HD molecule to occupy an unexpected position in the TAAT structure (**Fig. 2d** and **Fig. 5a**). However, in this case, the interaction with the DNA was less important and the structure was not well resolved in the electron density, as indicated by elevated B-factors (**Supplementary Fig. 6 and Supplementary Fig. 11c**). In contrast, in the other, likely more relevant configuration, the two WUS-HD molecules did not exhibit any protein-protein interface, but individually formed a stable protein-DNA complex with the TAAT probe (**Fig. 2d** and **Supplementary Fig. 7**).

Cooperative DNA-binding determines WUS-HD sequence specificity

Further analysis of the DNA-binding activity of WUS-HD by MST measurements clearly demonstrated a gain in binding affinity with increasing number of recognition motifs, indicating that the binding of multiple HD molecules per DNA molecule occurs in solution as well as in our crystal structures (**Supplementary Fig. 12a**). Although in this experimental setup, the derived K_d values are not directly comparable due to the variation in the number of binding sites, the increase in affinity from one (1xTGAA, K_d = 10.50 ± 2.30 µM) to two binding sites (2xTGAA, K_d = 0.30 ± 0.04 µM) was still higher than expected for two independent binding sites ($K_d \approx$ 2-4 µM). Thus, we hypothesized that this must be a cooperative effect due to favorable interactions between the protein molecules. Interestingly, the affinity of an ideal 2xTGAA DNA repeat was similar to that of the naturally occurring TGAA repeat probe from the *CLV1* locus (K_d = 0.27 ± 0.03 µM) (**Fig. 2a** and **Supplementary Fig. 9e**), suggesting the

binding of two WUS-HD molecules per DNA as seen in our TGAA crystal structure (**Fig. 2d**). Indeed, the *CLV1* derived sequence has two TGAA and one TGTA motif (**Fig. 2b**), demonstrating the importance of adenine at the 0 position, crucial for HD binding[23, 31]. To test the relevance of cooperativity for chromatin binding of WUS *in vivo*, we quantified reads of our the ChIP-seq data aligning to sequences containing one, two or three TGAA recognition motifs (**Supplementary Fig. 13a**). Consistent with the increase in binding affinity seen by MST, multiple TGAA repeat motifs were bound by WUS much more frequently compared to individual TGAA sequences, demonstrating that cooperative binding is a relevant mechanism for WUS chromatin interaction *in vivo*.

In order to assess the complex stoichiometry of WUS-HD/DNA complexes we performed MALS analysis with the same DNA probes containing tandem TGAA recognition motifs (**Supplementary Fig. 10b-e**). However, the determined molecular mass (MM_{calc}) for the complex fraction was always approximately 9 kDa lower than the expected theoretical molecular mass (MM_{theo}), if all recognition motifs were occupied. This suggested the absence of one WUS-HD monomer in the final protein-DNA complex and could be due to a dilution effect during gel filtration.

To unambiguously determine the number of binding events and to probe for cooperativity, we used isothermal titration calorimetry (ITC) to quantify the binding thermodynamics of WUS-HD with a 2xTGAA recognition motif (**Supplementary Fig. 12b**). In line with our expectations, binding of the 2xTGAA DNA could be fit best by a sequential binding model, indicating two binding events ($K_{d,1}$ = 1.24 ± 0.14 µM; $K_{d,2}$ = 0.82 ± 0.07 µM) with positive cooperativity and a protein to DNA stoichiometry of 2:1.

WUS-HD dimerization drives cooperative DNA-binding of direct repeat motifs

To mechanistically dissect the positive cooperativity for WUS-HD binding to atypical TGAA repeat sequences, we investigated the protein-protein interactions between the DNA-bound HD dimers observed in our crystal structures in more detail (**Fig. 5a**). Although the interaction surface area between the WUS-HD molecules was very small in all cases, covering only 2-6% (90-290 $Å^2$) of the solvent accessible surface area (SASA), a few hydrophobic residues (I66, F85 and F101) were notably more buried in both the TAAT and TGAA crystal structures (**Fig. 5a**). To functionally test the contribution of the dimerization interface to the DNA-binding activity of WUS, we independently substituted these residues with alanine. MST analysis revealed a reduced DNA-binding affinity of all three WUS-HD variants (**Fig. 5b**), supporting our hypothesis that high affinity DNA-binding to TGAA repeat probes requires WUS homodimerization. Interestingly, only positions I66 and F85, lining the interface of helices α2 and α3 of WUS-HD, are conserved within the WOX family, suggesting that F101 may represent a specific feature of WUS compared to other WOX members (**Fig. 1c**). Consistent with these findings, we observed that the F101A mutation only reduced binding affinity by a factor of about three (K_d = 0.80 ± 0.12 µM) compared to >20 (K_d = 8.20 ± 0.84 µM) and >20 (K_d = 6.81 ± 0.87 µM) for the I66A and the F85A mutations, respectively (**Fig. 5b** and **Supplementary Fig. 9o-q**).

To test whether these substitution alleles indeed modify the dimerization status rather than indirectly reducing DNA binding affinity by more globally affecting WUS-HD structure, we analyzed the interaction between the TGAA direct repeat probe and the mutants by EMSA experiments (**Supplementary Fig. 14**). In accordance with our MST results, all variants exhibited divergent DNA binding behavior compared to wild-type WUS-HD when probed with the 2xTGAA repeat sequence. Consistent with the

observations from the MST analysis, DNA binding of the I66A and F85A variants was largely impaired and no distinct band shifts were visible, suggesting that these conserved residues may play an important role for the overall fold of WUS-HD rather than only mediating dimerization (**Supplementary Fig. 14b,c**). In contrast, the F101A variant still bound DNA with reasonable affinity as observed in MST, but the resulting complex was predominantly monomeric, in comparison to the mostly dimeric form observed with wild-type WUS-HD (**Fig. 5b** and **Supplementary Fig. 14a,d**).

Therefore, these results confirmed that WUS-HD forms a homodimer upon DNA-binding, where the interaction between the monomers is scaffolded by DNA and limited to a few amino acid contacts with an important role for F101. In addition, the newly identified dimerization sites of WUS greatly contribute to the cooperative DNA-binding of tandemly arranged TGAA recognition motifs, such as the ones observed in the TGAA direct repeat or *CLV1* derived TGAA probes (**Fig. 2a,b**).

Since the arrangement of recognition motifs is likely to influence WUS-HD binding affinities for all probes (**Fig. 2a,b**), we further examined how WUS-HD DNA-binding depends on the orientation or spacing of two identical core recognition motifs using the TGAA interaction as a model. Interestingly, changing the relative position of the TGAA core recognition motif from a direct tandem repeat into an inverted (Tail-to-Tail) or everted (Head-to-Head) repeat configuration on opposite strands led to a drastic decrease in binding affinity by about ~10-fold (**Fig. 5c**). The affinity of the Head-to-Head sequence probe (K_d = 3.17 ± 0.35 µM) was similar to that of the naturally occurring G-Box probe (K_d = 3.78 ± 0.42 µM) from the *CLV1* locus, consistent with the observation that this sequence was bound with a lower affinity compared to the TGAA probe (**Fig. 2a,b**). This reduction was also observed when changing the orientation from a Head-to-Head arrangement to a Tail-to-Tail arrangement (K_d = 1.90 ± 0.20 µM)

(**Fig. 5c** and **Supplementary Fig. 9h,j**). To test whether these observations are relevant for WUS chromatin binding *in vivo*, we again mined our ChIP-seq data. In accordance with the MST results, the binding probabilities showed a clear correlation with the orientation of two TGAA motifs (**Supplementary Fig. 13b**). The direct TGAA repeat motif was bound significantly more often compared to the Head-to-Head and the Tail-to-Tail configuration, which both had similar read distributions, consistent with the binding affinities determined by MST. These results confirmed that high affinity binding of WUS-HD to direct TGAA repeat sequences is dependent on the protein-protein interactions observed in our crystal structure, and that these interactions are highly relevant *in vivo*.

To test this further, we analyzed direct tandem repeat sequences with variable spacing between the TGAA motifs by MST (**Fig. 5d**). In line with our hypothesis that interactions between neighboring WUS-HD molecules are required for high affinity binding, we observed a substantial reduction in DNA-binding affinity when we separated the TGAA motifs. Additional spacing by one nucleotide led to a reduction by ~7-fold (K_d = 2.21 ± 0.25 µM) and ~15-fold (K_d = 0.84 ± 0.12 µM) for 2xTGAA and 3xTGAA respectively (**Fig. 5d** and **Supplementary Fig. 9l,m**). Increasing the spacer length up to four nucleotides (K_d = 3.58 ± 0.44 µM) did not lead to a more pronounced effect. Notably, this effect was not observed when we introduced a spacer between motifs situated on different DNA strands of inverted and everted repeat probes (**Fig. 5c**), a motif arrangement that does not allow protein-protein interactions to begin with. Consistent with the observations from the MST analysis, the ChIP-seq data also showed a reduction in binding probability when two TGAA motifs were separated by an additional nucleotide (**Supplementary Fig. 13c**). Taken together with the cooperativity shown by ITC, these results strongly suggested that stabilizing protein-

protein interactions between WUS-HDs promote high affinity binding to DNA containing direct repeats of tandemly arranged TGAA recognition motifs.

Base specific contacts are crucial for WUS-HD DNA sequence specificity

Characterization of the WUS-HD DNA-binding specificity has shown that WUS-HD prefers atypical TGAA repeat sequences, while typical TAAT elements were bound less efficiently (**Fig. 2b** and **Fig. 3b**). Hence, we wanted to identify the mechanisms responsible for this behavior and test whether we could re-program the DNA-binding preferences of WUS from an atypical TGAA motif to a typical TAAT motif. In atypical HDs (e.g. Exd[30]), an arginine (R94 in WUS) reads out the guanine at position -1 of the DNA recognition sequence. However, in typical HDs this residue is commonly a lysine, which contacts the sugar-phosphate backbone (**Fig. 6a**). In addition, typical HDs (e.g. Antp[34] and En[23]) usually contain one or two positively changed residues at their N-terminal arm that specify an adenine at position -1 of the DNA recognition motif. In WUS these residues are T35 and S36, which were either not visible in the crystal structures of the WUS/DNA complexes or not involved in DNA contacts.

In order to test the contribution of these residues to the DNA binding preferences of WUS-HD, we therefore separately substituted T35 and S36 to arginines to promote the recognition of the typical TAAT motif via the adenine in the -1 position (**Fig. 6a**). In addition, R94 was substituted to lysine to abolish the interaction with guanine at position -1 of the atypical TGAA motif. To probe the DNA sequence specificity of these modified WUS-HD variants, EMSA experiments were performed, which clearly demonstrated a change in the DNA-binding behavior of WUS-HD to TGAA and TAAT sequence probes (**Fig. 6b**). Both the T35R and the S36R variant gave similar band shifts as wild-type (wt) WUS-HD when probed with the TGAA repeat sequence, suggesting no interference with binding of atypical recognition motifs. However, when

probed with the TAAT sequence the band shift pattern for these variants changed compared to the WUS wt and shifted to a predominantly monomeric complex, characteristic for typical HDs binding to the TAAT recognition motif. The R94 variant on the other hand, had no apparent influence on DNA-binding to the TAAT probe compared to WUS wt (**Fig. 6b**). In contrast, binding to the TGAA repeat probe was significantly impaired and no distinct band shifts were visible. A triple WUS-HD RRK mutant (T35R, S36R and R94K) was then analyzed for a synergetic effect on WUS DNA-binding specificity. To our surprise, combining all three mutations resulted in an intermediate binding to the TGAA probe, suggesting that the T35R and S36R mutations had restored the binding ability that had been diminished by the individual R94K mutation (**Fig. 6b**). The binding behavior to the TAAT probe was similar to the T35R and S36R mutants, demonstrating the preference for typical recognition motifs in this configuration.

To quantitatively delineate the observed changes in DNA-binding specificity, we analyzed the interaction between the TGAA and TAAT probes and the individual WUS-HD mutants by MST (**Fig. 6c**). Consistent with the observations from the EMSA experiments, the T35R (K_d = 0.54 ± 0.07 μM) and S36R (K_d = 0.32 ± 0.03 μM) variants had similar binding affinities to the TGAA probe as the wild-type (K_d = 0.27 ± 0.03 μM) (**Supplementary Fig. 9r-t**). Similarly, the ~40-fold decrease in binding affinity for the R94K (K_d = 11.00 ± 1.62 μM) variant agreed very well with the drastically impaired DNA-binding ability seen in EMSA (**Fig. 6b**,**c**). Interestingly, the same mutant enhanced binding to the TAAT probe by a factor of two (K_d = 4.59 ± 0.58 μM) compared to WUS-HD wt, while the T35R and the S36R mutations both increased binding affinity by a factor of six (K_d = 1.67 ± 0.20 μM and K_d = 1.67 ± 0.17 μM respectively) (**Fig. 6c** and **Supplementary Fig. 9u-w**). Collectively, the results from

the EMSA and MST experiments suggested that the high specificity of WUS-HD for the atypical TGAA repeat sequence was strongly dependent on R94, as substituting this residue to lysine significantly compromised DNA-binding affinity. In contrast, the introduction of arginines at the N-terminal arm (T35R and S36R) improved the recognition of TAAT sequences by WUS-HD.

Discussion

The HD transcription factor WUS plays key roles for plant development with at least four distinct and essential subfunctions: Firstly, regulation of stem cell identity in the shoot apical meristem (SAM), where it is required for noncell autonomous induction and maintenance of stem cell fate via stimulation of cytokinin and suppression of auxin signalling[2, 4, 6, 17, 28, 39]. Secondly, regulation of floral patterning through direct regulation of differentiation genes[16]. Thirdly, coordination of cell identity specification during female gametophyte development[40]. And lastly, cell differentiation during anther development[41]. Consistent with a broad functional portfolio, WUS has been shown to bind to a diverse set of DNA motifs in the regulatory regions of downstream genes involved in SAM development and signaling processes[16, 17, 18, 19, 20, 21, 28]. However, current *in vitro* DNA binding data have little predictive value when it comes to understanding WUS activity *in vivo,* since a clear correlation between the occurrence of specific DNA motifs with functional classes of response genes or activation or repression of targets has not emerged so far.

In the present study, we quantified the binding affinity of WUS-HD towards different known DNA sequences and performed structural and biochemical analyses to gain insight into the mechanism underlying its binding diversity. The crystal structure of WUS revealed a canonical HD fold with unique structural features compared to classical HD proteins found in metazoans, including elongated loop regions and a divergent mechanism for positioning the N-terminal arm. While most of these differences also apply on a sequence level to other WOX proteins, WUS is the only member where the specific insertion of a tyrosine residue at position 54 leads to a bulged-out helix or π-helix, typically correlated with functional sites in a protein[24, 25].

WUS uses a general mode of DNA-binding to all sequence motifs and in all cases the DNA is bound as dimer, even the classical TAAT motif usually only bound as monomer by all HD TFs analyzed so far. Interestingly, the DNA bound dimers have unique orientations relative to each other for all motifs and the underlying protein-protein interactions are only partially overlapping. This finding is supported by other studies, which showed that the relative binding orientation of protein dimers can vary when bound to different DNA sequences[42]. The crystal structures of WUS in complex with DNA sequences from the *AG* (TAAT) and *CLV1* (TGAA and G-Box) promotor regions also show that WUS-HD binds both typical and atypical recognition motifs with similar residues, although the amino acid configuration strongly prefers the atypical TGAA motif over the typical TAAT motif. For example, R94 in the recognition helix can establish two hydrogen bonds with a guanine base, which leads to a strong preference for the atypical TGAA repeat sequence observed in EMSA and MST experiments. Further experiments demonstrated that mutagenesis of three key residues, namely T35 and S36 in the N-terminal arm and R94 in the C-terminal recognition helix, could shift WUS-HD sequence specificity from TGAA to TAAT recognition. These findings for the first time provide a mechanistically founded starting point to explore the contribution of the different binding motif interactions for *in vivo* function of WUS.

Members of WUS and TALE subfamily of HD-TFs in plants have been shown to form homo- and heterodimers[18, 43]. Besides a downstream sequence (residues 134-208) important for dimerization[4, 18, 44], it was suggested that the HD also confers dimerization[44]. However, in this study MALS analysis of the isolated WUS-HD from *A. thaliana* clearly demonstrated that this domain is a monomer in solution in contrast to the full-length protein that exhibited dimerization in yeast two-hybrid and FRET assays[4, 18]. Consistently, previous analyses of the DNA-binding behavior of WUS-HD

to TAAT motifs have shown that it binds DNA as monomer at lower concentrations and as dimer at higher levels, suggesting that dimerization is dependent on the presence of DNA and higher protein concentration[21]. These findings are also consistent with the two DNA-bound WUS-HD molecules we observed in the TGAA, G-Box and TAAT crystal structures. While in the preferred arrangement on the TAAT motif we did not observe an interaction between the WUS-HD molecules, our structures and functional assays clearly identified the residues responsible for dimer formation for all other configurations. Interestingly, they do not match the suggestion made by Rodriguez and colleagues, who proposed that the interaction is dependent on G77[44]. Since their experiments relied on yeast two-hybrid assays only, it is likely that a more general conformational change caused by changing G77 to E was responsible for the loss of homotypic interaction, rather than specific disruption of the dimer interface.

The important role of DNA for homodimerization of the isolated WUS-HD was further strengthened by our MST analysis which showed that both arrangement and orientation of specific recognition motifs are critical determinants for high affinity binding of WUS-HD (**Fig. 7**). For example, in the TAAT structure the two HDs make intimate hydrophobic interactions that allow one monomer to occupy an unfavorable binding site, despite the overall binding affinity being in the low µM range. In contrast, high affinity binding of WUS is dependent on appropriately arranged sequence motifs in a direct tandem repeat, as seen in the TGAA structure, with the dimerization interface being governed by hydrophobic interactions involving only a few amino acids. Changing this binding code in the DNA motif results in a drastic decrease of WUS-HD affinity.

In general, dimerization of TFs synergistically enhances the DNA binding affinity relative to the interaction of each monomer alone[45]. Furthermore, dimerization

increases the effective length of the recognized DNA sequence, which helps to discriminate between functional regulatory elements and randomly occurring isolated motifs that will not lead to productive gene regulation. Notably, in the case of the TGAA homodimer, which showed the highest binding affinity, the HDs associate non-symmetrically via the C-terminal end of the recognition helix (α3), thereby establishing additional contacts with the DNA. In contrast, the only other known example of a HD homodimer, the complex of the paired (PAX) HD bound to DNA, shows a symmetric Head-to-Head arrangement[46] (**Supplementary Fig. 15a**). Here, the N-terminal arm of each HD contacts the N-terminal end of helix α2 through complementarity of shape and charge with the underlying residues (**Supplementary Fig. 15b**). The symmetrical contacts between the two HDs lead to cooperative binding of a palindromic DNA sequence composed of two inverted TAAT motifs with high affinity. Further analysis of the DNA-binding activity of WUS-HD by MST and ITC demonstrated a cooperative gain in binding affinity with increasing number of TGAA recognition motifs and indicated a stabilization of the DNA interaction upon dimerization. Therefore, the specific protein-DNA and protein-protein interactions that contribute to the overall docking geometry all contribute to the final sequence recognition specificity of WUS. Importantly, these biochemical properties were mirrored by the binding behavior of native WUS to chromatin in plant cells, suggesting that our findings hold promise to decipher the regulatory potential of WUS *in vivo*.

A recent study analyzed the global effect of cytosine methylation on TF binding[47] and showed that many TFs, particularly the extended HD family, prefer CpG methylated sequences. Structural analysis highlighted that the specificity for methylcytosine depends on direct hydrophobic interactions with the 5-methyl group. Interestingly, analysis of the DNA contacts in the crystal structures of WUS bound to the TGAA and

G-Box probes showed that the WUS-HD establishes hydrophobic contacts in a similar way to the methyl group of thymine. Furthermore, the bases at this position have a significant influence on the local DNA shape, with a narrow DNA minor groove width (MGW) in the unbound state correlating with high affinity DNA-binding of WUS by the insertion of the conserved R38 into the minor groove. These findings are consistent with other examples of HD-DNA interactions showing that local optima in the DNA structure are preferentially recognized[9, 36]. Besides DNA methylation, other epigenetic modifications, such as histone modifications, dynamically control gene expression and thereby play an important role in maintaining cellular identity[48]. In particular, it is generally believed that the methylation of TF binding sites is involved in cell proliferation and differentiation[47]. Emerging lines of evidence also indicate that the aforementioned epigenetic mechanisms play vital roles in meristem maintenance and termination[49].

Another important level of control is post-translational modifications (PTMs) of TFs that can affect protein-protein interactions or modulate DNA-binding affinities. So far more than 200 different types of PTMs have been identified, including acetylation, glycosylation and phosphorylation[50]. Numerous PTMs specifically modulate the interaction characteristics by modifying the electrostatic or structural properties of the protein[51]. Hence, PTM sites often correlate with protein interaction surfaces and frequently regulate key molecular processes. In line with these observations, the π-helix at the C-terminal end of helix α1 represents a potential PTM site for the regulation of WUS activity. The insertion of a tyrosine residue at position 54 is a WUS-specific feature within the WOX family that leads to a distortion at the end of helix α1. As π-helices are energetically disfavored they are usually associated with specific biological functions that are actively maintained during evolution[24, 25]. One possibility could be

that Y54 represents a regulatory switch, where phosphorylation of the hydroxyl-group would alter WUS DNA-binding specificity. A structural model suggests that R94 could be involved in the coordination of the phosphate group of pY54 (**Supplementary Fig. 16**). In addition, H91 and the DNA phosphate-backbone could stabilize this interaction, thereby altering the sequence specificity and the overall DNA-binding affinity.

In summary, the structural and biochemical analyses of the complex DNA-binding mechanism of WUS-HD presented here provide insight into the molecular basis of the underlying sequence recognition specificity of WUS and will help to dissect the complex regulatory network controlling stem cell fate in plant meristems. The crystal structures of WUS bound to different DNA probes determined in this study highlight important mechanistic details of specific sequence recognition. In particular, homodimerization of WUS upon DNA-binding represents one of the key determinants to achieve high affinity binding of specific regulatory elements (**Fig. 7**), despite the broad sequence specificity generally observed for HDs. Furthermore, this cooperative DNA-binding mechanism governed by the association of two HDs likely represents a general mechanism of the plant-specific WOX protein family to recognize direct repeat sequences of the atypical TGAA motif with high specificity. Nevertheless, further experiments are required to uncover the physiological role of WUS homodimerization for DNA-binding *in vivo* and its functional significance for gene regulation in the shoot meristem.

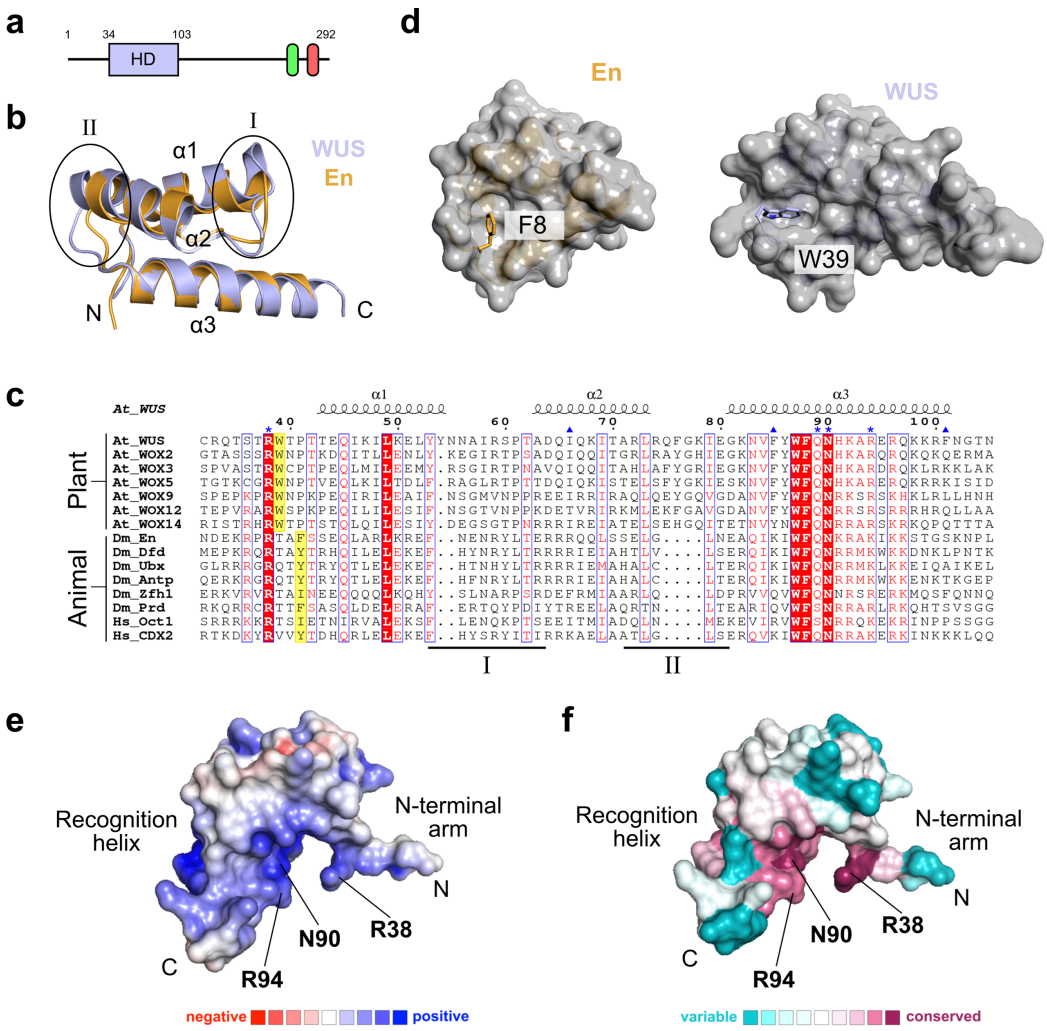


Fig. 1 Structure and conservation of the *A. thaliana* WUS-HD

a, Domain organization of WUS from *A. thaliana.* WUS contains an N-terminal homeodomain (HD; light blue) and two short linear motifs at its C-terminus, namely the WUS-box (green) and the EAR motif (red) respectively. The domain boundaries are given in residue numbers. **b**, Superimposition of WUS-HD structure (light blue) with HD from En (orange, PDB 3HDD[23]), illustrating the characteristic three-helix bundle fold of HDs. Encircled are loop regions I and II of WUS, which highlight structural differences compared to classical HDs. **c**, Multiple sequence alignment of representative HDs from plants and animals. The sequences of *Arabidopsis thaliana* (*At*), *Drosophila melanogaster* (*Dm*) and *Homo sapiens* (*Hs*) were aligned using Clustal Omega and visualized with ESPRIPT. Numbering and secondary structure assignment is according to *A. thaliana* WUS. Loop regions I and II are depicted below the sequences by black lines and the anchoring residue of the N-terminal arm is illustrated by yellow boxes. Highly conserved residues are highlighted (red boxes) and residues making direct DNA-base contact (*) and residues involved in the dimer interface (▲) are indicated. **d**, Comparison of different anchoring mechanisms of the N-terminal arm for En (orange, PDB 3HDD[23]) and WUS

(light blue). Surface representation of the HDs (grey) highlighting the hydrophobic pocket formed by helices α1 and α2. Amino acids responsible for fixing the N-terminal arm are shown as sticks. **e**, Electrostatic surface potential (red: negative, blue: positive, contoured at ± $5k_BT$) of WUS-HD. **f**, ConSurf analysis showing the degree of amino acid conservation (magenta: conserved, cyan: variable) mapped on to the surface of WUS-HD. Highly conserved amino acids typically involved in DNA base interactions are indicated.

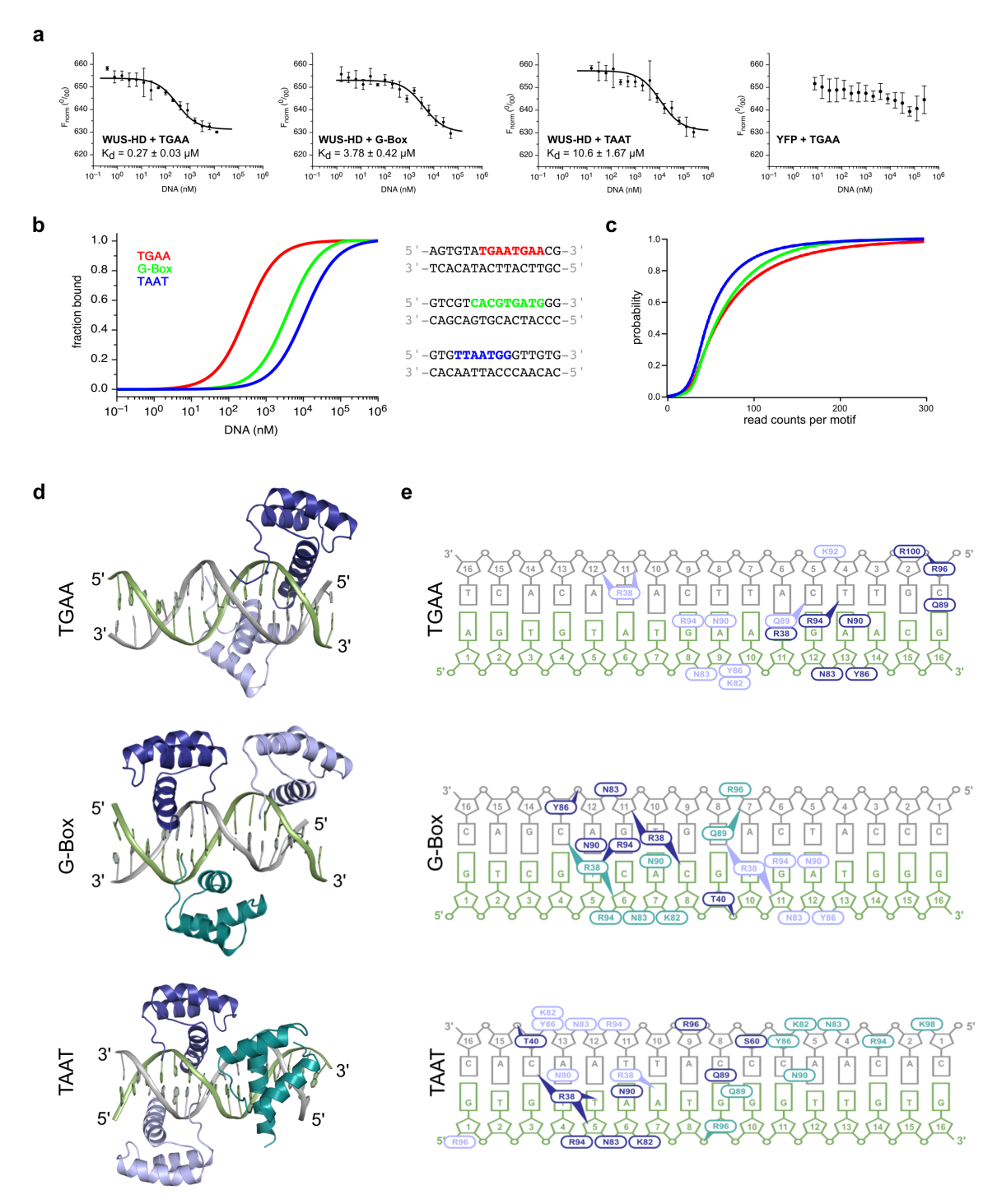


Fig. 2 Characterization of WUS-HD DNA-binding behavior *in vitro*, *in vivo* and by crystallography

a, DNA-binding affinity of YFP-WUS determined by MST for TGAA, G-Box and TAAT DNA. All measurements were performed in triplicates and the respective dissociation constant (K_d) is indicated. A control MST reaction was performed with YFP alone in the presence of TGAA DNA. **b**, Comparison of the fraction bound for WUS-HD binding to different sequences of a 16-

bp DNA fragment. The respective DNA recognition motifs are highlighted for TGAA (red), G-Box (green) and TAAT (blue). **c**, Analysis of WUS chromatin binding *in vivo* by ChIP-seq. Binding probabilities of WUS to TGAATGAA (red), TCACGTGA (G-Box, green) and TTAATGG (blue) containing chromatin regions. Curves shifted to the right indicate a higher probability of a given sequence element to be associated with chromatin of high WUS occupancy and hence high affinity binding. **d**, Overall structures of WUS-DNA complexes showing the mode of binding of two WUS-HD molecules per DNA for TGAA (top) and three WUS-HD molecules per DNA for G-Box (center) and TAAT (bottom). WUS-HDs are in teal, dark and light blue and DNA-strands are in grey and green. **e**, Schematic representation of DNA contact sites of WUS-HD for TGAA (top), G-Box (center) and TAAT (bottom). Ovals indicate amino acids that mainly contact a DNA base or a sugar-phosphate backbone moiety, and ovals with arrowheads specify amino acids that make multiple contacts with DNA bases and/or the sugar-phosphate backbone. Numbering of DNA bases is arbitrary starting from position 1 at each 5'-end.

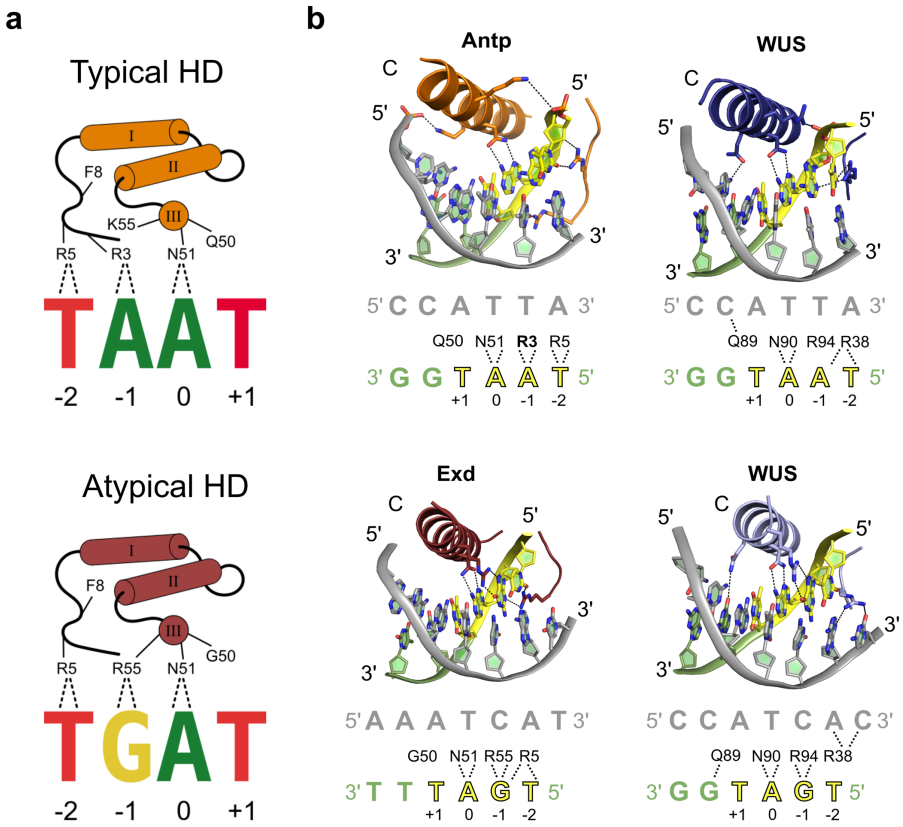


Fig. 3 Comparison of sequence specificity with typical and atypical HDs

a, Schematic representation of DNA sequence specificity for typical and atypical HDs. The DNA base-recognition details are depicted for a typical HD in orange (top) and an atypical HD in red (bottom). Note the specific read-out of the adenine at position 0 by N51, characteristic for HD proteins. The numbering of residues is according to the Antp-HD[34] **b**, DNA base-

recognition details of WUS compared to other HDs. Top, showing hydrogen bond interactions of Antp (orange, PDB 4XID[34]) and WUS with the same typical core DNA recognition motif (yellow). Bottom, showing hydrogen bond interactions of Exd (red, PDB 2R5Y[30]) and WUS with the same atypical core DNA recognition motif. Diagrams below each cartoon representation summarize DNA-base contacts made by each HD.

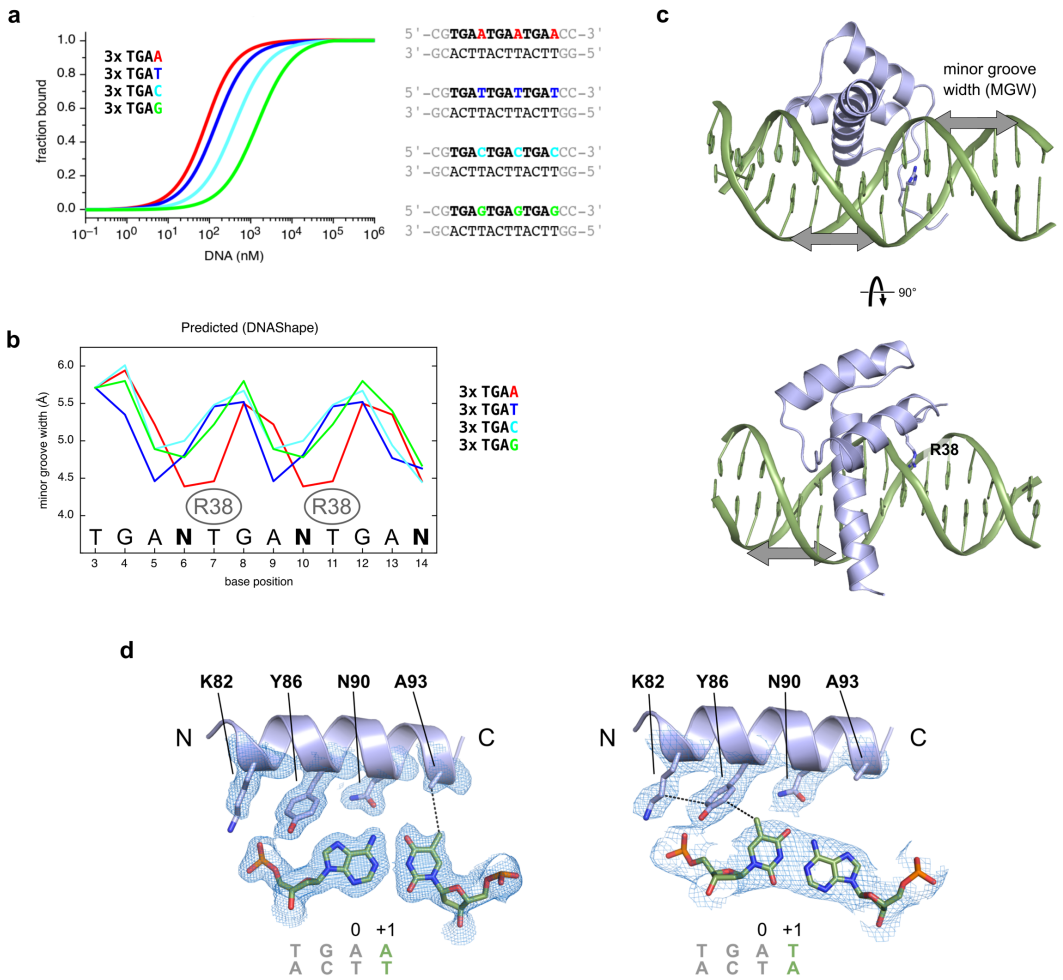


Fig. 4 Molecular basis for preference of the +1 recognition motif position

a, MST-analysis of WUS sequence recognition specificity for DNA position +1, based on an atypical recognition motif. The tandem recognition motif is in bold letters and position +1 is adenine (red), thymine (blue), cytosine (cyan) and guanine (green). **b**, Predicted minor groove width (MGW) profiles for atypical DNA sequences differing only in the +1 position of the WUS-HD recognition sequence. The color scheme is the same as in a) and the binding position of WUS R38 inserting into the minor groove is indicated. The DNA sequence is shown at the bottom, where N represents any of the four nucleotides (A,T,C,G). **c**, Structure of WUS (light blue) bound to DNA (green) highlighting the insertion of R38 into the minor groove. The minor groove width (MGW) is indicated by gray arrows. **d**, Structural basis of adenine (TGAA, left)

and thymine (G-Box, right) preference at the +1 position. Shown are the residues of the WUS recognition helix making hydrophobic contacts with the C5 methyl group of thymine. The conserved Asn90 residue is shown as a reference and the 2Fo-Fc electron density maps (blue mesh) are contoured at 1.0 σ.

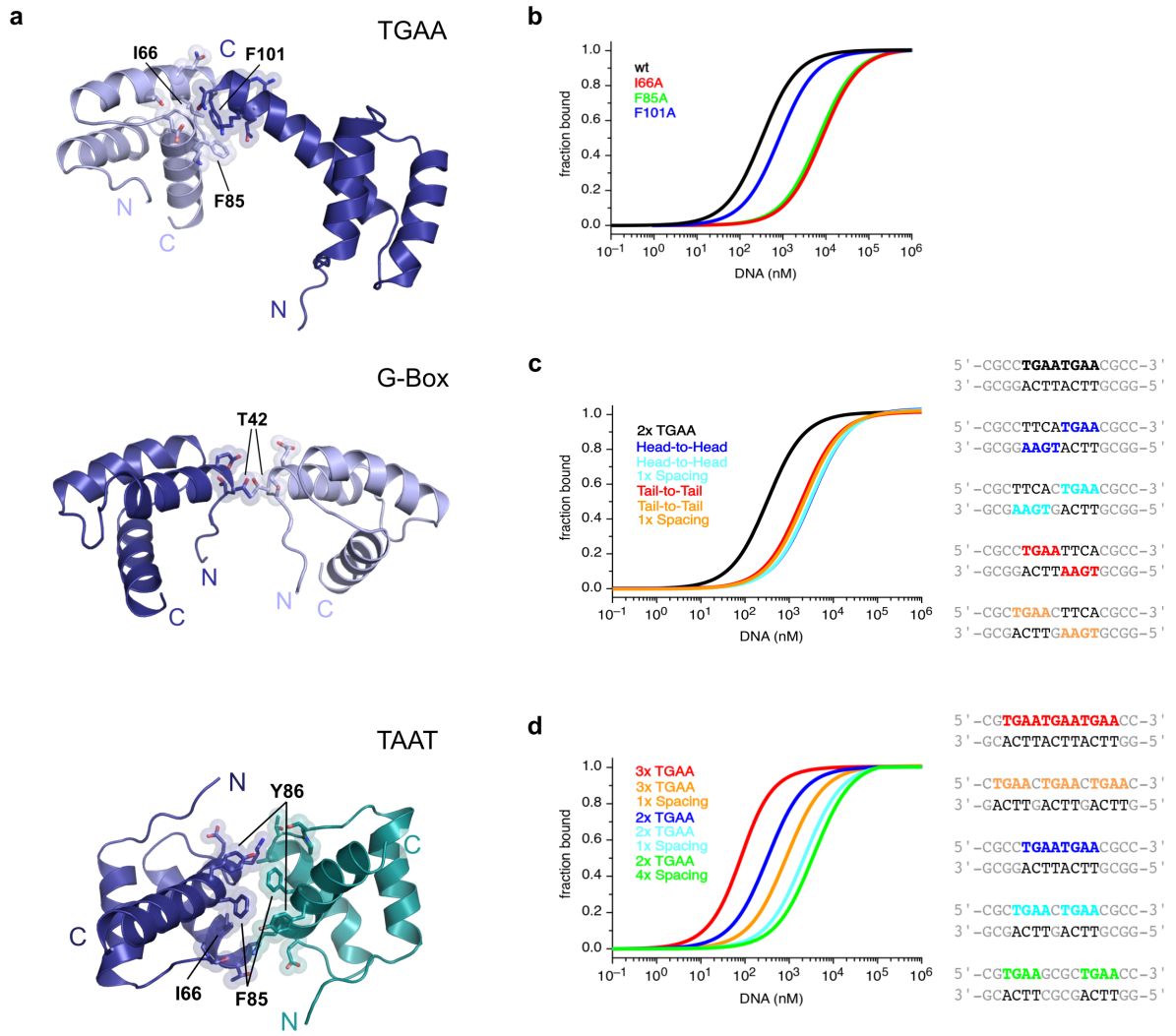


Fig. 5 DNA sequence specificity depends on WUS-HD dimerization

a, DNA-facilitated protein interactions between individual HDs of WUS bound to TGAA (top), G-Box (center) and TAAT (bottom). For clarity DNA was omitted and side chains mediating the protein-protein interface are shown. Amino acids contributing most to the buried surface area are indicated and colors schemes are related to Fig. 2. **b**, MST-analysis of the DNA-facilitated WUS dimerization interface. Single point mutants were introduced (I66A in red, F85A in green and F101A in blue) and binding was quantified for a 2xTGAA motif. **c**, MST-analysis of orientation preferences for WUS binding towards different arrangements of a 2xTGAA DNA motif. The Head-to-Head arrangement is in blue and the Tail-to-Tail arrangement is in red, whilst the same arrangements with a 1 bp spacer are shown in cyan and orange respectively.

As a reference, the binding towards the tandem repeat 2xTGAA DNA motif is shown in black. **d**, MST-analysis of spacing preferences for WUS DNA-binding activity. Binding affinity was measured for three TGAA recognition motifs (red) and with a 4 bp spacer (orange), and for two TGAA recognition motifs (blue), with a 1 bp spacer (cyan) and a 4 bp spacer (green).

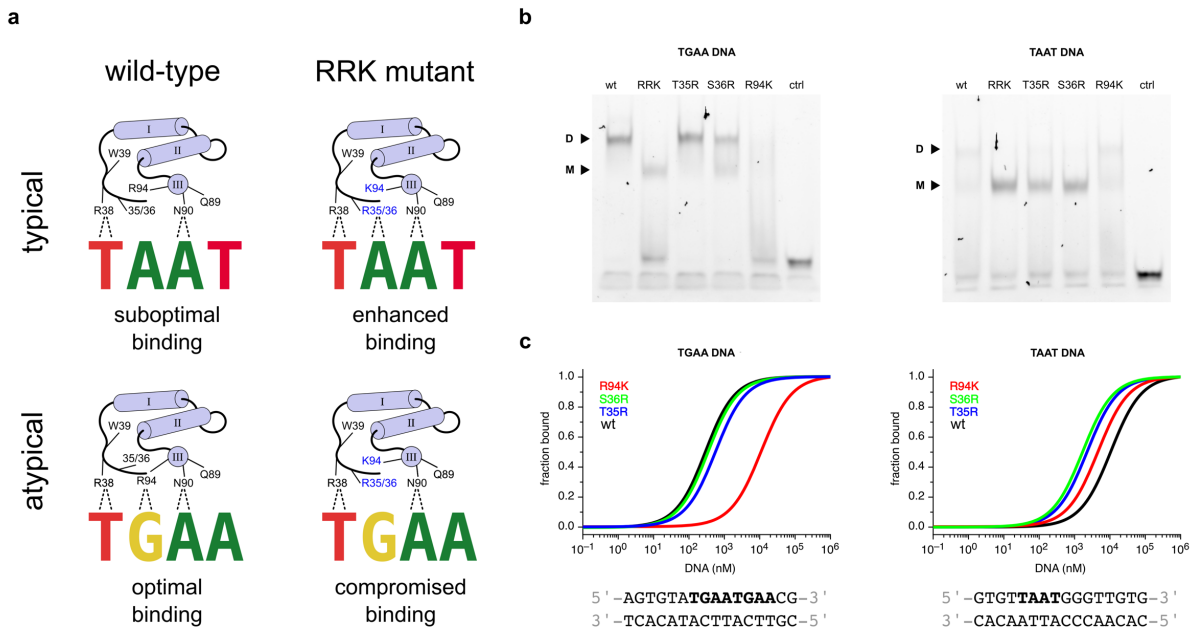


Fig. 6 Altered DNA specificity of the WUS-HD

a, Schematic representation showing the sequence specificity of WUS-HD wild-type (left) and the RRK mutant (right) for typical TAAT (top panel) and atypical TGAA recognition motifs (bottom panel). Hydrogen bond interactions involved in base-recognition are highlighted (dashed lines) and relevant residues are indicated. Altered residues in the RRK mutant are shown in blue. **b**, Electrophoretic mobility shift assays (EMSAs) of altered DNA specificity of WUS-HD for TGAA (left) and TAAT (right) probes. Monomer (M) and dimer (D) bound forms of WUS are indicated and the tested construct is given on top of the gel. **c**, Quantification of WUS-HD DNA-binding affinity for TGAA (left) and TAAT (right) DNA by MST for the single point mutants T35R (blue), S36R (green) and R94K (red). As a reference the DNA-binding affinity of WUS-HD wt is shown in black.

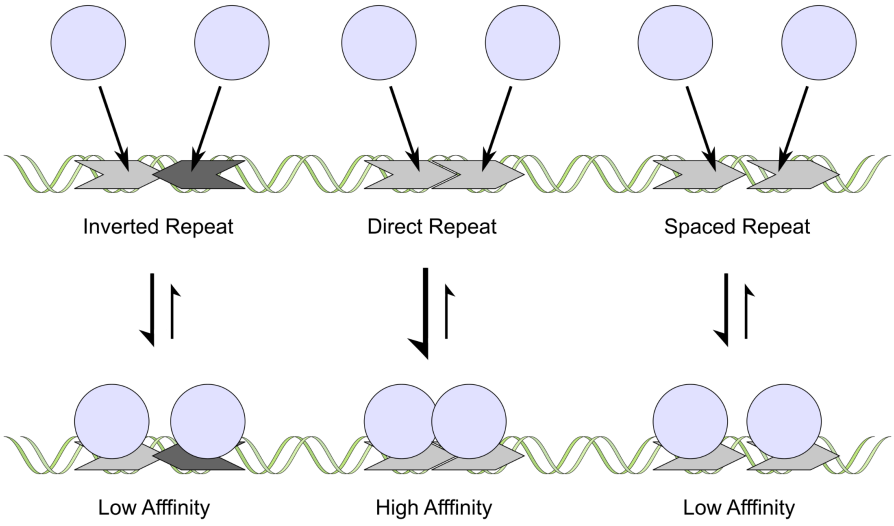
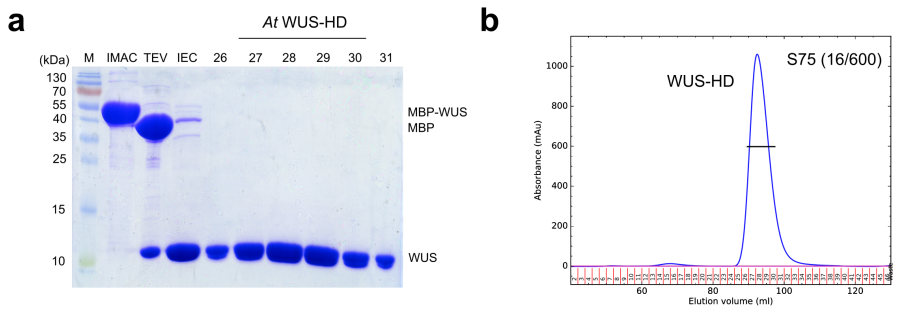


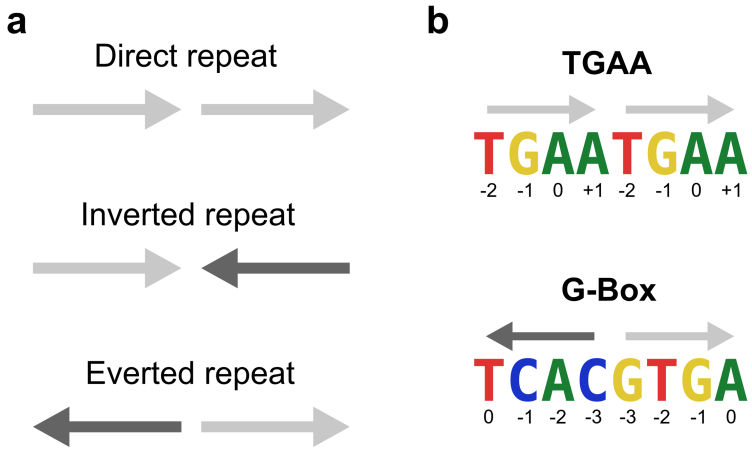
Fig. 7 DNA-binding behavior of WUS-HD

Schematic representation summarizing the complex DNA-binding behavior of the *At* WUS-HD to tandemly arranged recognition motifs. The WUS-HD (light blue) binds to specific DNA recognition sequences (grey), which can differ in their orientation and arrangement relative to each other. In the case of direct repeat sequences DNA-binding of the HD facilitates WUS-homodimerization via specific protein-protein interactions, leading to a cooperative gain in binding affinity. In contrast, inverted tandem repeats or direct repeats separated by a spacer do not allow protein-protein interactions to begin with and therefore are recognized with much lower affinity.



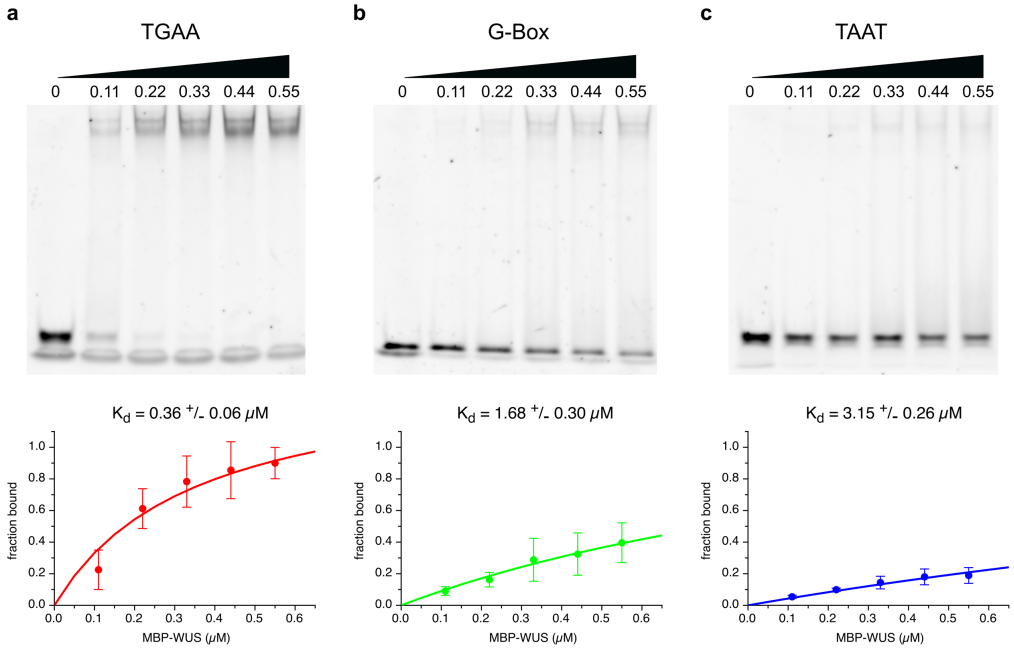
Supplementary Fig. 1 Purification and characterization of the WUS-HD

a, SDS-PAGE analysis of *At* WUS-HD purification after immobilized metal ion affinity chromatography (IMAC), TEV cleavage (TEV), ion exchange chromatography (IEC) and size-exclusion chromatography (SEC). The numbers on top of the panel correspond to peak fractions taken after gel filtration. **b,** SEC elution profile for WUS-HD. The blue line corresponds to absorbance readings at 280 nm and numbers represent collected fractions during SEC.



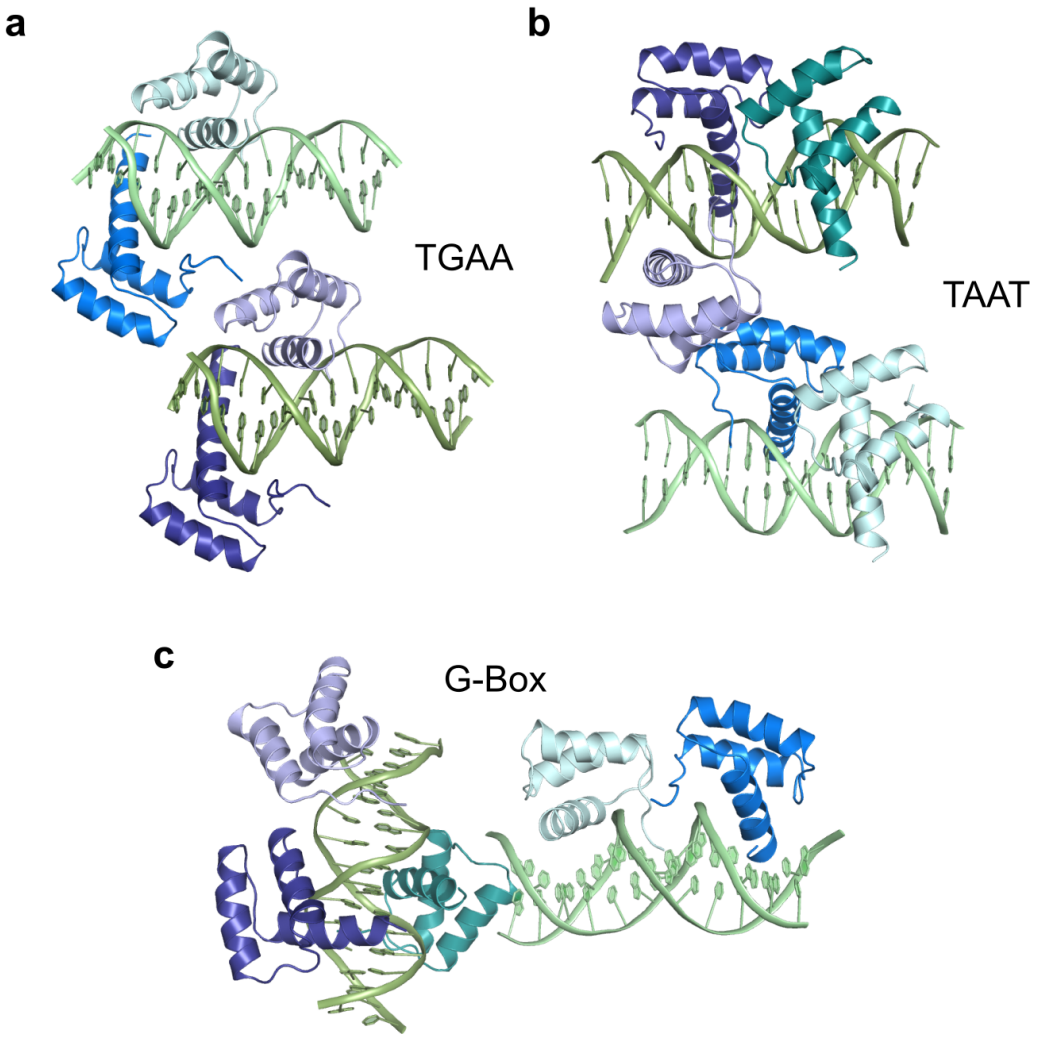
Supplementary Fig. 2 Orientation specificity of WUS DNA binding sites

a, Two DNA recognition elements can be oriented in three different configurations: direct (top), inverted (center) and everted (bottom) repeats. **b**, Direct repeat arrangement of the TGAA sequence (top) compared to the everted repeats in the G-box sequence (bottom). The recognition motif position is indicated below the sequences.



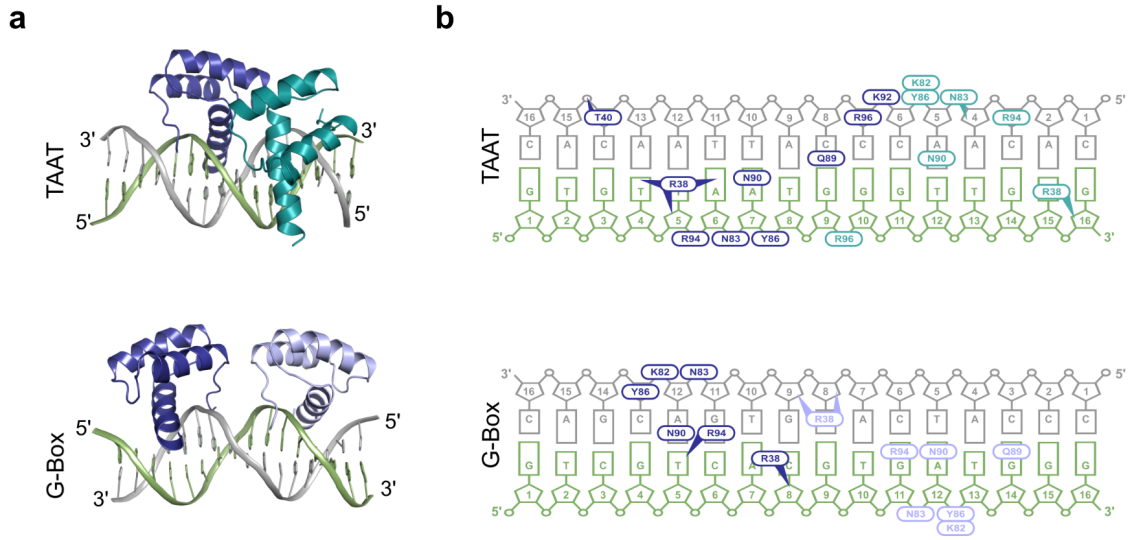
Supplementary Fig. 3 EMSAs probing the DNA-binding preferences of WUS-FL

EMSAs (top panel) and WUS-DNA saturation curves (lower panel) were performed using concentration series of WUS-FL probed with uniform levels of DNA containing TGAA (**a,** red), G-Box (**b,** green) and TAAT (**c,** blue). Protein concentrations (μM) are given on top of each gel and the calculated binding affinities are indicated below.



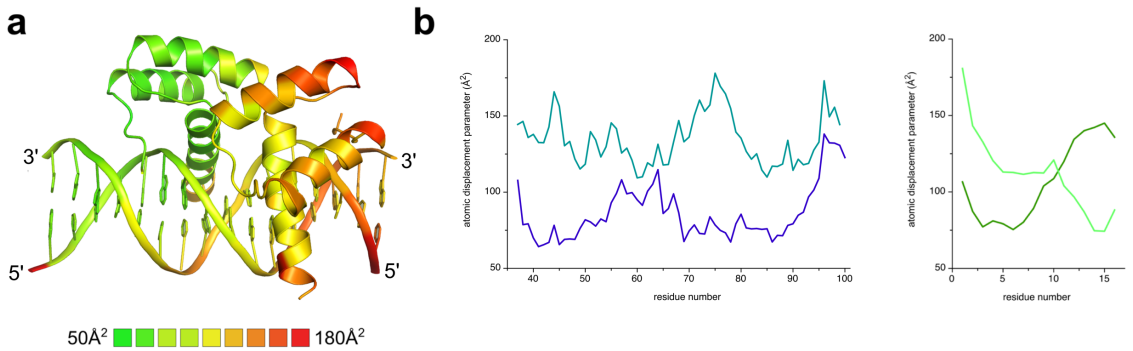
Supplementary Fig. 4 Crystal packing of the TGAA, TAAT and G-Box structures

a, The asymmetric unit of the TGAA structure contains two complexes consisting of two DNA molecules and four WUS-HD molecules. **b**, The asymmetric unit of the TAAT structure contains two complexes consisting of two DNA molecules and five WUS-HD molecules. **c**, The asymmetric unit of the G-Box structure contains two complexes consisting of two DNA molecules and five WUS-HD molecules. Colors schemes are related to Fig. 2, with DNA in shades of green and WUS-HD in shades of blue, and one complex is shown in lighter colors.



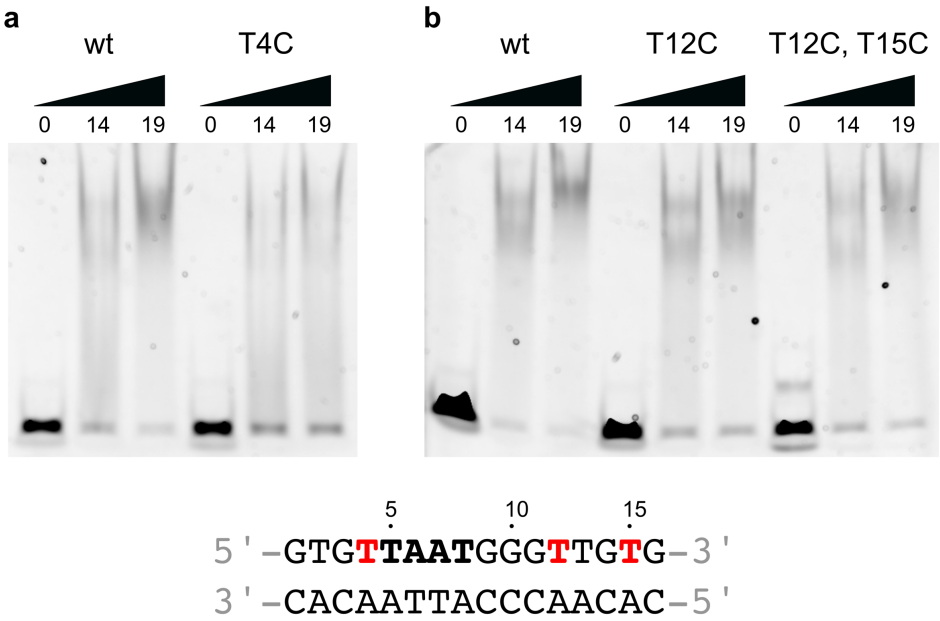
Supplementary Fig. 5 Additional HD binding in the TAAT and G-Box structures

a, Overall structures of WUS-DNA complexes showing two WUS molecules per DNA for TAAT (top) and G-Box (bottom). Colors of protein/DNA are the same as in Fig. 2. **b**, Summary of WUS-DNA interactions including the additional HD molecule. Order and colors are the same as in A and numbering and symbols are according to Fig. 2.



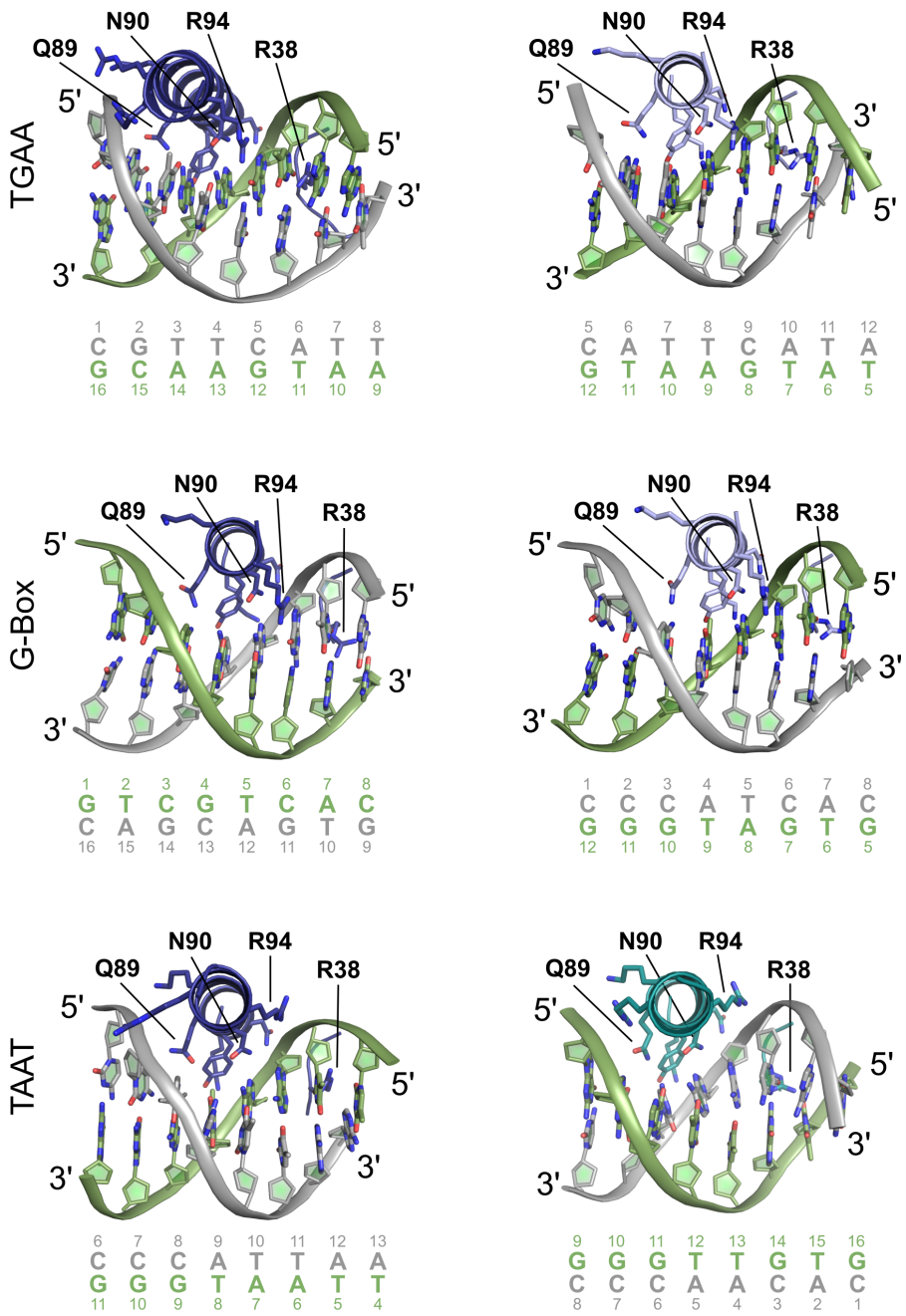
Supplementary Fig. 6 B-factor plot for the WUS-TAAT structure

a, Structure of WUS-HD bound to TAAT DNA colored according to residue average B-factor, scaled from 50 (green) to 180 (red) Å^2. **b**, Atomic displacement parameters plotted against the residue number of WUS-HD (left) and the TAAT DNA probe (right). Colors and numbering are the same as in Fig. 2.



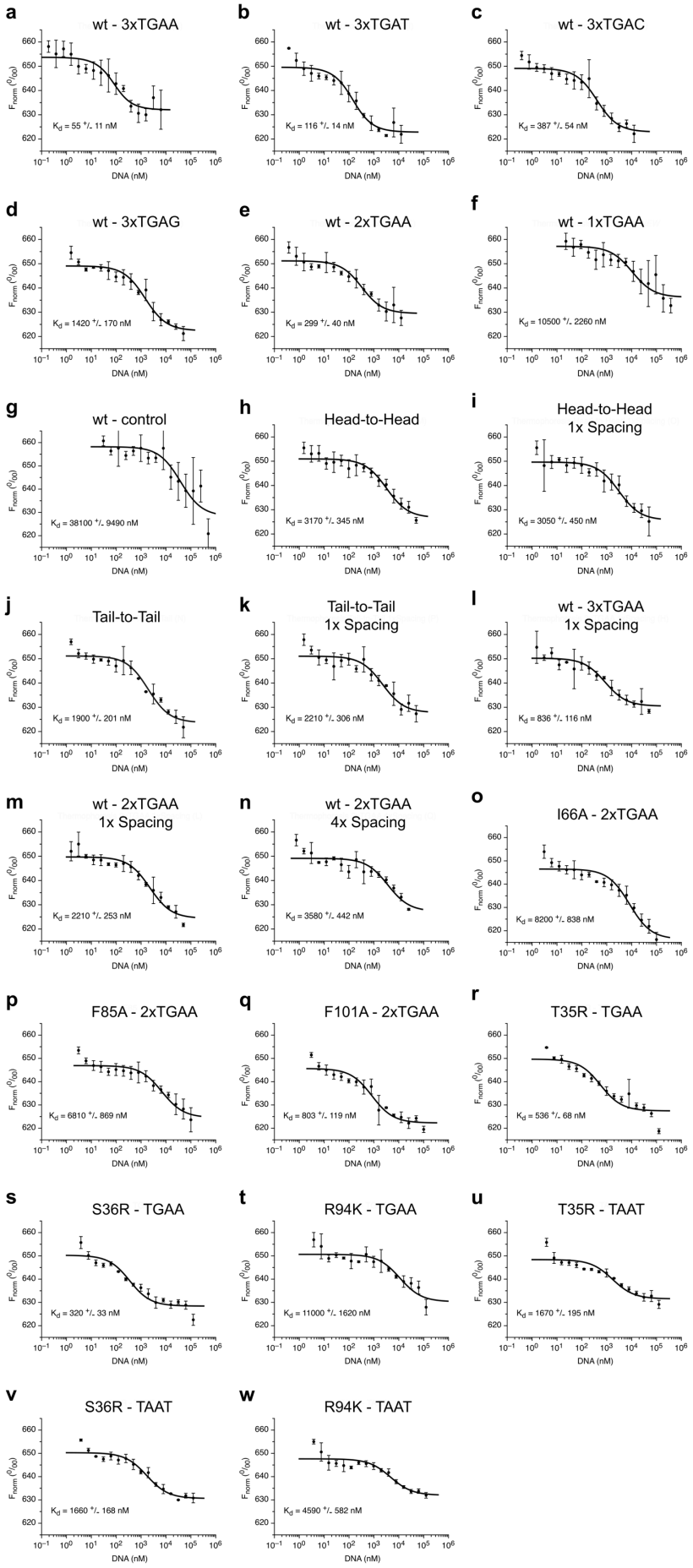
Supplementary Fig. 7 EMSAs probing WUS-HD configuration on the TAAT probe

Altered DNA-binding behavior of WUS-HD to a TAAT motif with specific mutations in flanking regions associated with additional protein-DNA contacts. The respective point mutations T4C (**a**), T12C and T15C (**b**) are highlighted in the sequence (red) at the bottom and the protein concentration (µM) is indicated at the top.



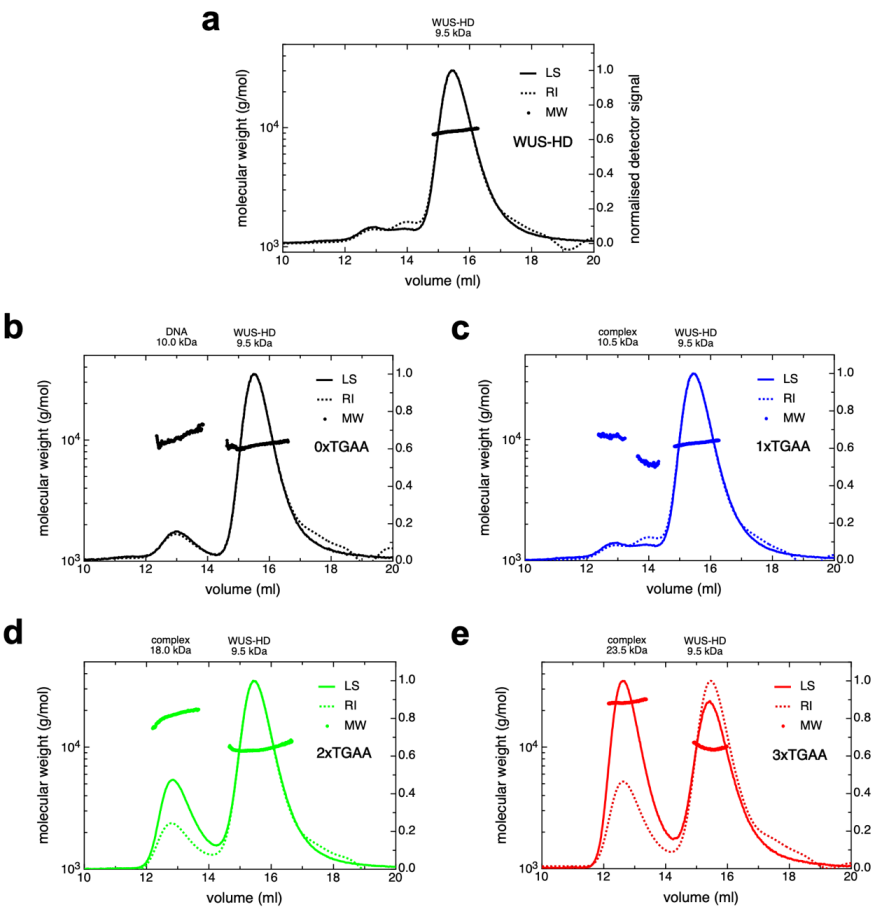
Supplementary Fig. 8 Detailed DNA interactions of WUS-HD

Detailed comparison of the DNA-contacting residues of each WUS-HD bound to TGAA (top), G-Box (center) and TAAT (bottom). Only the helix α3 and the N-terminal arm of WUS-HD involved in major-groove and minor-groove interactions, respectively, are shown for clarity. Colors and numbering of DNA bases are the same as in Fig. 2.


a
wt - 3xTGAA
Kd = 55 +/- 11 nM
b
wt - 3xTGAT
Kd = 116 +/- 14 nM
c
wt - 3xTGAC
Kd = 387 +/- 54 nM
d
wt - 3xTGAG
Kd = 1420 +/- 170 nM
e
wt - 2xTGAA
Kd = 299 +/- 40 nM
f
wt - 1xTGAA
Kd = 10500 +/- 2260 nM
g
wt - control
Kd = 38100 +/- 9490 nM
h
Head-to-Head
Kd = 3170 +/- 345 nM
i
Head-to-Head
1x Spacing
Kd = 3050 +/- 450 nM
j
Tail-to-Tail
Kd = 1900 +/- 201 nM
k
Tail-to-Tail
1x Spacing
Kd = 2210 +/- 306 nM
l
wt - 3xTGAA
1x Spacing
Kd = 836 +/- 116 nM
m
wt - 2xTGAA
1x Spacing
Kd = 2210 +/- 253 nM
n
wt - 2xTGAA
4x Spacing
Kd = 3580 +/- 442 nM
o
I66A - 2xTGAA
Kd = 8200 +/- 838 nM
p
F85A - 2xTGAA
Kd = 6810 +/- 869 nM
q
F101A - 2xTGAA
Kd = 803 +/- 119 nM
r
T35R - TGAA
Kd = 536 +/- 68 nM
s
S36R - TGAA
Kd = 320 +/- 33 nM
t
R94K - TGAA
Kd = 11000 +/- 1620 nM
u
T35R - TAAT
Kd = 1670 +/- 195 nM
v
S36R - TAAT
Kd = 1660 +/- 168 nM
w
R94K - TAAT
Kd = 4590 +/- 582 nM
Fnorm (‰)
DNA (nM)

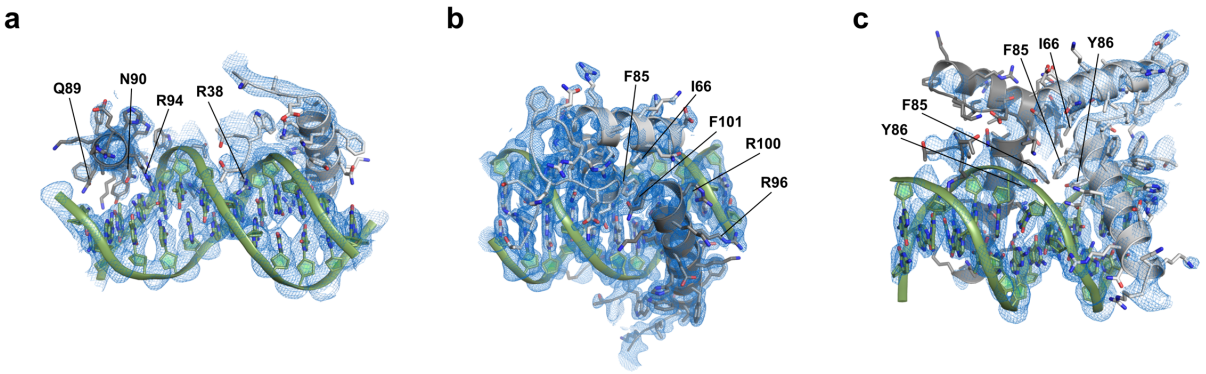
Supplementary Fig. 9 Summary of MST measurements

DNA-binding affinity of YFP-WUS wild-type determined by MST for (**a**) 3xTGAA, (**b**) 3xTGAT, (**c**) 3xTGAC, (**d**) 3xTGAG, (**e**) 2xTGAA, (**f**) 1xTGAA, (**g**) control without TGAA, (**h**) Head-to-Head arrangement of TGAA, (**i**) Head-to-Head arrangement of TGAA with 1x nucleotide spacer, (**j**) Tail-to-Tail arrangement of TGAA, (**k**) Tail-to-Tail arrangement of TGAA with 1x nucleotide spacer, (**l**) 3xTGAA with 1x nucleotide spacer, (**m**) 2xTGAA with 1x nucleotide spacer and (**n**) 2xTGAA with 4x nucleotide spacer. Binding affinity for 2xTGAA DNA of YFP-WUS with (**o**) I66A, (**p**) F85A and (**q**) F101A mutation. Binding affinity for TGAA DNA of YFP-WUS with (**r**) T35R, (**s**) S36R and (**t**) R94K mutation. Binding affinity for TAAT DNA of YFP-WUS with (**u**) T35R, (**v**) S36R and (**w**) R94K mutation.



Supplementary Fig. 10 MALS analysis of the WUS-DNA complex

SEC-MALS analysis of WUS-HD (**a**) and in the presence of control DNA (**b**), 1xTGAA (**c**), 2xTGAA (**d**) and 3xTGAA (**e**). The colored lines correspond to light scattering (LS) readings and the dashed lines correspond to the refractive index (RI) readings. Molar mass (MW) distributions across the peaks are shown as dots. The average calculated molecular weight for each peak is given on top of each panel.

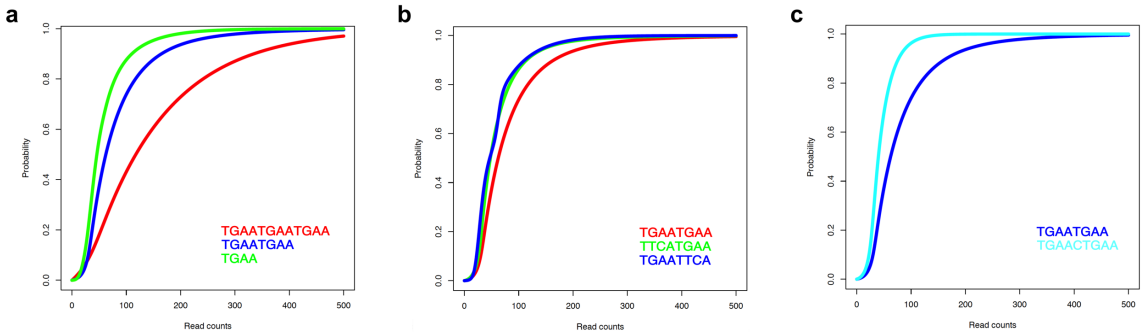


Supplementary Fig. 11 Electron density map of selected regions of WUS-DNA interactions

Close-up view of the WUS-DNA interaction highlighting DNA base-recognition in the G-Box complex (**a**) and showing the WUS dimerization interface in the TGAA (**b**) and TAAT complex (**c**). DNA is shown in green and WUS-HDs are depicted in white and gray. Relevant residues of WUS are indicated and the 2Fo-Fc electron density maps (blue mesh) are contoured at 1.0 σ.

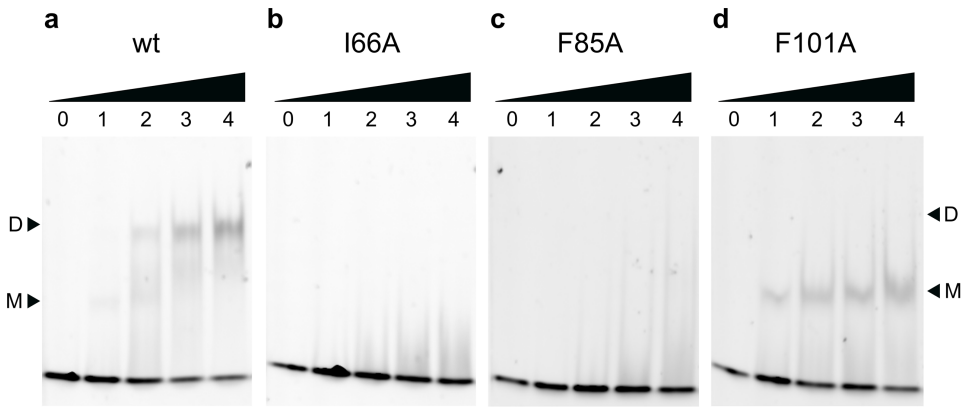
Supplementary Fig. 12 Binding stoichiometry of the WUS-DNA interaction

a, MST-analysis of WUS DNA-binding cooperativity. The number of recognition motifs is indicated: 3xTGAA (red), 2xTGAA (green), 1xTGAA (blue) and a control without TGAA (black). **b**, Quantification of the interaction between WUS-HD and 2xTGAA DNA by isothermal titration calorimetry (ITC).



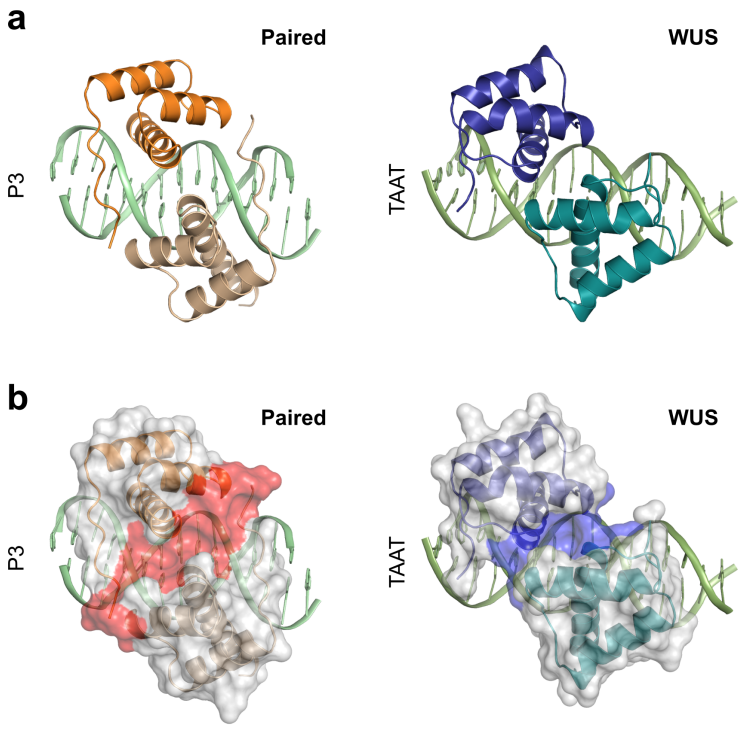
Supplementary Fig. 13 Analysis of WUS chromatin binding *in vivo* by ChIP-seq

a, Binding probabilities of WUS for 3xTGAA (red), 2xTGAA (blue) and 1xTGAA (green). **b**, Binding probabilities of WUS for 2xTGAA (red), Head-to-Head (green) and Tail-to-Tail (blue). **c**, Binding probabilities of WUS for 2xTGAA (blue) and 2xTGAA with 1xSpacing (light blue). Curves shifted to the right indicate a higher probability of a given sequence element to be associated with chromatin exhibiting high WUS occupancy and hence high affinity binding.



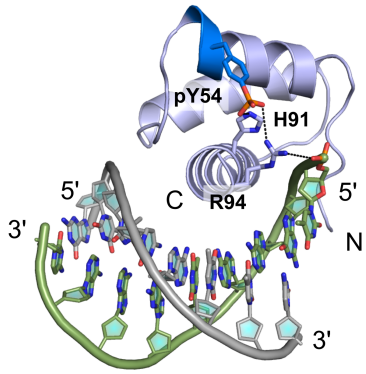
Supplementary Fig. 14 EMSAs probing the WUS dimerization interface.

Altered DNA-binding behavior to a 2xTGAA motif of WUS-HD wt (**a**), I66A (**b**), F85A (**c**) and F101 (**d**). Monomer (M) and dimer (D) bound forms of WUS are indicated and the protein concentration (µM) is given at the top.



Supplementary Fig. 15 Comparison of the HD dimer between Paired and WUS

a, Side-by-side comparison of the HD homodimers of Paired bound to P3 DNA sequence (left, PDB 1FJL[45]) and WUS bound to the TAAT DNA sequence (right). **b**, Same view as before now with surface representation (grey), highlighting the interaction interface formed between the two HD molecules of Paired (red) and WUS (blue).



Supplementary Fig. 16 Structural model of Y54 phosphorylation

Model of WUS-DNA complex with phosphorylated tyrosine (pY54) highlighting the hydrogen bond network of the phosphate group of pY54 from the π-helix (blue). Relevant residues of WUS involved in the interaction are indicated. WUS-HD is depicted in light blue and DNA-strands are shown in grey and green.

Table 1 Data collection and refinement statistics

	WUS-HD	WUS-HD + G-Box DNA	WUS-HD + TGAA DNA	WUS-HD + TAAT DNA
Data collection				
Space group	P 4_12_12	C 121	P 12_11	P 12_11
Cell dimensions				
a, *b*, *c* (Å)	43.12, 43.12, 82.01	123.43, 83.07, 85.65	55.74, 54.34, 75.56	82.66, 45.91, 87.54
α, β, γ (°)	90, 90, 90	90, 112.95, 90	90, 107.54, 90	90, 102.88, 90
Resolution (Å)	38.17 – 1.37 (1.42 – 1.37)	40.67 – 2.69 (2.79 – 2.69)	43.38 – 1.58 (1.63 – 1.58)	40.43 – 2.63 (2.72 – 2.63)
R_{pim}	0.008 (0.030)	0.052 (0.972)	0.026 (0.525)	0.038 (0.955)
$I / \sigma I$	61.1 (19.9)	11.0 (0.9)	12.1 (1.5)	11.6 (1.0)
Completeness (%)	99.9 (99.0)	98.1 (92.4)	98.8 (99.1)	99.4 (98.9)
Multiplicity	24.5 (16.5)	4.8 (4.7)	3.4 (2.6)	4.5 (4.6)
$CC_{1/2}$	0.999 (0.997)	0.998 (0.458)	0.999 (0.778)	0.998 (0.552)
Refinement				
Resolution (Å)	38.17 – 1.37	40.67 – 2.69	43.38 – 1.58	40.43 – 2.63
No. reflections	16816 (1628)	21855 (2051)	59049 (5890)	19403 (1907)
R_{work} / R_{free} (%)	18.3/21.1	19.5/25.1	18.7/22.7	23.7/26.3
No. atoms	655	3815	3810	4053
Protein	517	3815	3501	4053
Ligand/ion	14	-	1	-
Water	64	-	308	-
B-factors	22.13	94.48	38.68	105.49
Protein/DNA	20.03	94.48	38.41	105.49
Ligand/ion	52.19	-	57.99	-
Water	34.52	-	41.66	-
R.m.s. deviations				
Bond lengths (Å)	0.005	0.010	0.006	0.004
Bond angles (°)	0.71	1.11	0.84	0.74

Values in parentheses are for highest-resolution shell.

Supplementary Table 1. Oligonucleotides used for cloning and mutagenesis

Oligomer name	Oligomer sequence
Fwd_*At* WUS HD	GCT TCC ATG GGT CAG ACC AGC ACG AGG TGG AC
Rev_*At* WUS HD	GCT TCT CGA GTT AGG ATC CTC CGT TGA ATC TCT TCT TCT GAC G
Fwd_*At* WUS	ATT CGG ATC CAT GGA GCC GCC ACA GCA T
Rev_At WUS	TGT GCT CGA GCT AGT TCA GAC GTA GCT C
Fwd_*At* WUS HD I66A	GAT CAG GCG CAG AAG ATC ACT GCA AGG CTG

Rev_*At* WUS HD I66A	CTT CTG CGC CTG ATC GGC TGT TGG TGA CC
Fwd_*At* WUS HD F85A	GAA CGT CGC GTA CTG GTT CCA GAA CCA TAA GGC
Rev_*At* WUS HD F85A	CAG TAC GCG ACG TTC TTG CCC TCA ATC TTT C
Fwd_*At* WUS HD F101A	GAA GAG AGC GAA CGG AGG ATC CTA ACT CGA GTG
Rev_*At* WUS HD F101A	CCG TTC GCT CTC TTC TTC TGA CGC TCA CGA G
Fwd_*At* WUS HD T35R	TGT CGC CAG AGG AGC ACG AGG TGG A
Rev_*At* WUS HD T35R	TCC ACC TCG TGC TCC TCT GGC GAC A
Fwd_*At* WUS HD S36R	TGT CGC CAG ACC AGG ACG AGG TGG A
Rev_*At* WUS HD S36R	TCC ACC TCG TCC TGG TCT GGC GAC A
Fwd_*At* WUS HD R94K	CCA TAA GGC TAA GGA GCG TCA GA
Rev_*At* WUS HD R94K	TCT GAC GCT CCT TAG CCT TAT GG
Fwd_*At* WUS HD RRK	TGT CGC CAG AGG AGG ACG AGG TGG A
Rev_*At* WUS HD RRK	TCC ACC TCG TCC TCC TCT GGC GAC A

Supplementary Table 2. Oligonucleotides used for crystallization

Oligomer name	Oligomer sequence
TGAA fwd	AGTGTATGAATGAACG
TGAA rev	CGTTCATTCATACACT
G-Box fwd	GTCGTCACGTGATGGG
G-Box rev	CCCATCACGTGACGAC
TAAT fwd	AATGTGTTAATGGGTT
TAAT rev	AACCCATTAACACATT

Supplementary Table 3. Oligonucleotides used for MST

Oligomer name	Oligomer sequence
3xTGAA fwd	CGTGAATGAATGAACC
3xTGAA rev	GGTTCATTCATTCACG
2xTGAA fwd	CGCCTGAATGAACGCC
2xTGAA rev	GGCGTTCATTCAGGCG
1xTGAA fwd	CGCCGGTGAAGCCGCC
1xTGAA rev	GGCGGCTTCACCGGCG

Negative Control fwd	CGCCGGCGGCGCCGCC
Negative Control rev	GGCGGCGCCGCCGGCG
3xTGAT fwd	CGTGATTGATTGATCC
3xTGAT rev	GGATCAATCAATCACG
3xTGAC fwd	CGTGACTGACTGACCC
3xTGAC rev	GGGTCAGTCAGTCACG
3xTGAG fwd	CGTGAGTGAGTGAGCC
3xTGAG rev	GGCTCACTCACTCACG
TGAA Head-to-Head fwd	CGCCTTCATGAACGCC
TGAA Head-to-Head rev	GGCGTTCATGAAGGCG
TGAA Head-to-Head 1xSpacing fwd	CGCTTCACTGAACGCC
TGAA Head-to-Head 1xSpacing rev	GGCGTTCAGTGAAGCG
TGAA Tail-to-Tail fwd	CGCCTGAATTCACGCC
TGAA Tail-to-Tail rev	GGCGTGAATTCAGGCG
TGAA Tail-to-Tail 1xSpacing fwd	CGCTGAACTTCACGCC
TGAA Tail-to-Tail 1xSpacing rev	GGCGTGAAGTTCAGCG
3xTGAA 1xSpacing fwd	CTGAACTGAACTGAAC
3xTGAA 1xSpacing rev	GTTCAGTTCAGTTCAG
2xTGAA 1xSpacing fwd	CGCTGAACTGAACGCC
2xTGAA 1xSpacing rev	GGCGTTCAGTTCAGCG
2xTGAA 4xSpacing fwd	CGTGAAGCGCTGAACC
2xTGAA 4xSpacing rev	GGTTCAGCGCTTCACG

Supplementary Table 4. Oligonucleotides used for fluorescent EMSAs

Oligomer name	Oligomer sequence
f_TGAA fwd	CATCGTCGTTCATTCAATGGG
f_TGAA rev	CCCATTGAATGAACGACGATG
f_G-Box fwd	CATCGTCGTCACGTGATGGG
f_G-Box rev	CCCATCACGTGACGACGATG
f_TAAT fwd	CAATGTGTTAATGGGTTGTT
f_TAAT rev	AACAACCCATTAACACATTG
f_TAAT wt fwd	GTGTTAATGGGTTGTG
f_TAAT wt rev	CACAACCCATTAACAC
f_TAAT T4C fwd	GTGCTAATGGGTTGTG

f_TAAT T4C rev	CACAACCCATTAGCAC
f_TAAT T12C fwd	GTGTTAATGGGCTGTG
f_TAAT T12C rev	CACAGCCCATTAACAC
f_TAAT T12C, T15C fwd	GTGTTAATGGGCTGCG
f_TAAT T12C, T15C rev	CGCAGCCCATTAACAC

Methods

Cloning

The WUS-HD (34-103) was cloned into pETMBP[52] and pETYFP via NcoI/XhoI restriction sites. The pETYFP vector was generated by cloning the gene encoding YFP into pETM11 with a C-terminal StrepII-tag. The resulting YFP constructs are C-terminal fusions to a His_6-tagged YFP protein, with an additional C-terminal StrepII-tag. The resulting MBP constructs are C-terminal fusions to a His_6-tagged MBP protein with a cleavable tobacco etch virus (TEV) site. Point mutations in pETMBP_WUS-HD and pETYFP_WUS-HD were generated using the QuikChange system (Stratagene). To create the 6xHis-tagged MBP fusion of full-length WUS the coding sequence was amplified with Phusion DNA Polymerase (Thermo Fisher Scientific), digested with NcoI/XhoI and ligated into pMG210[53] previously digested with NcoI/SalI. All primers used for cloning are listed in Supplementary Table 1.

Protein production and purification

Protein expression was carried out in Rosetta2 (DE3) cells either by IPTG induction (1 mM final concentration) using LB medium or auto-induction based on the protocol by Studier[54]. Cells were harvested in cold lysis buffer (20 mM TRIS/HCl (pH 8.0), 150 mM NaCl, 10 mM imidazole, 0.02 % 1-thioglycerol, 1 mg/ml lysozyme, 1 mg/ml DNase, EDTA-free protease inhibitor cocktail (Roche)). The cell suspension was homogenized by four passes through a Microfluidizer (M1-10L, Microfluidics). The lysate was cleared by centrifugation at 50,000 x g and filtered through a 0.45 µm filter before application to a HisTrap FF column (GE Healthcare). The column was washed with IMAC buffer (20 mM TRIS/HCl (pH 8.0), 150 mM NaCl, 10 mM imidazole, 0.02 % 1-thioglycerol)

and the same buffer containing 1 M NaCl. The protein was eluted with IMAC buffer containing 330 mM imidazole and loaded onto a HiTrap SP HP column (GE Healthcare). The column was washed with ion exchange buffer (IEC) buffer (20 mM TRIS/HCl (pH 8.0), 150 mM NaCl, 1 mM DTT) and eluted with the same buffer containing 500 mM NaCl. The final eluate was subjected to size-exclusion chromatography (SEC) using a 16/60 Superdex 75 column (GE Healthcare) equilibrated in gel filtration buffer (20 mM TRIS/HCl (pH 8.0), 150 mM NaCl, 2 mM DTT). Finally, the protein was exchanged into storage buffer (20 mM TRIS/HCl (pH 8.0), 75 mM NaCl, 10 mM $MgCl_2$) using a PD-10 desalting column (GE Healthcare) and aliquots were snap-frozen in liquid nitrogen and stored at -80 °C. Purification of the MBP fusion constructs included the following adaptations: after IMAC purification an additional TEV cleavage (1:100) was performed overnight at 4 °C in gel filtration buffer for MBP WUS-HD and for MBP WUS-FL the size-exclusion step was omitted and the protein was concentrated directly after the desalting step.

Oligonucleotide assembly

Duplex DNAs were annealed using complementary single-strand HPLC purified DNAs (Eurofins). Lyophilized oligonucleotides were dissolved in crystallization buffer (20 mM TRIS/HCl (pH 8.0), 75 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT) or MST buffer (20 mM TRIS/HCl (pH 8.0), 75 mM NaCl, 10 mM $MgCl_2$, 0.05 % Tween-20) and mixed 1:1 with the complementary strand at a final concentration of 50-500 µM. The reaction mix was heated to 95 °C for 5 min and cooled over 75 cycles, each cycle lasting 1 min and decreasing the temperature by 1 K.

Protein crystallization and data collection

The WUS-HD alone (6-12 mg/ml) was crystallized in an in-house automated crystallization platform. The best crystals were obtained after 1-3 days at 18 °C in sitting drops containing 0.05 M MES (pH 6.0), 0.01 M $Mg(OAc)_2$ and 2.5 M $AmSO_4$. For each complex, the WUS-HD was first mixed with a solution of annealed, blunt-ended DNA duplex (16-bp) at a molar ratio of 1:0.6 (protein:DNA) and after 30 min on ice subjected to crystallization trials. The crystallization conditions for all complexes were optimized using an in house developed crystal screening kit for PEG conditions. Complexes were crystallized in sitting drops by vapor diffusion technique from solution containing different concentrations of various PEGs. Crystals with G-Box DNA were grown in 0.2 M LiOAc and 20 % PEG3350. The TGAA DNA crystals were grown in 0.1 M NaOAc (pH 4.6) and 8 % PEG4000. Crystals with TAAT DNA were grown in 0.1 M CHES (pH 9.5) and 20 % PEG8000. All crystals were cryoprotected in mother liquor containing 20 % (v/v) glycerol and flash-cooled in liquid nitrogen. Data sets were collected at the ESRF from individual crystals on beam-lines ID30-A3 (WUS-HD and WUS-HD + G-Box DNA) and ID23-1 (WUS-HD + TAAT DNA and WUS-HD + TGAA DNA) at 100 K. Oligonucleotides sequences are listed in Supplementary Table 2.

Structure determination and refinement

The WUS-HD alone crystallized in the tetragonal space group $P\,4_12_12$ with one molecule in the asymmetric unit. The cell parameters are a = 43.1 Å, b = 43.1 Å, c = 82.0 Å and α, β, γ = 90 °C. The WUS-HD structure was solved by molecular replacement (MR) as implemented in PHASER[55] using the HD of Engrailed (PDB code 3HDD[23]) as a search model. After several rounds of manual building in Coot[56] the structure was refined with Phenix[57]. The model quality was analyzed with

MOLPROBITY[58] giving Ramachandran statistics for the final model of 98.4% of residues in favored regions, 1.6 % in allowed regions and 0 % outliers.

All WUS-DNA complex structures were solved by molecular replacement (MR) as implemented in PHASER[55] using the WUS-HD as a search model. Structure refinement was performed with Phenix[57] and iterative model building in Coot[56]. WUS in complex with G-Box DNA crystallized in the monoclinic space group C 121 with seven molecules in the asymmetric unit. The cell parameters are a = 123.4 Å, b = 83.1 Å, c = 85.7 Å and α, β = 90 °C, γ = 113 °C. The model quality was analyzed with MOLPROBITY[58] giving Ramachandran statistics for the final model of 97.9 % of residues in favored regions, 1.7 % in allowed regions and 0.3 % outliers. WUS in complex with TGAA DNA crystallized in the monoclinic space group P $12_1$1 with six molecules in the asymmetric unit. The cell parameters are a = 55.7 Å, b = 54.3 Å, c = 75.6 Å and α, β = 90 °C, γ = 107.5 °C. The model quality was analyzed with MOLPROBITY[58] giving Ramachandran statistics for the final model of 100 % of residues in favored regions, 0 % in allowed regions and 0 % outliers. WUS in complex with TAAT DNA crystallized in the monoclinic space group P $12_1$1 with seven molecules in the asymmetric unit. The cell parameters are a = 82.7 Å, b = 45.9 Å, c = 87.5 Å and α, β = 90 °C, γ = 102.9 °C. The model quality was analyzed with MOLPROBITY[58] giving Ramachandran statistics for the final model of 99.3 % of residues in favored regions, 0.7 % in allowed regions and 0 % outliers. The data collection and refinement statistics for all structures are summarized in Table 1.

Multi-angle light scattering (MALS)

Experiments were performed using in-line size exclusion chromatography coupled to multi-angle light scattering (MALS) and differential refractive index (dRI)

measurements. The WUS-HD was injected on a Superdex 75 10/300 GL gel filtration column (GE Healthcare) equilibrated with gel filtration buffer (20 mM TRIS/HCl (pH 8.0), 75 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT). The chromatography system was coupled to a DAWN HELEOS II detector and Optilab T-rEX differential refractive index detector (both Wyatt Technology). Data analyses were performed with the ASTRA V software using a dn/dc value of 0.185 mg/ml for molar mass calculation.

Isothermal titration calorimetry (ITC)

All samples were dialyzed overnight at 4 °C against ITC buffer (20 mM TRIS/HCl (pH 8.0), 75 mM NaCl, 10 mM $MgCl_2$). ITC experiments of WUS-HD and a 16-bp DNA 2xTGAA oligonucleotide were performed using a PEAQ-ITC microcalorimeter (Malvern) at 20 °C. Titrations consisted of 13-19 injections of 2-3 µl aliquots of the titrant into the cell solution and 150 s intervals between injections. Typical concentrations used were 400-500 µM WUS-HD in the syringe and 20-40 µM DNA in the cell. Data evaluation was performed with the PEAQ-ITC analysis software.

Microscale thermophoresis (MST)

Binding affinities of WUS-HD to different DNA sequences was measured using microscale thermophoresis (MST). For MST measurements WUS-HD was prepared as a C-terminal fusion to YFP, taking advantage of the intrinsic fluorescence. The YFP-WUS fusion was diluted into MST buffer (20 mM TRIS/HCl (pH 8.0), 75 mM NaCl, 10 mM $MgCl_2$, 0.05 % Tween-20) prior to use. A dilution series of 16-bp target DNA in MST buffer was prepared and mixed 1:1 with the protein. Typical DNA concentrations used were 0.001-100 µM and a fixed concentration of 50 nM for YFP-WUS. Protein-

DNA binding reactions were loaded into Monolith NT.115 Premium Capillaries (NanoTemper Technologies) and the thermophoretic response was monitored with 30 % infrared (IR) laser intensity and 80 % MST power using a Monolith NT.115 instrument (NanoTemper Technologies) at 20 °C. The normalized change in fluorescence was plotted as a function of DNA concentration and analyzed using the MO.Affinity Analysis software (NanoTemper Technologies). Oligonucleotides sequences are listed in Supplementary Table 3.

Electrophoretic mobility shift assay (EMSA)

EMSAs were carried out as described in Brackmann *et al.*[53]. Each reaction contained 115 pmol WUS-HD and 200 fmol CY5-labeled probe in 10 mM TRIS/HCl (pH 8.0), 75 mM NaCl, 50 µg/ml Poly(dI-dC), 1 mM EDTA, 1 mM DTT, 5 % glycerol. Oligonucleotides sequences are listed in Supplementary Table 4.

Bioinformatics

Alignments were generated with Clustal Omega[59] and visualized with ESPRIPT[60]. Surface representations of conserved residues were generated using the ConSurf server[61]. All structural figures were prepared with PyMOL[62]. Electrostatic surface potentials were calculated with APBS[63] integrated in PyMOL. The atomic displacement parameters (residue average) were calculated with BAVERAGE from the CCP4 package[64]. Superimpositions were calculated with GESAMT[65] from the CCP4-package. The interaction parameters and surface area were calculated with PISA[66] from the CCP4 package.

As a *in vivo* affinity measure for native WUS protein to chromatin containing the three binding sites studied here, we analyzed ChIP-seq data from Arabidopsis seedlings with ectopically induced WUS expression. As a proxy for affinity we used the number of reads aligned to regions containing WUS binding sequences represented as specific k-mer. For this purpose, we created BED files with using the of Arabidopsis thaliana (TAIR10) genome and k-mers of interest. We the identified regions with known WUS binding sites using R-package GenomicRanges v1.32.6 (ref. 67) and defined a window of +- 25 bp around the binding sequence. The resulting genomic ranges were transformed into GTF file and used for counting aligned reads from WUS ChIP-seq data (GEO accession GSE122611) with featureCount v1.6.3 (ref. 68). The obtained counts per k-mer regions were visualized by empirical cumulative probability density using ecdf function in R and smoothed using plogspline function from R-package logspline v2.1.15 (ref. 69).

References

1. Greb T, Lohmann JU. Plant Stem Cells. *Curr Biol* **26**, R816-821 (2016).

2. Laux T, Mayer KF, Berger J, Jurgens G. The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. *Development* **122**, 87-96 (1996).

3. Sarkar AK*, et al.* Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. *Nature* **446**, 811-814 (2007).

4. Daum G, Medzihradszky A, Suzaki T, Lohmann JU. A mechanistic framework for noncell autonomous stem cell induction in Arabidopsis. *Proc Natl Acad Sci U S A* **111**, 14619-14624 (2014).

5. Pi L*, et al.* Organizer-Derived WOX5 Signal Maintains Root Columella Stem Cells through Chromatin-Mediated Repression of CDF4 Expression. *Dev Cell* **33**, 576-588 (2015).

6. Yadav RK, Perales M, Gruel J, Girke T, Jonsson H, Reddy GV. WUSCHEL protein movement mediates stem cell homeostasis in the Arabidopsis shoot apex. *Genes Dev* **25**, 2025-2030 (2011).

7. Wunderlich Z, Mirny LA. Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet* **25**, 434-440 (2009).

8. Yesudhas D, Batool M, Anwar MA, Panneerselvam S, Choi S. Proteins Recognizing DNA: Structural Uniqueness and Versatility of DNA-Binding Domains in Stem Cell Transcription Factors. *Genes (Basel)* **8**, (2017).

9. Morgunova E, Taipale J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol* **47**, 1-8 (2017).

10. Jolma A*, et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384-388 (2015).

11. Burglin TR, Affolter M. Homeodomain proteins: an update. *Chromosoma* **125**, 497-521 (2016).

12. Pick L, Heffer A. Hox gene evolution: multiple mechanisms contributing to evolutionary novelties. *Ann N Y Acad Sci* **1256**, 15-32 (2012).

13. Dolzblasz A*, et al.* Stem Cell Regulation by Arabidopsis WOX Genes. *Mol Plant* **9**, 1028-1039 (2016).

14. Mallo M, Alonso CR. The regulation of Hox gene expression during animal development. *Development* **140**, 3951-3963 (2013).

15. Hueber SD, Lohmann I. Shaping segments: Hox gene function in the genomic age. *Bioessays* **30**, 965-979 (2008).

16. Lohmann JU, *et al.* A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell* **105**, 793-803 (2001).

17. Leibfried A, *et al.* WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. *Nature* **438**, 1172-1175 (2005).

18. Busch W, *et al.* Transcriptional control of a plant stem cell niche. *Dev Cell* **18**, 849-861 (2010).

19. O'Malley RC, *et al.* Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **166**, 1598 (2016).

20. Yadav RK, *et al.* Plant stem cell maintenance involves direct transcriptional repression of differentiation program. *Mol Syst Biol* **9**, 654 (2013).

21. Perales M, Rodriguez K, Snipes S, Yadav RK, Diaz-Mendoza M, Reddy GV. Threshold-dependent transcriptional discrimination underlies stem cell homeostasis. *Proc Natl Acad Sci U S A* **113**, E6298-E6306 (2016).

22. Ikeda M, Mitsuda N, Ohme-Takagi M. Arabidopsis WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. *Plant Cell* **21**, 3493-3505 (2009).

23. Fraenkel E, Rould MA, Chambers KA, Pabo CO. Engrailed homeodomain-DNA complex at 2.2 A resolution: a detailed view of the interface and comparison with other engrailed structures. *J Mol Biol* **284**, 351-361 (1998).

24. Cooley RB, Arp DJ, Karplus PA. Evolutionary origin of a secondary structure: pi-helices as cryptic but widespread insertional variations of alpha-helices that enhance protein functionality. *J Mol Biol* **404**, 232-246 (2010).

25. Kumar P, Bansal M. Dissecting pi-helices: sequence, structure and function. *FEBS J* **282**, 4415-4432 (2015).

26. Berger MF, *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266-1276 (2008).

27. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**, 1277-1289 (2008).

28. Ma Y, *et al.* WUSCHEL acts as an auxin response rheostat to maintain apical stem cells in Arabidopsis. *Nat Commun* **10**, 5093 (2019).

29. Ades SE, Sauer RT. Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. *Biochemistry* **34**, 14601-14608 (1995).

30. Joshi R, *et al.* Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530-543 (2007).

31. Wolberger C, Vershon AK, Liu B, Johnson AD, Pabo CO. Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**, 517-528 (1991).

32. Hanes SD, Brent R. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell* **57**, 1275-1283 (1989).

33. Treisman J, Gonczy P, Vashishtha M, Harris E, Desplan C. A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* **59**, 553-562 (1989).

34. Zandarashvili L, Nguyen D, Anderson KM, White MA, Gorenstein DG, Iwahara J. Entropic Enhancement of Protein-DNA Affinity by Oxygen-to-Sulfur Substitution in DNA Phosphate. *Biophys J* **109**, 1026-1037 (2015).

35. Ekker SC, von Kessler DP, Beachy PA. Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J* **11**, 4059-4072 (1992).

36. Zeiske T*, et al.* Intrinsic DNA Shape Accounts for Affinity Differences between Hox-Cofactor Binding Sites. *Cell Rep* **24**, 2221-2230 (2018).

37. Zhou T*, et al.* DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res* **41**, W56-62 (2013).

38. Ivarie R. Thymine methyls and DNA-protein interactions. *Nucleic Acids Res* **15**, 9975-9983 (1987).

39. Pfeiffer A, Wenzl C, Lohmann JU. Beyond flexibility: controlling stem cells in an ever changing environment. *Curr Opin Plant Biol* **35**, 117-123 (2017).

40. Gross-Hardt R, Lenhard M, Laux T. WUSCHEL signaling functions in interregional communication during Arabidopsis ovule development. *Genes Dev* **16**, 1129-1138 (2002).

41. Deyhle F, Sarkar AK, Tucker EJ, Laux T. WUSCHEL regulates cell differentiation during anther development. *Dev Biol* **302**, 154-159 (2007).

42. Siggers T, Gordan R. Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res* **42**, 2099-2111 (2014).

43. Nagasaki H, Matsuoka M, Sato Y. Members of TALE and WUS subfamilies of homeodomain proteins with potentially important functions in development form dimers within each subfamily in rice. *Genes Genet Syst* **80**, 261-267 (2005).

44. Rodriguez K, Perales M, Snipes S, Yadav RK, Diaz-Mendoza M, Reddy GV. DNA-dependent homodimerization, sub-cellular partitioning, and protein destabilization control WUSCHEL levels and spatial patterning. *Proc Natl Acad Sci U S A* **113**, E6307-e6315 (2016).

45. Georges AB, Benayoun BA, Caburet S, Veitia RA. Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? *FASEB J* **24**, 346-356 (2010).

46. Wilson DS, Guenther B, Desplan C, Kuriyan J. High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell* **82**, 709-719 (1995).

47. Yin Y*, et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).

48. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500 (2016).

49. Cao X, He Z, Guo L, Liu X. Epigenetic Mechanisms Are Critical for the Regulation of WUSCHEL Expression in Floral Meristems. *Plant Physiol* **168**, 1189-1196 (2015).

50. Deribe YL, Pawson T, Dikic I. Post-translational modifications in signal integration. *Nat Struct Mol Biol* **17**, 666-672 (2010).

51. Seet BT, Dikic I, Zhou MM, Pawson T. Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* **7**, 473-483 (2006).

52. Bogomolovas J, Simon B, Sattler M, Stier G. Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. *Protein Expr Purif* **64**, 16-23 (2009).

53. Brackmann K*, et al.* Spatial specificity of auxin responses coordinates wood formation. *Nat Commun* **9**, 875 (2018).

54. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-234 (2005).

55. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

56. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

57. Adams PD*, et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).

58. Chen VB*, et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

59. Sievers F*, et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

60. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324 (2014).

61. Landau M, *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299-302 (2005).

62. Schrödinger L. The PyMOL Molecular Graphics System.). Version 2.0 edn.

63. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041 (2001).

64. Winn MD, *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242 (2011).

65. Krissinel E. Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem* **1**, 76-85 (2012).

66. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007).

67. Lawrence M, *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**:e1003118 (2013).

68. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **1**; 923-30 (2014).

69. Kooperberg C. logspline: Routines for Logspline Density Estimation. R package version 2.1.15; https://CRAN.R-project.org/package=logspline (2019).