The Tramtrack group BTB protein domains are Arthropoda-specific multimerization modules

Artem N. Bonchuk[1,2,*], Konstantin I. Balagurov[1], Konstantin M. Boyko[1,3,4], Nikolai N. Sluchanko[3], Anastasia M. Khrustaleva[1†], Anna D. Burtseva[1,3,4], Natalia V. Kuzmina[7], Olga V. Arkova[2], Karina K. Khalisova[1], Vladimir O. Popov[3,6], Pavel G. Georgiev[1,*]

[1]Department of the Control of Genetic Processes, Institute of Gene Biology Russian Academy of Sciences, 34/5 Vavilov St., Moscow 119334, Russia;
[2]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilov St., Moscow 119334, Russia;
[3]Bach Institute of Biochemistry, Research Center of Biotechnology Russian Academy of Sciences, Leninsky pr-t, 33, bld. 2, Moscow 119071, Russia
[4]Moscow Institute of Physics and Technology, Institutsky lane 9, Dolgoprudny, Moscow region, 141700, Russia
[5]Department of Drosophila Molecular Genetics, Institute of Gene Biology Russian Academy of Sciences, 34/5 Vavilov St., Moscow 119334, Russia;
[6]Complex of NBICS Technologies, National Research Center "Kurchatov Institute", Akad. Kurchatova sqr, 1, Moscow 123182, Russia
[7]A.N. Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Moscow, Russia;

* To whom correspondence should be addressed. Tel: +7(499)135-60-89 Fax: +7(499)135-41-05; E-mail: georgiev_p@mail.ru (P.G.G.), bonchuk_a@genebiology.ru (A.N.B.)

Abstract

BTB (Bric-a-brack, Tramtrack and Broad Complex) is a diverse group of protein-protein interaction domains found within various eukaryotic proteins. In metazoans, many DNA-binding transcription factors contain a dimerizing BTB domain with a characteristic N-terminal extension. The Tramtrack group (TTK) is a distinct type of BTB domain, which can multimerize by an unknown mechanism. Here, we demonstrated that the TTK-type BTB domains are found only in Arthropods and have undergone lineage-specific expansion in modern insects. The *Drosophila* genome encodes 24 transcription factors with TTK-type BTB domains, whereas only four have non-TTK-type BTB domains. A combination of multi-angle laser light scattering (MALS), molecular modeling, small-angle X-ray scattering (SAXS), and microscopy revealed that the TTK-type BTB domains assemble into a hexameric structure consisting of three canonical BTB dimers. The dimers are connected through a previously uncharacterized interface that involves the formation of a β-sheet by β-strands of neighboring dimers. Such architecture was further supported by site-directed mutagenesis. Yeast two-hybrid analysis revealed that the TTK-type BTB domains have an unusually broad potential for heteromeric associations. According to the mutagenesis data, such associations occur mostly between different homodimers through the dimer-dimer interaction interface, which is highly similar among members of the TTK-type BTBs. Thus, the TTK-type BTB domains are a structurally and functionally distinct group of protein domains specific to Arthropodan transcription factors.

1. Introduction

BTB, also known as POZ (Pox-virus and zinc-finger), is an evolutionarily conserved domain that was originally found in the *Drosophila* proteins bric-a-brac, Tramtrack and broad complex [1-4]. BTB proteins have been identified in poxviruses and eukaryotes, and have various functions including regulation of transcription, chromatin remodeling, cytoskeletal function, ion transport, and ubiquitination/degradation of proteins [5-7]. BTB domains are 100–120 aa in size and the core structure consists of five α-helices and three β-strands [8, 9]. In addition to this core, different subclasses of BTB proteins include N- and C-terminal BTB extension regions that facilitate protein-specific functions. As a result, the BTB fold is a versatile scaffold that participates in a variety of family-specific protein-protein interactions.

In metazoans, BTB domains of transcription factors contain an amino-terminal extension that enables homodimerization [8, 10]. Most members of this family contain C2H2-type zinc fingers (BTB-C2H2). In humans, 156 genes are predicted to encode BTB domain-containing proteins, of which 49 possess between 2 and 14 C2H2 domains [11]. BTB-C2H2 proteins act as classical transcription factors, binding to chromatin and participating in the regulation of transcription. In insects, some transcription factors contain BTB domains in combination with other DNA-binding domains such as Pipsqueak (PSQ, BTB PSQ) [12, 13] and helix-turn-helix (HTH) [14]. Several BTB-containing transcription factors contain poorly characterized FLYWCH (named after characteristic sequence motif) domains that may be involved in the interaction with either DNA, RNA, or other proteins [15-17].

In mammals, BTB-C2H2 transcription factors are required for the development of lymphocytes, fertility, skeletal morphogenesis, and neurological development [18]. The well-characterized BTB domains form tightly intertwined dimers and possess a peptide-binding groove, which is responsible for the interaction with various transcription factors and co-repressor complexes [10, 19-21]. In most cases, mammalian C2H2 transcription factors have BTB domains which exclusively form homodimers. An exception to this rule is the C2H2 protein Miz-1, whose BTB domain forms tetramers [22]. Several BTB domains are capable of heterodimer formation [23]. The structural basis for heterodimerization has been studied using chimeric Bcl6/Miz-1 and Miz-1/NAC1 assembly [24]; however, there is evidence that formation of such heterodimers *in vivo* is prevented by co-translational dimer assembly and the quality-control protein degradation machinery [25-27].

In *Drosophila*, a number of C2H2 proteins with a BTB domain have been described, some containing a typical BTB domain, such as CP190 and CG6792 [28]. The most well-studied of these, CP190, has four C2H2 zinc finger domains that appear to be involved in protein-protein interactions rather than DNA binding [29, 30]. The N-terminal BTB/POZ domain of CP190 forms stable homodimers [21, 29, 31-33]. CP190 is required for the activity of housekeeping promoters and insulators [34-38]. CG6792 is similar to mammalian BTB-C2H2 proteins, has seven DNA-binding zinc fingers, and is involved in wing development [28].

Most of the well-characterized *Drosophila* BTB transcription factors, including GAF, Mod(mdg4), LOLA, Broad-complex (BR-C), Batman, Pipsqueak, and Bric-a-brac (Bab), have TTK-type BTB domains [1, 31]. Proteins from this group often have important functions in transcription regulation, development, and chromosome architecture [18]. Many TTK-group proteins, such as BR-C, TTK and Bab, are critical regulators of development that function as transcriptional repressors [18, 39-42]. Several have been implicated in chromatin architectural function, acting as a component of a chromatin insulator complex (Mod(mdg4)) or recruiting chromatin remodeling complexes (GAF, Pipsqueak) [14, 43, 44]. Recently, it

was shown that GAF is a pioneer factor that acts as a stable mitotic bookmarker during zygotic genome activation during *Drosophila* embryogenesis [45, 46].

TTK-type BTB domains contain a highly conserved N-terminal FxLRWN motif, where x is a hydrophilic residue [31]. Several TTK-type BTB domains can selectively interact with each other and form multimers [31]. Although there are more than 30 crystal structures of non-TTK BTB domains that form stable homodimers [5, 8, 21, 47, 48], the structures of the TTK-type BTB domains and the structural basis for their multimerization remain unknown.

In this study, we investigated the presence of TTK-type and canonical BTB domains in transcription factors from various phylogenetic groups of Bilateria. The TTK-type BTB domains were found only in Arthropodan transcription factors and underwent lineage-specific expansion in insects. Using an integrative structural biology approach, we built a structural model of the TTK-group BTB hexameric assembly, validating it using MALS, SAXS, microscopy techniques, and site-directed mutagenesis. Finally, we found an unusual potential for these domains to form heteromultimers, which likely involves a dimer-dimer interaction interface.

2. Results

2.1. Most transcription factor BTB domains are TTK-type in *Drosophila melanogaster*

The *Drosophila melanogaster* genome contains 28 genes encoding transcription factors with BTB domains (Figure 1). Many genes encode several BTB-containing protein isoforms, which differ in their C-terminal sequences. At two exceptional loci, *mod(mdg4)* and *lola*, multiple isoforms are formed by *trans*-splicing, the mechanism of which is still unclear [49-51]. The *mod(mdg4)* locus encodes at least 30 isoforms, most of which have FLYWCH domains [16]. Seventeen of the 20 isoforms produced by the *lola* locus contain different C-terminal C2H2 zinc-fingers [52-54].

A characteristic feature of the TTK-type BTB domain is the presence of the FxLRWN motif at the N-terminus. Comparison of the amino acid sequence of BTBs from 28 *Drosophila* BTB-containing transcription factors showed that 24 BTB domains contain the characteristic “TTK motif” (Figures 2A and 2B). The TTK-type BTB domains are usually located at N-termini of proteins, however, in two proteins (CG6118, bric-a-brac2), the BTB domains are located in the middle of the protein (Figure 1). The Batman protein is the only one of these proteins to contain only a single BTB domain. Only four transcription factors (CP190, CG6792, CG15725 and Ken) have BTB domains without a TTK motif (Figure 1).

2.2. The TTK-type BTB domains display a broad heteromeric interaction propensity

The characteristic feature of the TTK-type BTB domains is their ability to selectively interact with each other [31]. We tested interaction of all TTK-type BTB domains with BTB domains from Mod(mdg4), Lola, Chinmo, CG8924 and GAF in a yeast two-hybrid assay (Y2H; Figure 2C, Supplementary Table S2 and Supplementary Figure S1). We found that all BTB domains can interact with themselves. 23 out of 24 TTK-type BTB domains also interacted with at least one of GAF, Lola, or Mod(mdg4) BTB domains. GAF, Mod(mdg4), Lola and CG8924 all interacted with at least 10 of the 24 TTK-type BTB domains, while Chinmo interacted with 18 TTK-type BTB domains (Figure 2C, Supplementary Table S2). The homology

between interacting domains was mostly in the range 35–47%, but could be as low as 25%. No obvious relationship between homology and heteromeric interaction ability was found (Supplementary Table S3).

We also tested interaction between four non-TTK BTB domains (CP190, Ken, CG15275 and CG6792 (dPLZF)) and did not find heteromeric interaction between them (Supplementary Table S4). The BTBs of CP190 and CG6792 formed homodimers *in vitro,* like all other classical BTB domains (Supplementary Figure S2). The Ken and CG15725 BTB domains were insoluble after bacterial expression. Using Y2H, we detected only a small number of interactions between the non-TTK BTB domains of CP190 and Ken with the TTK-type BTB domains: CP190 interacted with GAF and Ribbon, while Ken interacted with Batman (Supplementary Table S5 and Supplementary Figure S3). These results confirm that the TTK-type domains are functionally distinct from classical BTB domains.

2.3. Integrative structural biology reveals hexameric assembly of the TTK-type BTB domains

To test the ability of TTK-type BTB domains to multimerize and shed light on their possible structure, we screened all TTK-type BTB domains from *Drosophila melanogaster* for possibility of expression and purification at high concentrations for subsequent structural analysis.

We expressed all 24 TTK-type *Drosophila* BTB domains as TEV-cleavable thioredoxin fusions. Out of 24 BTBs, only ten domains were soluble and stable after TEV-cleavage. Eight BTB domains appeared as a single peak on size exclusion chromatography (SEC), with elution volumes roughly corresponding to an octamer (apparent Mw 90–150 kDa). Batman BTB eluted as a dimer (Supplementary Figure S4), while CG32121 BTB formed multiple peaks. Unfortunately, most of these BTBs had low solubility (~1 mg/mL) and, therefore, were not suitable for structural studies.

Only three BTB domains (CG6765$^{1-133}$, CG32121$^{1-147}$ and LOLA$^{1-120}$) were soluble at concentrations above 5 mg/mL. CG32121$^{1-147}$ was excluded from further analysis due to its unstable oligomeric state.

The SEC profile for the representative BTB domain of LOLA had a single symmetric peak, the position of which remained unchanged even upon 10-fold dilution (Fig. 3A). Since the apparent Mw of 113 kDa determined relative to protein standards did not allow us to unequivocally determine the oligomeric state of LOLA (the predicted monomer mass is 15.25 kDa), it was analyzed using SEC-MALS, which provides the absolute Mw. The Mw determined by this method was 86.3 kDa, which was 5.7 times larger than the predicted monomeric Mw, and therefore most likely to correspond to a hexameric species. Significantly, the monodispersity of the sample (Mw/Mn = 1.000) supported the unchanged position on the elution profile upon dilution, and together indicated that the observed oligomer is stable. Consistent with this, we observed that MBP-fused LOLA also had a single symmetrical peak on the SEC profile, and its MALS-derived Mw of 315 kDa suggested a monodisperse (Mw/Mn=1.000) 5.6-mer (Fig. 3B).

In order to use X-ray crystallography for structure determination, we performed multiple crystallization trials, which were unsuccessful. Thus, we utilized an AlphaFold multimer implementation [55, 56] to model LOLA and CG6765 hexamers. Both hexamers consisted of three canonical BTB dimers with extensive hydrophobic molecular contacts forming a β-sheet between the β4-strands (corresponding to B3 according to Stogios et al [5]; Fig. 3C, 3D). Interdimeric interactions were strengthened by contacts between α2 and the β2/β3 sheet, with the β3 strand formed by residues conserved only within the TTK group (Figure 2A). The N-terminal TTK-motif is located within the first β-strand and is involved in dimer stabilization. We also modelled a hexamer of the Mod(mdg4) BTB domain as we have model systems to

study the function of this protein in flies [17, 57-59]. The model had an architecture similar to those of the LOLA and CG6765 hexamers (Supplementary Figure S5). The dimer-dimer interaction interface is also conserved in Batman, which did not form multimers larger than dimer, according to SEC data. We therefore modelled a Batman BTB dimer and compared its structure with a LOLA dimer within the hexamer. We found no steric hindrance obstructing hexamer formation in Batman, although some hexamer-stabilizing interactions were absent in this case, including the substitution of F42, which stabilized the LOLA hexamer, to T42 (Fig. S6A). From this observation, we suggest that the core β-sheet is not sufficient for stable hexamer formation, and further stabilizing contacts are required, which may determine the specificity of the interaction. The dimer-dimer interaction interface was significantly different in the model of Chinmo BTB, which can still form multimers, and in Ribbon and CG15812. Unfortunately we could not test the multimerization of the latter two domains due to strong aggregation; however, the AlphaFold model suggested that the β-strand involved in the dimer-dimer interaction remained unchanged.

Finally, to verify the AlphaFold models for LOLA and CG6765, we applied SAXS and electron microscopy. SAXS-derived structural parameters are listed in Table 1. Estimated molecular weights for LOLA and CG6765 BTB domains roughly correspond to hexamers and are in agreement with SEC-MALS data. Then, we evaluated the AlphaFold models with different stoichiometry (from dimers to octamers) using approximation of experimental SAXS data with curves calculated from models using CRYSOL [60] (Figures S6B, S6C). For both CG6765 and LOLA BTBs, the theoretical scattering for hexameric models agreed best with the experimental data (χ^2 values 1.7 and 3.6, respectively), while the fits from alternative oligomeric assemblies predicted by AlphaFold were inadequate (Fig. S6B, S6C). The models of 7-mers contained largely disconnected hexamers and an additional subunit and were not considered. The models of 5-mers provided second-best fits to the experimental SAXS profiles, but were disregarded due to symmetry considerations and also because they appeared to represent hexamers lacking one of the subunits. We believe that, while the presence of incomplete assemblies *in vitro* cannot be excluded, they do not represent the main oligomeric state of TTK-group BTB domains. Thus, SAXS data suggest that the prevalent oligomer of LOLA and CG6765 is the hexamer.

Finally, we utilized two complementary microscopy methods – atomic force microscopy (AFM) and negative-staining EM. To this end, we expressed the LOLA BTB domain fused to MBP (40 kDa), which resulted in a monomer molecular weight of 56 kDa, which was suitable for study by electron microscopy. MALS revealed the homogeneity of the sample with the assembly molecular weight of 315 kDa corresponding to a 5.6-mer. Both microscopy techniques gave a variety of particle sizes, which seemed to result from partial protein degradation and aggregation under experimental conditions. However, negative-staining EM revealed characteristic 2D classes which could be attributed to hexameric particles (Figure 3E). These flower-like hexameric structures had a diameter of about 200 Å, with a central particle of approximately 80 Å and six smaller surrounding particles of approximately 60 Å each. Assuming that the MBP domains do not stably interact with each other, we suggest that the central particle is the LOLA hexamer. In agreement with this, the size of the AlphaFold model of the LOLA hexamer is approximately 90 Å, which is comparable to the size of the central particle in the 2D classes.

Likewise, AFM revealed particles with the same symmetry, consisting of six subunits and with a size similar to that observed in negative stain micrographs (Figure 3F).

2.4. Testing the effect of amino acid substitutions on multimerization of the Mod(mdg4), CG6765, and Lola BTB domains

To confirm the role of residues in the newly identified dimer-dimer interaction interface, we introduced several substitutions to the BTB domains of LOLA, CG6765 and Mod(mdg4) and tested their influence on multimerization. We substituted conserved hydrophobic residues involved in the inter-subunit β-sheet interface either to alanine or to lysine (to strengthen the effect through electrostatic repulsion). The following substitutions were tested: LOLA$^{I70A;I72A}$; LOLA$^{I70K;I72K}$; Mod(mdg4)$^{I71A;F73A}$; Mod(mdg4)$^{I71K;F73K}$; CG6765$^{Y84K;V86K}$. All mutant BTB domains were unstable after cleaving from thioredoxin, therefore only thioredoxin fusion proteins were studied *in vitro*. In general, all mutations led to a strong decrease in protein solubility and appearance of aggregates in the SEC profiles, suggesting a general impact on protein folding (Figures 4A, 4B and 4C, Supplementary Figures S7, S8 and S9). Only a small fraction of hexamers was present, along with a dimer fraction of comparable intensity. A large peak of proteolytic fragments was also visible, suggesting protein misfolding and degradation (see Supplementary Figures S7, S8 and S9). In a Y2H assay, the alanine-substituted LOLA and Mod(mdg4) proteins were able to self-associate (Figure 4D, Supplementary Figure S10). Mod(mdg4)$^{I71K;F73K}$ could only interact with wild-type BTB, whereas LOLA$^{I70K;I72K}$ did not interact even with wild-type BTB and is therefore likely misfolded.

According to the modeling, the TTK motif facilitates the formation of specific homodimers resembling classical BTB dimers, which are subsequently assembled into hexamers through a dimer-dimer interaction interface. To study the role of the TTK motif in specific dimerization and subsequent hexamerization of the BTB domains, we performed several substitutions of key amino acids in the BTB domains of Mod(mdg4) and LOLA. According to the modeling, conserved aromatic and hydrophobic residues (F7 and L9 in LOLA) form hydrophobic contacts with the β5 strand of the other subunit in the dimer. A conserved positively charged residue (R10 in LOLA) likely forms a salt bridge with a glutamate residue of the adjacent subunit, which is conserved in the TTK group (E92 in LOLA). A conserved tryptophan residue (W11 in LOLA) fits into a hydrophobic cavity of the adjacent subunit. While F7A or L9A substitutions in the TTK motif of LOLA had no effect on multimerization, a W11A/N12S double mutation resulted in a complete loss of the hexamer, shifting the peak on the SEC profile to the size of the dimer (Figures 5A and 5C). LOLA$^{W11A;N12S}$ was less soluble than the wild-type protein, which prevented SAXS experiments. These non-classical dimers are presumably formed through the hexamerization interface described above.

Since Mod(mdg4) BTB is extremely prone to aggregation *in vitro*, all SEC experiments were performed with thioredoxin fusion proteins. We introduced the W11A/N12S mutation in the Mod(mdg4) BTB domain (Figure 5A and 5B), and observed a similar dramatic change on the elution profile of the thioredoxin-tagged domain, with the main peak corresponding to degradation products, and small fractions corresponding to the hexamer and the dimer.

We then substituted all six residues in the conserved N-terminal sequence element to ASASAS, to preserve the order of hydrophobic/hydrophilic residues in LOLA (F7A/C8S/L9A/R10S/W11A/N12S, hereafter LOLA^{Nmut}) and Mod(mdg4) (F7A/L9A/C10S/W11A/N12S hereafter Mod(mdg4)^{Nmut}). While LOLA^{Nmut} was dimeric like LOLA$^{W11A;N12S}$, Mod(mdg4)^{Nmut} restored the ability to multimerize (Figure 4).

However, the solubility of this protein was very low even as a thioredoxin fusion, therefore the non-multimerized domains were likely completely insoluble and were absent in preparations.

To test if residues of the TTK motif have the same effect in other TTK-type BTB domains, we introduced similar substitutions in the BTB domains of CG6765 and MAMO. Surprisingly, in contrast to LOLA and Mod(mdg4), a double substitution of the conserved tryptophan and its adjacent residue (CG6765$^{W10A;D11S}$ and MAMO$^{W10A;N11S}$) did not result in a loss of multimer formation. Only substitution of all six conserved N-terminal residues in CG6765 (Y6A/H7S/L8A/K9S/W10A/D11S) led to a shift of the SEC elution profile, with the main peak corresponding to a monomer (Figures 5A and S11). In MAMO, even the substitution of all conserved residues (F6A/L8A/C9S/W10A/N11S, hereafter MAMO^{Nmut}) did not result in a loss of multimerization (Supplementary Figure S11). These data suggest that the contribution of the TTK motif to dimer formation varies between different BTB domains.

2.5. The dimer-dimer interface is critical for most heteromeric interactions between TTK-type BTB domains.

It is likely that interactions between different TTK-type BTB domains (heteromeric interactions) occur due to the association of different homodimers through the interdimeric interfaces described above, which are highly similar in different members of this family (Figure 2A). We used a Y2H assay to test the effect of mutations within the dimer-dimer interaction interface on the ability of BTBs to selectively interact with each other.

We screened LOLA$^{I70A;I72A}$ and Mod(mdg4)$^{I71A;F73A}$ for their ability to interact with all TTK-type BTB domains (Figure 6A, Supplementary Table S6, Supplementary Figure S10). Most interactions were affected by these substitutions. The effect was even stronger with lysine substitutions: Mod(mdg4)$^{I71K;F73K}$ retained only the ability to interact with the wild-type Mod(mdg4) BTB domain, which likely occurs through the intact dimerization interface. Since LOLA$^{I70K;I72K}$ lost even the ability to self-interact, it is likely misfolded and was therefore excluded from analysis of heteromeric interactions. These results demonstrate the importance of the dimer-dimer interface for specific heteromeric association between TTK-type BTB domains.

We then screened LOLA^{Nmut} and Mod(mdg4)^{Nmut} against all TTK-type BTB domains to investigate the possible formation of heterodimers with an altered TTK motif. The change in interaction pattern of both domains was significantly smaller than the change following mutations to the dimer-dimer interface (Figure 6B, Supplementary Table S7). We further studied the interaction of LOLA^{Nmut} with MAMO^{Nmut} in detail using Y2H, since this interaction was not affected by mutations in the dimer-dimer interface. Both LOLA^{Nmut} and MAMO^{Nmut} retained the ability to form heteromers with both wild-type BTB domains, but the interaction was lost between mutants (Figure 6C).

Taken together, our data suggest that selective heteromultimerization of TTK-type BTB domains occurs mostly through dimer-dimer interaction via the conserved interface.

2.6. TTK-type BTB domains are specific to Arthropoda

In *D. melanogaster*, most transcription factor BTB domains are TTK-type. We therefore investigated when these domains emerged over the course of evolution and how widespread they are in animals. We built the Hidden Markov Model (HMM) profile “TTK motif” based on sequence alignment of TTK-type

BTBs from 14 Dipteran species (Figure 2B) and performed a search within the main groups of Arthropoda and several other Metazoan species (including basal groups such as Onychophora, Tardigrada, and Nematoda). TTK-type BTB domains were not found beyond outside of Arthropoda. The most basal clades in which they emerged were Crustaceans and Arachnoidea (Figure 7A). This agrees with a recent study which traced the origin of GAF and Mod(mdg4) insulator proteins to ancestral groups of insects related to modern *Protura* and *Plecopthera* [61]. TTK-type BTB domains underwent lineage-specific expansion in modern phylogenetic groups of insects, to the most extent in Diptera and Hymenoptera. In these species, TTK-type BTB domains almost completely replace transcription factors with classical dimeric BTB domains, as in *D. melanogaster* (Figure 7A). An ortholog search of the 11 best-studied *Drosophila* TTK-type BTB proteins revealed that the oldest proteins with TTK-type BTBs are LOLA, bric-a-brac2 and Tramtrack, for which orthologs were found in the most basal clades (Figure 7B). The same search for four non-TTK *Drosophila* BTB proteins revealed that CP190 and CG6792 (a *Drosophila* PLZF homologue) are the oldest: a CP190 ortholog was found in Nematodes and CG6792 orthologs are present in almost all higher Metazoan taxa (Figure 7C). Notably, each non-TTK *Drosophila* BTB-C2H2 protein has only one isoform, whereas many factors with a TTK-type BTB have multiple isoforms, which typically differ in their DNA-binding domain, resulting in a wide diversity of these factors.

Most BTB proteins also contain C2H2 or HTH DNA-binding domains. With the exception of Arachnida, in all examined Arthropoda species, the FLYWCH and HTH domains are found only in combination with the TTK-type BTB domains (Figure 8A). In the human proteome, BTB-containing transcription factors almost exclusively use C2H2 zinc-fingers as DNA-binding domains. Interestingly, TTK-type BTBs are usually associated with one or two C2H2 domains (only MAMO has five C2H2 domains), while mammalian transcription factors commonly have BTB domains in combination with arrays of many (on average, five) C2H2 domains (Figure 8B).

Taken together, our results suggest that TTK-type BTBs are associated with a specific group of domains in transcription factors of Arthropoda.

3. Discussion

The TTK-type BTB domains form a distinct group of BTB domains specific to Arthropodan transcription factors. Here, we studied the mechanism of multimer formation of TTK-type BTB domains using an integrative structural biology approach. Molecular modeling, SAXS, MALS and microscopy techniques revealed that these domains form hexamers consisting of three dimers assembled through a novel dimer-dimer interaction interface. A similar surface (based on the β3-strand) was recently found to be involved in the Miz-1 BTB interaction with the HUWE-1 protein [62] and was earlier was implicated in Miz-1 tetramer formation [22]. Notably, the dimer-dimer interaction in Miz-1 involves conformational flexibility in this region, and monomers within each dimer are non-equivalent, with only one containing the β3-strand. Our models predict that dimers of TTK-type BTB domains consist of identical monomers; however, further structural studies will be required to elucidate precise details of this interface. Single amino acid substitutions further confirmed that hydrophobic residues at the dimer-dimer interface are critical for multimer formation. The characteristic conserved N-terminal sequence FxLRWN forms the first β-strand and was found to be important for specific dimer formation rather than for higher-order oligomerization as previously suggested [31]. TTK-type BTB domains possess an unusually wide

potential for heteromeric interactions despite a rather low sequence homology. Heteromultimerization likely occurs through the interaction of different dimers, since the dimer-dimer interface is highly similar in different TTK-type BTB domains. This is a unique property of this class of BTB domains, as classical BTB domains form almost exclusively homodimers.

The expansion of the TTK-type BTB domains in insects suggests an important but poorly-understood functional role for their ability to form homo- and hetero-multimers. Most of the transcription factors of the TTK group bind to short degenerate DNA motifs. Thus, the multimerization of proteins via the BTB domain can increase the binding affinity for several motifs located in close vicinity. Such a mechanism was previously demonstrated for the pioneer and bookmarking protein GAF that is stably associated with chromatin [45, 63]. GAF has only one zinc finger domain that binds to a GAGAG motif. It was shown that BTB oligomerization mediates strong co-operative binding of GAF to multiple sites but inhibits binding to a single motif [64, 65].

Heteromultimerization of BTB domains can allow the formation of complexes of several TTK-type proteins on chromatin. The most striking example is the Batman protein, which consists only of the BTB domain [66]. Batman forms complexes with Pipsqueak and GAF by interacting with their BTBs [67, 68]. Batman also interacts with the BAB1 BTB domain, and these proteins cooperate in the control of sex combs on male tarsa [69]. Since the Batman BTB can be recruited to chromatin only through interaction with other DNA-binding BTB proteins, it does not form homomultimers that would prevent it from interaction with other BTBs. Thus, TTK-type BTB transcription factors may constitute a regulatory network controlling common target genes.

In this study, we have shown that TTK-type BTB domains are Arthropod-specific and underwent lineage-specific expansion in modern insects. The *Drosophila* proteome contains 24 transcription factors with TTK-type BTB domains. Moreover, non-TTK-type BTB domains are found in only four *Drosophila* transcription factors, suggesting that the TTK-type BTB domain confers some evolutionary-important benefits. The most basal clades in which TTK-type BTBs can be traced are Crustacea and Arachnoidea. The only such proteins present in both taxa are Tramtrack orthologs, suggesting that it is the oldest member of the family.

In human and mouse, BTB-C2H2 proteins are encoded by at least 49 genes that are important regulators of development and commonly function as sequence-specific repressors of gene expression. In contrast to Arthropoda TTK-type BTB proteins, the mammalian BTB-C2H2 proteins usually have many C2H2 domains that recognize extended and specific DNA motifs. Despite the fact that mammals have more highly-specific DNA-binding proteins, many of the TTK-type BTB transcription factors in Arthropods have multiple isoforms with different DNA-binding or protein interaction domains which, along with the ability to heteromultimerize, can result in a large diversity of complexes possessing a broad range of DNA-binding specificities.

In conclusion, TTK-type BTB proteins form a structurally and functionally distinct group of Arthropod key regulatory factors with unique functions in the process of cell differentiation and transcriptional regulation. What functional role is played by the ability of these transcription factors to form various combinations of heterologous complexes remains an open and very interesting question.

4. Methods

4.1. Bioinformatics

The search for orthologs for 24 *D. melanogaster* proteins with BTB domains containing ttk sequences was carried out in Metazoa, with the exception of Chordata, using the OrthoDB database [70]. There were no records for Mod(mdg4), GAF, and LOLA in OrthoDB, and orthologs of Fruitless were found only in Diptera. In order to fill up the data obtained from OrthoDB, orthologs for these proteins were searched in Ensembl using biomaRt [71]. The total number of identified orthologs was 3027. Amino acid sequences of orthologs of 23 ttk proteins of *D. melanogaster* in 14 Diptera species (five *Drosophila* species, two *Aedes* species, two *Anopheles* species, one *Culex* and fore flies) were aligned in MUSCLE [72]. The motif of the ttk domain in the resulting alignment was isolated using the EM algorithm by positional weight matrix (PWM) (at no limit motif E-value threshold and motif size 6-10 aa (Figure 2A)) in MEME (MEME SUITE) [73], then in FIMO (MEME SUITE) we searched for the motif in the obtained database of orthologous sequences for all Metazoa (the threshold FDR value was 0.001). In addition, for the training set for 14 Diptera species, an hmm profile for the ttk domain was generated in MEGA-X [74], and it was searched using HMMSEARCH (HmmerWeb version 2.41.1 [75]) in the UniProtKB database at E-value = 10. Data on full protein sequences (n = 259, excluding isoforms) deposited in UniProtKB were combined with previously obtained ttk-containing sequences of orthologs from OrthoDB and Ensembl (n = 2208). In addition, in the large taxa closest to Arthropoda (several phyla of Protostomes: Tardigrada, Onychophora, Priapulida, Kinorhyncha, Loricifera, Nematoda, Annelida, Mollusca), the search for the ttk sequence motif in blastp and hmm-profile in HMMSEARCH were carried out; in the representatives of the above taxa, no proteins containing the ttk motif were found in non-redundant DBs.

Phylogenetic classification was reconstructed in accordance with the NCBI Taxonomy Database using taxize [76], visualization of the phylogenetic tree is implemented in ggtree [77]. In the most basal clades (Crustacea, Arachnida, Myriapoda), the presence of the corresponding orthologs was predicted manually in blastp, due to the low similarity of amino acid sequences. We considered as orthologs, in addition to the previously annotated ones, only those proteins for which homology was observed in the structure, in both the BTB domain (including the presence of the ttk sequence) and the DNA binding domain. The presence/absence of some orthologs in individual orders of insects was also checked in blastp and HMMSEARCH (HmmerWeb version 2.41.1). In addition, the previously published data by Pauli et al. [61] were used for Mod (mdg4) and GAF.

The domain structure of the detected orthologs of ttk-containing proteins was studied in 37 reference species - representatives of 16 Hexapoda orders, two orders of Crustaceans and Arachnids, and one - Myriapoda (Figure 5A). The search for domains in the orthologous sequences was performed using the Pfam database in Batch CD-Search (CDD / SPARCLE NCBI (Lu, Wang et al., 2020)), the threshold E-value was taken to be 0.5. The detection results were filtered: incomplete and overlapping domains were removed. In sum, after filtering, the total number of domains found was 1055; the analyzed domains belonged to 26 PFAM families.

For 16 species with well-annotated proteomes belonging to the phyla Arthropoda, Nematoda, Annelida, as well as for *Homo sapiens*, the sequences of all proteins containing BTB domains in their proteomes were searched using InterProScan in the Pfam protein domain database by the generalized hmm -BTB domain profile. In the set of the found sequences, the search for the ttk-domain motif in FIMO was carried out as described above, the ttk-containing proteins were filtered, and the remaining

sequences were analyzed in Batch CD-Search in order to identify their domain structure. In total, 3028 domains belonging to 99 PFAM families were found in the sequences of BTB proteins without the ttk motif in 16 species.

4.2. Plasmids and cloning

cDNAs of BTB domains were PCR-amplified using corresponding primers (Supplementary Table S1) and cloned into modified pET32a(+) vector (Novagen) encoding TEV protease cleavage site after 6xHis-tag and Thioredoxin, and into pGAD424 and pGBT9 vectors (Clontech) in frame with GAL4 Activation or DNA-binding domains, respectively. PCR-directed mutagenesis was used to create constructs expressing mutant BTBs using mutagenic primers (Supplementary Table S1).

4.3. Protein expression and purification

BL21(DE3) cells transformed with a construct expressing BTB domain fused with TEV-cleavable 6xHis-Thioredoxin were grown in 1 L of LB media to an D600 of 1.0 at 37 °C and then induced with 1 mM IPTG at 18 °C overnight. Cells were disrupted by sonication in buffer A (30 mM HEPES (pH 7.5), 400 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol, 0.1% NP40, 10 mM imidazole) containing 1 mM PMSF and Calbiochem Complete Protease Inhibitor Cocktail VII (1 μL/ml). After centrifugation, lysate was applied to a Ni-NTA column, and after washing with buffer B (30 mM HEPES (pH 7.5), 400 mM NaCl, 5 mM β-mercaptoethanol, 30 mM imidazole) was eluted with 300 mM imidazole. For cleavage of the 6x-His-thioredoxin-tag, 6x-His-tagged TEV protease was added at a molar ratio of 1:50 directly to the eluted protein and the mixture was incubated for 2 hours at room temperature, then dialyzed against buffer A without NP-40 and applied to a Ni-NTA column. Flow-through was collected; dialyzed against 20 mM Tris-HCl (pH 7.4), 50 mM NaCl and 1 mM DTT; and then applied to a SOURCE15Q 4.6/100 column (GE Healthcare) and eluted with a 50-500 mM linear gradient of NaCl. Fractions containing protein were concentrated, frozen in liquid nitrogen, and stored at -70 °C. Size-exclusion chromatography was performed in 20 mM Tris-HCl (pH 7.4), 200 mM NaCl and 1 mM DTT using Superdex S200 10/300GL column (GE Healthcare).

4.4. SEC-MALS

Size-exclusion chromatography with multi-angle light scattering (SEC-MALS) detection was used to determine absolute Mw for LOLA and MBP-LOLA samples. Protein samples (1-5 mg/ml) were loaded individually on a Superdex 200 Increase 10/300 column (GE Healthcare), and the elution profiles were obtained using a tandem of sequentially connected UV-Vis Prostar 335 (Varian, Australia) and miniDAWN detectors (Wyatt Technology, USA). The column was pre-equilibrated with filtered (0.1 μm) and degassed 20 mM Tris-HCl buffer, pH 7.6, containing 200 mM NaCl and 5 mM β-mercaptoethanol and was operated at a 0.8 ml/min flow rate. Data were processed in ASTRA 8.0 (Wyatt Technology, USA) using dn/dc equal to 0.185 and extinction coefficients ε(0.1%) at 280 nm equal to 0.80 ml/(mg cm) and 1.38 ml/(mg cm) for LOLA and MBP-LOLA, respectively. Additionally, apparent Mw values were determined from column calibration with standard proteins. Data were processed and presented using Origin 9.0 (Originlab, Northampton, MA, USA).

4.5. Negative-staining EM and AFM

Negative staining EM was used to get insight into a low-resolution structure of the MBP-LOLA chimera. Copper grids supported by a thin layer carbon film (Lacey C only 300 mesh, TED PELLA, USA) were used in the experiment. The grids were glow-discharged for 30 s with a PELCO easyGlow system (Ted Pella, USA) using standard protocol. Twenty µL of the protein (10 mg/ml) in 20 mM Tris-HCl (pH 7.4), 200 mM NaCl and 1 mM DTT buffer was applied to a grid and stained with 1 % uranyl acetate in ethanol. The negative stained grid was examined using JEM-2100 (JEOL, Japan) equipped with a Direct Electron DE-20 (Gatan, USA) camera. The microscope operated at 200 kV with a pixel size of 2.5 Å, the total dose was 80 $e^{-}/Å^{2}$. Data collection was performed using SerialEM software [78].

Three hundred micrographs of the negatively stained MBP-LOLA were processed in CryoSPARC (v.3.3.0) [79]. Particle picking was performed after CTF-correction using Blob picker with a particle diameter in range 180-250 Å. Several rounds of 2D classification with subsequent template picking with a particle diameter of 200 Å were performed. Visual analysis of thirty 2D classes containing ~20 000 particles in total allowed to identify few classes with an apparent 6-fold symmetry.

AFM experiments were performed using a Multimode Nanoscope V AFM (Bruker, USA) in a tapping mode in air. For the sample preparation, a 10 µL aliquot of MBP-Lola protein solution in an appropriate concentration of 0.006 µM was deposited onto freshly cleaved mica, incubated for 2 minutes, then gently rinsed with Milli-Q water and dried by nitrogen. AFM measurements were made using ultrasharp silicon NSG30_SS tips (Tipsnano, Estonia) having a spring constant of 40 N/m and a radius of curvature of 2 nm. Scan rate was typically 2 Hz. Image processing was performed using the FemtoScan Online software (Advanced Technologies Center, Russia).

4.6. Yeast two-hybrid assay

Yeast two-hybrid assay was performed as described [31].

4.7. SAXS data collection and analysis

Synchrotron radiation X-ray scattering data were collected using standard procedures on the BM29 BioSAXS beamline at the ESRF (Grenoble, France) as described previously [80]. Data analysis was performed using ATSAS software package [81]. Approximation of the experimental scattering profiles using calculated scattering curves was performed with CRYSOL [60]. The molecular mass (MM) of the protein was calculated using several algorithms implemented in ATSAS package [81].

Data accessibility

SAXS data have been deposited in the Small Angle Scattering Biological Data Bank (www.sasbdb.org) under accession codes SASDP59 (merged data for $LOLA^{1-120}$ at 1.0 mg/ml and 3.0 mg/ml), SASDP49 ($CG6765^{1-133}$ at 1.5 mg/ml). Atomic models (both native and exactly corresponding to expression constructs) and reports of SAXS approximation are provided as Supplemental files. Results of bioinformatic analysis are provided as Supplemental tables.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Russian Science Foundation – project 19-74-10099 to A.B. (expression and purification of proteins and their mutants), project 19-74-30026 to P.G. (analysis of protein-protein interactions), and by Ministry of Science and Higher Education of the Russian Federation — grants 075-15-2019-1661 (structural and bioinformatic analysis) and 075-15-2021-1354 (X-ray scattering experiments). Funding for open access charge: Ministry of Science and Higher Education of the Russian Federation and Russian Science Foundation.

Authors' contributions

A.N.B.: conceptualization, formal analysis, investigation, project administration, visualization, resources, software, methodology, validation and writing—original draft; K.I.B.: molecular cloning, protein interaction assays; K.M.B.: electron microscopy data analysis, molecular modelling, writing—original draft; N.N.S.: SEC-MALS, SAXS, editing; A.D.B.: negative staining electron microscopy; N.V.K. ; O.V.A.: protein expression and purification; A.M.K.: bioinformatic analysis, software and writing—original draft; K.Y.K.: protein interaction assays; V.O.P.: conceptualization, data curation, project administration; P.G.G.: conceptualization, data curation, project administration, funding acquisition, and writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Acknowledgements

We are grateful to Dr. Alexander Kuklin (Joint Institute of Nuclear Research, Dubna) for help in SAXS data collection.

Figure legends

Figure 1. Schematic representation of Drosophila transcription factors with BTB domains. Only one representative isoform is shown for each gene.

Figure 2. Characterization of the TTK-type BTB domains. (A) Multiple sequence alignment of BTB domains from *Drosophila* transcription factors and a few human BTB domains with known 3D structures. **(B)** HMM-profile model for the TTK motif obtained for 14 Diptera species. **(C)** Testing of the GAF, Mod(mdg4), and LOLA BTB domains for interaction with all TTK-type BTB domains found in *Drosophila melanogaster*. Original data are shown in Supplementary Table S3.

Figure 3. An integrative biology approach reveals the hexameric assembly of BTB domains. SEC-MALS data for LOLA **(A)** and MBP-LOLA **(B)** showing the chromatographic peaks with the Mw distributions across each peak. Average Mw values in kDa, polydispersity index (Mw/Mn) and the formally calculated oligomeric state are shown. For comparison, apparent Mw values determined from column calibration are also presented. Note that tenfold dilution of LOLA (black and red curves in panel A) did not cause any shift of the chromatographic peak, indicating stability of the observed oligomer. Alphafold2-derived models of LOLA **(C)** and CG6765 **(D)** oligomers and their fits to the experimental SAXS data. Positions of the TTK motif and dimer-dimer interaction interface are shown with arrows. **(E,F)** Microscopic studies of MBP-LOLA1-120 fusion protein. **(E)** 2D classes with apparent 6-fold symmetry

revealed by negative-stain cryo-electron microscopy. **(F)** A set of hexameric particles revealed by atomic force microscopy. Tentative MBP subunits located on the periphery of the LOLA hexamer are numbered.

Figure 4. Testing the impact of single amino-acid substitutions in dimer-dimer interaction interface on the oligomerization status of TTK-type BTB domains. Close-up views of AlphaFold-derived molecular models of dimer-dimer interaction interfaces and effect of single amino-acid substitutions on the oligomerization status of BTB domains of Mod(mdg4) **(A)**, LOLA **(B)**, and CG6765 **(C)** studied with Superdex S200 size-exclusion chromatography. Amino acids subjected to mutagenesis are shown in red. Positions of molecular weight markers are shown below. Molecular weights of monomers are shown in brackets. Further details are shown in Supplementary Figures S7, S8, and S9. **(D)** The self-interaction ability studied with a yeast two-hybrid assay. Growth assay plates are shown in Supplementary Figure S10.

Figure 5. Testing the impact of single amino-acid substitutions in the TTK motif on the oligomerization status of TTK-type BTB domains. (A) Summary of single amino-acid substitution effects on multimerization studied with size-exclusion chromatography. Dimerization ability in yeast two-hybrid assay was used as a control for correct folding. Yeast growth assay plates are shown in Supplementary Figure S1. **(B)** The structure of the region surrounding the TTK motif in the LOLA protein according to AlphaFold modelling. **(C)** Effect of single amino-acid substitutions on the oligomerization status of BTB domains of LOLA (left) and Mod(mdg4) (right) studied with Superdex S200 size-exclusion chromatography. Positions of molecular weight markers are shown below. Molecular weights of monomers are shown in brackets.

Figure 6. The effect of TTK motif and hexamerization interface mutations on heteromeric interactions. Summary of alterations in heteromeric interactions between LOLA/Mod(mdg4) BTB domains and other TTK-type BTBs induced by substitution of residues at the dimer-dimer interaction interface **(A)** and all six residues in the N-terminal TTK motif **(B)**. Complete data are shown in Supplementary Tables S6 and S7 and Supplementary Figures S1 and S10. **(C)** Effect of substitution of all six residues in the N-terminal conserved sequence on interaction between LOLA and MAMO BTB domains in the yeast two-hybrid assay. AD stands for GAL4 activation domain and BD for GAL4 DNA-binding domain. Assay plates without histidine are shown on the left, the summary of interactions is shown on the right.

Figure 7. (A) Phylogenetic analysis of the distribution of the main DNA and protein interaction domain types in TTK-type BTB proteins in proteomes of representatives of Hexapoda, Crustacea, Myriapoda, and Arachnida. Each type of BTB protein domain architecture is shown as a bar segment. Total search results are shown in Supplementary Table S8. The orthologs of several *Drosophila* BTB domains of TTK-type **(B)** and non-TTK-type **(C)** in proteomes of key taxa in major Arthropod and other Ecdysozoa phylogenetic groups. The phylogenetic relationships among taxa are according to NCBI Taxonomy Database. Azure blue – the ortholog is absent in the taxa, dark blue – the ortholog is present in the taxa, dark cyan – orthologs are not uniquely defined.

Figure 8. (A) Distribution of major DNA interaction domain types in non-TTK (left) and TTK-containing (right) BTB proteins in the proteomes of 16 Metazoan species. Each type of BTB-associated protein domain is denoted as a bar segment. Total search results are shown in Supplementary Tables S8 (TTK BTB domains) and S9 (non-TTK BTB domains). **(B)** Representation of non-TTK and TTK-containing BTB proteins with less than three (dark cyan) and three or more (dark blue) C2H2 domains in *Drosophila melanogaster* and *Homo sapiens* proteomes.

Table 1. SAXS-derived structural parameters for CG6765 and LOLA BTB domains. Rg is radius of gyration, Dmax – maximum dimension of the particles, Vp – Porod volume – volume of the particles. The sample of CG6765$^{1-133}$ with concentration of 1.5 mg/ml provided sufficiently high quality scattering data, which were used for all fitting experiments, whereas data obtained for LOLA$^{1-120}$ at concentrations 1.0 and 3.0 mg/ml were merged to obtain the necessary quality. Samples with higher concentrations exhibited signs of aggregation and were excluded from analysis.

Polypeptide	Sample concentration, mg/ml	Rg, nm	Dmax, nm	Vp, nm3	Estimated molecular weight, kDa	Molecular weight of the monomer, kDa
CG6765$^{1-133}$	1.5	3.70	12.73	166	90.5-116.5	15.3
LOLA$^{1-120}$	merged data for 1.0 and 3.0 mg/ml	4.1	14.5	189	108-136	15.2

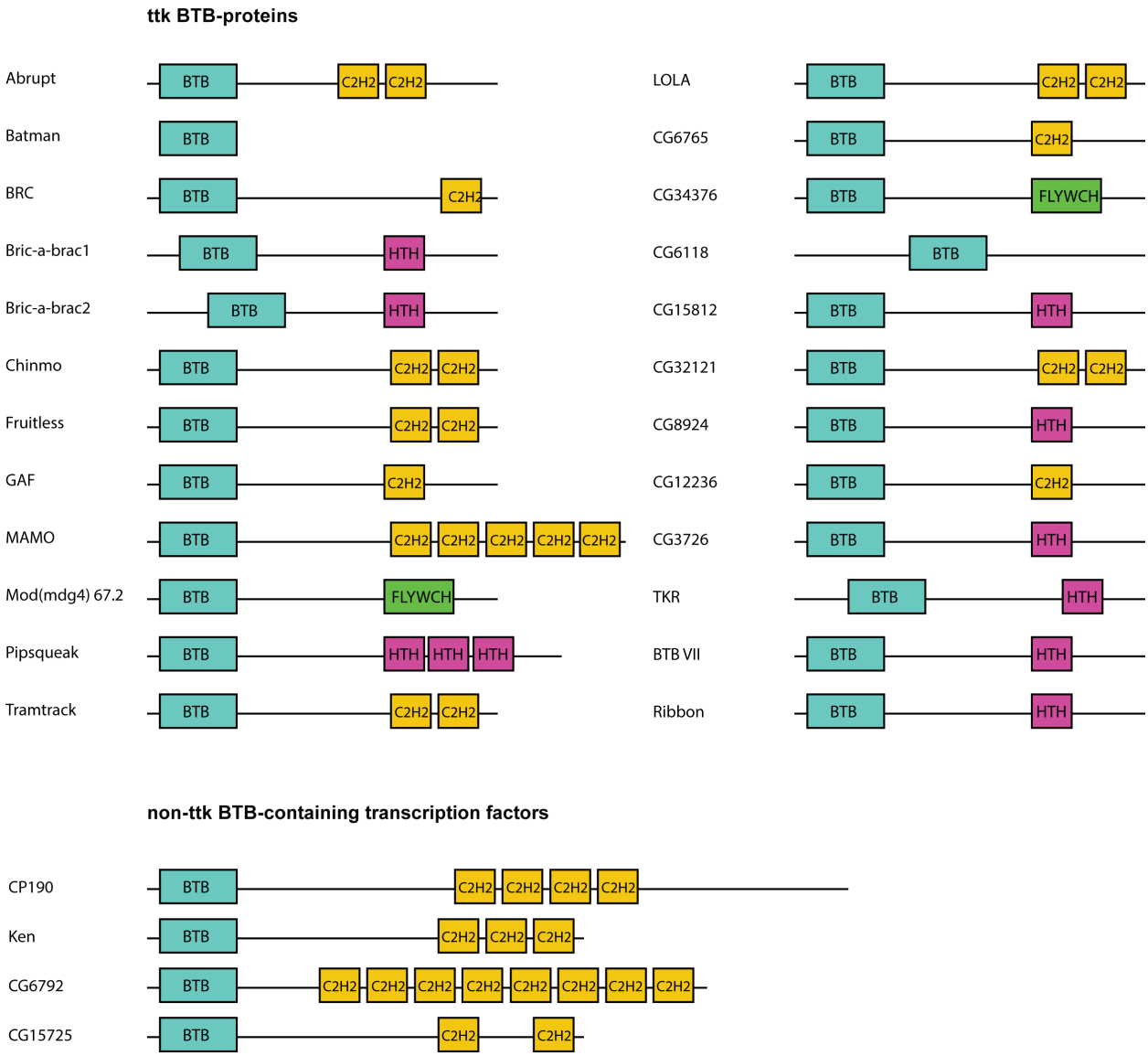


Figure 1. Schematic representation of Drosophila transcription factors with BTB domains. Only one representative isoform is shown for each gene.

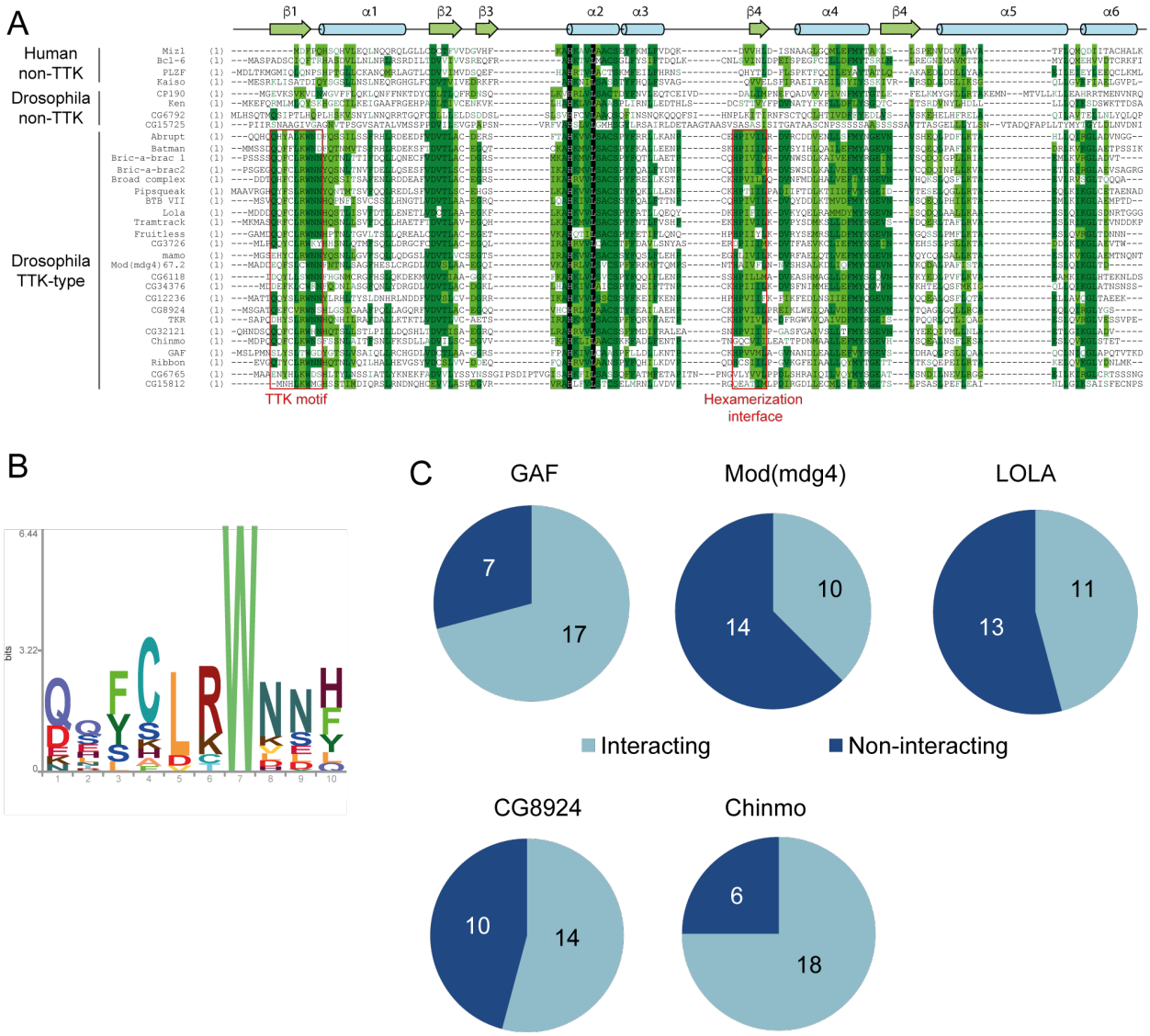


Figure 2. Characterization of the TTK-type BTB domains. (A) Multiple sequence alignment of BTB domains from *Drosophila* transcription factors and a few human BTB domains with known 3D structures. **(B)** HMM-profile model for the TTK motif obtained for 14 Diptera species. **(C)** Testing of the GAF, Mod(mdg4), and LOLA BTB domains for interaction with all TTK-type BTB domains found in *Drosophila melanogaster*. Original data are shown in Supplementary Table S3.

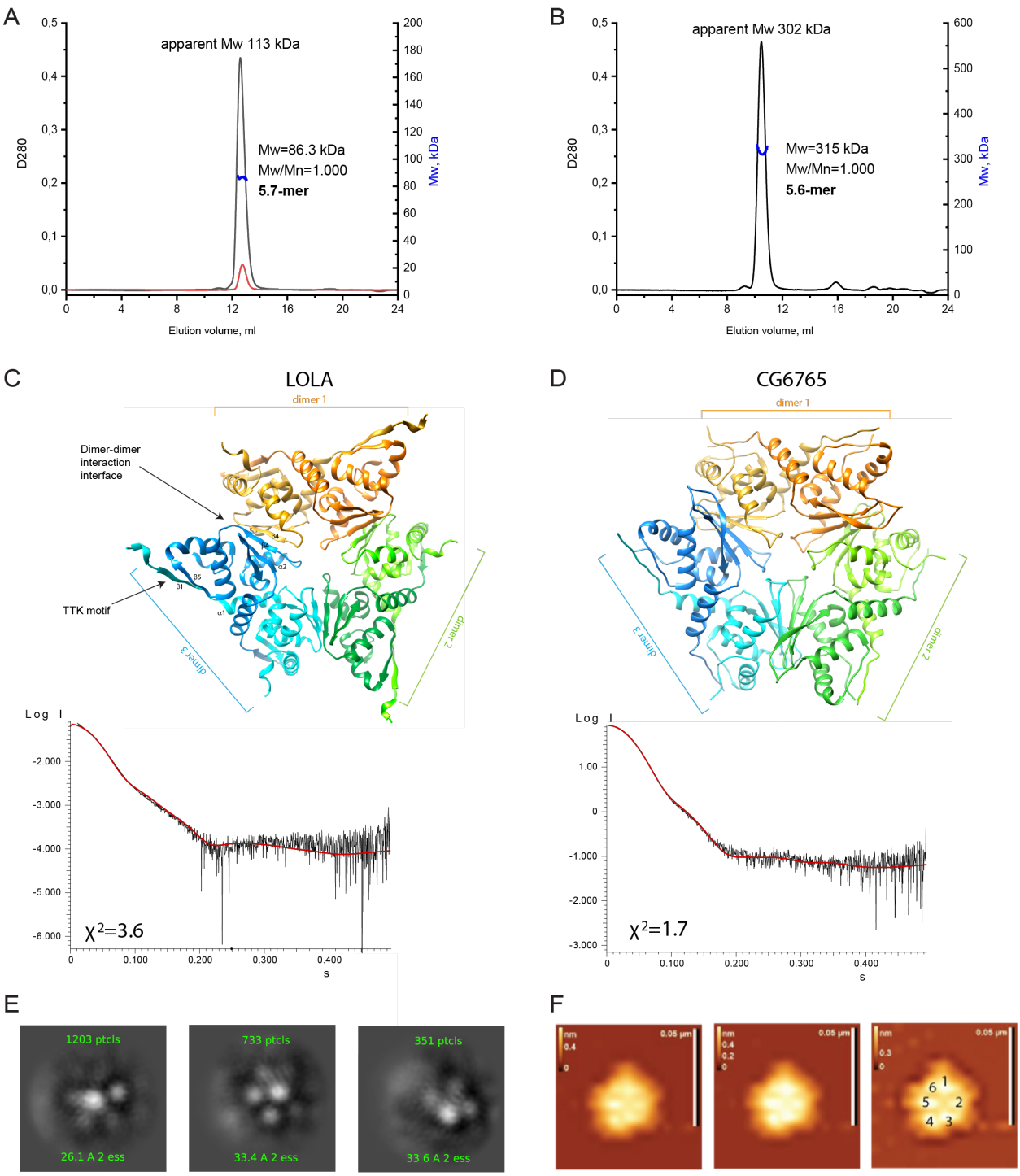
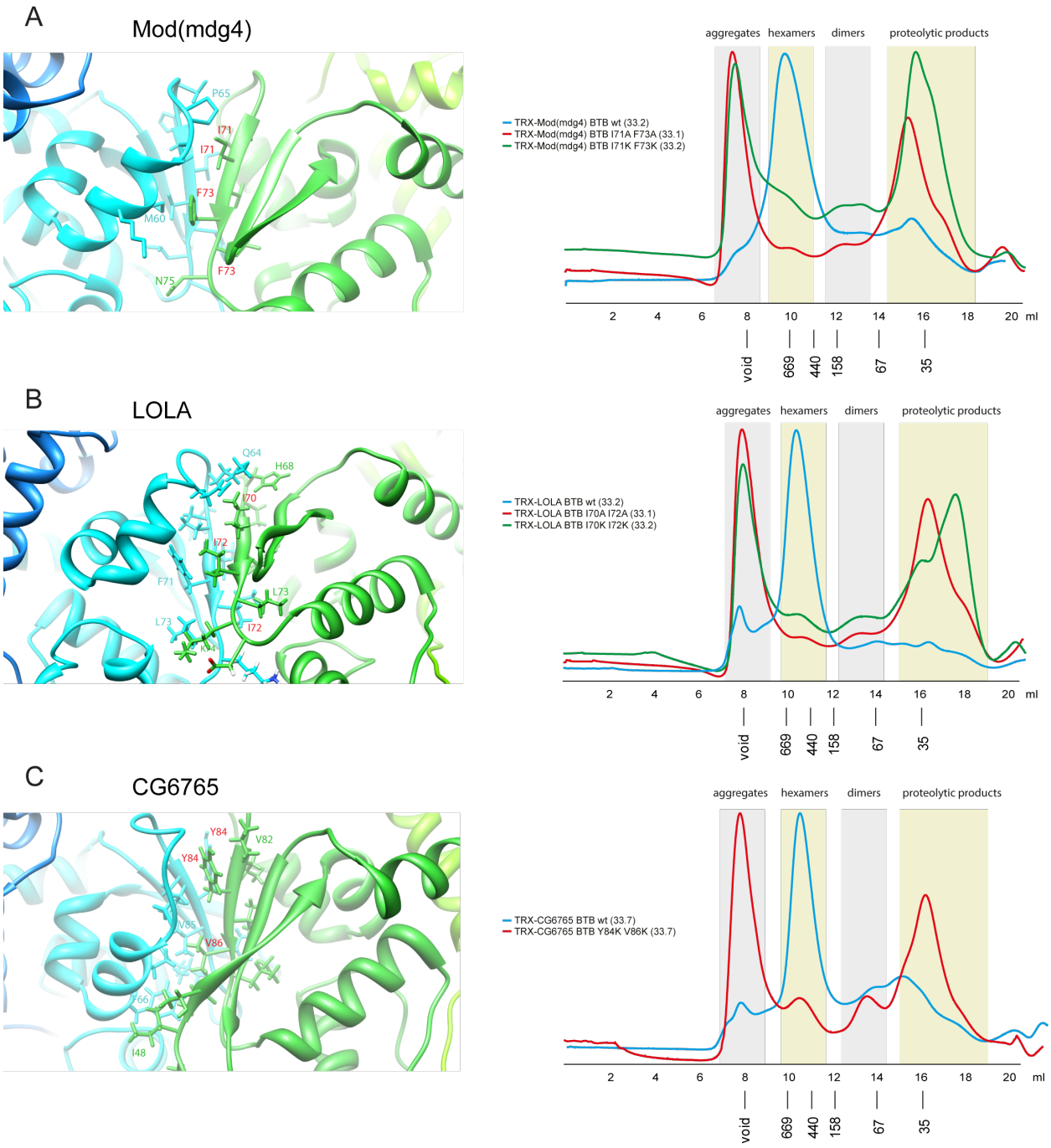


Figure 3. An integrative biology approach reveals the hexameric assembly of BTB domains. SEC-MALS data for LOLA **(A)** and MBP-LOLA **(B)** showing the chromatographic peaks with the Mw distributions across each peak. Average Mw values in kDa, polydispersity index (Mw/Mn) and the formally calculated oligomeric state are shown. For comparison, apparent Mw values determined from column calibration are also presented. Note that tenfold dilution of LOLA (black and red curves in panel A) did not cause any shift of the chromatographic peak, indicating stability of the observed oligomer. Alphafold2-derived models of LOLA **(C)** and CG6765 **(D)** oligomers and their fits to the experimental SAXS data. Positions of the TTK motif and dimer-dimer interaction interface are shown with arrows. **(E,F)** Microscopic studies of MBP-LOLA1-120 fusion protein. **(E)** 2D classes with apparent 6-fold symmetry revealed by negative-stain cryo-electron microscopy. **(F)** A set of hexameric particles revealed by atomic force microscopy. Tentative MBP subunits located on the periphery of the LOLA hexamer are numbered.



D

	LOLA$^{I70A;I72A}$	LOLA$^{I70K;I72K}$	MOD$^{I71A;F73A}$	MOD$^{I71K;F73K}$
Self-interaction control	+	–	+	–
Interaction with wild-type BTB	+	–	+	+

Figure 4. Testing the impact of single amino-acid substitutions in dimer-dimer interaction interface on the oligomerization status of TTK-type BTB domains. Close-up views of AlphaFold-derived molecular models of dimer-dimer interaction interfaces and effect of single amino-acid substitutions on the oligomerization status of BTB domains of Mod(mdg4) **(A)**, LOLA **(B)**, and CG6765 **(C)** studied with Superdex S200 size-exclusion chromatography. Amino acids subjected to mutagenesis are shown in red. Positions of molecular weight markers are shown below. Molecular weights of monomers are shown in brackets. Further details are shown in Supplementary Figures S7, S8, and S9. **(D)** The self-interaction ability studied with a yeast two-hybrid assay. Growth assay plates are shown in Supplementary Figure S10.

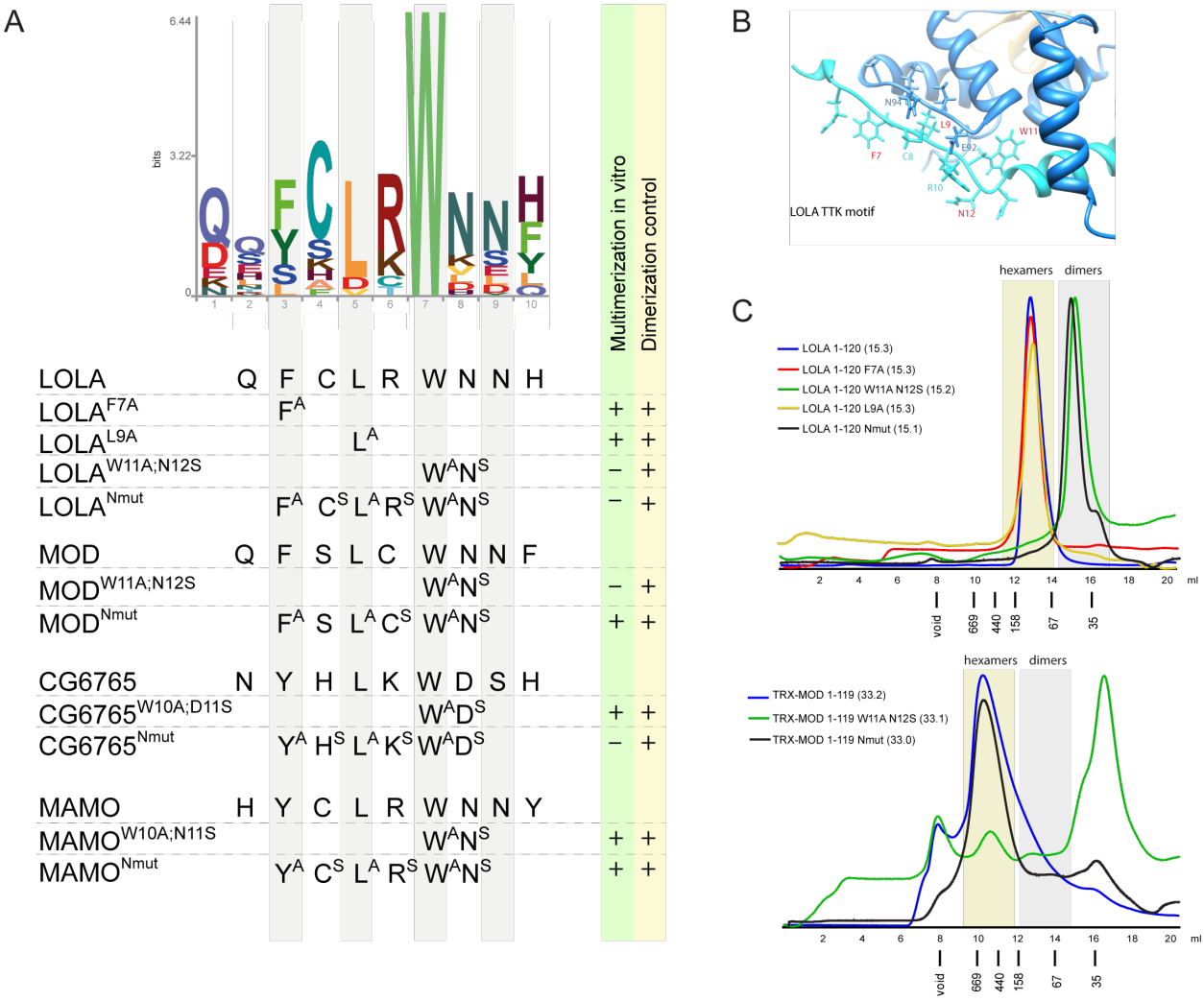
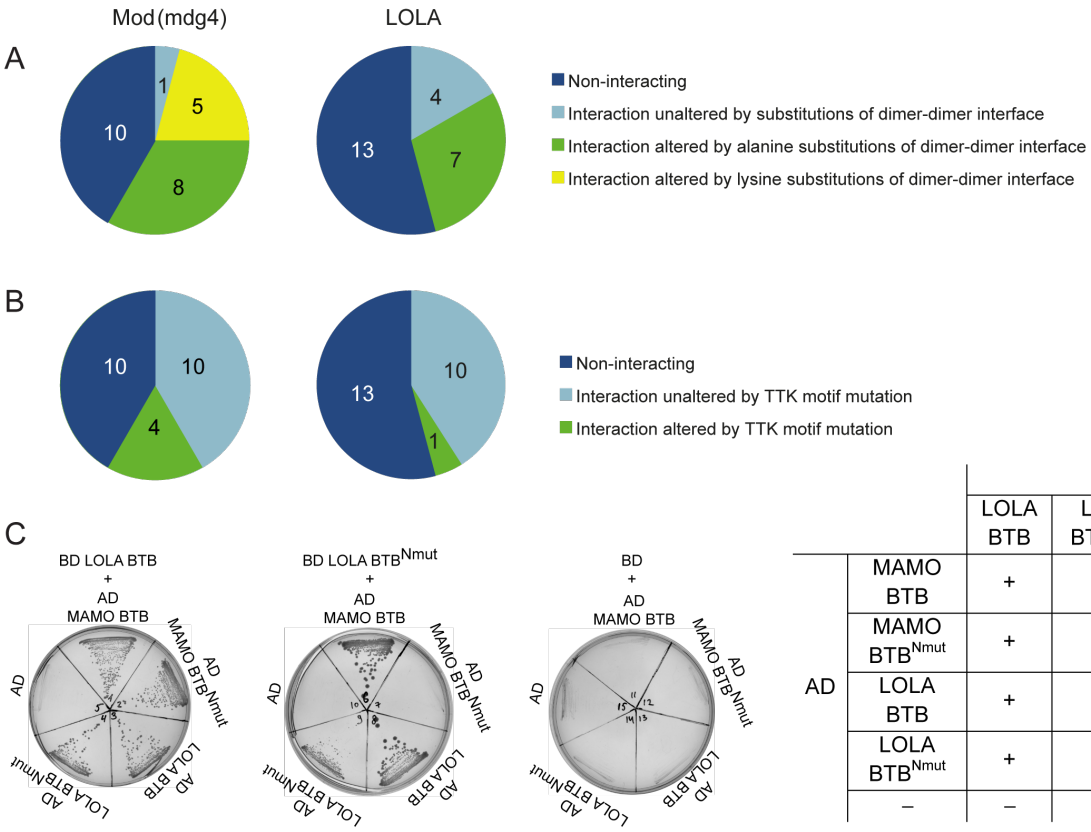


Figure 5. Testing the impact of single amino-acid substitutions in the TTK motif on the oligomerization status of TTK-type BTB domains. (A) Summary of single amino-acid substitution effects on multimerization studied with size-exclusion chromatography. Dimerization ability in yeast two-hybrid assay was used as a control for correct folding. Yeast growth assay plates are shown in Supplementary Figure S1. **(B)** The structure of the region surrounding the TTK motif in the LOLA protein according to AlphaFold modelling. **(C)** Effect of single amino-acid substitutions on the oligomerization status of BTB domains of LOLA (left) and Mod(mdg4) (right) studied with Superdex S200 size-exclusion chromatography. Positions of molecular weight markers are shown below. Molecular weights of monomers are shown in brackets.



		BD		
		LOLA BTB	LOLA BTB^{Nmut}	–
AD	MAMO BTB	+	+	–
	MAMO BTB^{Nmut}	+	–	–
	LOLA BTB	+	+	–
	LOLA BTB^{Nmut}	+	+	–
	–	–	–	–

Figure 6. The effect of TTK motif and hexamerization interface mutations on heteromeric interactions. Summary of alterations in heteromeric interactions between LOLA/Mod(mdg4) BTB domains and other TTK-type BTBs induced by substitution of residues at the dimer-dimer interaction interface **(A)** and all six residues in the N-terminal TTK motif **(B)**. Complete data are shown in Supplementary Tables S6 and S7 and Supplementary Figures S1 and S10. **(C)** Effect of substitution of all six residues in the N-terminal conserved sequence on interaction between LOLA and MAMO BTB domains in the yeast two-hybrid assay. AD stands for GAL4 activation domain and BD for GAL4 DNA-binding domain. Assay plates without histidine are shown on the left, the summary of interactions is shown on the right.

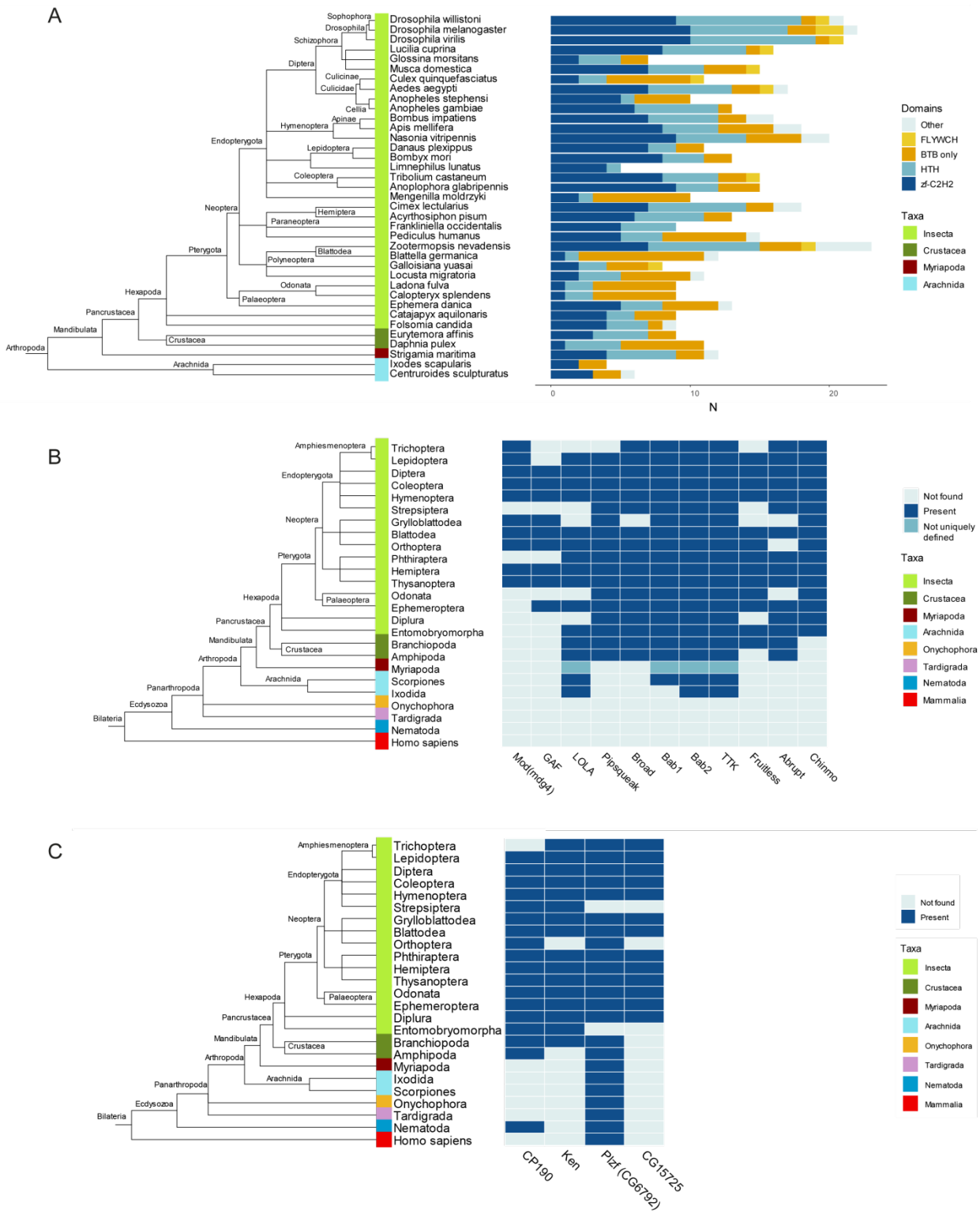


Figure 7. (A) Phylogenetic analysis of the distribution of the main DNA and protein interaction domain types in TTK-type BTB proteins in proteomes of representatives of Hexapoda, Crustacea, Myriapoda, and Arachnida. Each type of BTB protein domain architecture is shown as a bar segment. Total search results are shown in Supplementary Table S8. The orthologs of several *Drosophila* BTB domains of TTK-type **(B)** and non-TTK-type **(C)** in proteomes of key taxa in major Arthropod and other Ecdysozoa phylogenetic groups. The phylogenetic relationships among taxa are according to NCBI Taxonomy Database. Azure blue – the ortholog is absent in the taxa, dark blue – the ortholog is present in the taxa, dark cyan – orthologs are not uniquely defined.

Figure 8. (A) Distribution of major DNA interaction domain types in non-TTK (left) and TTK-containing (right) BTB proteins in the proteomes of 16 Metazoan species. Each type of BTB-associated protein domain is denoted as a bar segment. Total search results are shown in Supplementary Tables S8 (TTK BTB domains) and S9 (non-TTK BTB domains). **(B)** Representation of non-TTK and TTK-containing BTB proteins with less than three (dark cyan) and three or more (dark blue) C2H2 domains in *Drosophila melanogaster* and *Homo sapiens* proteomes.

References

1 Zollman, S., Godt, D., Prive, G. G., Couderc, J. L., Laski, F. A. 1994 The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. *Proc Natl Acad Sci U S A*. **91**, 10717-10721. (10.1073/pnas.91.22.10717)
2 Godt, D., Couderc, J. L., Cramton, S. E., Laski, F. A. 1993 Pattern formation in the limbs of Drosophila: bric a brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development*. **119**, 799-812.
3 DiBello, P. R., Withers, D. A., Bayer, C. A., Fristrom, J. W., Guild, G. M. 1991 The Drosophila Broad-Complex encodes a family of related proteins containing zinc fingers. *Genetics*. **129**, 385-397. (10.1093/genetics/129.2.385)
4 Harrison, S. D., Travers, A. A. 1990 The tramtrack gene encodes a Drosophila finger protein that interacts with the ftz transcriptional regulatory region and shows a novel embryonic expression pattern. *EMBO J*. **9**, 207-216.
5 Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K., Prive, G. G. 2005 Sequence and structural analysis of BTB domain proteins. *Genome Biol.* **6**, R82.
6 Perez-Torrado, R., Yamada, D., Defossez, P. A. 2006 Born to bind: the BTB protein-protein interaction domain. *Bioessays*. **28**, 1194-1202. (10.1002/bies.20500)
7 Mathew, R., Seiler, M. P., Scanlon, S. T., Mao, A. P., Constantinides, M. G., Bertozzi-Villa, C., Singer, J. D., Bendelac, A. 2012 BTB-ZF factors recruit the E3 ligase cullin 3 to regulate lymphoid effector programs. *Nature*. **491**, 618-621. (10.1038/nature11548)
8 Ahmad, K. F., Engel, C. K., Prive, G. G. 1998 Crystal structure of the BTB domain from PLZF. *Proc Natl Acad Sci U S A*. **95**, 12123-12128. (10.1073/pnas.95.21.12123)
9 Li, X., Peng, H., Schultz, D. C., Lopez-Guisa, J. M., Rauscher, F. J., 3rd, Marmorstein, R. 1999 Structure-function studies of the BTB/POZ transcriptional repression domain from the promyelocytic leukemia zinc finger oncoprotein. *Cancer Res*. **59**, 5275-5282.
10 Ahmad, K. F., Melnick, A., Lax, S., Bouchard, D., Liu, J., Kiang, C. L., Mayer, S., Takahashi, S., Licht, J. D., Prive, G. G. 2003 Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol Cell*. **12**, 1551-1564. (10.1016/s1097-2765(03)00454-4)
11 Siggs, O. M., Beutler, B. 2012 The BTB-ZF transcription factors. *Cell Cycle*. **11**, 3358-3369. (10.4161/cc.21277)
12 Siegmund, T., Lehmann, M. 2002 The Drosophila Pipsqueak protein defines a new family of helix-turn-helix DNA-binding proteins. *Dev Genes Evol*. **212**, 152-157. (10.1007/s00427-002-0219-2)
13 Horowitz, H., Berg, C. A. 1996 The Drosophila pipsqueak gene encodes a nuclear BTB-domain-containing protein required early in oogenesis. *Development*. **122**, 1859-1871.
14 Huang, D. H., Chang, Y. L., Yang, C. C., Pan, I. C., King, B. 2002 pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. *Mol Cell Biol*. **22**, 6261-6271. (10.1128/MCB.22.17.6261-6271.2002)
15 Beaster-Jones, L., Okkema, P. G. 2004 DNA binding and in vivo function of C.elegans PEB-1 require a conserved FLYWCH motif. *J Mol Biol*. **339**, 695-706. (10.1016/j.jmb.2004.04.030)
16 Buchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G., Dorn, R. 2000 Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila. *Genetics*. **155**, 141-157. (10.1093/genetics/155.1.141)
17 Melnikova, L., Kostyuchenko, M., Molodina, V., Parshikov, A., Georgiev, P., Golovnin, A. 2017 Multiple interactions are involved in a highly specific association of the Mod(mdg4)-67.2 isoform with the Su(Hw) sites in Drosophila. *Open Biol*. **7**, (170150 [pii] 10.1098/rsob.170150

rsob.170150 [pii])
18 Chaharbakhshi, E., Jemc, J. C. 2016 Broad-complex, tramtrack, and bric-a-brac (BTB) proteins: Critical regulators of development. *Genesis (New York, N.Y. : 2000)*. **54**, 505-518. (10.1002/dvg.22964)
19 Zacharchenko, T., Wright, S. 2021 Functionalization of the BCL6 BTB domain into a noncovalent crystallization chaperone. *IUCrJ*. **8**, 154-160. (10.1107/S2052252520015754)
20 Ghetu, A. F., Corcoran, C. M., Cerchietti, L., Bardwell, V. J., Melnick, A., Prive, G. G. 2008 Structure of a BCOR corepressor peptide in complex with the BCL6 BTB domain dimer. *Mol Cell*. **29**, 384-391. (10.1016/j.molcel.2007.12.026)
21 Vogelmann, J., Le Gall, A., Dejardin, S., Allemand, F., Gamot, A., Labesse, G., Cuvier, O., Negre, N., Cohen-Gonsaud, M., Margeat, E.*, et al.* 2014 Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the Drosophila genome. *PLoS Genet*. **10**, e1004544. (10.1371/journal.pgen.1004544)
22 Stead, M. A., Trinh, C. H., Garnett, J. A., Carr, S. B., Baron, A. J., Edwards, T. A., Wright, S. C. 2007 A beta-sheet interaction interface directs the tetramerisation of the Miz-1 POZ domain. *J. Mol. Biol.* **373**, 820-826.
23 Olivieri, D., Paramanathan, S., Bardet, A. F., Hess, D., Smallwood, S. A., Elling, U., Betschinger, J. 2021 The BTB-domain transcription factor ZBTB2 recruits chromatin remodelers and a histone chaperone during the exit from pluripotency. *J Biol Chem*. **297**, 100947. (10.1016/j.jbc.2021.100947)
24 Stead, M. A., Wright, S. C. 2014 Structures of heterodimeric POZ domains of Miz1/BCL6 and Miz1/NAC1. *Acta Crystallogr F Struct Biol Commun*. **70**, 1591-1596. (10.1107/S2053230X14023449)
25 Bertolini, M., Fenzl, K., Kats, I., Wruck, F., Tippmann, F., Schmitt, J., Auburger, J. J., Tans, S., Bukau, B., Kramer, G. 2021 Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science*. **371**, 57-64. (10.1126/science.abc7151)
26 Mena, E. L., Kjolby, R. A. S., Saxton, R. A., Werner, A., Lew, B. G., Boyle, J. M., Harland, R., Rape, M. 2018 Dimerization quality control ensures neuronal development and survival. *Science*. **362**, (10.1126/science.aap8236)
27 Mena, E. L., Jevtic, P., Greber, B. J., Gee, C. L., Lew, B. G., Akopian, D., Nogales, E., Kuriyan, J., Rape, M. 2020 Structural basis for dimerization quality control. *Nature*. **586**, 452-456. (10.1038/s41586-020-2636-7)
28 Maeng, O., Son, W., Chung, J., Lee, K. S., Lee, Y. H., Yoo, O. J., Cha, G. H., Paik, S. G. 2012 The BTB/POZ-ZF transcription factor dPLZF is involved in Ras/ERK signaling during Drosophila wing development. *Mol Cells*. **33**, 457-463. (10.1007/s10059-012-2179-3)
29 Oliver, D., Sheehan, B., South, H., Akbari, O., Pai, C. Y. 2010 The chromosomal association/dissociation of the chromatin insulator protein Cp190 of Drosophila melanogaster is mediated by the BTB/POZ domain and two acidic regions. *BMC Cell Biol*. **11**, 101. (10.1186/1471-2121-11-101)
30 Erez, N., Israitel, L., Bitman-Lotan, E., Wong, W. H., Raz, G., Cornelio-Parra, D. V., Danial, S., Flint Brodsly, N., Belova, E., Maksimenko, O.*, et al.* 2021 A Non-stop identity complex (NIC) supervises enterocyte identity and protects from premature aging. *Elife*. **10**, (10.7554/eLife.62312)
31 Bonchuk, A., Denisov, S., Georgiev, P., Maksimenko, O. 2011 Drosophila BTB/POZ domains of "ttk group" can form multimers and selectively interact with each other. *J Mol Biol*. **412**, 423-436. (10.1016/j.jmb.2011.07.052)
32 Plevock, K. M., Galletta, B. J., Slep, K. C., Rusan, N. M. 2015 Newly Characterized Region of CP190 Associates with Microtubules and Mediates Proper Spindle Morphology in Drosophila Stem Cells. *PLoS One*. **10**, e0144174. (10.1371/journal.pone.0144174)

33 Sabirov, M., Popovich, A., Boyko, K., Nikolaeva, A., Kyrchanova, O., Maksimenko, O., Popov, V., Georgiev, P., Bonchuk, A. 2021 Mechanisms of CP190 Interaction with Architectural Proteins in Drosophila Melanogaster. *Int J Mol Sci*. **22**, (10.3390/ijms222212400)
34 Sabirov, M., Kyrchanova, O., Pokholkova, G. V., Bonchuk, A., Klimenko, N., Belova, E., Zhimulev, I. F., Maksimenko, O., Georgiev, P. 2021 Mechanism and functional role of the interaction between CP190 and the architectural protein Pita in Drosophila melanogaster. *Epigenetics Chromatin*. **14**, 16. (10.1186/s13072-021-00391-x)
35 Ahanger, S. H., Shouche, Y. S., Mishra, R. K. 2013 Functional sub-division of the Drosophila genome via chromatin looping: the emerging importance of CP190. *Nucleus*. **4**, 115-122. (10.4161/nucl.23389)
36 Melnikova, L., Kostyuchenko, M., Molodina, V., Parshikov, A., Georgiev, P., Golovnin, A. 2018 Interactions between BTB domain of CP190 and two adjacent regions in Su(Hw) are required for the insulator complex formation. *Chromosoma*. **127**, 59-71. (10.1007/s00412-017-0645-6)
37 Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G. D., Becker, P. B., Renkawitz, R. 2009 Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J*. **28**, 877-888. (10.1038/emboj.2009.34)
38 Savitsky, M., Kim, M., Kravchuk, O., Schwartz, Y. B. 2016 Distinct Roles of Chromatin Insulator Proteins in Control of the Drosophila Bithorax Complex. *Genetics*. **202**, 601-617. (10.1534/genetics.115.179309)
39 Weyers, J. J., Milutinovich, A. B., Takeda, Y., Jemc, J. C., Van Doren, M. 2011 A genetic screen for mutations affecting gonad formation in Drosophila reveals a role for the slit/robo pathway. *Dev Biol*. **353**, 217-228. (10.1016/j.ydbio.2011.02.023)
40 Bradley, P. L., Andrew, D. J. 2001 ribbon encodes a novel BTB/POZ protein required for directed cell migration in Drosophila melanogaster. *Development*. **128**, 3001-3015.
41 Silva, D., Olsen, K. W., Bednarz, M. N., Droste, A., Lenkeit, C. P., Chaharbakhshi, E., Temple-Wood, E. R., Jemc, J. C. 2016 Regulation of Gonad Morphogenesis in Drosophila melanogaster by BTB Family Transcription Factors. *PLoS One*. **11**, e0167283. (10.1371/journal.pone.0167283)
42 Mukai, M., Hayashi, Y., Kitadate, Y., Shigenobu, S., Arita, K., Kobayashi, S. 2007 MAMO, a maternal BTB/POZ-Zn-finger protein enriched in germline progenitors is required for the production of functional eggs in Drosophila. *Mech Dev*. **124**, 570-583. (10.1016/j.mod.2007.05.001)
43 Ghosh, D., Gerasimova, T. I., Corces, V. G. 2001 Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *EMBO J*. **20**, 2518-2527. (10.1093/emboj/20.10.2518)
44 Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A. V., Moresco, J. J., Blokhina, T., Wolle, D., Aoki, T., Ryabykh, V., Yates, J. R., 3rd*, et al.* 2017 The GAGA factor regulatory network: Identification of GAGA factor associated proteins. *PLoS One*. **12**, e0173602. (10.1371/journal.pone.0173602)
45 Bellec, M., Dufourt, J., Hunt, G., Lenden-Hasse, H., Trullo, A., Zine El Aabidine, A., Lamarque, M., Gaskill, M. M., Faure-Gautron, H., Mannervik, M.*, et al.* 2022 The control of transcriptional memory by stable mitotic bookmarking. *Nat Commun*. **13**, 1176. (10.1038/s41467-022-28855-y)
46 Tang, X., Li, T., Liu, S., Wisniewski, J., Zheng, Q., Rong, Y., Lavis, L. D., Wu, C. 2022 Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. *Nat Struct Mol Biol*. **29**, 665-676. (10.1038/s41594-022-00800-z)
47 Stogios, P. J., Chen, L., Prive, G. G. 2007 Crystal structure of the BTB domain from the LRF/ZBTB7 transcriptional regulator. *Protein Sci*. **16**, 336-342. (10.1110/ps.062660907)

48 Stogios, P. J., Cuesta-Seijo, J. A., Chen, L., Pomroy, N. C., Prive, G. G. 2010 Insights into strand exchange in BTB domain dimers from the crystal structures of FAZF and Miz1. *J Mol Biol*. **400**, 983-997. (10.1016/j.jmb.2010.05.028)
49 Gao, J. L., Fan, Y. J., Wang, X. Y., Zhang, Y., Pu, J., Li, L., Shao, W., Zhan, S., Hao, J., Xu, Y. Z. 2015 A conserved intronic U1 snRNP-binding sequence promotes trans-splicing in Drosophila. *Genes Dev*. **29**, 760-771. (10.1101/gad.258863.115)
50 Tikhonov, M., Utkina, M., Maksimenko, O., Georgiev, P. 2018 Conserved sequences in the Drosophila mod(mdg4) intron promote poly(A)-independent transcription termination and trans-splicing. *Nucleic Acids Res*. **46**, 10608-10618. (10.1093/nar/gky716)
51 McManus, C. J., Duff, M. O., Eipper-Mains, J., Graveley, B. R. 2010 Global analysis of trans-splicing in Drosophila. *Proc Natl Acad Sci U S A*. **107**, 12975-12979. (10.1073/pnas.1007586107)
52 Ohsako, T., Horiuchi, T., Matsuo, T., Komaya, S., Aigaki, T. 2003 Drosophila lola encodes a family of BTB-transcription regulators with highly variable C-terminal domains containing zinc finger motifs. *Gene*. **311**, 59-69. (10.1016/s0378-1119(03)00554-7)
53 Zhang, W., Wang, Y., Long, J., Girton, J., Johansen, J., Johansen, K. M. 2003 A developmentally regulated splice variant from the complex lola locus encoding multiple different zinc finger domain proteins interacts with the chromosomal kinase JIL-1. *J Biol Chem*. **278**, 11696-11704. (10.1074/jbc.M213269200)
54 Horiuchi, T., Giniger, E., Aigaki, T. 2003 Alternative trans-splicing of constant and variable exons of a Drosophila axon guidance gene, lola. *Genes Dev*. **17**, 2496-2501. (10.1101/gad.1137303)
55 Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M. 2022 ColabFold: making protein folding accessible to all. *Nat Methods*. **19**, 679-682. (10.1038/s41592-022-01488-1)
56 Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J.*, et al.* 2022 Protein complex prediction with AlphaFold-Multimer. *bioRxiv* *https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2*. (10.1101/2021.10.04.463034)
57 Golovnin, A., Mazur, A., Kopantseva, M., Kurshakova, M., Gulak, P. V., Gilmore, B., Whitfield, W. G., Geyer, P., Pirrotta, V., Georgiev, P. 2007 Integrity of the Mod(mdg4)-67.2 BTB domain is critical to insulator function in Drosophila melanogaster. *Mol Cell Biol*. **27**, 963-974. (10.1128/MCB.00795-06)
58 Wilkins, R. C., Lis, J. T. 1997 Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. *Nucleic Acids Res*. **25**, 3963-3968. (10.1093/nar/25.20.3963)
59 Chetverina, D., Erokhin, M., Schedl, P. 2021 GAGA factor: a multifunctional pioneering chromatin protein. *Cell Mol Life Sci*. **78**, 4125-4141. (10.1007/s00018-021-03776-z)
60 Svergun, D. I., Barberato, C., Koch, M. H. J. 1995 CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates *J. Appl. Cryst.* **28**, 768-773.
61 Pauli, T., Vedder, L., Dowling, D., Petersen, M., Meusemann, K., Donath, A., Peters, R. S., Podsiadlowski, L., Mayer, C., Liu, S.*, et al.* 2016 Transcriptomic data from panarthropods shed new light on the evolution of insulator binding proteins in insects : Insect insulator proteins. *BMC Genomics*. **17**, 861. (10.1186/s12864-016-3205-1)
62 Orth, B., Sander, B., Moglich, A., Diederichs, K., Eilers, M., Lorenz, S. 2021 Identification of an atypical interaction site in the BTB domain of the MYC-interacting zinc-finger protein 1. *Structure*. **29**, 1230-1240 e1235. (10.1016/j.str.2021.06.005)
63 Gaskill, M. M., Gibson, T. J., Larson, E. D., Harrison, M. M. 2021 GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. *Elife*. **10**, (10.7554/eLife.66668)

64 Katsani, K. R., Hajibagheri, M. A., Verrijzer, C. P. 1999 Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J*. **18**, 698-708. (10.1093/emboj/18.3.698)
65 Espinas, M. L., Jimenez-Garcia, E., Vaquero, A., Canudas, S., Bernues, J., Azorin, F. 1999 The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. *J Biol Chem*. **274**, 16461-16469. (10.1074/jbc.274.23.16461)
66 Mishra, K., Chopra, V. S., Srinivasan, A., Mishra, R. K. 2003 Trl-GAGA directly interacts with lola like and both are part of the repressive complex of Polycomb group of genes. *Mech Dev*. **120**, 681-689. (10.1016/s0925-4773(03)00046-7)
67 Faucheux, M., Roignant, J. Y., Netter, S., Charollais, J., Antoniewski, C., Theodore, L. 2003 batman Interacts with polycomb and trithorax group genes and encodes a BTB/POZ protein that is included in a complex containing GAGA factor. *Mol Cell Biol*. **23**, 1181-1195. (10.1128/MCB.23.4.1181-1195.2003)
68 Schwendemann, A., Lehmann, M. 2002 Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. *Proc Natl Acad Sci U S A*. **99**, 12883-12888. (10.1073/pnas.202341499)
69 Gibert, J. M., Peronnet, F., Schlotterer, C. 2007 Phenotypic plasticity in Drosophila pigmentation caused by temperature sensitivity of a chromatin regulator network. *PLoS Genet*. **3**, e30. (10.1371/journal.pgen.0030030)
70 Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simao, F. A., Zdobnov, E. M. 2019 OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. **47**, D807-D811. (10.1093/nar/gky1053)
71 Durinck, S., Spellman, P. T., Birney, E., Huber, W. 2009 Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. **4**, 1184-1191. (10.1038/nprot.2009.97)
72 Edgar, R. C. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32**, 1792-1797. (10.1093/nar/gkh340)
73 Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., Noble, W. S. 2009 MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. **37**, W202-208. (10.1093/nar/gkp335)
74 Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. 2018 MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. **35**, 1547-1549. (10.1093/molbev/msy096)
75 Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., Finn, R. D. 2018 HMMER web server: 2018 update. *Nucleic Acids Res*. **46**, W200-W204. (10.1093/nar/gky448)
76 Chamberlain, S. A., Szocs, E. 2013 taxize: taxonomic search and retrieval in R. *F1000Res*. **2**, 191. (10.12688/f1000research.2-191.v2)
77 Yu, G. 2020 Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinformatics*. **69**, e96. (10.1002/cpbi.96)
78 Mastronarde, D. N. 2005 Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*. **152**, 36-51. (10.1016/j.jsb.2005.07.007)
79 Punjani, A., Rubinstein, J. L., Fleet, D. J., Brubaker, M. A. 2017 cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*. **14**, 290-296. (10.1038/nmeth.4169)
80 Bonchuk, A., Kamalyan, S., Mariasina, S., Boyko, K., Popov, V., Maksimenko, O., Georgiev, P. 2020 N-terminal domain of the architectural protein CTCF has similar structural organization and ability to self-association in bilaterian organisms. *Sci Rep*. **10**, 2677. (10.1038/s41598-020-59459-5)

81 Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M.*, et al.* 2017 ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr*. **50**, 1212-1225. (10.1107/S1600576717007786)