

## Distant sequence regions of JBP1 contribute to J-DNA binding

Ida de Vries<sup>1</sup>, Danique Ammerlaan<sup>1</sup>, Tatjana Heidebrecht<sup>1</sup>, Patrick H. N. Celie<sup>1</sup>, Daan P. Geerke<sup>2</sup>, Robbie P. Joosten<sup>1</sup> and Anastassis Perrakis<sup>1,\*</sup>

<sup>1</sup> Oncode Institute and Division of Biochemistry, Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, The Netherlands.

<sup>2</sup> Department of Chemistry and Pharmaceutical Sciences, Amsterdam Institute of Molecular and Life Sciences (AIMMS) and Amsterdam Center for Multiscale Modeling (ACMM), Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, The Netherlands.

\* To whom correspondence should be addressed. Email: a.perrakis@nki.nl

### ABSTRACT

Base-J ( $\beta$ -D-Glucopyranosyloxymethyluracil) is a modified DNA nucleotide that replaces 1% of thymine in kinetoplastid flagellates. The biosynthesis and maintenance of base-J depends on the base-J Binding Protein 1 (JBP1), that has a thymidine hydroxylase domain (THD) and a J-DNA binding domain (JDBD). How the THD synergizes with the JDBD to hydroxylate thymine in specific genomic sites, maintaining base-J during semi-conservative DNA replication, remains unclear. Here we present a crystal structure of the JDBD including a previously disordered DNA-contacting loop and use it as starting point for Molecular Dynamics (MD) simulations and computational docking studies to propose recognition models for JDBD binding to J-DNA. These models guided mutagenesis experiments, providing additional data for docking, which reveals a binding mode for JDBD onto J-DNA. This model, together with the crystallographic structure of the TET2 JBP1-homologue in complex with DNA and the AlphaFold model of full-length JBP1, allowed us to hypothesize that the flexible JBP1 N-terminus contributes to DNA-binding, which we confirmed experimentally. A high-resolution JBP1:J-DNA complex, which must involve conformational changes, would however need to be determined experimentally to further understand this unique underlying molecular mechanism that ensures replication of epigenetic information.

## INTRODUCTION

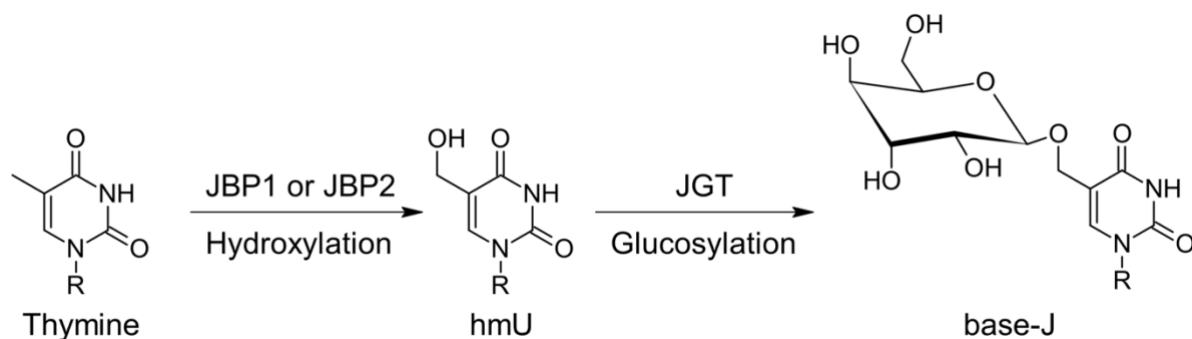
The modified nucleotide  $\beta$ -D-Glucopyranosyloxymethyluracil (base-J) replaces 1% of the thymine nucleotides in kinetoplastid flagellates (1). 99% of base-J is found in telomeric repeats, mainly in GGGTTA repeats wherein the second thymine is modified to base-J (2–4). The remaining 1% of base-J is involved in transcription. In *Leishmania*, base-J has been shown to be involved in transcription termination (5, 6), while in *Trypanosoma* base-J is marking transcription initiation (7). The biosynthesis of base-J occurs in two steps (Figure 1). In the first step, the methyl group of a thymine is hydroxylated by JBP1 or JBP2 (8, 9). Both have a  $\beta$ -oxoglutarate and  $\text{Fe}^{2+}$  thymidine hydroxylase domain (THD), which is a functionally divergent homologue also present in human TET proteins, an important finding that established the TET-JBP family (10–12). In the second step, the base J-associated glucosyltransferase (JGT) adds a sugar moiety to the hydroxymethyluracil (hmU) intermediate, resulting in base-J (13–15). More recently another protein bearing a JDBD, JBP3, has been identified and shown to be involved in transcription termination in trypanosomes and *Leishmania* (16, 17).

JBP1 specifically recognizes base-J containing double stranded DNA (J-DNA) (18, 19) through a J-DNA binding domain (JDBD) which has been structurally characterized (20). In the context of JBP1, JDBD recognizes base-J and then probably invokes a conformational change in JBP1 (21, 22), enabling the THD to initiate the biosynthesis of base-J on a thymine 13 base pairs downstream the complementary DNA strand (J+13' position) (23). This synergy between recognition of a pre-existing base-J in the parental strand, and hydroxylation of a thymidine in the daughter strand, is likely important for replication of the base-J epigenetic marker.

It has been previously shown that the C2 and C3 hydroxyl groups of the sugar moiety of base-J interact with the non-bridging phosphoryl oxygen of the nucleotide that is positioned before base-J (J-1 position), ensuring that the sugar moiety is positioned perpendicular towards the major groove of the J-DNA helix (24). That would leave the C4 and C6 hydroxyl groups free for recognition of J-DNA by the JDBD (20). Structure-based mutagenesis and functional experiments have shown Asp525 in the JDBD to be adequate in explaining the discrimination between J-DNA and standard DNA oligonucleotides. The D525A mutation results in ~1,000 times worse binding to J-DNA and 10 times better binding to regular DNA, abolishing all specificity. This strongly indicates that Asp525 forms hydrogen bonds with the free C4 and C6 hydroxyl groups of base-J, as both are found to be essential in forming the JBP1:J-DNA complex. Further mutation studies showed that residues Lys518, Lys522 and Arg532 are important for the affinity of JBP1 towards both J-DNA and regular DNA (20). This indicates that these residues are not involved in the specificity for J-DNA, but are important in formation of a JBP1:DNA complex. Combining these conclusions with the crystal structure of the binding domain (PDB-ID: 2XSE) and with results from hydrogen deuterium exchange rate (HDX-MS) analyses, a manually built model for the JDBD:J-DNA complex has been proposed and validated by small angle X-ray scattering experiments (20). Notably, the crystal structure of the JDBD had a missing loop of eight residues (529–537) between helices  $\alpha$ 4 and  $\alpha$ 5, including the Arg532 that is important for DNA binding. This loop had been modelled by hand, but had not been observed in the electron density maps of the structure.

The crystallization and structure determination of JDBD was used as a demonstration project during the Cold Spring Harbor Laboratory (CHSL) course. JDBD is one of the proteins that were

provided to the students, and they were asked to crystallize it, collect diffraction data, and determine the structure. Remarkably, in one of numerous datasets that have been obtained during the course, the missing density for the loop 529-537 was observed. Motivated by this new structure in which we were able to build that missing loop, we performed molecular dynamics (MD) simulations in combination with docking studies and determined a new and informative structure model representative for the JDBD:J-DNA complex, which was further refined by mutagenesis experiments and validated with small angle X-ray scattering (SAXS) data. The recent availability of AlphaFold structure prediction of the full-length JBP1 protein (25, 26) together with the X-ray structure of the TET2 homologue of the THD in complex with DNA (27) and the JBP1-JDBD:J-DNA model from this study, allowed to further understand how JBP1 binds J-DNA, and to demonstrate that the N-terminal helical region participates in this binding.



**Figure 1:** Biosynthesis of base-J ( $\beta$ -D-Glucopyranosyloxymethyluracil). Thymine is hydroxylated by JBP1 or JBP2 to form hmU (hydroxymethyluracil), which is glucosylated by JGT to result in the modified nucleotide base-J. R indicates the rest of the nucleotide of thymine, which is located in a DNA strand.

## MATERIAL AND METHODS

### Macromolecule production and crystallization

Protein expression, crystallization, crystal harvest and X-ray crystallographic experiment was performed as described before (20). Details are shown in Table 1.

Table 1: JDBD crystallization details

<b>Source organism</b>	<i>Leishmania Tarentolae</i>
<b>Expression host</b>	BL21(DE3)T1 <sup>R</sup>
<b>Crystallization method</b>	Hanging drop
<b>Plate type</b>	MRC 2drop
<b>Temperature (K)</b>	277
<b>Protein concentration</b>	20 mg/ml
<b>Buffer composition of protein solution</b>	20 mM HEPES pH 7.5, 140 mM NaCl, 1 mM TCEP
<b>Composition of reservoir solution</b>	15-17% Peg 6000, 0.1M Sodium iodide or 15-17% Peg, 0.2M potassium nitrate
<b>Volume and ratio of drop</b>	100 nl, 1:1
<b>Volume of reservoir</b>	150 µl

### X-ray crystallography procedures

Crystallographic details are shown in Table 2. The images were processed and scaled at the 17-ID-1 (AMX) beamline of the National Synchrotron Light Source II at Brookhaven (28). The integrated data were scaled in CCP4i2 (29, 30), using AIMLESS (31). The space group P6<sub>1</sub>22 was selected and diffraction data extended to 1.95 Å. The structure of JDBD (PDB-ID: 2XSE) was used as search model for molecular replacement. Next, the model was refined in multiple iterative cycles between REFMAC (32) and manual modelling in COOT 0.8.9.2 (33). Within REFMAC, a single TLS body was refined for 15 cycles with starting B-factors fixed to a value of 20.0 Å<sup>2</sup>, jelly body restraints with sigma 0.01 and maximum distance of 4.2 were set. Typically, 30 cycles were run using riding hydrogens. Optimization of the model based on the PDB-REDO webserver (34) was used to provide additional optimized refinement parameters: simple solvent scaling was used with explicit solvent mask with custom parameters (1.0 increase VDW radius of non-ion atoms, 0.9 increase ionic radius of potential ion atoms, shrink the mask area by 0.9 after calculation). Double conformations were modelled for Ser408, Glu447, Glu488, Met494 and Met546. The final structure was validated using MolProbity (35). The structure has been deposited to the Protein Data Bank (PDB) (36) as entry 8BBM.

Table 2: JDBD data collection and processing details. Values for the outer shell are given in parentheses.

Diffraction source	
<b>Wavelength (Å)</b>	0.98
<b>Space group</b>	P 6 <sub>1</sub> 2 2
<b>a, b, c (Å)</b>	68.29, 68.29, 185.84
<b>α, β, γ (°)</b>	90.0, 90.0, 120.0
<b>Resolution range (Å)</b>	59.2 – 1.95 (2.00 – 1.95)
<b>Total No. of reflections</b>	753607 (53016)
<b>No. of unique reflections</b>	19605 (1327)
<b>Completeness (%)</b>	100 (100)
<b>Multiplicity</b>	38.4 (40.0)
$\langle I \rangle / \sigma(I)$	17.5 (1.0)
$R_{\text{r.i.m.}}$	0.157 (5.033)
<b>Overall B factor from Wilson plot (Å<sup>2</sup>)</b>	32.3
<b>No. of reflections, working set, test set</b>	18522, 1017
<b>Final <math>R_{\text{cryst}}</math>, <math>R_{\text{free}}</math></b>	0.202, 0.252
No. of non-H atoms	
<b>Protein</b>	1440
<b>Water</b>	114
R.m.s. deviations / rmsZ Z-score	
<b>Bonds</b>	0.014 / 0.69
<b>Angles (°)</b>	1.62 / 0.76
Average B factors (Å <sup>2</sup> )	
<b>Protein</b>	55
<b>Water</b>	54
<b>Ramachandran plot favoured/outliers (%)<sup>†</sup></b>	99.4 / 0.0
<b>Ramachandran Z-score</b>	-0.98 ± 0.72
<b>Rotamers favoured/poor (%)<sup>†</sup></b>	96.1 / 1.3
<b>Clashscore (%-ile)<sup>†</sup></b>	100
<b>MolProbity score (%-ile)<sup>†</sup></b>	100

<sup>†</sup> As reported by MolProbity (35)

### Molecular dynamics simulations of JDBD

*Protocol.* Two independent MD simulations of the JDBD were performed using the GROMACS-2020.2 software (37) and timesteps of 2 fs. The Amber99SB-ILDN force field (38) with periodic boundary conditions was used to describe the system. The topology file was generated from the new JDBD crystal structure (PDB-ID: 8BBM) using the GROMACS  *pdb2gmx* tool, after which the protein was placed in a dodecahedral periodic box and energy minimized with the steepest descent method. Subsequently, 10,341 TIP3 water molecules (39) and 4 Cl<sup>-</sup> ions were added to solvate the system and neutralize the total charge. Another energy minimization using the steepest descent method was performed. Initial atomic velocities were randomly assigned following a Maxwell-Boltzmann distribution (using different

random seeds for the two independent simulations). The structure was heated till 300 K in three subsequent *NVT* simulations at 100, 200 and 300 K with position restraint force constants of 10000, 5000 and 50 kJ/mol/nm<sup>2</sup> respectively, to keep the protein backbone C $\alpha$  atoms positionally restrained with harmonic potentials. Subsequently an *NpT* simulation was performed (without the use of position restraints) at 1 atm and 300 K for 1 ns, writing coordinates every step, to equilibrate the structure. Next, the MD production run of 100 ns (in which coordinates were written out every five steps) was performed and analyzed. The LINCS algorithm (40) was used to constrain bond lengths to their zero-energy value using a single iteration and with the highest order in the expansion of the constraint coupling matrix set to 4. The Langevin integrator was used and set to a friction coefficient for each particle at mass/0.1 ps. The Berendsen barostat (41) was used in order to keep the pressure close to 1 atm, using a coupling time of 0.5 ps and an isothermal compressibility of 4.5 x 10<sup>-5</sup> bar<sup>-1</sup>. Short-range electrostatic and Van der Waals interactions were evaluated every time step with a distance cut-off of 0.9 nm. The smooth particle mesh Ewald method (42) was used to evaluate long-range electrostatic interactions with a grid spacing of 0.125 nm. Center-of-mass motion was removed every 10 time steps.

**Analyses.** The output trajectories of the MD production runs were processed using the GROMACS-2020.2 *trjconv* tool, keeping every set of coordinates written to disc. Trajectories were visualized using VMD 1.9.3 (43). C $\alpha$  atom positional root-mean-square deviations (RMSDs) and fluctuations (RMSFs) with respect to the corresponding thermally equilibrated structure were calculated over time or averaged over time per residue, respectively, using GROMACS-2020.2 tools *gmx rms* and *gmx rmsf*, respectively. Plots of these time series or values were made using Seaborn 0.11.1 (44). Subsequently both simulation trajectories were reduced using the GROMACS-2020.2 *trjconv* tool to timesteps of 250 ps. All solvent molecules and ions were removed from these structures to reduce file sizes. The JDBD structures from these reduced trajectories were clustered separately with Bitclust 0.11 (45) utilizing Daura's algorithm (46). The maximum number of clusters to generate was set at 10 clusters per simulation with a RMSD cut-off for pairwise comparisons set at 2.25 Å (45).

### Docking base-J containing DNA to JBP1 with HADDOCK

*Implementation parameters for base-J.* Topology parameters and charges for the base-J residue were generated and implemented in the HADDOCK2.4 webserver (47). For the other DNA nucleotides and the protein, the standard topology parameters and charges defined by default through this webserver were used.

*Generation of docking input files for base-J containing DNA.* A starting structure for a 20-mer DNA strand containing the sequence 5'-CAGAAGGCAGCJGCAACAAG-3' was created in Pymol using the Build menu (48) in which base-J was built as a regular thymine. Subsequently, this structure was aligned with the coordinates of the structure model published by Heidebrecht *et al.* (2011) (20) in Coot 0.8.9.2 (33) and the coordinates were saved. Using a text editor, the thymine residue was replaced by base-J. The resulting file was inspected with YASARA version 20.8.23 (49) and the phosphate backbone bonds with the residues before and after base-J were added. The structure was energy minimized in vacuum

followed by energy minimization in explicit TIP3P water molecules, both using the Amber99 force field in YASARA (49).

*Docking protocol.* The J-DNA strand was docked to JDBD using the HADDOCK2.4 web server (47). To guide the docking, explicit restraints were added for the distance range between Asp525 OD1 and OD2 with both base-J hydrogen atoms HO4 and HO6 (1.7-3.5 Å). Furthermore, Lys518, Lys522, and Arg532 were manually defined as active residues and Ser515, Arg517, Lys518, Val521, Lys524, Phe528, Lys535 were manually marked as passive residues of JDBD, because mutation studies indicated involvement in formation of the protein:DNA complex for these residues (20). For J-DNA, base-J was manually defined as active residue and the passive residues were defined by the default settings in HADDOCK. The default docking protocol of the HADDOCK2.4 webserver was utilized and the HADDOCK scoring and clustering was used to rank the output structures (47). J-DNA was docked to several JDBD structures, namely: i) the new crystal structure; ii) all central protein structure models of the three most occupied clusters for each MD simulation trajectory that were obtained from clustering; and iii) the new crystal structure, with Asn455 and Arg448 defined as additional active residues in JDBD. The resulting models were visualized using CCP4mg (50) and superposed based on the positions of the backbone atoms of J-DNA, in such a way that the positions of base-J align.

*Comparison with experimental SAXS and HDX-MS data.* For each best-scoring model (cluster1\_1.pdb) of the JDBD:J-DNA complex obtained by HADDOCK protocols ii or iii (see previous section), the SAXS curve was calculated and plotted against the experimentally obtained SAXS curve (20) using CRYSTAL from the ATSAS suite (51). Additionally, the best-scoring models (cluster1\_1.pdb) of the docking models obtained by HADDOCK protocol iii (see previous section), were compared against the HDX-MS data (20) by mapping the HDX-MS scores onto the protein structures.

## Protein assays

*Protein expression and purification.* Wild-type (wt) 6xhis-JBP-JDBD protein (*Leishmania Tarentolae*) was expressed in *E. coli* and purified as described before (20). Single point mutations E437A, H440A, R448A and N455A were introduced in the pETNKL-his-JBP1 JDBD plasmid using a modified QuickChange site-directed mutagenesis method (Agilent). Forward and reverse mutagenesis primers containing the single point mutations were purchased from ThermoFisher Scientific: (E437A: 5'-CAT GTG AGC CCA TGC TTT ACC AAC CAG TTC CAG C-3', H440A: 5'-GGG TTC AGA GCC AGC ATG GCA GCC CAT TCT TTA CCA AC-3', R448A: 5'-CCA CAG GAA GTC TTT AGC TTC CGG GTT CAG AGC C-3', N455A: 5'-CAG AGT TCA TTT CAG ACT GGG CTT TCC ACA GGA AGT CTT TAC GTT C-3'). For the mutants, each of the plasmid DNA strands was first amplified in two separate PCR reactions (5 cycles each), one reaction with the forward mutagenesis primer (5' to 3') and a second with the reverse mutagenesis primer. Both reactions were combined and the PCR was continued for another 15 cycles to obtain double-stranded plasmids comprising the single base substitutions. Plasmids were sequence-verified to confirm the presence of the mutations. Two constructs encoding N-terminal truncations of JBP1 lacking the first 22 amino acid residues (pETNKL-his-3C-JBP123-827, Δ23-JBP1)

or the first 37 residues (pETNKL-his-3C-JBP138-827, Δ38-JBP1) were created using Ligation Independent Cloning (52). Δ23-JBP1 and Δ38-JBP1 proteins were expressed and purified as described for full-length wt-JBP1 (20).

*Fluorescence polarization assays.* Fluorescence polarization assays were performed as described before (20). Serial dilutions of wt-JBP1, Δ38-JBP1, Δ23-JBP1, the JDBD and point-mutated JDBD proteins were prepared on ice in 20 mM HEPES pH 7.5, 140 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM TCEP, 1 mg/ml chicken ovalbumin, 0.05 % TWEEN 20 and supplemented with either 1 nM of TAMRA-labelled J-DNA (5'-GGCAGCJGCAACAA-3') or T-DNA (5'-GGCAGCTGCAACAA-3'). Fluorescence polarization was measured in a PHERAstar FS plate reader (BMG Labtech). Data were analyzed and plotted using Prism (Graphpad).

*Stability of JDBD mutants.* The thermal stability of wt-JDBD and mutant proteins was measured in the Prometheus NT.48 nanoDSF (nano differential scanning fluorescence) instrument (NanoTemper Technologies). For each protein, the thermal stability was determined at two concentrations, 0.5 mg ml<sup>-1</sup> and 0.25 mg ml<sup>-1</sup> in 20 mM HEPES/HCl at pH 7.5, 140 mM NaCl and 1 mM TCEP. Protein unfolding (ratio of fluorescence at 330 and 350 nm) and aggregation (scattering) was assessed using a temperature slope of 1 °C/min in a range from 20 °C to 90 °C, with an excitation power of 30 %. All measurements were performed in duplo.

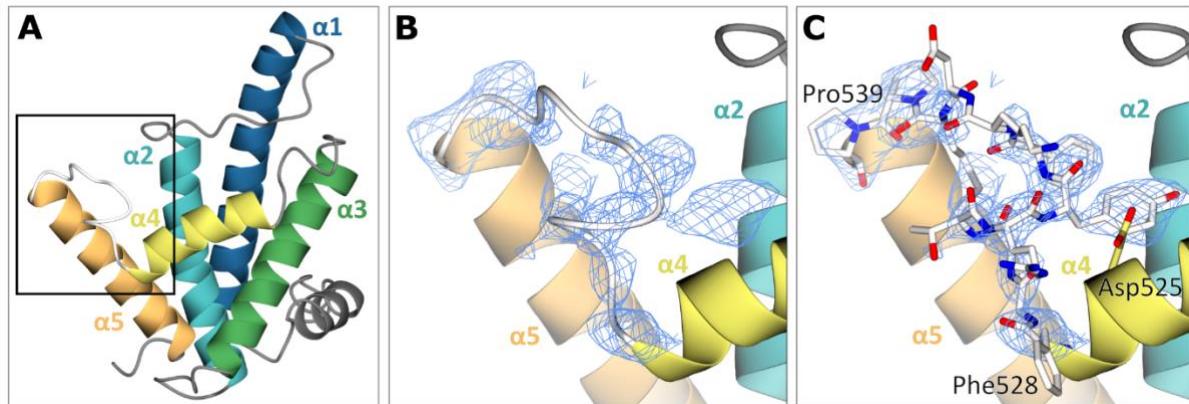
### Modelling of the JBP1:J-DNA complex using AlphaFold

The AlphaFold2.0 structure of the full-length JBP1 was downloaded (entry Q9U6M1) from the AlphaFold protein structure database (25, 26). To create a model of the THD domain, residues 1-39, and 366-576 (low confidence and JDBD) were removed from the model. Then, the PDB-REDO (34) crystallographic structure of a TET2 homologue of the hydroxylase domain (PDB-ID: 5DEU) was aligned to the THD of this model and the coordinates of loop 113-121 in the AlphaFold model were replaced by the coordinates of the loop 1291-1299 of the TET2 homologue. The DNA of the TET2 homologue was used to model THD:J-DNA binding in the AlphaFold structure. The 5HC in the DNA was mutated to hmU (5MU) and the DNA was adjusted to a 25-mer J-DNA with sequence 5'-TCGATTJGTTCATAGACTAATACGT-3'. This THD model in complex with the 25mer J-DNA was energy minimized in vacuum, followed by energy minimization in water using the Amber99 force field (53) and the TIP3P water model in the YASARA software package (49). The ligands oxoglutarate and Fe<sup>2+</sup> were added to the THD of the newly generated model based on their respective positioning in the TET2 homologous structure model. Next, the JDBD:JDNA model obtained from docking protocol iii (see above) was aligned with this model, in such way that the J-DNA backbone atoms align and the base-J position overlaps. The JDBD was added to the model, and the whole complex was energy minimized in vacuum and water once more. This finally resulted in a JBP1:J-DNA model containing a 25mer J-DNA in complex with the JDBD and THD-JBP1 including the oxoglutarate and Fe<sup>2+</sup> ligands in the THD.

## RESULTS AND DISCUSSION

### A new crystal structure for JDBD with an ordered loop

The previously published crystal structure of JDBD (20) (PDB-ID: 2XSE), was missing residues 529-537. It has been noticed that the electron density maps from a data set collected by students following the CSHL 2018 course on X-Ray Methods in Structural Biology, showed rather clear density for that loop. As this loop follows the  $\alpha$ 4 recognition helix and harbors the Arg532 residue which affects DNA binding (20), the diffraction data from the CSHL 2018 course were re-processed, and the model was rebuilt and refined. The final structure was validated using MolProbity (35) and it belongs to the 100th percentile, with a Ramachandran Z-score (54) of  $-0.98 \pm 0.72$ . The new structure at a resolution of 1.95 Å adopts a highly similar conformation as 2XSE, which had been refined at a nearly identical resolution (1.90 Å). The space group is identical ( $P6_122$ ) and the cell dimensions do not differ substantially (a, b, c: 67.34, 67.34, 186.76 Å for 2XSE vs. a, b, c: 68.29, 68.29, 185.84 Å for 8BBM). However, the new crystal structure contains all residues from 392 to 561 (Figure 2A), including the 529-537 loop which is well resolved (Figure 2B and C). The B factors in the loop range between 53.5-125.2 (mean 91.6), compared to a mean of 65.3 for the JDBD structure excluding the five  $\alpha$ -helices. There are no crystal contacts observed for this loop, and there are no substantial differences in crystal properties overall. Thus, it remains unclear why the 529-537 loop is ordered in the new structure and not in previous experiments.



**Figure 2:** Crystal structure of JDBD with an ordered loop. A) New crystal structure of JDBD (PDB-ID: 8BBM) with the black box indicating the zoom-in area of B and C. B and C) Electron density ( $2mF_o - DF_c$  map contoured at 1rms) observed for loop 528-539 in the new crystal structure (PDB-ID: 8BBM); the 528-539 loop is shown as ribbon (B) and with all atoms (C).

### Investigating the JDBD interaction with J-DNA by computational modeling

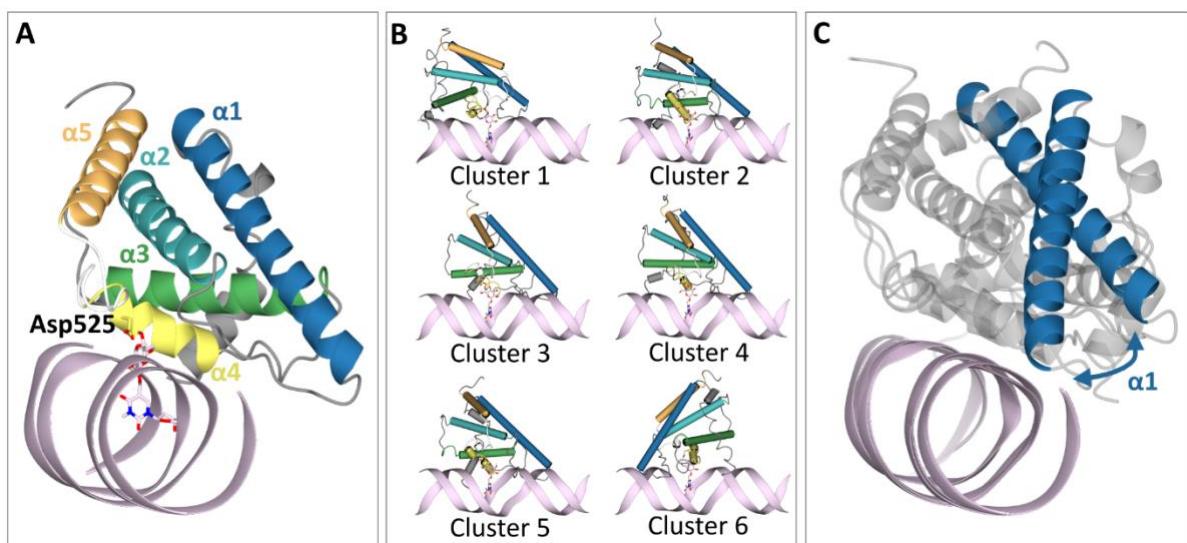
We then wanted to use our new crystal structure of JDBD for docking to a J-DNA model, to investigate if that would lead to new insights in the possible binding mode of JDBD to J-DNA. While previously we have done such studies manually, here we decided to follow a computational approach. In the manually

created model (20), base-J is oriented perpendicular towards the major groove of the DNA and the interactions of the O2 and O3 hydroxyl groups of base-J with the non-bridging phosphoryl oxygen at J-1 in J-DNA are modelled, as first observed by Grover *et al.* (2007) (24). To now be able to create a computational model, parameters to model base-J and J-DNA were generated and implemented in HADDOCK2.4 (47). In the docking protocol, JDBD was docked onto a 20-mer J-DNA helix. The Asp525 residue was restrained to be in the proximity of the C4 and C6 hydroxyl groups of the sugar moiety of base-J in both complexes, based on its importance for recognizing base-J.

The docking study using JDBD resulted in three HADDOCK clusters with docked structures, of which the most occupied cluster contains 93% of the generated models. The best scoring model is shown in Figure 3A. The JDBD crystal structure was then used to start MD simulations to explore the behavior and flexibility of this protein domain in solution. Two parallel simulations were run, in which both the Ca RMSD and RMSF values were comparable (Supplemental Figure S1). The overall fold of the protein is maintained as seen by clustering the obtained structures (Supplemental Figure S2), with, as expected, more flexibility for loops and termini compared to residues located in the helices (Supplemental Figure S1B). As in both MD simulations most obtained conformations are represented by the three most occupied clusters (93% coverage in simulation run1, 64% coverage in run2, Supplemental Figure S3), we assume that the center structures of JDBD represented in these six clusters provide a representative set of structures for the behavior and flexibility of JDBD in solution.

The six selected representative JDBD models from clustering the MD trajectories were used to investigate if protein flexibility would affect the JDBD:J-DNA model proposed by HADDOCK. In the obtained complexes, the overall positioning of JDBD along the J-DNA strand is consistent (Figure 3B, Supplemental Figure S4A-E). The  $\alpha$ 1 helix is positioned towards the J-DNA, but has a different orientation in Cluster 1, compared to Clusters 2, 3, 4 and 5 (Figure 33C). The model obtained using Cluster 6 as JDBD template shows a mirrored protein orientation with respect to the J-DNA compared to the models obtained using the centers of Clusters 1-5 as protein template (Figure 3B, Supplemental Figure S4F). While this mirrored orientation is preferred in the majority of the HADDOCK models (69%), the remaining 31% of the structure models obtained from the docking run with Cluster 6 is similar to the models in Clusters 1-5.

To validate the docking models, the obtained structures were compared to experimental SAXS and HDX-MS data presented in previous work (20). The experimental SAXS curves fit well to the calculated curve for the models obtained from docking, as indicated by the  $\chi^2$  values (Supplemental Figure S5). Furthermore, the HDX-MS data were mapped onto the docking output (Supplemental Figure S6), demonstrating that the  $\alpha$ 4 helix (which contains the Asp525 residue) and the  $\alpha$ 1 helix (which is positioned towards J-DNA as in the docking models) show the most pronounced HDX-MS reduction.



**Figure 3:** Docking models of JDBD in complex with J-DNA. A) Docking output of docking J-DNA (pink) onto JDBD new crystal structure with base-J in pink and Asp525 in yellow cylinders. B) Docking output using the center structures of the six most occupied clusters obtained from clustering MD simulations as protein template. JDBD is shown as tubes; Clusters 1, 2 and 3 originate from MD simulation run 1, Clusters 4, 5 and 6 from MD simulation run 2. C) The blue arrow is illustrating the flexible orientation of the  $\alpha 1$  helix perpendicular to the J-DNA as observed in the JDBD:J-DNA complexes obtained from docking. The overlay is generated from the best scoring structure of the most occupied cluster obtained from docking with Cluster 1 and Cluster 2 as protein template.

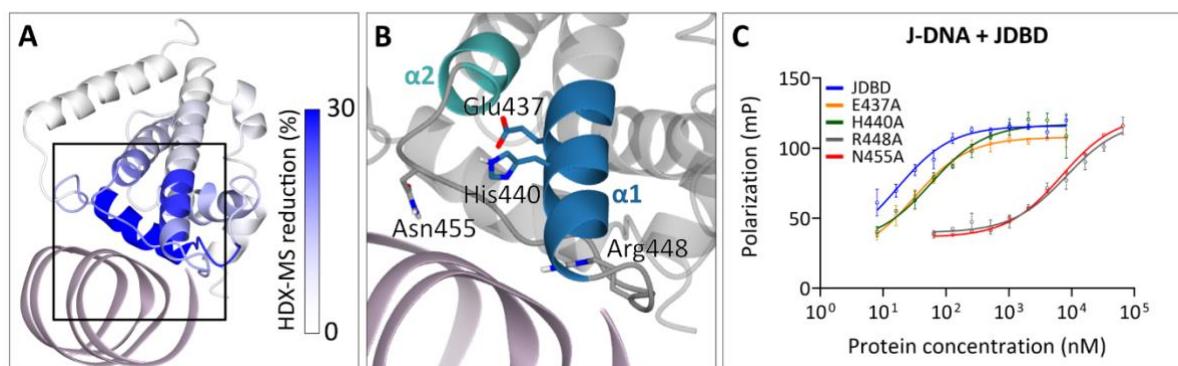
The interaction of base-J with Asp525 as well as the known intramolecular interaction with the phosphate group at J-1 that were used as explicit HADDOCK restraints, and are respected in the models (Supplemental Figure S7). Residues Lys518 and Lys522, which were previously suggested to be involved in complex formation (20), do not form direct interactions with J-DNA in the obtained models. On the other hand, in the complexes with JDBD Cluster 1, 2 and 4, Arg532 seemingly forms hydrogen bonds with the non-bridging phosphoryl oxygen of the nucleic acid two base pairs that are further along the opposite strand (J+2' position) (Supplemental Figure S7A, B and D). In Clusters 3 and 5 this arginine is positioned such that hydrogen bond formation with the non-bridging phosphoryl oxygen of the nucleic acid base pair at position J+3' can be possible (Supplemental Figure S7C and E). However, the distance between the NH hydrogen of Arg532 and the non-bridging phosphoryl oxygen in the DNA is too long ( $\geq 2.5 \text{ \AA}$ ) to form direct hydrogen bonds in all models. Because a docking model is a static representation of the JDBD:J-DNA complex, it is likely that hydrogen bonds or salt bridges are formed with the phosphoryl group at the J+2' and/or the J+3' position to stabilize the negative charges in the J-DNA backbone.

#### Computational modeling provides new insight into JDBD mediated J-DNA recognition

Having verified that our previous hypotheses about the JDBD interaction with J-DNA were basically correct, we examined the models for novel insights focusing on charge conserved residues in the  $\alpha 1$ - $\alpha 2$  region that are interacting with J-DNA. Based on multiple sequence alignment of JBP1 in 28 different

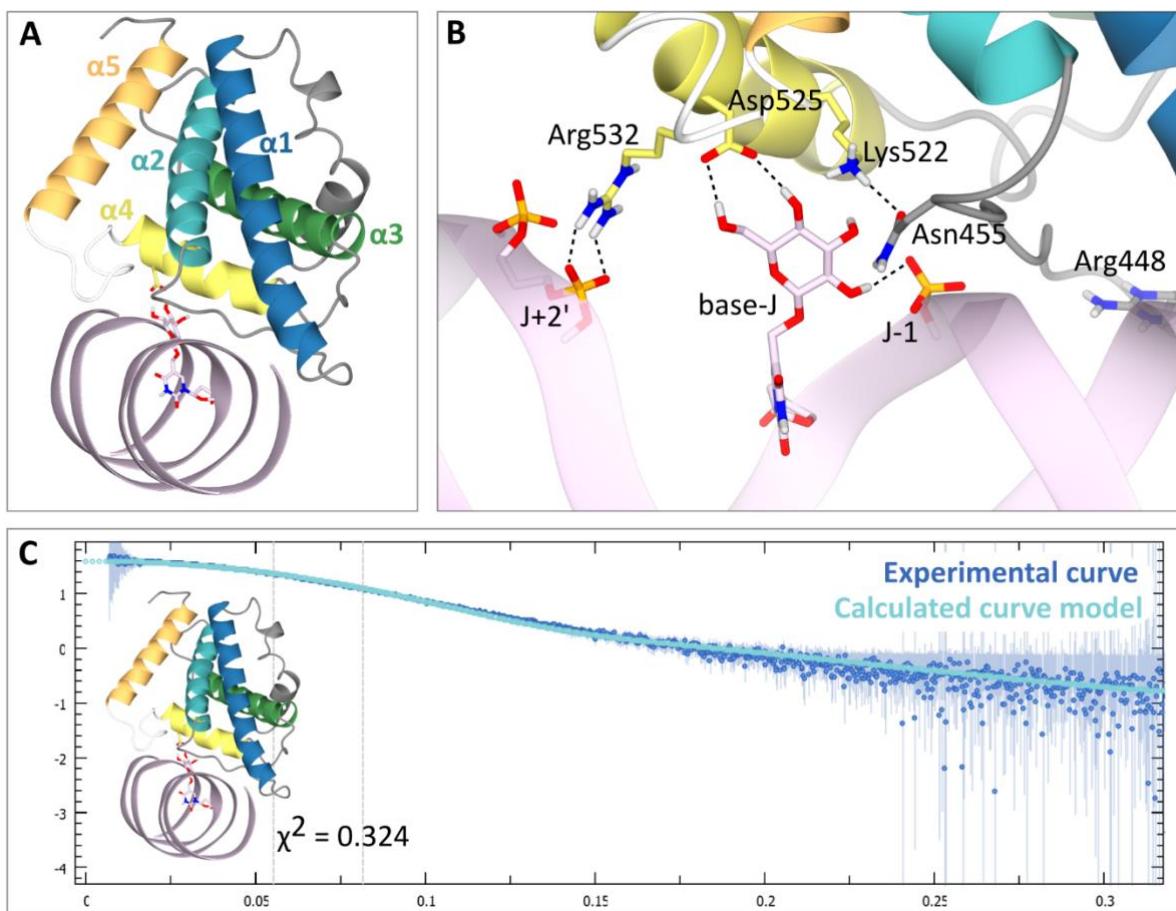
species (Supplemental Figure S8), Glu437, His440, Arg448 and Asn455 were selected for site-directed mutagenesis. Arg448 and Asn455 are located in the  $\alpha$ 1- $\alpha$ 2 loop, fully conserved and in proximity to the J-DNA when the  $\alpha$ 1 helix is oriented towards the DNA. Both residues are also located within the part that shows HDX-MS reduction (9% for peptide 442-463) (Figure 4A and B). The negative charge of residue 437 is conserved in 93% of the species and 86% of residues present at position 440 are positively charged (Supplemental Figure S8). Both Glu437 and His440 are pointing towards the DNA and are also positioned within the region that shows high HDX-MS reduction (23% for the peptide 434-441) (Figure 4A).

The single point mutants JDBD-E437A, -H440A, -R448A and -N455A were produced. For mutant E437A the expression was substantially reduced compared to wt-JDBD and the other JDBD mutants, suggesting that this mutation could also affect the overall stability of the protein which could hamper binding to DNA. To address this issue, the thermal stability of wt- and mutant JDBD proteins was analyzed at two protein concentrations using nanoDSF. The E437A mutant showed about 9-12°C reduction in melting temperature compared to that of wt-JDBD and the other mutants (Supplemental Table S2). With all mutant proteins in hand, binding to both J-DNA and T-DNA was measured by fluorescence polarization. The wt-JDBD showed an affinity of 17nM. Substitution of E437 and H440 only resulted in a minor difference in affinity for J-DNA (30nM and 57nM respectively; Figure 4C and Supplemental Table S3), indicating that these residues have a negligible contribution to the stabilization of the JDBD:J-DNA complex. In contrast, R448A showed a ~500 fold reduction in affinity to J-DNA (~8 $\mu$ M) and N455A a ~400 fold reduction (~7 $\mu$ M). The affinities of all four mutants and wt-JDBD towards T-DNA were all similar (>100 $\mu$ M, Supplemental Figure S9, Supplemental Table S3), indicating that mutation of Arg448 and Asn455 residues induces a loss in discrimination towards J-DNA (20).



**Figure 4:** Arg448 and Asn455, located in the  $\alpha$ 1 helix of JDBD are important for J-DNA recognition. A) JDBD:J-DNA complex Cluster 2 colored by HDX-MS reduction that was mapped onto the complex. J-DNA is colored pink and the black box is indicating the zoom-in area of B. B) Zoom-in on cluster 2 indicating residues 434-463 of the protein with non-transparent colors and residues that were selected for mutagenesis studies indicated. The rest of the protein is colored in transparent grey and J-DNA in pink. C) Polarization curves for wt-JDBD and single point mutants with J-DNA.

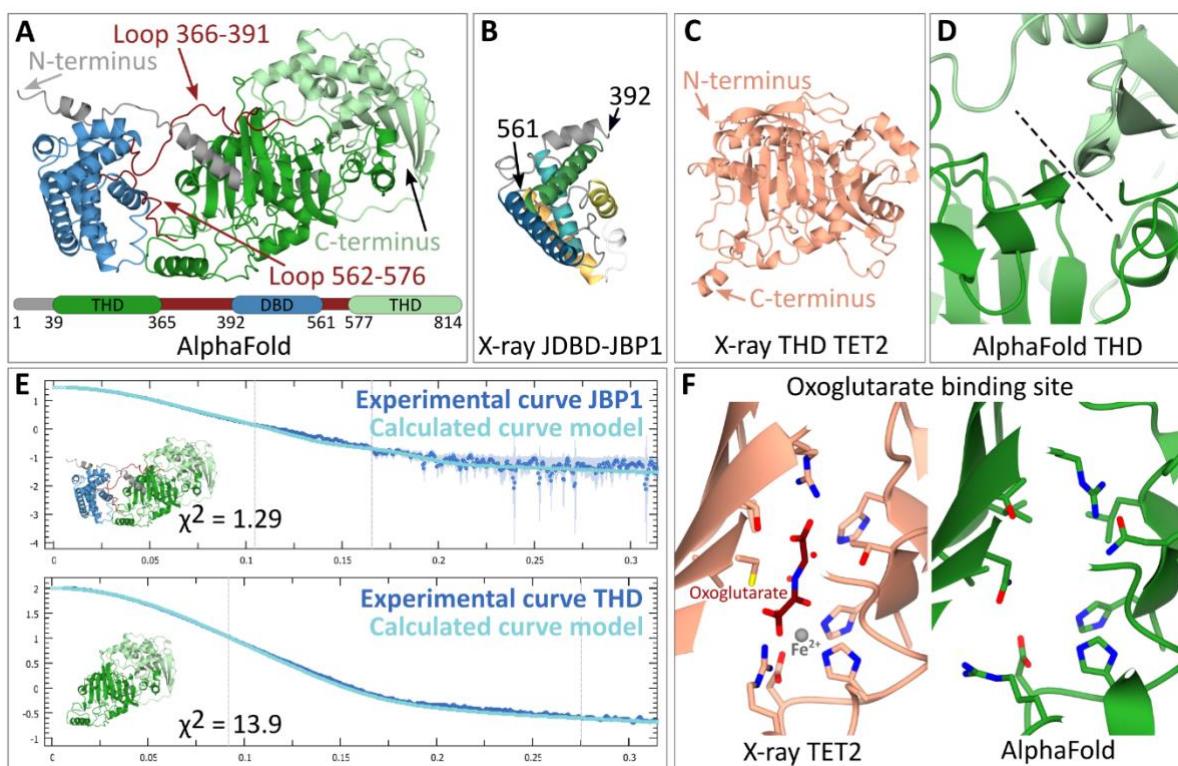
Based on these new experimental insights, residues Arg448 and Asn455 were included as additional active residues in HADDOCK. The results show one clearly most occupied cluster, that contains 79% of the generated structure models (Supplemental Table S4). The best scoring structure in this cluster (cluster1\_1.pdb) (Figure 5A and B) fits the SAXS curve obtained previously (20) (Figure 5C). In addition to the interactions found in our initial docking experiments, Arg532 forms hydrogen bonds with the non-bridging phosphoryl oxygens at position J+2'. Interestingly, Asn455 forms a hydrogen bond with Lys522, which previously was found to be an important residue for the JDBD:J-DNA complex formation (20). This interaction supports Asn455 in forming an additional hydrogen bond with the J-DNA backbone to stabilize the negative charge. The residues Lys518 and Arg447 are also contributing in stabilizing the negative DNA charge distribution by orienting towards the DNA backbone and forming salt bridges with phosphate oxygen atoms. Such interactions are also observed for Arg448 and Arg517, of which the latter was previously suggested to be relevant for complex formation (20).



**Figure 5:** A docking model of the JDBD:J-DNA complex incorporating all available data. A) Overall docking model; J-DNA is colored in pink; base-J is shown as pink cylinders, and Asp525 is highlighted as yellow cylinders. B) Binding site of this model with residues Lys522, Asp525 and Arg532 as yellow cylinders, Asn455 is highlighted as grey cylinders. Base-J is shown as pink cylinders and hydrogen bonds are indicated as black dotted lines. For clarity residues 534 till 537 are not shown (in the white loop). C) Experimental SAXS curve of JDBD:J-DNA (blue) compared to the calculated curve of the obtained docking model (cyan).

## New insights through AI-based modeling of JBP1

The AlphaFold structure prediction model (26) of JBP1 (entry Q9U6M1) (Figure 6A) is a source of new information, that became available during the course of this study. In this predicted model, JDBD shows a similar overall fold when compared to our new crystal structure (Figure 6B), with a noticeable difference in the loop between the  $\alpha$ 4 and  $\alpha$ 5 helix, that we found ordered in the new JDBD crystal structure. The THD of the AlphaFold model was compared to the hydroxylase domain of the human TET2 homologue (Figure 6C and D), which looks similar based on secondary structure elements (RMS difference of 0.35 over 120 selected Ca atoms). The AlphaFold model also shows that  $\beta$ -sheets are continuing between the N- and the C-terminal parts of the THD (Figure 6D) This observation confirms our previous results describing the THD domain (22), which is formed between the N- and C-terminal parts of JBP1 and retained enzymatic activity *in vitro*. The AlphaFold model fits the SAXS curve that was obtained previously for the full-length JBP1 protein and the THD only (Figure 6E). In addition, the residues involved in  $\text{Fe}^{2+}$  and oxoglutarate binding are in similar conformations as in the corresponding binding domain TET2 homologue (Figure 6F). These observations suggest that the AlphaFold prediction provides a reliable description of the JDBD and THD in JBP1 that can be used for further investigation.



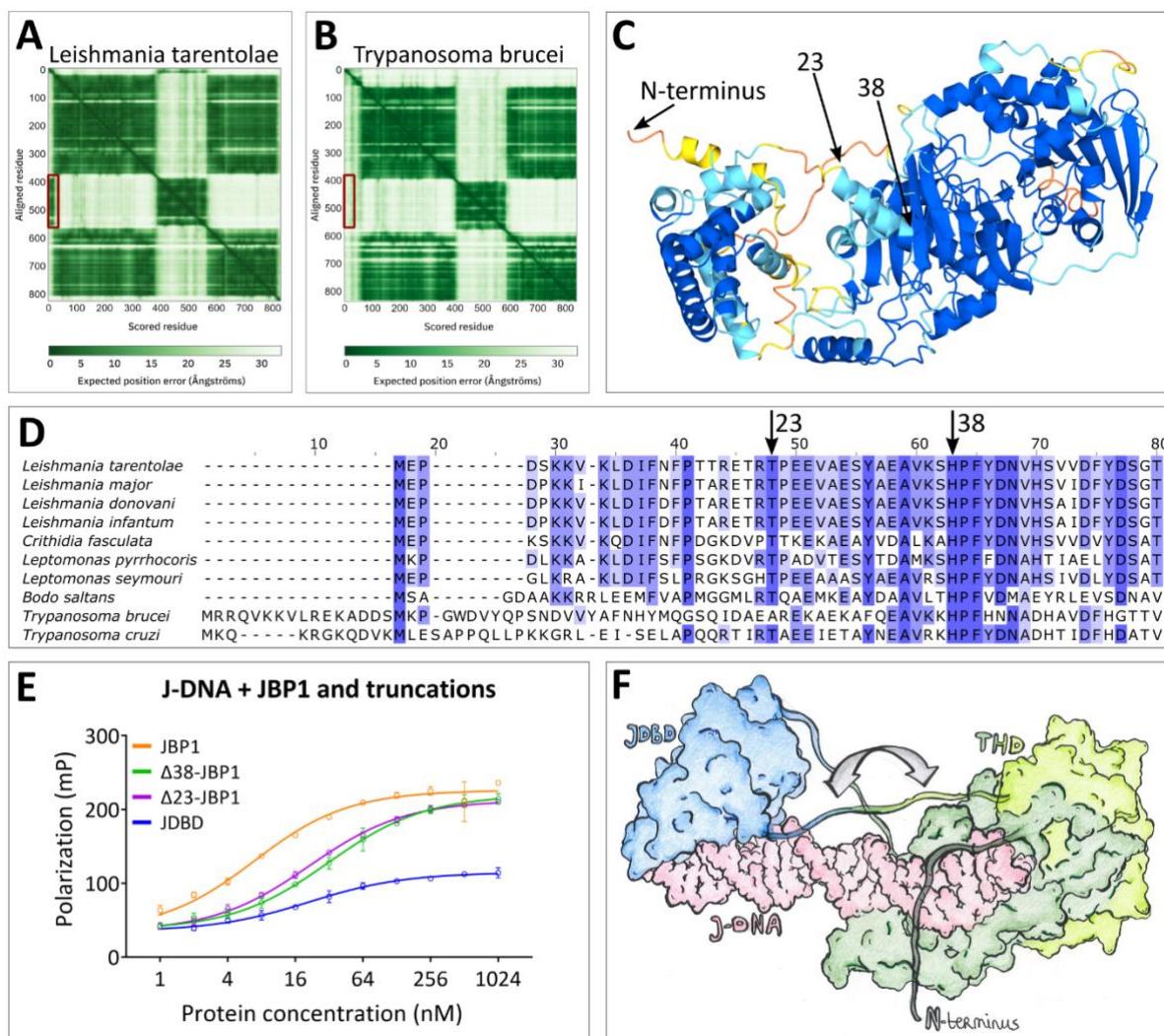
**Figure 6.** The JDBD and THD in the AlphaFold model of JBP1 in *Leishmania Tarentolae* are accurately predicted based on comparison with experimental data. A) Structure of the AlphaFold model of JBP1, the different domains and flexible regions are indicated by coloring and arrows. B) X-ray structure of the JDBD as obtained in this work has the same fold as the JDBD in the AlphaFold model. C) The THD of the JBP1 AlphaFold model adopts a similar fold compared to TET2 homologue D) The  $\beta$ -sheets between the N- and C-terminus of the THD are contiguous. The black dotted line indicates the interface. E) Both the JPB1 model and the THD only are in agreement with the

SAXS data obtained previously. F) The oxoglutarate binding site in the THD of the AlphaFold model is predicted properly compared to the homologues structure of TET2.

The predicted aligned error (PAE) plot of the AlphaFold JBP1 model is consistent with the flexibility between the THD and the JDBD (Figure 7A), which we previously suggested based on experimental data (22). Interestingly, the PAEs for the N-terminal ~40 residues of JBP1 suggest that the N-terminal residues are reliably modelled in proximity to the JDBD. This PAE pattern stands for *Leishmania*, *Cryptosporidium*, and *Leptomonas* species but not for *Trypanosoma* (Figure 7B and Supplemental Figure S10). The AlphaFold model for *Leishmania tarentolae* predicts two short helices in the N-terminal region, connected by a linker. The per-residue confidence scores (pLDDT) range between “low” and “confident” (Figure 7C). In addition, the N-terminal region has positively charged residues that could in principle mediate DNA binding (Figure 7D). Finally, we have previously shown that the JDBD binds to J-DNA with an affinity about three times less than full-length JBP1 (20). These observations and previous findings, prompted us to investigate if the JBP1 N-terminus plays a role in DNA recognition.

To this end, we decided to study the J-DNA binding affinity of JBP1 truncated after the first predicted N-terminal helix (at position 23, Δ23-JBP1) and after both N-terminal helices (at position 38, Δ38-JBP1). The truncated mutants were expressed and purified, and their affinity for J-DNA was studied by fluorescence polarization (Figure 7E). While wt-JBP1 binds DNA with an affinity of 6.4 nM, Δ23-JBP1 binds J-DNA three times weaker (20.8 nM). Removal of all 38 N-terminal residues results in a further decrease in affinity to 31.2 nM (Supplemental Table S5). These findings suggest that the N-terminal residues are involved in J-DNA binding. Notably, the affinity of the JDBD alone on J-DNA is 20.7 nM, which is similar to the N-terminally truncated JBP1. The derived free energy of binding ( $\Delta G$ ) of the JDBD alone or Δ23-JBP1 is about  $-48 \text{ kJ mol}^{-1}$ , while the  $\Delta G$  for wt-JBP1 is  $-45 \text{ kJ mol}^{-1}$ . Thus, the difference in binding free energies between JDBD and the truncated mutants is relatively small ( $3 \text{ kJ mol}^{-1}$  or less).

Based on all these data, we attempted to create a molecular model of JBP1 bound to DNA. First, we used the THD of the AlphaFold model and the X-ray model of a TET2 homologue in complex with DNA, to create a model of the THD of JBP1 in complex with J-DNA. The THD active site was bound to a thymidine 13 nucleotides downstream base-J, according to the preferred offset between an existing and a new base-J (23). Then, the JDBD:J-DNA complex we obtained above, was added to this model by aligning the corresponding DNA regions. However, several modelling attempts did not yield a model that fits with the SAXS data (Supplemental Figure S11). During the modelling process, it also became apparent that the N-terminal region in the exact conformation of the AlphaFold model does not allow DNA binding by the JDBD. Based on these findings and as illustrated in Figure 7F, we propose that while the N-terminal region of JBP1 wraps around DNA, the combined binding of the THD and JDBD to J-DNA is accompanied by structural rearrangements which likely include DNA bending to allow simultaneous binding of the JDBD and THD-mediated hydroxylation of a thymidine that is located 13 nucleotides downstream in the opposing DNA strand.



**Figure 7.** An N-terminal sequence of JBP1 contributes to DNA binding. A, B) The PAE plot for the model of JBP1 from (A) *Leishmania Tarentolae* and (B) *Trypanosoma brucei* from the AlphaFold database (entries Q9U6M and P86937, respectively); the PAE regions indicative of the association of the N-terminal sequence with the JBD are marked with a red box. C) The AlphaFold model of the full length JBP1 protein colored by pLDDT values; orange and yellow indicate very low and low confidence respectively, while light blue and blue indicate confidence and high confidence regions respectively. D) Indicative alignment of JBP1 in different species. The N-terminus of JBP1 is highly conserved between *Leishmania* species. E) Polarization curves of JBP1 and truncation mutants binding to J-DNA. F) Drawing of the model of the JBD and THD of JBP1 in complex with J-DNA. The binding sites of base-J and the T base that should be hydroxylated to hmU, are located 13 base pairs apart in opposing strands. The grey arrow indicates conformational changes that need to occur to be able to provide a convincing JBP1:J-DNA model. JBD is colored blue and is based on the JBD:J-DNA docking model, THD is colored green (C-terminus dark, N-terminus light green) and originated from the AlphaFold model, J-DNA is colored pink and was constructed manually based on the human TET2 homologue structure (PDB-ID: 5DEU).

## CONCLUSIONS

A computational modeling study helped to identify two new residues in the JDBD of JBP1, Arg448 and Asn455, that are involved in J-DNA binding. After experimental validation by mutational analysis, these residues were added as additional guidelines in docking. This allowed for the construction of a new, improved model for the JDBD:J-DNA complex, that was compatible with SAXS data. Further, the AlphaFold model for full-length JBP1, showed that the individual domains (JDBD and THD) are predicted confidently. This model allowed us to hypothesize that the N-terminus of JBP1 for many *Trypanosomatidae* species (including *Leishmania*) is involved in DNA binding, which was confirmed by mutational analysis. The AlphaFold model, in combination with the X-ray structure model of TET2 in complex with DNA and our JDBD:J-DNA docking model, did now allow to construct a reliable model structure for full length JBP1 bound to J-DNA. Future experiments are required to obtain a structural model for such a complex, which must involve conformational changes between the JDBD and THD domains of JBP1. Such model would be essential to provide more detailed insights in the molecular mechanism of the unique biochemistry of base-J synthesis that facilitates replication of epigenetic information.

## ACCESSION NUMBERS

Atomic coordinates and structure factors for the reported JDBD crystal structure has been deposited with the Protein Data bank under accession number 8BBM.

## SUPPLEMENTARY DATA

Supplementary Data are available as one PDF formatted file.

## ACKNOWLEDGEMENT

We thank Piet Borst for extensive discussions and suggestions to improve the text of this manuscript, Rosa A. Luijink for providing example scripts for the MD simulations and its analyses, and Alexander Fish for guidance while performing the fluorescence polarization assays. We thank the NKI's Research High Performance Computing for the computational infrastructure for the MD calculations. Additionally, we thank Alexandre M.J.J. Bonvin and Maxim Petoukhob for implementing the base-J parameters in the HADDOCK2.4 webserver, and for the advises regarding modelling using CORAL, respectively. All students from the Cold Spring Harbor Laboratory class of 2018 are acknowledged for their work on JDBD crystallization and structure modeling.

## FUNDING

This work was supported by a TOP grant of the Nederlandse Organisatie voor Wetenschappelijk Onderzoek to AP [grant number 714.014.002] and by an institutional grant of the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport.

## CONFLICT OF INTEREST

None declared.

## REFERENCES

1. Gommers-Ampt,J.H., Van Leeuwen,F., de Beer,A.L.J., Vliegenthart,J.F.G., Dizdaroglu,M., Kowalak,J.A., Crain,P.F. and Borst,P. (1993)  $\beta$ -d-glucosyl-hydroxymethyluracil: A novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell*, **75**, 1129–1136.
2. Genest,P.-A., ter Riet,B., Cijssouw,T., van Luenen,H.G.A.M. and Borst,P. (2007) Telomeric localization of the modified DNA base J in the genome of the protozoan parasite *Leishmania*. *Nucleic Acids Research*, **35**, 2116–2124.
3. van Leeuwen,F., Wijsman,E.R., Kuyl-Yeheskiely,E., van der Marel,G.A., van Boom,J.H. and Borst,P. (1996) The Telomeric GGGTTA Repeats of *Trypanosoma Brucei* Contain the Hypermodified Base J in Both Strands. *Nucleic Acids Research*, **24**, 2476–2482.
4. Leeuwen,F. van, Wijsman,E.R., Kieft,R., Marel,G.A. van der, Boom,J.H. van and Borst,P. (1997) Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev.*, **11**, 3232–3241.
5. van Luenen,H.G.A.M., Farris,C., Jan,S., Genest,P.-A., Tripathi,P., Velds,A., Kerkhoven,R.M., Nieuwland,M., Haydock,A., Ramasamy,G., *et al.* (2012) Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in *Leishmania*. *Cell*, **150**, 909–921.
6. Reynolds,D., Cliffe,L., Förstner,K.U., Hon,C.-C., Siegel,T.N. and Sabatini,R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania* major and *Trypanosoma brucei*. *Nucleic Acids Research*, **42**, 9717–9729.
7. Cliffe,L.J., Siegel,T.N., Marshall,M., Cross,G.A.M. and Sabatini,R. (2010) Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei*. *Nucleic Acids Research*, **38**, 3923–3935.
8. Yu,Z., Genest,P.-A., ter Riet,B., Sweeney,K., DiPaolo,C., Kieft,R., Christodoulou,E., Perrakis,A., Simmons,J.M., Hausinger,R.P., *et al.* (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Research*, **35**, 2107–2115.
9. DiPaolo,C., Kieft,R., Cross,M. and Sabatini,R. (2005) Regulation of Trypanosome DNA Glycosylation by a SWI2/SNF2-like Protein. *Molecular Cell*, **17**, 441–451.
10. Ito,S., D'Alessio,A.C., Taranova,O.V., Hong,K., Sowers,L.C. and Zhang,Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
11. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L., *et al.* (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, **324**, 930–935.
12. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base, 5-hydroxymethylcytosine is present in brain and enriched in Purkinje neurons. *Science*, **324**, 929–930.
13. Bullard,W., Lopes da Rosa-Spiegler,J., Liu,S., Wang,Y. and Sabatini,R. (2014) Identification of the Glucosyltransferase That Converts Hydroxymethyluracil to Base J in the Trypanosomatid Genome. *Journal of Biological Chemistry*, **289**, 20273–20282.

14. Sekar,A., Merritt,C., Baugh,L., Stuart,K. and Myler,P.J. (2014) Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, **196**, 9–11.
15. Iyer,L.M., Zhang,D., Maxwell Burroughs,A. and Aravind,L. (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Research*, **41**, 7635–7655.
16. Jensen,B.C., Phan,I.Q., McDonald,J.R., Sur,A., Gillespie,M.A., Ranish,J.A., Parsons,M. and Myler,P.J. (2021) Chromatin-Associated Protein Complexes Link DNA Base J and Transcription Termination in *Leishmania*. *mSphere*, **6**, e01204-20.
17. Kieft,R., Zhang,Y., Marand,A.P., Moran,J.D., Bridger,R., Wells,L., Schmitz,R.J. and Sabatini,R. (2020) Identification of a novel base J binding protein complex involved in RNA polymerase II transcription termination in trypanosomes. *PLOS Genetics*, **16**, e1008390.
18. Cross,M., Kieft,R., Sabatini,R., Wilm,M., de Kort,M., van der Marel,G.A., van Boom,J.H., van Leeuwen,F. and Borst,P. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *The EMBO Journal*, **18**, 6573–6581.
19. Sabatini,R., Meeuwenoord,N., van Boom,J.H. and Borst,P. (2002) Recognition of Base J in Duplex DNA by J-binding Protein. *Journal of Biological Chemistry*, **277**, 958–966.
20. Heidebrecht,T., Christodoulou,E., Chalmers,M.J., Jan,S., ter Riet,B., Grover,R.K., Joosten,R.P., Littler,D., van Luenen,H., Griffin,P.R., *et al.* (2011) The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic Acids Research*, **39**, 5715–5728.
21. Heidebrecht,T., Fish,A., Borst,P. and Perrakis,A. (2012) Binding of the J-binding protein to DNA containing glucosylated hmU (base J) or 5-hmC; evidence for a rapid conformational change upon DNA binding.
22. Adamopoulos,A., Heidebrecht,T., Roosendaal,J., Touw,W.G., Phan,I.Q., Beijnen,J. and Perrakis,X.A. (2019) The domain architecture of the protozoan protein J-DNA– binding protein 1 suggests synergy between base J DNA binding and thymidine hydroxylase activity.
23. Genest,P.-A., Baugh,L., Taipale,A., Zhao,W., Jan,S., van Luenen,H.G.A.M., Korlach,J., Clark,T., Luong,K., Boitano,M., *et al.* (2015) Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Research*, **43**, 2102–2115.
24. Grover,R.K., Pond,S.J.K., Cui,Q., Subramaniam,P., Case,D.A., Millar,D.P. and Wentworth,P. (2007) O-Glycoside Orientation Is an Essential Aspect of Base J Recognition by the Kinetoplastid DNA-Binding Protein JBP1. *Angew. Chem. Int. Ed.*, **46**, 2839–2843.
25. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
26. Tunyasuvunakool,K., Adler,J., Wu,Z., Green,T., Zielinski,M., Žídek,A., Bridgland,A., Cowie,A., Meyer,C., Laydon,A., *et al.* (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, **596**, 590–596.

27. Hu,L., Lu,J., Cheng,J., Rao,Q., Li,Z., Hou,H., Lou,Z., Zhang,L., Li,W., Gong,W., *et al.* (2015) Structural insight into substrate preference for TET-mediated oxidation. *Nature*, **527**, 118–122.
28. Fuchs,M., Sweet,R., Berman,L., Bhogadi,D., Hendrickson,W., Chubar,O., Yang,L., McSweeney,S. and Schneider,D. (2014) NSLS-II MX beamlines FMX for micro-crystallography & AMX for highly automated MX. *Acta Cryst*, **A70**, C1733–C1733.
29. Potterton,L., Agirre,J., Ballard,C., Cowtan,K., Dodson,E., Evans,P.R., Jenkins,H.T., Keegan,R., Krissinel,E., Stevenson,K., *et al.* (2018) CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Cryst D*, **74**, 68–84.
30. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G.W., McCoy,A., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Cryst D*, **67**, 235–242.
31. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Cryst D*, **69**, 1204–1214.
32. Murshudov,G.N., Skubák,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst D*, **67**, 355–367.
33. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486–501.
34. Joosten,R.P., Long,F., Murshudov,G.N. and Perrakis,A. (2014) The PDB\_REDQ server for macromolecular structure model optimization. *IUCrJ*, **1**, 213–220.
35. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12–21.
36. Berman,H.M., Battistuz,T., Bhat,T.N., Bluhm,W.F., Bourne,P.E., Burkhardt,K., Feng,Z., Gilliland,G.L., Iype,L., Jain,S., *et al.* (2002) The Protein Data Bank. *Acta Cryst D*, **58**, 899–907.
37. Lindahl, Abraham, Hess and van der Spoel (2020) GROMACS 2020.2 Source code. 10.5281/zenodo.3773801.
38. Lindorff-Larsen,K., Piana,S., Palmo,K., Maragakis,P., Klepeis,J.L., Dror,R.O. and Shaw,D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.
39. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
40. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, **18**, 1463–1472.
41. Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., DiNola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.

42. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.

43. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, **14**, 33–38.

44. Michael Waskom, Maoz Gelbart, Olga Botvinnik, Joel Ostblom, Paul Hobson, Saulius Lukauskas, David C Gempertline, Tom Augspurger, Yaroslav Halchenko, Jordi Warmenhoven, *et al.* (2020) mwaskom/seaborn: v0.11.1 (December 2020). 10.5281/zenodo.4379347.

45. González-Alemán,R., Hernández-Castillo,D., Rodríguez-Serradet,A., Caballero,J., Hernández-Rodríguez,E.W. and Montero-Cabrera,L. (2019) BitClust: Fast Geometrical Clustering of Long Molecular Dynamics Simulations. *J. Chem. Inf. Model.*, **60**, 444–448.

46. Daura,X., Gademann,K., Jaun,B., Seebach,D., Gunsteren,W.F. van and Mark,A.E. (1999) Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition*, **38**, 236–240.

47. van Zundert,G.C.P., Rodrigues,J.P.G.L.M., Trellet,M., Schmitz,C., Kastritis,P.L., Karaca,E., Melquiond,A.S.J., van Dijk,M., de Vries,S.J. and Bonvin,A.M.J.J. (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, **428**, 720–725.

48. Schrödinger (LLC) PyMOL 2.4.0. <https://pymol.org/>.

49. Krieger,E. and Vriend,G. (2014) YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics*, **30**, 2981–2982.

50. McNicholas,S., Potterton,E., Wilson,K.S. and Noble,M.E.M. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Cryst D*, **67**, 386–394.

51. Manalastas-Cantos,K., Konarev,P.V., Hajizadeh,N.R., Kikhney,A.G., Petoukhov,M.V., Molodenskiy,D.S., Panjkovich,A., Mertens,H.D.T., Gruzinov,A., Borges,C., *et al.* (2021) ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Cryst*, **54**, 343–355.

52. Luna-Vargas,M.P.A., Christodoulou,E., Alfieri,A., van Dijk,W.J., Stadnik,M., Hibbert,R.G., Sahtoe,D.D., Clerici,M., Marco,V.D., Littler,D., *et al.* (2011) Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases. *Journal of Structural Biology*, **175**, 113–119.

53. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry*, **25**, 1157–1174.

54. Sobolev,O.V., Afonine,P.V., Moriarty,N.W., Hekkelman,M.L., Joosten,R.P., Perrakis,A. and Adams,P.D. (2020) A Global Ramachandran Score Identifies Protein Structures with Unlikely Stereochemistry. *Structure*, **28**, 1249-1258.e2.