The domain architecture of the protozoan protein J-DNA–binding protein 1 suggests synergy between base J DNA binding and thymidine hydroxylase activity

Athanassios Adamopoulos¹, Tatjana Heidebrecht¹, Jeroen Roosendaal⁵, Wouter G. Touw⁴, Isabelle Q. Phan⁶, Jos Beijnen¹, and Anastassios Perrakis¹²

From the Departments of Biochemistry and Clinical Pharmacology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands and Seattle Structural Genomics Center for Infectious Disease, Center for Global Infectious Disease Research, Seattle Children’s Research Institute, Seattle, Washington 98109

Edited by Qi-Qun Tang

J-DNA–binding protein 1 (JBP1) contributes to the biosynthesis and maintenance of base J (β-d-glucosyl-hydroxymethyluracil), an epigenetic modification of thymidine (T) confined to pathogenic protozoa such as Trypanosoma and Leishmania. JBP1 has two known functional domains: an N-terminal T hydroxylase (TH) domain homologous to the 5-methylcytosine hydroxylase domain in TET proteins and a J-DNA–binding domain (JDBD) that resides in the middle of JBP1. Here, we show that removing JDBD from JBP1 results in a soluble protein (Δ-JDBD) that resides in the middle of JBP1. We postulate that this Δ-JDBD domain retains TH activity in vitro but displays a 15-fold lower apparent rate of hydroxylation compared with JBP1. Small-angle X-ray scattering (SAXS) experiments on JBP1 and JDBD in the presence or absence of J-DNA and on Δ-JDBD enabled us to generate low-resolution three-dimensional models. We conclude that Δ-JDBD, and not the N-terminal region of JBP1 alone, is a distinct folding unit. Our SAXS-based model supports the notion that binding of JDBD specifically to J-DNA can facilitate T hydroxylation 12–14 bp downstream on the complementary strand of the J-recognition site. We postulate that insertion of the JDBD module into the Δ-JDBD scaffold during evolution provided a mechanism that synergized J recognition and T hydroxylation, ensuring inheritance of base J in specific sequence patterns following DNA replication in kinetoplastid parasites.

Base J (β-d-glucosyl-hydroxymethyluracil) is a modified base that replaces 1% of thymine in kinetoplastid flagellates (1), such as Trypanosoma, Leishmania, and Crithidia. In Leishmania, 99% of J is located in telomeric repeats (2–4), whereas 1% is in internal chromosomal positions where transcription starts (5) or stops (6).

Bio- synthesis of J occurs in two steps. First, the 5-methyl group of specific Ts in the genome is hydroxylated, forming hydroxymethyluracil (hmU). A second, a glucose molecule is transferred to hmU, resulting in J (7–9). The first step is catalyzed by both J-DNA–binding proteins 1 and 2 (JBP1 and JBP2), which have a distinct thymidine hydroxylase (TH) domain in their N terminus (10). The hydroxylation of T in oligonucleotides depends on the presence of Fe(II) and 2-oxoglutarate (11). The discovery of the hydroxylase function of JBP1 has led to the discovery of the function of the mammalian TET family of enzymes that convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (hmC) (12) and have crucial roles in epigenetic regulation through modification of 5mC to hmC. Several structures of TET and TET-like hydroxylase domains have been determined (13–15), including in complex with 5mC and hmC, providing significant insight into 5mC hydroxylation. However, the limited sequence similarity between TET and JBP, underlined by large deletions and insertions in the TH fold, makes it impossible to deduce the structures of JBP1 and JBP2 from that of TET.

JBP1, but not JBP2, specifically recognizes base J in DNA (16). This recognition is mediated by a short ~150-residue domain in the middle of JBP1, the J-DNA–binding domain (JDBD) (17). JDBD recognizes J-DNA with high affinity (~10 nM) and remarkable specificity over normal DNA (~10,000-fold). The structure of JDBD revealed a novel variant of the helix–turn–helix domain, with an unusually elongated turn between the recognition and the supporting helices (17). Importantly, we have shown that a single residue (Asp-525) in the recognition helix is almost entirely responsible for the specificity toward J-DNA, as the D525A mutation abrogated specificity toward J-DNA both in vitro and in vivo (17).

This work was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek Grant 714.014.002. Work from the Seattle Structural Genomics Center for Infectious Disease was supported by NIAID, National Institutes of Health, Department of Health and Human Services, under Contract HHSN272201700059C. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains Figs. S1–S6 and Tables S1–S3.

¹ Both authors contributed equally to this work. The first two authors should be regarded as joint first authors.

‡ To whom correspondence should be addressed. E-mail: a.perrakis@nki.nl.

The abbreviations used are: hmU, hydroxymethyluracil; JBP, J-DNA–binding protein; TH, thymidine hydroxylase; 5mC, 5-methylcytosine; hmC, 5-hydroxymethylcytosine; JDBD, J-DNA–binding domain; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; PDB, Protein Data Bank; Vₚ, volume of correlation; Rₓ, radius of gyration; CHCT, C-terminal helical domain; Acr, anti-CRISPR; HOMedU, 5-hydroxymethyl-2′-deoxyuridine; Dₓₘₐₓ, maximum distance; r.m.s.d., root mean square deviation; TET, Ten-Eleven Translocation.
**Domain architecture of JBP1**

JBP1 recognizes and binds J-DNA in two steps (18). Pre-steady-state kinetic data revealed that the initial binding of JBP1 to glucosylated DNA is very fast and followed by a second, much slower, and concentration-independent step. From this observation and small-angle neutron scattering experiments, we inferred that JBP1 undergoes a conformational change upon binding to DNA and postulated that this may allow the hydroxylase domain of JBP1 to make contact with the DNA and hydroxylate Ts in spatial proximity.

From what we know about the mechanism of J biosynthesis, it follows that the highly restricted distribution of base J must be codetermined by the thymidine hydroxylases (JBP1 and JBP2) that catalyze the initial step in J synthesis. Using single-molecule real-time sequencing of DNA segments inserted into plasmids grown in *Leishmania* (19), it has been shown that J modification usually occurs near G-rich sequences potentially capable of forming G-quadruplexes and at pairs of Ts on opposite DNA strands separated by 12 nucleotides. That led Genest *et al.* (19) to propose a model in which JBP2 is responsible for initial J synthesis; then JBP1 binds to pre-existing J and hydroxylates another T that typically resides 13 bp downstream (but not upstream) on the complementary DNA strand. This model provides a mechanism explaining how JBP1 can maintain existing J following DNA replication.

On the basis of these results, we developed the conformational change model presented in Ref. 18. We postulated that J binding docks JBDB on duplex DNA, allowing the TH domain to come into contact with a T that is preferably 13 bp downstream (but not upstream) on the complementary strand and hydroxylate it. To further test this model, we set out to understand the domain organization of JBP1 and describe its three-dimensional organization alone and in relation with DNA.

Here, we present new deletion mutants and hydroxylase activity data of JBP1, resulting in a new definition for the TH domain. We also studied the domain architecture of JBP1 by small-angle X-ray scattering (SAXS). Developments in SAXS, namely improved software and hardware, have established SAXS as a powerful tool for the analysis of molecular structures, including multidomain, flexible molecules (20–23). Coupling size-exclusion chromatography (SEC) to SAXS allows separation of complexes from the constituent partners, degradation products, and eventual contaminants, allowing the determination of particle size and shape of macromolecules. Collecting and analyzing SAXS data on different JBP1 deletion mutants and their complexes with J-DNA allowed us to generate low-resolution three-dimensional models of JBP1 and its complex with DNA. Our data suggest that synergy between the TH and JBDB domains (that is likely a recurrent fold in nature, not solely confined to JBP1 orthologues) is an evolutionary adaptation, crucial for replicating epigenomic information in kinetoplastids.

**Results**

**The N- and C-terminal regions of JBP1 behave as a single folding unit**

Countless previous attempts to truncate JBP1 either C-terminally or N-terminally (before or after the JBDB), to obtain the N-terminal TH domain or a putative C-terminal domain, had invariably failed to yield soluble protein in our hands. The crystal structure of JBDB showed that the N and C termini of this domain are in proximity (Fig. 1A). As JBDB is in the middle of the JBP1 sequence (Fig. 1B), we decided to test the possibility that the JBDB is an insertion domain into a “TH domain” fold that spans the rest of the JBP1 sequence. To validate this hypothesis, we replaced the JBDB domain with a connecting linker. Remarkably, the resulting protein (JBP11–382/561–882; ∆-JBDB) was expressed well in soluble form and could be purified in good amounts (Fig. 1C). Incubation with 3C protease over a period of 13–36 h resulted in the protein chain being cleaved in two, and these bands could be observed on SDS-PAGE (Fig. 1C). However, when we run the cleaved protein on an SEC column in the absence of detergent, the elution profile showed a single symmetric peak (Fig. 1D) of approximately the same molecular weight as ∆-JBDB, whereas the SDS-PAGE analysis of the eluted fractions confirmed that this peak has both bands present. This strongly suggests that the two polypeptides behave as one protein, indicating a strong interaction between the two termini. This indicates that both termini should be part of the same folding unit, which is the folding unit necessary to provide a functional catalytic site containing the TH activity. This experiment led us to revisit our previous view of JBP1 with an N-terminal TH domain followed by the JBDB domain and a mysterious C-terminal region. We now hypothesized that two folding units compose JBP1: the ∆-JBDB and the JBDB that is “inserted” in the ∆-JBDB folding scaffold.

**JBP1, JBDB, and ∆-JBDB are all well-folded globular domains in solution**

To further characterize ∆-JBDB in solution, we performed small-angle X-ray scattering experiments on JBP1, JBDB, and ∆-JBDB. All samples were injected in an SEC column, and the SAXS profile (as well as the absorption spectrum) was analyzed under flow. All samples eluted as single peaks from the SEC column, and the SAXS curves from the absorption peak region were averaged to obtain a scattering curve for each component (Fig. 2A and Table S1). Standard analysis tools from the ATSAS (24) and ScÅtter (20) suites were used to obtain model-independent parameters (Table 1). To confirm that all samples were properly folded, we performed a dimensionless Kratky plot analysis (Fig. 2B). This plot allows comparison of the shape of particles independently of their size and shows that both JBP1 and ∆-JBDB have a similar profile with a maximum close to 1.104, characteristic of compacted and folded molecules. The JBDB peak is shifted slightly right and upward, suggesting that JBDB is less globular but compacted (which agrees with the crystal structure shape). The pair distribution function (Fig. 2C) is compatible with this analysis, and together they confirm our interpretation of the biochemical experiments, suggesting that ∆-JBDB is a stable single domain.
H9004 - JDBD is a catalytically active domain that has thymidine hydroxylase activity. We first established an MS-based assay to measure TH activity of JBP1 in vitro. We used purified proteins with a 14-mer.

Figure 1. Domain organization of JBP1. A, cartoon representation of JDBD crystal structure, showing the characteristic helical bouquet fold containing the helix-turn-helix motif. Asp-525 is shown as a stick. B, constructs used in this study: full-length JBP1, Δ-JDBD-3C, and Δ-JDBD. C, SDS-PAGE showing the overexpressed constructs in the following order: marker; JDBD, JBP1, Δ-JDBD-3C, and Δ-JDBD-3C products after 13- and 36-h incubation with 3C protease; marker; SEC elution fractions of Δ-JDBD-3C after incubation with 3C protease. D, chromatogram of the SEC for JBP1, Δ-JBP1, and Δ-JDBD-3C. Δ-JBP1 before and after cleavage with 3C protease are identical. The N- (N-ter) and C-terminal (C-ter) regions of Δ-JDBD-3C elute in the same peak (see C above), suggesting that there is a strong interaction between them. mAu, milli-absorbance units.

Figure 2. Folding of JBP1, Δ-JDBD, JDBD, and 23 J-DNA. A, experimental scattering curves of all components. B, dimensionless Kratky plot; the 1.104 maximum for an ideal compacted molecule is shown as the intersection between the two gray lines. C, normalized pair distribution function for all components.

Δ-JDBD is a catalytically active domain that has thymidine hydroxylase activity

We first established an MS-based assay to measure TH activity of JBP1 in vitro. We used purified proteins with a 14-mer.
oligonucleotide in conditions similar to those reported previously (11) to convert T to hmU, which was measured by quantitative LC-MS after converting the oligonucleotide to nucleosides (see “Experimental procedures” for details). This activity was fully dependent on the presence of the cofactor 2-oxoglutarate and showed a modest but appreciable dependence on Fe^{2+}, ascorbic acid as a reducing agent, and buffer degassing (Fig. S1). We then monitored the rate of catalysis over time for both JBP1 and Δ-JDBD. Δ-JDBD was clearly active but showed a catalytic rate of about 17-times lower than that of WT JBP1 (Fig. 3).

These experiments clearly establish that Δ-JDBD is a well-folded active TH domain, which is not disrupted by splicing out the JDBD domain. We thus decided to characterize the relative domain architecture between Δ-JDBD and JDBD in the context of the JBP1 protein.

**Modeling of JBP1 as a two-domain (Δ-JDBD and JDBD) molecule shows flexibility for JDBD**

We first further compared JBP1 with Δ-JDBD by examining the Porod–Debye plot (Fig. 4A). The presence of a plateau for Δ-JDBD indicates that it forms a distinct particle with sharp scattering contrast. This feature is not present in JBP1, indicating a more diffuse scattering contrast for the full-length JBP1. The same is observed when the packing densities for both Δ-JDBD and JBP1 were calculated: Δ-JDBD has a packing density of 0.91 g cm\(^{-3}\) compared with 0.80 g cm\(^{-3}\) for JBP1. The observed diffusion in scattering contrast and the reduced packing density in the WT, full-length JBP1 compared with Δ-JBP1 suggest that the JDBD domain is flexible with respect to the Δ-JDBD scaffold.

To validate this hypothesis, we decided to create *ab initio* three-dimensional models based on the SAXS data. As we have SAXS data for Δ-JDBD and JDBD alone as well as for both of them together (JBP1), we decided it is more appropriate to model them with the procedures developed for macromolecular complexes and multidomain proteins (24, 25). The program MONSA from the ATSAS suite (24) identified five clusters (Fig. 4B), all of which have JDBD located on the same end of the elongated Δ-JDBD scaffold but in various positions around the long axis of the Δ-JDBD domain (Fig. 4C). If Δ-JDBD is viewed as an ellipsoid, the JDBD is consistently positioned toward one half of the ellipsoid but adopts multiple conformations around the long axis of the ellipsoid. This analysis is compatible with the model-independent analysis of the SAXS data and strengthens our previous hypothesis that the JDBD domain is flexible with respect to the Δ-JDBD scaffold.

**Binding of JBP1 to J-DNA leads to reduced flexibility of the JDBD domain**

An examination of the individual models showed that the JDBD domain adopts multiple conformations with respect to the Δ-JDBD scaffold. Clustering analysis with the program DAMCLUST from the ATSAS suite (24) identified five clusters (Fig. 4B), all of which have JDBD located on the same end of the elongated Δ-JDBD scaffold but in various positions around the long axis of the Δ-JDBD domain (Fig. 4C). If Δ-JDBD is viewed as an ellipsoid, the JDBD is consistently positioned toward one half of the ellipsoid but adopts multiple conformations around the long axis of the ellipsoid. This analysis is compatible with the model-independent analysis of the SAXS data and strengthens our previous hypothesis that the JDBD domain is flexible with respect to the Δ-JDBD scaffold.

**Table 1**

| Monomers     | \(\text{R}_{\text{g, Guinier}}\) | \(\text{R}_{\text{g, real space}}\) | \(D_{\text{max}}\) | \(\text{MW}_{\text{theor}}\) | \(V_{\text{Porod}}\) | \(V_{\text{Porod}}\) | \(V_{\text{r}}\) |
|--------------|-----------------|-----------------|---------------|-----------------|-----------------|-----------------|---------------|
| JBP1         | 34.2 ± 0.2      | 35.3            | 120           | 93,403          | 167,838         | 98.7            | 610.1         |
| Δ-JDBD       | 31 ± 0.2        | 30.8            | 99            | 72,118          | 129,990         | 76.4            | 505.5         |
| JDBD         | 21.8 ± 0.9      | 21.8            | 71            | 21,285          | 48,240          | 28.3            | 260.9         |
| J-23-DNA     | 22.5 ± 0.1      | 23.4            | 73            | 14,241          | 27,627          | 16.2            | 201.7         |
| J-15-DNA     | 15.8 ± 0.3      | 16.1            | 51            | 9,301           | 17,658          | 10.3            | 134.0         |
| Complexes    |                 |                 |               |                 |                 |                 |               |
| JBP1:J-23-DNA| 40.98 ± 1       | 40.8            | 141           | 107,644         | 182,075         | 107.1           | 685.8         |
| JBP1:J-15-DNA| 35.6 ± 0.1      | 36              | 120           | 102,704         | 176,907         | 104             | 641.9         |
| JDBD:J-23-DNA| 25.2 ± 0.3      | 25.33           | 86            | 35,526          | 70,587          | 41.5            | 306.7         |

**Figure 3. The apparent catalytic rates of JBP1 and Δ-JDBD.** The linear parts of the reaction curves (5–60 min for JBP1; 10–240 min for Δ-JDBD) are fitted with a line to estimate the rates.

We have previously shown that JBP1 and J-DNA complex formation is accompanied by a conformational change (18). Our new data allow us to formulate the hypothesis that this conformational change might be the ordering of the JDBD: when JBP1 binds to DNA, JDBD might adopt a more defined conformation with respect to the Δ-JBP1 scaffold. To test this hypothesis, we used the SEC-SAXS data on JBP1 in complex with 23-mer J-DNA (J-23-DNA). SEC-SAXS data were first collected for J-23-DNA, which had the expected parameters for an elongated molecule (Table 1 and Table S1). The complex between J-23-DNA and JBP1 eluted from the SAXS column as a single peak and was confirmed by the

**Table 2**

| Species   | Rg (Å) | I (Da Å⁻¹) | D (Å) | Mw (kDa) | V (Å³) |
|-----------|--------|------------|-------|----------|-------|
| JBP1      | 3.42   | 3.53       | 120   | 93.403   | 167.838 |
| Δ-JDBD    | 3.10   | 3.08       | 99    | 72.118   | 129.990 |
| JDBD      | 2.18   | 2.18       | 71    | 21.285   | 48.240  |
| J-23-DNA  | 2.25   | 2.34       | 73    | 14.241   | 27.627  |
| J-15-DNA  | 1.58   | 1.61       | 51    | 9.301    | 17.658  |

| Species   | Rg (Å) | I (Da Å⁻¹) | D (Å) | Mw (kDa) | V (Å³) |
|-----------|--------|------------|-------|----------|-------|
| JBP1:J-23-DNA | 4.098 | 4.08        | 141   | 107.644  | 182.075 |
| JBP1:J-15-DNA | 3.56  | 3.6         | 120   | 102.704  | 176.907 |
| JDBD:J-23-DNA | 2.52  | 2.533       | 86    | 35.526   | 70.587  |
280/260 nm absorption ratio; the averaged SAXS profile from the elution peak is shown in Fig. 4A, model-independent parameters are in Table 1, and SEC details are in Table S1.

Although visual inspection of the scattering intensity for JBP1 alone and the complex suggests that they are very similar (Fig. 5A), plotting the intensity ratio of the two data sets shows that the molecular form factors for the two data sets have prominent differences, as we observe strong features throughout the curve (Fig. 5B). Analysis of the data also shows that an increase in radius of gyration ($R_g$) (Fig. 5C) was accompanied by an increase in maximum distance ($D_{max}$), suggesting that J-DNA binds away from the center of mass (Table 1). The volume of correlation ($V_c$) is higher in the J-DNA–bound state, similar to the Porod volume that also increases by 14,000 Å$^3$ in the presence of J-23-DNA. Finally, examination of the dimensionless Kratky plot reveals a shift away from the Guinier–Kratky point (1.104), indicating that, upon J-DNA binding, JBP1 has a more elongated shape (Fig. 5D). All these data establish that the complex of JBP1 and J-23-DNA is formed and that the J-23-DNA binds away from the JBP1 center of mass, resulting in a more elongated particle.

To visualize the relative position of J-DNA in the complex, we again used the program MONSA. We defined two phases, J-23-DNA and JBP1, and calculated 20 models consistent with the scattering data for JBP1, J-23-DNA, and the JBP1:J-23-DNA complex. Cluster analysis with DAMCLUST resulted in four major clusters (Fig. 6). In all clusters, J-DNA is located toward one end of the JBP1 ellipsoid, away from its center of mass, compatible with the positioning of the JDBD domain. In contrast with the two-body modeling of JDBD and /H9004-JBP1, the clusters are rather similar, suggesting that J-23-DNA binds in similar conformations. As previous SAXS data on the complex between JDBD and J-DNA suggested that this is a rigid complex without conformational flexibility (17), this postulates that this JDBD:J-23-DNA rigid complex is now in one conformation with respect to /H9004. In other words, this analysis is compatible with the hypothesis that the JDBD becomes ordered upon J-DNA binding. To confirm these findings, we repeated the procedure with 15-mer J-DNA (J-15-DNA); the results (Table 1 and Fig. S2) lead to the same conclusions.

We then proceeded to test our hypothesis for the ordering of the JDBD upon complex formation using a different modeling approach. For this approach, we treated the JDBD complex with J-DNA as a rigid body (17) and defined two different phases: Δ-JBP1 and the JDBD:J-23-DNA complex. We again used MONSA for calculating 20 models consistent with the scattering data for Δ-JBP1 and the JDBD:J-23-DNA complex. Cluster analysis with DAMCLUST resulted in three clusters (Fig. S3). Consistent with the previous analysis, these clusters are fairly similar, confirming the reduced flexibility of the complex of JBP1 with J-DNA, and again show the DNA and JDBD located toward one end of the complex.
As we have an atomic model for the JDBD:J-23-DNA complex (17), we were able to place it in the respective dummy atom model using SUPCOMB (26). Thus, we created a pseudoatomic hybrid model where the \text/H9004-JDBD (for which we do not have an atomic model) is shown as the dummy atom model, and JDBD and J-23-DNA are all-atom models. Using CRYSOL (27), we evaluated the fit of both the dummy atom reconstruction and the pseudoatomic hybrid model against the SAXS curve for JBP1:J-23-DNA. The dummy atom reconstructions for the two most populated clusters, with 10 and six members, respectively, show the best fit to the experimental data (\(\chi^2=2.32\) and \(\chi^2=1.68\), respectively). However, the pseudoatomic hybrid model corresponding to the most populous cluster shows a considerably better fit to the experimental data (\(\chi^2=6.87\)) compared with the second cluster (\(\chi^2=38.03\)). Thus, we consider this model as the most likely interpretation of our experimental data for the complex of full-length JBP1 with J-DNA (Fig. 7).

This pseudoatomic model now shows the position of base J and the most likely orientation of the DNA. Interestingly, in this model, the T base 13 bp away from J in the complementary strand makes contact with the \text/H9004-JDBD domain that contains the TH activity.

**Discussion**

The discovery that JBP1 has an N-terminal TH domain sequence signature, which likely functions as a thymidine hydroxylase (10), was an important finding for the field of J biosynthesis. Perhaps more remarkably, this sparked a revolution for the study of methylcytosine conversion to hydroxymethylcytosine (12). Together with subsequent experimental proof of the TH activity hypothesis (11), these findings established the notion that JBP1 consists of an N-terminal TH domain followed by a J-DNA-binding domain (17) and a C-terminal sequence, which has received little attention. Here, we show that the JDBD should be seen as an insertion domain within a single TH domain that spans the N-terminal and C-terminal sequence regions of JBP1 and behaves as a single folding unit.
in solution, Δ-JDBD. We note that a JBP1 construct spanning residues 1–451, which has been previously reported to have hydroxylase activity (28), as well as numerous other constructs of the N-terminal region from a variety of species, did not yield soluble protein in our hands. The sole exception to that rule has been the Δ-JDBD domain. Remarkably, this new folding unit is functional as a thymidine hydroxylase in a new enzymatic activity assay that we developed, having an apparent catalytic rate about 17 times slower compared with full-length JBP1. We suggest that this lower rate is explained by the inability of Δ-JDBD to bind to DNA, bringing it in proximity to its T base substrates.

We have previously shown that binding of JBP1 to J-DNA is followed by a conformational change of JBP1 (18). Here, we extend this model, providing data that this conformational change represents a transition of JDBD: although JDBD is flexible with respect to the Δ-JDBD scaffold in the absence of J-DNA, it becomes ordered in the presence of J-DNA. This is in agreement with our previous observation from small-angle neutron scattering data (18) that the protein apparent $R_g$ is reduced upon complexation with J-DNA. This is consistent with JDBD sampling a more defined conformational space and reducing the apparent size of the protein particle.

The complex between JDBD and J-DNA is a well-defined rigid structure, as shown by our previous structural analysis (17) and current data. Our current analysis of Δ-JDBD and the JBP1 complex with J-DNA shows these to also be fairly rigid. These allowed us to propose a pseudoatomic hybrid model, showing the orientation of J-DNA with respect to the Δ-JDBD scaffold in the absence of J-DNA, it becomes ordered in the presence of J-DNA. This is in agreement with our previous observation from small-angle neutron scattering data (18) that the protein apparent $R_g$ is reduced upon complexation with J-DNA. This is consistent with JDBD sampling a more defined conformational space and reducing the apparent size of the protein particle.

The complex between JDBD and J-DNA is a well-defined rigid structure, as shown by our previous structural analysis (17) and current data. Our current analysis of Δ-JDBD and the JBP1 complex with J-DNA shows these to also be fairly rigid. These allowed us to propose a pseudoatomic hybrid model, showing the orientation of J-DNA with respect to the Δ-JDBD scaffold in the absence of J-DNA, it becomes ordered in the presence of J-DNA. This is in agreement with our previous observation from small-angle neutron scattering data (18) that the protein apparent $R_g$ is reduced upon complexation with J-DNA. This is consistent with JDBD sampling a more defined conformational space and reducing the apparent size of the protein particle.

From an evolutionary perspective, it is reasonable to presume that the thymidine hydroxylation activity to make hydroxymethyluracil precedes the glucosylation step to make J. We hypothesize that the last evolutionary step was the acquisition of J-binding activity to guide the TH activity to areas of
pre-existing J to replicate that epigenetic marker in kinetoplastids. As we show here, JDBD has likely been acquired by JB1 through an insertion event that did not disturb the TH scaffold. Based on these observations, one would expect to find JDBD homologues, with or without specificity for J-DNA binding, in additional proteins. As sequence searches in public databases do not reveal clear homologues of the JDBD outside the context of JB1 orthologues, we performed structural similarity searches using Dali (29) (see “Experimental procedures” for details). These searches revealed two new structural homologues in addition to MogR (30), which we have described previously (17). The closest structural homologue of JDBD is AcrF3, belonging to a family of proteins produced by bacteriophages to inactivate the CRISPR-Cas bacterial immune system (31); the other homologue is a C-terminal helical domain (CHCT) of the chromatin-remodeling protein CHD1 (32).

Although JDBD, MogR, and CHCT clearly have a conserved DNA recognition structural domain for preventing DNA recognition or might present an extreme example of repurposing a DNA-recognition structural domain, this positive patch is entirely missing in the AcrF3 protein.

**Expression and purification of recombinant proteins**

All constructs of JB1, JDBD, and ∆-JDBD were inserted in the NIKI-LIC-1.1 vector (36) and produced as soluble proteins in Escherichia coli. BL21 (DE3) T1R cells were used for protein overexpression. Protein production was induced with isopropyl 1-thio-β-D-galactopyranoside at 15 °C for 16–18 h. Cell lysis was performed in buffer A (20 mM HEPES/NaOH (pH 7.5), 350 mM NaCl, and 1 mM tris(2-carboxyethyl)phosphine) containing 10 mM imidazole. The lysate was bound to Ni-chelating Sepharose beads in batch mode, and elution was performed in buffer A containing 400 mM imidazole. Affinity tags were removed by 3C protease cleavage overnight at 4 °C, and the sample was applied to an S75 16/60 gel filtration column.

**In vitro hydroxylation assays**

The thymidine hydroxylase activity assay was carried out in a reaction buffer containing 50 mM HEPES/NaOH (pH7.6), 50 mM NaCl, 8 mM ascorbic acid, 4 mM 2-oxoglutarate, 1 mM Fe_{2}SO_{4}, 1 mM ADP, 20 μg ml⁻¹ BSA, and 0.5 mM DTT. The buffer was made anaerobic by degassing it with argon for 1 h at 4 °C. The reaction was carried out at 37 °C in a total volume of 50 μl, including 4 μM protein and 15 μM 14-mer dsDNA (CAGCAGCTGACACA). Upon completion of the reaction at the indicated time points, samples were stored at −20 °C for further processing.

**Sample preparation for MS**

Aliquots of 20 μl of the reaction mixtures were placed in 1.5-ml reaction tubes and incubated at 95 °C for 3 min followed by rapid cooling on ice to denature the dsDNA to single-stranded oligonucleotides. In each tube, 4 units of nucleotase P1 (Sigma-Aldrich) were added together with 100 μl of digestion buffer containing 0.04 mM deferoxamine mesylate, 3.25 mM ammonium acetate (pH 5.0), and 0.5 mM zinc chloride). The samples were incubated at 65 °C for 10 min to convert the oligonucleotides into single nucleotides. We then added 20 μl of Trizma base (pH 8.5) and 4 units of alkaline phosphatase (Roche Applied Science) and vigorously mixed for ~10 s. Samples were incubated at 37 °C (heating block) for 1 h to allow for nucleotide to nucleoside conversion, after which we added 20 μl of 300 mM ammonium acetate (pH 5.0) and evaporated to dryness at 40 °C in a TurboVap LV (Biotoptech, Uppsala, Sweden). Finally, we added 50 μl of 5 mM ammonium acetate in water/acetonitrile (2:98, v/v) and vigorously mixed for ~1 min.

**Experimental procedures**

**Cloning of ∆-JDBD JB1 and ∆-JDBD-3C JB1**

A JB1 synthetic gene encoding the sequence for Leishmania tarentolae (17) was used as the template for all constructs in this study. Primers to delete the JDBD domain and to replace it with the 3C protease DNA sequence were designed using Protein-CDD software (35). To create the ∆-JDBD construct, the JDBD domain was deleted using mutagenesis PCR and primers Del_fw (5′-CTC GTC TGG GTG TCT CTG AAA CCT CTC AGG AAA AAC GTG CTA ACT GGC TG-3′) and Del_rev (5′-CAG CCA GTT AGC ACC TTT TTG CTT GAG GGT TCT AGA GAA ACC CAG ACG AG-3′). The generated construct contained the N-terminal part, residues 1–392, followed by the C-terminal part, residues 564–827, without any connecting linker in between. To generate the ∆-JDBD-3C construct, a 3C protease cleavage site was introduced using primers Del-DB-3C_fw (5′-CTC GTC TGG GTG TCT CTG AAA CCC TGG TGC TGT TTC AGG GCC GTT CTC ACG AAA-3′) and Del-DB-3C_rev (5′-CAG TTA GCA GTG TTT TCG TGA GAC GGG CCC TGA AAC AGC ACT TCC AGG GTT TCA GAG AAA-3′) (the DNA sequence for the 3C protease cleavage site is highlighted in bold).
**Measurement of hmU by MS**

Oligonucleotide HmU content was analyzed as the released amount of 5’-hydroxymethyl-2’-deoxyuridine (HOMedU) after sample processing. A reference standard of HOMedU (Santa Cruz Biotechnology, Inc., Dallas, TX) was used to prepare calibration standards for HOMedU sample quantification.

For quantification, the HPLC-MS/MS system consisted of a QTRAP 5500 tandem mass spectrometer (Sciex, Framingham, MA) coupled to an HPLC Acquity I-Class pump (Waters). The HPLC system was equipped with an FTN I-Class autosampler and I-Class column oven (Waters). Data acquisition was performed using Analyst 1.6.2, software (Sciex).

The HPLC-MS/MS system was based on a previously developed method to quantify decitabine DNA incorporation (37). This assay was modified to allow for HOMedU quantification in the positive electrospray ionization mode by using the following m/z transition: 257.0 → 124.0. The remaining settings of the method were unchanged.

**Preparation of J-DNA and JDBD:J-DNA and JBP1:J-DNA complexes**

J-DNA oligos (38) were mixed with their complementary strand and annealed as described previously (18). Briefly, the hmC-containing oligonucleotide and the complementary strand were dissolved in water to a concentration of 100 μM and then heat-annealed. The double-stranded oligonucleotide was then glucosylated by the T4 phase β-glucosyltransferase (T4-BGT) from New England Biolabs according to the manufacturer’s instructions. To create the protein-J-DNA complexes, 1 mg ml⁻¹ JBP1 was mixed with J-DNA at a 1:1.1 molar ratio and then concentrated with Amicon concentrators. The same procedure was used for making the JDBD-J-DNA subcomplexes. The sequences used in this study were J-23-DNA (TGCATTGTTCATAGACTAATAC) and J-15-DNA (TAGAACCCGJAACCAT).

**SEC-SAXS data collection and analysis**

Synchrotron X-ray data for all components were collected on a Pilatus 1M detector at the European Synchrotron Radiation Facility (ESRF) beamline BM29 (39). About 40 μl of each sample, at a concentration 3–10 mg ml⁻¹, was loaded onto a Superose-6 column (Table S1). The flow rate for SAXS data collection was 0.2 ml min⁻¹, and a scattering profile was integrated every second. Frames for each data set were selected based on the examination of the size-exclusion profile together with the calculated \( R_g \) and \( D_{max} \) values. At least 20 frames for each data set were selected, scaled, and averaged using PRIMUS following the standard procedures (Table S1). For JDBD-J-23-DNA and JBP1-J-15-DNA, frames were analyzed with DATASW (40).

**Model-independent analysis of SAXS data**

SAXS data analysis was performed using the PRIMUS (24) and ScAtter (41) software packages. The forward scattering, \( I(0) \), was evaluated using the Guinier approximation (8) assuming the formula \( I(q) = I(0) \exp(-q^2R_g^2/3) \) for a very small range of momentum transfer values \( qR_g < 1.3 \). Calculation of the pair distribution function (Fig. S4) and \( D_{max} \) was performed using GNOM (42). The \( R_g \) was estimated by Guinier approximation (Fig. S5). The molecular mass was calculated using the Porod volume, the \( Q_R \) method, and the SAXSMoW2 webserver (25, 43, 44). Ambiguity of all data sets was measured with AMBIMETER (45, 46). The useful range for each data set was determined by SHANUM (47) analysis prior to proceeding to ab initio modeling.

**Ab initio modeling using SAXS data**

Molecular envelopes of ab initio created models were made for all the components using DAMMIN (48). Ten individual models were created for each component, and averaging was performed using DAMAVER. Fitting of atomic-resolution structures to molecular envelopes was performed using SUPCOMB (26).

To resolve the relative positions of individual subunits of the JBP1 structure, a volumetric analysis was performed using the program MONSA, an extension of the DAMMIN algorithm (45), following an approach similar to the one described previously (41). MONSA allows ab initio modeling of macromolecular complexes by fitting simultaneously multiple experimental data sets. The search volume is defined as a sphere with radius equal to half the \( D_{max} \) of the complex of study. A minimization algorithm based on simulated annealing fits the experimental data sets of each component (phase), and all components together should fit to the experimental data set of the corresponding complex. In the case of a multicomponent complex, each different phase is assigned a different contrast (1 for protein, 2 for nucleic acid, or 0 for solvent). In our approach, we treat as a phase different components (protein and DNA) but also the two distinct folding units that make JBP1 as we established in this work (Δ-JDBD and JDBD). At least 20 individual runs for each complex (particle) were created. The online version of MONSA was used for all models generated in this study.

In the case of a two-body modeling that consists of two components \( X_y \), \( X \) denotes the component with the larger mass, and \( y \) denotes the component with the smaller mass. In our experiments, \( X_y \) would thus be JBP1-J-23-DNA or Δ-JDBD: JDBD. This modeling assumes that each component (phase) does not undergo conformational changes upon complex formation or that the difference at that resolution is negligible.

Examination of the parameters stored in the log file for each MONSA run shows the fitting and the calculated \( R_g \) values for each component from each individual MONSA run. To evaluate each MONSA modeling run, we compared the \( R_g \) derived from the experimental data set for each component with the calculated \( R_g \) for each phase calculated for each MONSA model and stored in the log file. Runs with calculated \( R_g \) values that differ significantly from the experimental data—derived \( R_g \) were excluded from the analysis.

**Aligning models generated by MONSA**

To compare MONSA models, we wanted to visualize the different position of each phase \( y \) relative to \( X \) in a common reference framework. We therefore chose to first align all phases \( X_y \) from each individual run. Then the transformation matrix of each \( X_y \) component was used to also transform the corresponding \( y \) phase. The chosen procedure highlights the possible relative positions of \( y \), relative to \( X_y \). This is preferable.
Domain architecture of JBP1

to aligning all complexes $X_i y_i$ to each other, which only yields the average relative position.

For that, we developed a script in Python (available upon request) that performs the following operations. DAMAVER is used to align all models of $X_i$, the component that has the larger mass. DAMAVER superimposes all models on each other, averages them, selects the best model based on the normalized spatial discrepancy metric as the reference model ($X_{ref}$), and aligns all models to $X_{ref}$. DAMAVER creates a new PDB file for each $X_i$ in the new aligned position that also contains the transformation matrix ($T_i$) that aligns each model ($X_i$) with the reference ($X_{ref}$). Then, the transformation matrix $T_i$ is used to transform every $y_i$ phase. Finally, DAMCLUST is used to cluster the transformed $y_i$ phases, effectively creating clusters that have the $y_i$ component in similar orientations. These clusters define the conformational clusters of the $X_i y_i$ complex.

Each cluster is presented as the average model of all $X_i$ components and the average of the $y_i$ components for each cluster. Singleton clusters were excluded from our analysis. To evaluate the clustered models, we examined their fit to the experimental data based on $\chi^2$ analysis.

To fit atomic-resolution structures on molecular envelopes generated by either DAMMIN or MONSA, the program SUPCOMB (26) was used (Fig. S6). To calculate the $\chi^2$ values of the hybrid dummy atoms model of $\Delta$-JBP1 and the atomic model of JDBD1-23-DNA, we transformed the hybrid model using CRYSO and compared it with the experimental data of the complex.

Structure similarity searches

Structure similarity searches were carried out by Dali searches against the whole PDB (29). Dali returned new hits compared with previous searches (17). We then inspected the top hits manually. The recognition helix, the supporting helix, two of the other helices, and the connectivity we have described for the helical bouquet fold of JDBD (17) were present in the top one (self), two, three, and five hits. Hit 5 ($Z$-score, 5.2; r.m.s.d., 3.0; 10% identity over 84 aligned residues) is the motility gene receptor MogR that we have previously identified as a JDBD component and the average of the transformation units throughout the genome of Trypanosoma brucei. Nucleic Acids Res. 35, 3923–3935 CrossRef Medline

References

1. Gommers-Ampt, J. H., Van Leeuwen, F., de Beer, A. L., Vliegenthart, J. F., 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 CrossRef Medline

2. van Leeuwen, F., Wijsman, E. R., Kieft, R., van der Marel, G. A., van Boom, J. H., and Borst, P. (1997) Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. Genes Dev. 11, 3232–3241 CrossRef Medline

3. van Leeuwen, F., Wijsman, E. R., Kuyl-Yeheskiely, I., 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 Res. 24, 2476–2482 CrossRef Medline

4. Genest, P.-A., Ter Riet, B., Cijouw, T., van Luemen, H. G., and Borst, P. (2007) Telomeric localization of the modified DNA base J in the genome of the protozoan parasite Leishmania. Nucleic Acids Res. 35, 2116–2124 CrossRef Medline

5. Cliffe, L. J., Siegel, T. N., Marshall, M., Cross, G. A., 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 Res. 38, 3923–3935 CrossRef Medline

6. van Luemen, H. G., Farris, C., Jan, S., Genest, P.-A., Tripathi, P., Velds, A., Kerkhoven, R. M., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., van Steensel, B., et al. (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. Cell 150, 909–921 CrossRef Medline

7. 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. J. Biol. Chem. 289, 20273–20282 CrossRef Medline

8. Iyer, L. M., Zhang, D., Burroughs, A. M., and Aravind, L. (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res. 41, 7635–7655 CrossRef Medline

9. 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. Mol. Biochem. Parasitol. 196, 9–11 CrossRef Medline

10. Yu, Z., Genest, P.-A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J. M., Hausinger, R. P., van Luemen, H. G., Rigden, D. J., Sabatini, R., and Borst, P. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res. 35, 2107–2115 CrossRef Medline

11. Cliffe, L. J., Hirsch, G., Wang, J., Ekanevaye, D., Bullard, W., Hu, M., Yang, Y., and Sabatini, R. (2012) JBP1 and JBP2 proteins are $\text{Fe}^{2+}$/2-oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. J. Biol. Chem. 287, 19886–19895 CrossRef Medline

12. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Nature Genet. 41, 930–935 Cell

13. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P., and Xu, Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell 155, 1545–1555 CrossRef Medline

14. Hashimoto, H., Pai, E. J., Zhang, X., Saleh, L., Fu, Z.-Q., Dai, N., Corrêa, I. R., Jr., Zheng, Y., and Cheng, X. (2014) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature 506, 391–395 CrossRef Medline

Author contributions—A. A., I. Q. P., and A. P. conceptualization; A. A. data curation; A. A., T. H., and J. R. formal analysis; A. A. and A. P. validation; A. A. and T. H. visualization; A. A., J. R., W. G. T., J. B., and A. P. methodology; A. A. writing-original draft; T. H., J. R., and A. P. investigation; T. H. and A. P. project administration; W. G. T. software; J. B. and A. P. supervision; A. P. funding acquisition; A. P. writing-review and editing.

Acknowledgments—We thank Piet Borst for extensive discussions and suggestions to improve the text of this manuscript and Piet Borst and Henri van Luemen for helping to establish the JBP1 activity assays.
15. Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., Liu, M., Sun, C., Yin, X., Li, J., Tan, X., et al. (2015) Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118–122 CrossRef Medline

16. Cross, M., Kieft, R., Sabatini, R., Wilm, M., de Kort, M., van der Marek, 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. *EMBO J.* **18**, 6573–6581 CrossRef Medline

17. Heidebrecht, T., Christodoulou, E., Chalmers, M. J., Ian, S., Ter Riet, B., Grover, R. K., Joosteen, R. P., Littler, D., van Luenen, H., Griffin, P. R., Wentworth, P., Jr., Borst, P., and Perrakis, A. (2011) The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic Acids Res.* **39**, 5715–5728 CrossRef Medline

18. Heidebrecht, T., Fish, A., van Castelmur, E., Johnson, K. A., Zaccal, G., 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. *J. Am. Chem. Soc.* **134**, 13357–13365 CrossRef Medline

19. Genest, P.-A., Baugh, L., Taipale, A., Zhao, W., Jan, S., van Luenen, H. G., Korklach, J., Clark, T., Luong, K., Boitano, M., Turner, S., Myler, P. J., and Borst, P. (2015) Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Res.* **43**, 2102–2115 CrossRef Medline

20. Rambo, R. P., and Tainer, J. A. (2010) Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr. Opin. Struct. Biol.* **20**, 128–137 CrossRef Medline

21. Petoukhov, M. V., and Svergun, D. I. (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.* **17**, 562–571 CrossRef Medline

22. Bizien, T., Durand, D., Roblina, P., Thureau, A., Vachette, P., and Pérez, J. (2016) A brief survey of state-of-the-art BioSAXS. *Protein Pept. Lett.* **23**, 217–231 CrossRef Medline

23. Schindler, C. E. M., de Vries, S. J., Sasse, A., and Zacharias, M. (2016) SAXS data alone can generate high-quality models of protein–protein complexes. *Structure* **24**, 1387–1397 CrossRef Medline

24. Petoukhov, M. V., Franc, D., Shkumatov, A. V., Tria, G., Kiknhey, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 CrossRef Medline

25. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–571 CrossRef Medline

26. Kozin, M. B., and Svergun, D. I. (2011) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **43**, 33–41 CrossRef Medline

27. Svergun, D. I., C. B., and, J. K. M. H. (1995) CRYSOL—a program to determine atomic coordinates. *J. Appl. Crystallogr.* **28**, 33–41 CrossRef Medline

28. Cliffe, L. J., Kieft, R., Southern, T., Birkeland, S. R., Marshall, M., Sweeney, K., and Sabatini, R. (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res.* **37**, 1452–1462 CrossRef Medline

29. Cliffe, L. J., Kieft, R., Southern, T., Birkeland, S. R., Marshall, M., Sweeney, K., and Sabatini, R. (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res.* **37**, 1452–1462 CrossRef Medline

30. Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res.* **44**, W351–W355 CrossRef Medline

31. Shen, A., Higgins, D. E., and Panne, D. (2009) Recognition of AT-rich DNA binding sites by the MogR repressor. *Structure* **17**, 769–777 CrossRef Medline

32. Wang, X., Yao, D., Xu, J.-G., Li, A.-R., Xu, J., Fu, P., Zhou, Y., and Zhu, Y. (2016) Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nat. Struct. Mol. Biol.* **23**, 868–870 CrossRef Medline

33. Mohanty, B., Holder, S., Silva, A. P., Mackay, J. P., and Ryan, D. P. (2016) The Chromatin remodelling protein CHD1 contains a previously unrecon-