7250–7260 *Nucleic Acids Research, 2018, Vol. 46, No. 14* *Published online 26 June 2018*
doi: 10.1093/nar/gky582

Convergent evolution of integration site selection upstream of tRNA genes by yeast and amoeba retrotransposons

Eva Kling, Thomas Spaller, Jana Schiefner, Doreen Bönisch and Thomas Winckler*

Pharmaceutical Biology, Institute of Pharmacy, Friedrich Schiller University Jena, Germany

Received April 25, 2018; Revised June 12, 2018; Editorial Decision June 13, 2018; Accepted June 15, 2018

ABSTRACT

Transposable elements amplify in genomes as selfish DNA elements and challenge host fitness because their intrinsic integration steps during mobilization can compromise genome integrity. In gene-dense genomes, transposable elements are notably under selection to avoid insertional mutagenesis of host protein-coding genes. We describe an example of convergent evolution in the distantly related amoebozoan *Dictyostelium discoideum* and the yeast *Saccharomyces cerevisiae*, in which the *D. discoideum* retrotransposon DGLT-A and the yeast Ty3 element developed different mechanisms to facilitate position-specific integration at similar sites upstream of tRNA genes. Transcription of tRNA genes by RNA polymerase III requires the transcription factor complexes TFIIIB and TFIIIC. Whereas Ty3 recognizes tRNA genes mainly through interactions of its integrase with TFIIIB subunits, the DGLT-A-encoded ribonuclease H contacts TFIIIC subunit Tfc4 at an interface that covers tetratricopeptide repeats (TPRs) 7 and 8. A major function of this interface is to connect TFIIIC subcomplexes τA and τB and to facilitate TFIIIB assembly. During the initiation of tRNA gene transcription τB is displaced from τA, which transiently exposes the TPR 7/8 surface of Tfc4 on τA. We propose that the DGLT-A intasome uses this binding site to obtain access to genomic DNA for integration during tRNA gene transcription.

INTRODUCTION

Parasitism of genomes by mobile elements is a general phenomenon in nature. Transposable elements represent selfish DNA that evolved strategies to invade genomes and maintain functional source elements (1,2). With the exception of rare horizontal transfer, transposable elements are stuck in the genomes they colonize and must therefore develop mechanisms to avoid excessive damage of host fitness to prevent their own eradication. The yeast *Saccharomyces cerevisiae* and the amoebozoan *Dictyostelium discoideum* have in common that they have gene-dense and haploid genomes in which short intergenic regions offer limited space for non-hazardous insertion of mobile elements. In such compact genomes, mobile elements are under particular selection pressure to avoid insertional host gene mutagenesis and therefore evolved position-specific integration. Among distinct modes of position-specific integration (3), the most prominent examples in yeast and amoeba genomes are the integration into heterochromatin and the active targeting of regions near genes transcribed by RNA polymerase III (Pol III) (4,5).

In the *D. discoideum* genome, the Ty3/gypsy-type long-terminal repeat (LTR) retrotransposon Skipper-1 is exclusively found in the centromeric heterochromatin (6,7). This targeting is most likely mediated by binding of Skipper's chromo domain to histone H3 lysine-9 methylation marks that are enriched in the heterochromatin regions of *D. discoideum* centromers (7,8). Probably because *Saccharomyces cerevisiae* lacks histone H3 lysine-9 methylation, Skipper-like chromo domain-containing retrotransposons are lacking in the *S. cerevisiae* genome; yet targeting to heterochromatin is achieved by alternative mechanisms such as direct binding of Ty5 integrase to the structural component protein of heterochromatin, Sir4 (9).

Pol III-transcribed genes are hotspots for integration by retrotransposons in both the *S. cerevisiae* and *D. discoideum* genomes (10–12). Because interspersed Pol III-transcribed tRNA genes may repress the expression of nearby Pol II genes (13–15), considerable distances are maintained between the 5′ ends of Pol III genes and Pol II genes even in the compact yeast genome. It was hypothesized that upstream regions of tRNA genes present 'safe havens' for mobile elements to integrate without causing damage to adjacent Pol II genes (16). Furthermore, tRNA genes are distributed in multiple copies throughout the genome, which makes them ideal targets for integration because it allows mobile elements to spread in genomes with low risk of compromising the integrity of protein-coding genes if they stay in close

*To whom correspondence should be addressed. Tel: +49 3641 949841; Fax: +49 3641 949842; Email: t.winckler@uni-jena.de

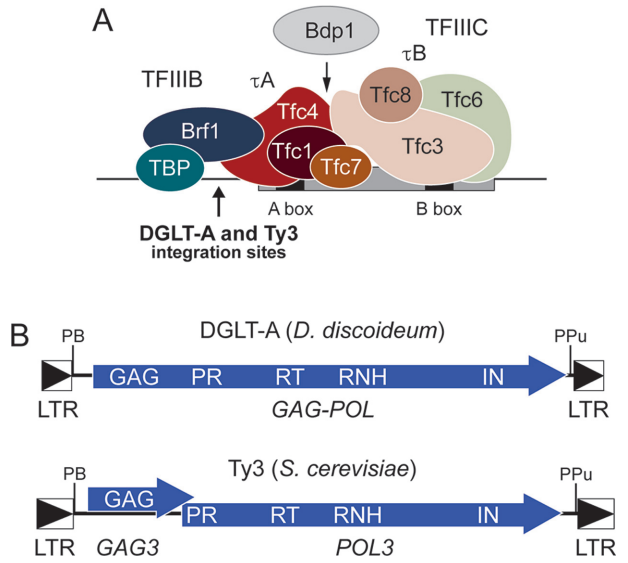

Figure 1. Topology of the RNA polymerase III transcription complex and the LTR retrotransposons DGLT-A and Ty3. (**A**) Model of the TFIIIC/TFIIIB complex based on experiments in *S. cerevisiae* (26,27). TFIIIC consists of the subunits Tfc1 (τ95), Tfc3 (τ138), Tfc4 (τ131), Tfc6 (τ91), Tfc7 (τ55), and Tfc8 (τ60). The factor is assembled in two subcomplexes, τA and τB, which are mainly connected by the interaction of Tfc3 with Tfc4. Brf1, a subunit of TFIIIB, is recruited by interacting with Tfc4 and subsequently incorporates TBP into the complex. Bdp1 interacts with the complex only transiently during transcription initiation by binding to the same interface on Tfc4 that is bound by Tfc3. Note that Tfc4 and Tfc1 of the τA subcomplex are the only subunits of TFIIIC currently identified in the *D. discoideum* genome based on sequence homology. The intragenic promoter elements (A box and B box) of a tRNA gene are indicated. The preferred integration sites of DGLT-A and Ty3 are located ∼15 bp upstream of the first nucleotide of a mature tRNA. (**B**) Schematic presentations of DGLT-A and Ty3. In DGLT-A, one single ORF (*GAG POL*) encodes the entire protein machinery, whereas in Ty3 the *GAG3* and *POL3* genes are organized in two overlapping reading frames and translated as Gag3 and Gag3-Pol3 fusion protein by a +1 frameshift. LTR: long terminal repeat; GAG: group-specific antigen; PR: protease; RT: reverse transcriptase; RNH: ribonuclease H; IN: integrase; PB, tRNA primer binding site; PPu, polypurine tract.

proximity to tRNA genes. In *S. cerevisiae*, the copia-like Ty1 and the gypsy-like Ty3 LTR retrotransposons use different mechanisms to locate integration sites within a window of ∼700 bp (Ty1) or ∼15 bp (Ty3) upstream of tRNA genes (17,18). In *D. discoideum*, the *Dictyostelium* gypsy-like transposable element A (DGLT-A) is found in similar positions upstream of tRNA genes as the Ty3 element in yeast (5). The non-LTR retrotransposons of the TRE5 and TRE3 families have independently invented position-specific integration ∼50 bp upstream and ∼100 bp downstream of tRNA genes, respectively (5,19). Thus, it seems that convergent evolution has led to similar solutions of the problem to amplify in compact genomes without causing excessive damage to the host.

The transcription of tRNA genes by Pol III is regulated by the transcription factor complexes TFIIIB and TFIIIC (20). TFIIIC recognizes bipartite internal promoters of tRNA genes known as A and B boxes (Figure 1A). The B box is the primary recognition element for TFIIIC (21,22), whereas the A box ensures the proper positioning of the factor relative to the transcription start site (23). DNA-bound TFIIIC recruits TFIIIB to the transcription start site, and subsequent binding of Pol III initiates transcription (24). TFIIIB is a three-subunit complex consisting of TATA-binding protein (TBP), TFIIB-related factor 1 (Brf1), and TFIIIB double prime (Bdp1), although the latter is incorporated only after assembly of the complete Brf1/TBP-TFIIIC complex (25) and may interact only transiently to help recruit Pol III (26). The yeast TFIIIC is composed of six subunits, which are arranged in two subcomplexes, τA and τB (27). Subcomplex τA contains the proteins Tfc4 (also known as τ131), Tfc1 (τ95) and Tfc7 (τ55), whereas τB consists of Tfc3 (τ138), Tfc6 (τ91) and Tfc8 (τ60) (Figure 1A).

Most of the affinity of TFIIIC to tRNA genes is mediated by the binding of Tfc3 to the B box (21), whereas assembly of TFIIIB on DNA-bound TFIIIC starts with the binding of Brf1 through complex interactions with Tfc4 (28–32). The recruitment of TBP into the Pol III transcription preinitiation complex results from the interaction of Brf1 with DNA-bound TFIIIC, and the incorporation of Bdp1 into the TBP/Brf1–TFIIIC–DNA complex is likely facilitated by cumulative conformational changes in the Tfc4–Brf1 complex that promote the interaction of Bdp1 with the central TPR arrays of Tfc4 (29,30,33,34). Based on cross-linking and structural analyses, τB is displaced from the TBP/Brf1-τA complex during the initiation of transcription by the competition of Bdp1 with Tfc3 for a binding site involving TPR 8 of Tfc4 in the τA complex (26).

In the genome of *D. discoideum*, 18 out of 20 genomic DGLT-A insertions are located within a window of 13–18 bp upstream of tRNA genes (7). This target site preference is strikingly similar to the Ty3 element of the yeast *S. cerevisiae* (18). DGLT-A and Ty3 have a comparable overall structure (Figure 1B), except that the coding sequence of the group-specific antigen (GAG) domain of DGLT-A is fused in-frame with the polyprotein gene that codes for protease (PR), reverse transcriptase (RT), ribonuclease H (RNH), and integrase (IN). The single open reading frame of DGLT-A codes for 1437 amino acids. The RT/RNH and IN core domains of DGLT-A display 35% and 26% amino acid sequence identity with the respective domains of the Ty3 POL3 protein (Supplementary Figure S1). The recognition of integration sites by Ty3 is mediated by direct interaction of the IN protein with the TBP/Brf1 subunits of TFIIIB (35,36). Ty3 IN also interacts with the subunit Tfc1 of TFIIIC, which affects the orientation of integration rather than target site recognition itself (37). Considering the similarity of DGLT-A and Ty3, we examined whether both elements use similar molecular mechanisms to identify integration sites, despite the long evolutionary distance between social amoebae and yeasts. Using a yeast two-hybrid system we showed that DGLT-A and Ty3 use different contacts to Pol III transcription factors to localize integration sites.

MATERIALS AND METHODS

Vector construction

A list of PCR primers used for vector construction is provided in Supplementary Table S1. The plasmids pGBKT7 and pACT2 were used in the yeast two-hybrid assays. DGLT-A DNA fragments for expression of BD- and

AD fusion proteins in yeast were amplified by PCR using genomic DNA of *D. discoideum* strain AX2 as a template. Isolation of the Tfc1- and Tfc4-coding regions was based on cDNA prepared from RNA of growing *D. discoideum* cells. For pull-down experiments, proteins were expressed in BL21(DE3) bacteria from pET33b-derived plasmids. To construct these plasmids, codon-adapted Tfc4, RNH and IN-NED DNAs were generated at Eurofins Genomics. First, a DNA fragment containing a hexahistidine (his_6) tag followed by a 3xFLAG tag, a HindIII restriction site and the DGLT-A RNH-coding region (amino acids 733–884) was constructed. A second synthetic gene contained a his_6 tag followed by the coding region of GFP, a HindIII site and a sequence coding for Tfc4(446–608). Both fusion genes were designed as NcoI/NotI fragments and cloned into pET33b. Plasmids for the expression of the other his_6-3xFLAG- or his_6-GFP-tagged proteins were constructed by replacing HindIII/NotI fragments in the respective pET33b plasmids (Supplementary Table S1).

Yeast two-hybrid assays

Plasmids were transformed into the *S. cerevisiae* strain AH109 as described (38), and the transformants were selected on SD minimal medium with dropout supplements (–trp/–leu). Twenty randomly selected clones from each transformation were streaked onto selection plates containing SD minimal medium and dropout supplements (–trp/–leu/–his/–ade) to test for protein-protein interactions. Protein interactions were scored as positive if at least 80% of the analyzed single clones grew on –trp/–leu/–his/–ade plates. One representative clone from each transformation is shown in the figures. For each protein pair analyzed, plasmid rescue from transformed yeast cells was performed. Identities of plasmid inserts were confirmed by restriction analysis and DNA sequencing.

Pull-down assays

His_6-3xFLAG- and his_6-GFP-tagged proteins were expressed in BL21(DE3). The bacteria were suspended in buffer A (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and the cells were lysed by sonification. Insoluble material was removed by centrifugation, and the extracts were applied to HiTrap® Chelating High Performance columns purchased from Sigma-Aldrich. Metal chelate chromatography was performed in buffer A at room temperature. After washing with buffer A containing 10 mM imidazole, the his_6-tagged proteins were eluted with buffer A containing 500 mM imidazole. Recombinant proteins were dialyzed against buffer B (20 mM Tris-HCl pH 7.5) and stored in aliquots at −20°C. For pull-down experiments, 20 μl of anti-FLAG® M2 magnetic beads (Sigma-Aldrich) were blocked overnight in 200 μl of binding buffer (buffer B including 4% BSA or 1% gelatin) at room temperature. The beads were then washed twice with binding buffer and incubated with FLAG-tagged proteins in 50 μl of binding buffer for 1.5 h at 4°C. After removal of unbound FLAG-tagged protein, 50 μl of the GFP-tagged protein partners were added, and the mixtures were further incubated for 2 h at 4°C. The beads were subsequently washed twice with washing buffer A (20 mM Tris–HCl, 100 mM NaCl pH 7.5). An additional washing step was performed in washing buffer B (20 mM Tris–HCl, 200 mM NaCl, 0.05% Tween 80). Elution was performed by boiling the beads in loading buffer for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Proteins were separated on 12.5% acrylamide gels and blotted onto nitrocellulose membranes (Amersham™ Protan™ 0,45 μM NC, GE Healthcare Life Science). FLAG-tagged proteins were detected using an anti-FLAG tag polyclonal antibody conjugated to horseradish peroxidase (Abcam ab2493). GFP fusion proteins were stained using the GFP (D5.1) XP® Rabbit mAb purchased from Cell Signaling Technology. A secondary anti-rabbit antibody conjugated to horseradish peroxidase was purchased from Cell Signaling Technology.

Alignments and modelling of RT/RNH and Tfc4 TPR structures

Conserved core domains of the Ty3 POL3 and DGLT-A ORF protein were determined by searching the Conserved Domain Database at NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Alignments of these domains were conducted using CLUSTAL X (39). The RT/RNH sequence of Ty3, as defined by the alignment shown in Supplementary Figure S1, was then used as template to build a structure model of the corresponding region of DGLT-A RT/RNH region using SWISS-MODEL (https://swissmodel.expasy.org) based on the solved crystal structure of the Ty3 RT/RNH region (PDB entry 4OL8) (40). Sequence identity in the region used to build the DGLT-A RT/RNH model was 35.5%. The models are visualized using MacPyMol version 1.8.

To build a structure model of the central TPR arrays of *D. discoideum*, the sequences of *D. discoideum* and yeast Tfc4 were compared using the Conserved Domain Database and subsequent alignment of the predicted TRP regions using CLUSTAL X. Next, a model of the *D. discoideum* Tfc4 TPR region was generated using SWISS-MODEL and the solved crystal structure of yeast Tfc4 (PDB 5AIO) (26) as template. *D. discoideum* Tfc4 was 20.7% identical to yeast Tfc4 (PDB 5AIO) in this region. The pictures were drawn using MacPyMol version 1.8.

RESULTS

DGTL-A RNH and IN-NTD proteins bind to TFIIIC subunit Tfc4

We used yeast two-hybrid assays to identify interactions between domains of the DGLT-A polyprotein and either TFIIIB or TFIIIC subunits of the Pol III transcription complex. We first cloned the individual domains of the DGLT-A polyprotein as indicated in Figure 2A. The group-specific antigen-like domain (GAG) was covered by amino acids 2–236 of the DGLT-A polyprotein, the protease (PR) was covered by amino acids 231–425, the reverse transcriptase (RT) was covered by amino acids 419–742, and the ribonuclease H (RNH) was covered by amino acids 733–879. The integrase protein (IN) was separated in three parts based on an alignment with Ty3 IN (Supplementary Figure S1):

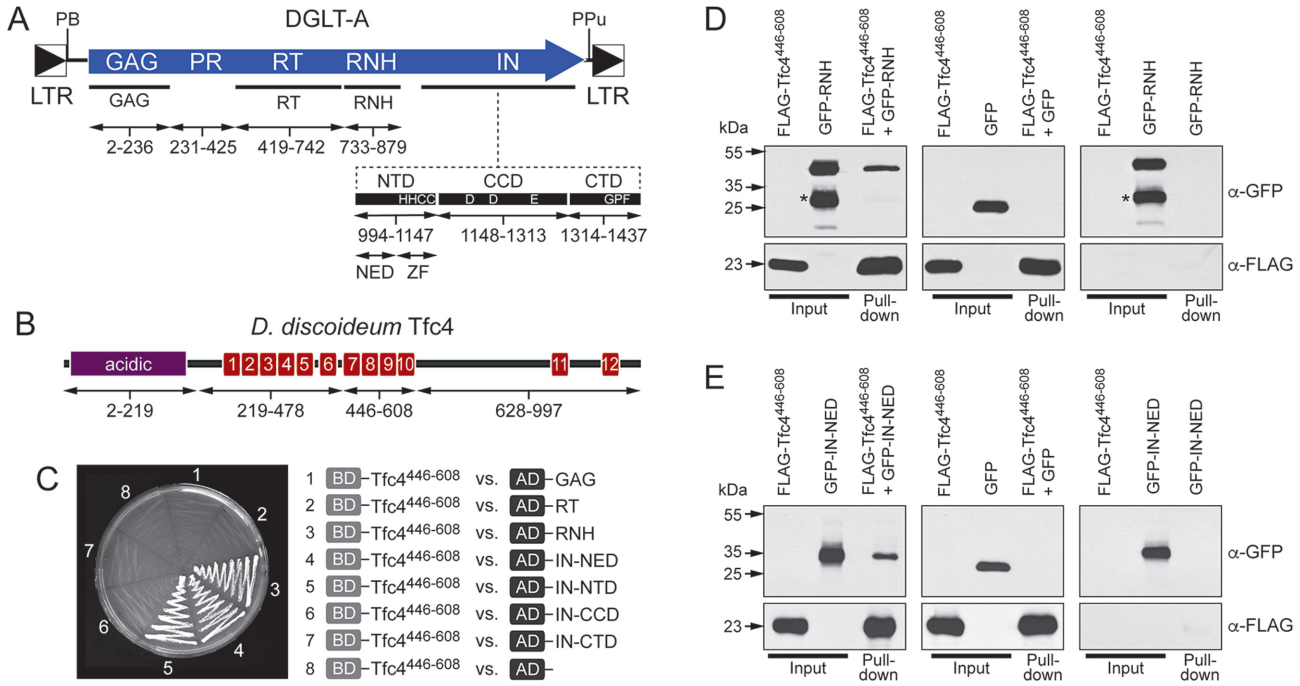

Figure 2. Interaction of DGLT-A proteins and Tfc4. (**A**) Amino acid positions of the DGLT-A GAG-POL polyprotein are indicated as inserted into yeast two-hybrid vectors. The DGLT-A integrase (IN) is characterized by a catalytic core domain (CCD; DGLT-A$^{1148–1313}$) containing a DDE catalytic triad, an N-terminal domain (NTD; DGLT-A$^{994–1147}$) containing a typical HHCC zinc finger-like motif, and a C-terminal extension containing a conserved GPY/F motif. For yeast two-hybrid experiments the IN-NTD was further split into the N-terminal extension domain (NED; DGLT-A$^{994–1067}$) and the zinc finger (ZF) region (DGLT-A$^{1068–1147}$). See Supplementary Figure S1B for an alignment of DGLT-A and Ty3 IN sequences. (**B**) Schematic presentation of *D. discoideum* Tfc4. Positions of the TPRs are indicated by red boxes with numbers. The amino acid positions Tfc4 fragments used in yeast two-hybrid screenings are listed. (**C**) Results of yeast two-hybrid experiments testing BD-fused TPR 7–10 of Tfc4 (Tfc4$^{446–608}$) against AD-fused DGLT-A proteins as indicated. (**D**) Pull-down experiment testing the binding of FLAG-tagged Tfc4 TPR 7–10 (FLAG-Tfc4$^{446–608}$) to GFP-tagged DGLT-A RNH (GFP-RNH). Both proteins were expressed in bacteria and used as purified proteins. Input refers to the purified proteins and is used as markers. Pull-down left panel: FLAG- Tfc4$^{446–608}$ was bound to anti-FLAG antibodies immobilized on magnetic beads. The loaded beads were incubated with GFP- RNH (DGLT-A$^{733–879}$). Pull-down middle panel: FLAG-Tfc4$^{446–608}$ was immobilized on anti-FLAG beads and incubated with GFP (first negative control). Pull-down right panel: Beads carrying anti-FLAG antibodies were incubated with GFP-RNH without prior loading of the FLAG-tagged bait protein (second negative control). After western blotting, the proteins were first stained with anti-GFP antibodies and then with anti-FLAG antibodies. The asterisks indicate prominent degradation products of GFP-RNH. (E) Similar pull-down experiment using FLAG- Tfc4$^{446–608}$ as bait and GFP-tagged IN-NED (DGLT-A$^{994–1067}$) as the binding partner.

the amino-terminal domain (IN-NTD; 994–1147), the catalytic core domain (IN-CCD; 1148–1313), and the carboxy-terminal domain (IN-CTD; 1314–1437). All DGLT-A proteins were fused to either the GAL4 DNA binding domain (BD) or the GAL4 transactivating domain (AD).

Because DGLT-A is similar to Ty3 in both its overall structure and amino acid sequences of catalytic core domains (Supplementary Figure S1) and Ty3 uses the interaction of IN with TFIIIB subunits to determine integration sites, we first tested DGLT-A proteins for interactions with *D. discoideum* TFIIIB subunits TBP, Bdp1 and Brf1 (the latter was divided into the TFIIB-like N-terminal domain and the Pol III-specific C-terminal domain). In this screening, we did not detect any interaction of DGLT-A proteins with TFIIIB subunits, irrespective of whether the tested proteins were fused to BD or AD (data not shown). Next, we wanted to determine whether DGLT-A interacts with TFIIIC subunits located close to the integration site. However, only the orthologs of the most conserved subunits of the TFIIIC subcomplex τA, Tfc1 and Tfc4, could be identified in the *D. discoideum* reference genome by homology searches using either the yeast or human TFIIIC subunits as queries.

Dictyostelium discoideum Tfc1 (dictyBase gene ID DDB_G0289935) is an 865 amino acid protein whose only significant sequence similarity with orthologous proteins is the amino-terminal Tau95 domain. To perform yeast-two-hybrid screenings against DGLT-A proteins, we divided the Tfc1 protein into an N-terminal portion covering amino acids 22–401 and a C-terminal portion spanning amino acids 420–865. Tfc1 fused to BD and AD was tested against all DGLT-A-derived polyprotein domains, but we did not detect any interactions with DGLT-A proteins (data not shown).

Dictyostelium discoideum Tfc4 (dictyBase ID DDB_G0278321) consists of 997 amino acids and has an acidic N-terminal domain and two arrays of TPR motifs, which are the determining features in orthologous Tfc4 proteins (26,29) (Figure 2B). To examine whether Tfc4 interacts with DGLT-A proteins in the yeast two-hybrid system, we divided the protein into four portions as presented in Figure 2B. In initial experiments, we observed a robust interaction of the RNH domain of DGLT-A (DGLT-A$^{733–879}$) fused to AD with a BD fusion of TFC4$^{446–608}$, which contains TPRs 7–10 (Figure 2C).

Interestingly, the IN-NTD (DGLT-A$^{994–1147}$) fused to AD also interacted with BD-fused TFC4$^{446–608}$ (Figure 2C). No binding of RNH or IN-NTD to other parts of the Tfc4 protein was observed (data not shown). Similarly, no binding of other DGLT-A proteins, including the IN-CCD and IN-CTD, to Tfc4 was detected. Notably, the interactions of DGLT-A RNH or IN-NTD with Tfc4 were not observed when the Tfc4 part was fused to AD instead of BD. However, interactions are not always observed in the two-hybrid assay in each of the two potential directions, likely because the generation of artificial fusion proteins leads to steric retraction of interacting surfaces or forces instability of the fusion proteins.

To confirm the direct binding of TFC4$^{446–608}$ to DGLT-A RNH, we performed pull-down experiments with purified, bacterially expressed proteins. GFP-tagged RNH (DGLT-A$^{733–879}$) was mixed with FLAG-tagged Tfc4$^{446–608}$, and the complexes were precipitated with magnetic beads carrying anti-FLAG antibodies. As shown in Figure 2D, GFP-tagged DGLT-A RNH was precipitated with FLAG-tagged TFC4$^{446–608}$ but did not bind to immobilized GFP or to unloaded beads used as negative controls.

We next attempted to confirm the binding of DGLT-A IN-NTD to TFC4$^{446–608}$ in pull-down experiments using bacterially expressed proteins, but the GFP-tagged IN-NTD protein could not be expressed in a soluble form. In most retroviruses, the IN-NTD consists of an ∼50 amino acid sequence containing a zinc finger-like HHCC motif (41). This motif is also present in the Ty3 and DGLT-A IN proteins (Supplementary Figure S1). In Ty3 the IN-NTD is further extended at the N-terminal end by ∼70 amino acids; this sequence is referred to as the N-terminal extension domain (IN-NED). This sequence contributes to the position-specific integration of Ty3 by interacting with TFIIIC subunit Tfc1 (37). A similar IN-NED can be postulated in the DGLT-A IN-NTD, although the similarity to Ty3 in this region is limited (Supplementary Figure S1). Notably, the interaction of TFC4$^{446–608}$ with IN-NTD of DGLT-A was retained in the yeast two-hybrid system when the IN-NTD was reduced to the N-terminal IN-NED sequence (Figure 2C). Fortunately, a bacterially expressed GFP-IN-NED fusion protein was soluble and could be used in pull-down experiments with FLAG-tagged Tfc4$^{446–608}$ as bait. As shown in Figure 2E, a robust interaction between the partners was observed, suggesting that IN-NED provides a platform to interact with Tfc4 in *D. discoideum* DGLT-A, whereas the analogous platform in the yeast Ty3 element mediates binding to Tfc1.

DGLT-A RNH and IN-NTD have overlapping binding sites on Tfc4

Tfc4 proteins are characterized by two central TPR arrays that provide platforms for interactions with Tfc3 to facilitate τA/τB complex formation (26) and for the recruitment of TFIIIB subunits Brf1 and Bdp1 (28–30,34,42). In silico secondary structure analysis of yeast Tfc4 at first predicted two central TPR arrays covering TPRs 1–5 and 6–9, respectively (43). However, a crystal structure of the central TPR arrays of yeast Tfc4 was recently published and unexpectedly determined ten instead of nine TPRs. Indeed, a previously proposed linker between the two TPR arrays is actually an extended helix as part of TPR 5 (26) (Figure 3A). In addition, yeast Tfc4 contains a previously undetected 'ring domain' between TPR 6 and TPR 7 (26) (Figure 3A). The Tfc4 proteins of *S. cerevisiae* and *D. discoideum* are 21% identical in the TPR 1–10 region, which allowed us to generate a structure-based model of the central TPR arrays of *D. discoideum* Tfc4 based on the crystal structure of the *S. cerevisiae* Tfc4 protein. The model revealed the presence of ten TPRs as in the yeast protein, but *D. discoideum* Tfc4 apparently lacks the helix extension of TPR 5 and the ring domain (Figure 3B, Supplementary Figure S2).

Based on the modeled TPR arrays of *D. discoideum* Tfc4, binding sites of DGLT-A RNH and IN-NTD on Tfc4 are located on the right TPR arm comprising TPRs 7–10. We performed a TPR deletion analysis and mapped the binding of the DGLT-A RNH domain to TPRs 7 and 8 (Figure 3C). The RNH bound to TPR 7 and TPR 8 separately, suggesting that both TPRs provide similar platforms for this interaction albeit limited sequence conservation. The binding of IN-NTD to Tfc4 also mapped to the TPR 7/8 region, but this interaction required the entire TPR 7/8 fragment and binding was not observed on either TPR 7 or TPR 8 (Figure 3D). This finding suggested that the binding site of IN-NTD on Tfc4 may be located at the interface of TPRs 7/8, but in the proximity of the binding site of RNH on TPR 7 and TPR 8.

DGLT-A RNH supports multiple protein interactions

RNH is an integral part of the RT domain of LTR retrotransposons and retroviruses. The structure of the RT-RNH domain of DGLT-A can be modeled based on the recently solved structure of the Ty3 RT/RNH domain (40). The structures of DGLT-A and Ty3 RT/RNH domains are strikingly similar (Supplementary Figure S3). The DGLT-A RNH contains the central five-stranded β-sheet layer of retroviral RNH proteins and two extended α-helices (Figure 4A). The compact nature of the RNH domain hampered further mapping of surfaces that interact with Tfc4. In the initial two-hybrid screenings, we used a fragment of the DGLT-A polyprotein spanning amino acids 733–879. Comparison of this sequence with the model of the RT-RNH domain of DGLT-A enabled the refinement of the RNH core to amino acids 759–879, whereas amino acids 733–758 actually represented helix H of the thumb region of the RT domain (40). We cloned fragment 733–758 separately and further divided the RNH core region into fragments covering amino acids 759–800, 801–846 and 846–879. We tested the binding of these AD fusion proteins to Tfc4$^{446–608}$ (TPR 7–10) in the yeast two-hybrid assay (Figure 4B). The RT helix (733–758) was not involved in binding to Tfc4$^{446–608}$, but we detected a bipartite binding region located at the N-terminal and C-terminal parts of the RNH core (Figure 4B). These binding platforms seemed to cover parts of DGLT-A$^{759–846}$ and DGLT-A$^{801–879}$ of RNH, revealing an extended surface covering most of RNH. As shown in Figure 4C and D, the isolated TPR 7 and TPR 8 of Tfc4 bound to fragments of RNH comparable to the entire TPR 7–10 array, which underscores the apparent independent binding of RNH to either TPR 7 or TPR 8.

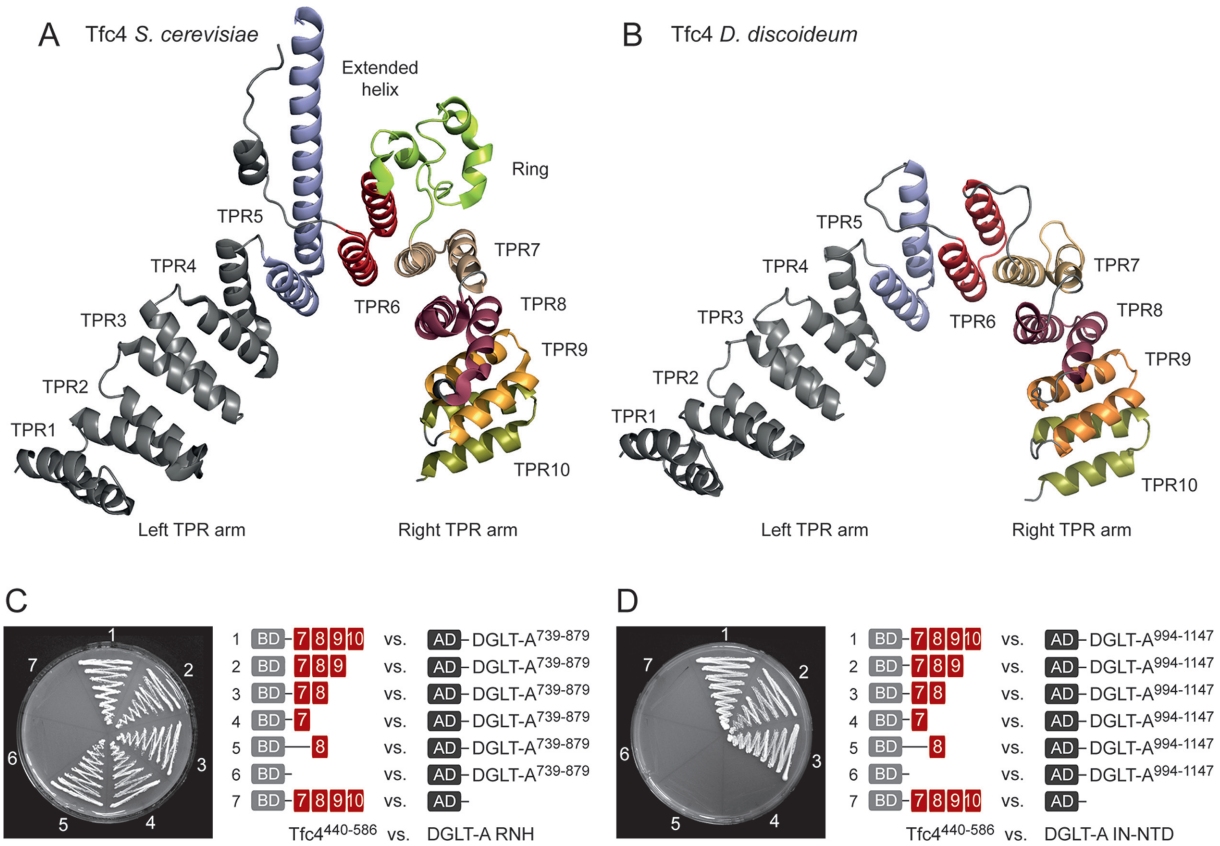

Figure 3. Interaction of DGLT-A RNH and IN-NTD with TFIIIC subunit Tfc4. (**A**) Solved structure of the central TPR arrays (TPRs 1–10) of yeast τTfc4. The picture was generated with PyMol based on PDB entry 5AIO (26). (**B**) Model of the central TPR region of *D. discoideum* Tfc4. The model was generated using the solved structure of the yeast Tfc4 TPRs 1–10 (PDB 5AIO) (26) as template. (**C**, **D**) Yeast two-hybrid assay testing different deletions of the right TPR arm (TPRs 7–10) against DGLT-A RNH (C) and DGLT-A IN-NTD (D).

In addition to binding to Tfc4, the DGLT-A RNH provides a platform for protein interactions within other parts of the DGLT-A polyprotein. In yeast two-hybrid assays, we observed a robust binding of RNH to IN-NED (Figure 5A). This interaction could be confirmed by *in vitro* pull-down experiments using bacterially expressed proteins (Figure 5B). Furthermore, DGLT-A RNH may form dimers, as suggested by the robust RNH–RNH interaction in yeast two-hybrid assays (Figure 5C) and *in vitro* pull-down experiments (Figure 5D). IN-NED also provided signals in two-hybrid experiments when tested against itself, indicating that IN-NED may also provide a platform for IN homodimerization (data not shown).

Taken together, the observations described above suggested that the compact DGLT-A RNH domain supports interactions with itself, part of the integrase (IN-NED) and Tfc4. Whereas binding of Tfc4 and IN-NED required a similar bipartite landscape on RNH (compare Figures 4B and 5A), homodimerization of RNH only required contact at the interface of RNH fragments DGLT-A$^{759–800}$ and DGLT-A$^{801–846}$ (compare Figure 5A and C). We considered that RNH dimerization may occur on the central five stranded β-sheet of RNH, whereas interaction of RNH with Tfc4 or IN-NED may involve the helical and loop region of RNH. To evaluate the feasibility of this hypothesis, we designed an RNH mutant in which three charged amino acids at the surface of the β-sheets were replaced by alanine (E771A, K831A, E864A; Figure 6A). In yeast two-hybrid assays, we tested the triple mutant of RNH for binding to wild-type RNH, IN-NED and Tfc4. Notably, the binding of wild-type RNH to the RNH triple mutant was completely abolished (Figure 6B). This result was not due to the poor expression of the RNH mutant in yeast cells, as the interaction of the RNH triple mutant with both IN-NED and Tfc4 was retained (Figure 6B). Thus, the interaction of RNH with IN-NED and Tfc4 requires a surface that is different from the RNH homodimerization platform.

DISCUSSION

Ty3 and DGLT-A use different mechanisms to interact with the Pol III transcription complex

There are numerous reports that non-randomly integrating retroviruses and retrotransposons determine integration sites by using interactions of their IN proteins with local chromatin-associated host factors. Examples include the accumulation of HIV-1 integrations in protein-coding regions mediated by transcriptional coactivator lens epithelium-derived growth factor (LEDGF) (44,45), tethering of retrotransposon Ty5 to heterochromatin by Sir4 protein in *S. cerevisiae* (9), targeting of *Schizosaccharomyces pombe* Tf1 to Pol II promoters by interaction of Tf1 IN with transcrip-

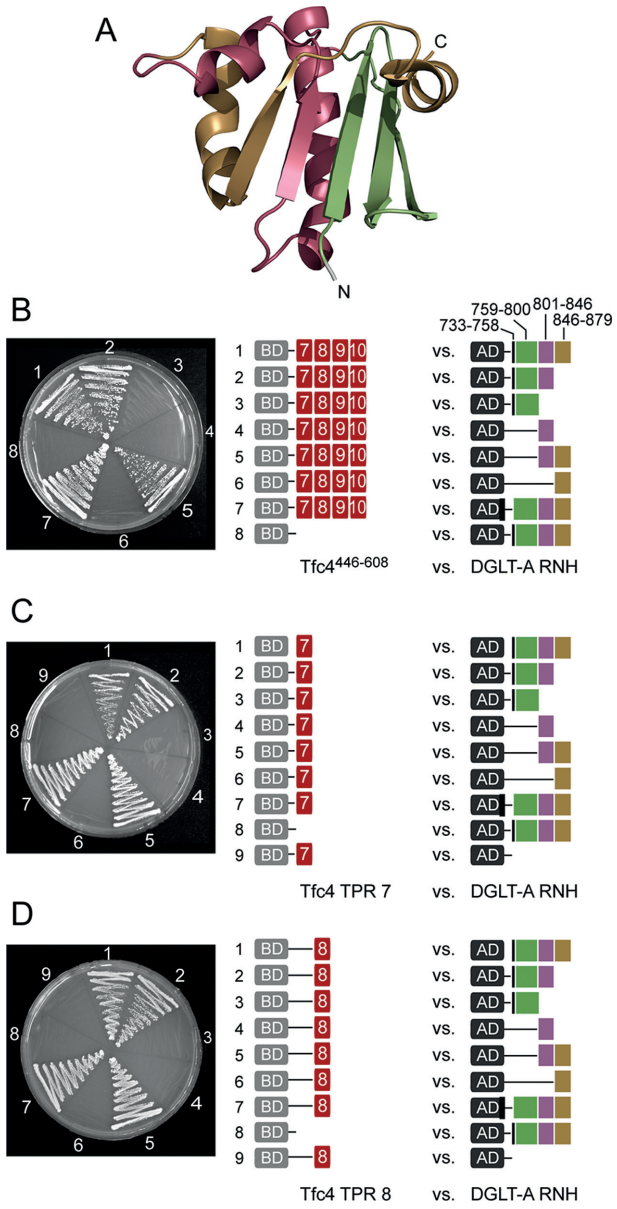

Figure 4. Mapping of the DGLT-A RNH surface that binds to Tfc4 TPRs 7–10. (**A**) The model of the DGLT-A RNH domain was generated using the solved structure of RT-RNH of Ty3 (PDB entry 4OL8) (40) as template. The color code represents RNH fragments cloned for separate testing in the yeast two-hybrid system. (**B**) Yeast two-hybrid experiment testing for the interaction of the right TPR arrays of *D. discoideum* Tfc4 (TPRs7–10) with different parts of DGLT-A RNH. Note that positions 733–758 are localized outside of the RNH core domain and belong to the RT thumb region. (**C**, **D**) Similar yeast two-hybrid experiments using only TPR 7 and TPR 8 as baits.

tional regulator Atf1 (46) and probably also Sap1 (47), or integration site selection by yeast Ty1 upstream of tRNA genes by interaction with Pol III subunits (48,49). Recognition of TFIIIB by Ty3 requires contacts of the IN protein with the TBP/Brf1 subunits (35,36), but the additional recognition of TFIIIC subunit Tfc1 by Ty3 IN involves only the IN-NED (37).

Considering that DGLT-A and Ty3 show a similar integration preference ∼15 bp upstream of a tRNA coding sequence, it was surprising that, within the limitations of negative results obtained in yeast two-hybrid assays, we did not detect any protein interactions between DGLT-A and TFIIIB subunits or TFIIIC subunit Tfc1. Instead, we determined that the IN-NED of DGLT-A interacts with Tfc4 instead of Tfc1. Further we have, for the first time to our knowledge, identified an RNH domain of a retrotransposon as an interaction partner of a host protein – likely contributing to position-specific integration. For the moment, we were unable to evaluate whether the interaction of DGLT-A proteins with Tfc4 is in fact required and sufficient for the selection of integration sites upstream of tRNA genes because we lack a retrotransposition-competent DGLT-A element to design retrotransposition assays in *D. discoideum* cells.

The DGLT-A preintegration complex may compete with τA/τB interaction during the initiation of Pol III transcription

We showed that the TFIIIC subunit Tfc4 is the primary binding partner of two different DGLT-A proteins, RNH and IN-NTD. These proteins bind to overlapping sites within the TPR 7/8 region of Tfc4, which is a binding hotspot for protein interactions that facilitates the assembly of TFIIIC subcomplexes τA and τB by the interaction of Tfc4 with Tfc3 and as well as recruitment of TFIIIB subunits Brf1 and Bdp1 to the DNA-bound TFIIIC complex (26,28–30,34,42). Yeast Brf1 binds independently to both isolated TPR arms of Tfc4, with higher affinity for the isolated TPR 7–10 region than for the isolated TPR 1–5 region (28,42). However, the affinity of Brf1 for TPR 1–5 is considerably higher in the presence of amino acid sequences in the N-terminus of TPR 1–5 (42), suggesting that the incorporation of Brf1 into the TFIIIC-DNA complex is accomplished by the initial interaction of Brf1 with the right TPR arm, followed by a conformational change in the complex that shifts Brf1 to its high-affinity binding site on the left TPR arm (42). This makes the TPR 7–10 region available for subsequent interactions with Bdp1 (30,34,50). Mutations of either D468 or L469 in TPR 8 of yeast Tfc4 resulted in strongly reduced binding to both Brf1 and Bdp1 (26,30,34) and completely abolished the interaction between Tfc4 and Tfc3 (i.e., τA/τB subcomplex formation) (26). Intriguingly, the TPR 7/8 region is the primary binding site for DGLT-A proteins to *D. discoideum* Tfc4 and the amino acids D468/L469 positions in yeast Tfc4 (TPR 8) are conserved in the *D. discoideum* Tfc4 protein (D486 and L487; Supplementary Figure S2). Additionally, a similar motif is present at the N-terminus of helix A of TPR 7 (D447/L448), which may explain why DGLT-A RNH is able to bind TPR7 and TPR8 separately. Male et al. (26) suggested that L469 of yeast Tfc4 TPR 8 is oriented toward the neighboring helix of TPR 7, whereas D468 may cover the inner groove formed by TPR 8 and may contribute to a protein-interaction platform. Interestingly, modeling of *D. discoideum* Tfc4 TPR arrays predicted that L448 in TPR 7 and L487 in TPR 8 are both oriented toward helices of neighboring TPRs and residues D447 of TPR 7 and D486 of TPR 8 may indeed be oriented toward the predicted inner groove (data not shown). This implies that both TPR 7 and

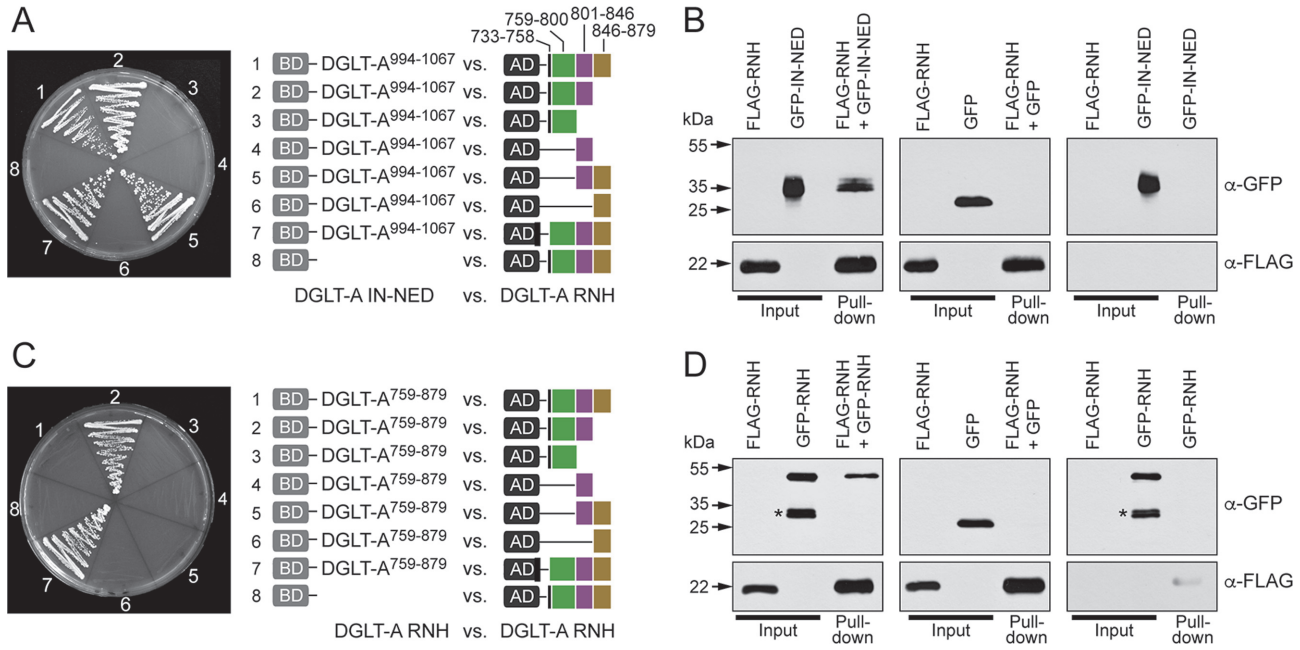

Figure 5. Protein interactions within the DGLT-A element. (**A**) Yeast two-hybrid experiment localizing the surface on DGLT-A RNH required for binding to DGLT-A IN-NED. A comparison of IN-NTD with the corresponding part of yeast Ty3 IN-NTD is provided in Supplementary Figure S1. (**B**) Pull-down experiment to confirm the binding of DGLT-A IN-NED to DGLT-A RNH. Proteins were expressed in bacteria and used as purified proteins. Input refers to the purified proteins and is used as markers. Pull-down left panel: FLAG-tagged DGLT-A RNH (FLAG-DGLT-$A^{733–879}$) was immobilized on anti-FLAG antibodies bound to magnetic beads. The loaded beads were incubated with GFP-tagged DGLT-A IN-NED (GFP-IN-NED containing DGLT-$A^{994–1067}$). Pull-down middle panel: FLAG- RNH (DGLT-$A^{733–879}$) was bound to anti-FLAG beads and incubated with GFP (first negative control). Pull-down right panel: Beads carrying anti-FLAG antibodies were incubated with GFP-IN-NED (DGLT-$A^{994–1067}$) without prior loading of the FLAG-tagged bait protein (second negative control). The blotted proteins were first stained with anti-GFP antibodies and then with anti-FLAG antibodies. (**C**) Yeast two-hybrid experiment investigating dimerization of DGLT-A RNH. (**D**) Pull-down experiment testing GFP-tagged RNH against FLAG-tagged RNH (DGLT-$A^{733–879}$). The asterisks indicate the prominent degradation products of GFP-RNH.

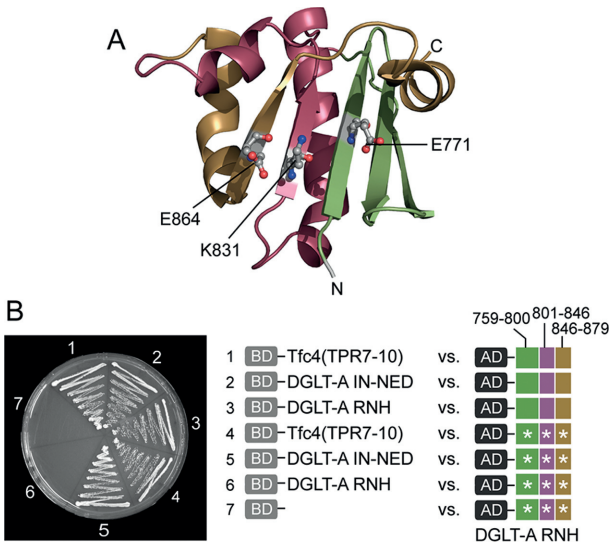

Figure 6. Comparison of interaction platforms on DGLT-A RNH. (**A**) The model of the RNH domain of DGLT-A is based on the solved structure of RT-RNH of Ty3 (PDB entry 4OL8). Mutations E771A, K831A, E864A at the surface of the β-sheets are presented as sticks. (**B**) The wild-type and triple mutant of DGLT-A RNH were tested for binding to Tfc4 (TPRs 7–10; Tfc4$^{440–586}$), DGLT-A IN-NED (DGLT-$A^{994–1067}$) and DGLT-A RNH (DGLT-$A^{759–879}$). White asterisks indicate the triple mutant of RNH.

TPR 8 may in fact contribute to protein interactions during assembly of the Pol III transcription complex.

Considering these observations, we propose a model suggesting that convergent evolution resulted in the development of different molecular mechanisms that Ty3 and DGLT-A employ to solve the same problem: to find safe integration sites in compact genomes by targeting to a region closely upstream of tRNA genes. Both Ty3 and DGLT-A target tRNA gene during the initiation of transcription after DNA-bound TFIIIC has recruited the TBP/Brf1 heterodimer to the 5′ end of the tRNA gene. Ty3 then uses direct binding of Ty3 IN with the TBP/Brf1 heterodimer to recognize tRNA genes integration sites (35,36), whereas interaction of Ty3 IN with TFIIIC subunit Tfc1 only affects the orientation of the integrated Ty3 copy relative to the targeted tRNA gene (37) (Figure 7A). In contrast, DGLT-A neither interacts with TFIIIB or Tfc1, but targets the Tfc4 subunit of TFIIIC at a time when the assembly of TFIIIB subunits TBP and Brf1 into the transcription complex is accomplished and incoming Bdp1 transiently displaces τB from the transcription complex by binding to the TPR 8 region of Tfc4. We speculate that the DGLT-A intasome competes with Bdp1 and τB for the binding site covering TPR 8 on Tfc4 and uses the interaction with τA to obtain access to genomic DNA for integration during tRNA gene transcription (Figure 7B).

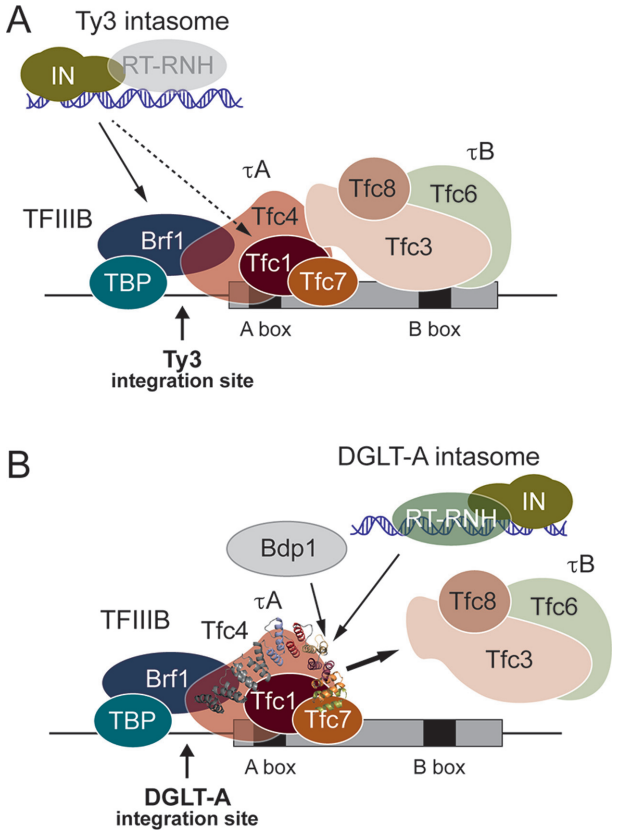

Figure 7. Model of integration site selection by Ty3 and DGLT-A. The τB complex contains the B box binding activity (Tfc3) and mediates the binding of TFIIIC (τA/τB complex) to tRNA genes. The assembly of TFIIIB is mediated by DNA-bound τA by interactions between TFC4 and TFIIIB subunit Brf1. Subsequently, TBP is incorporated primarily by interaction with Brf1 (25). (**A**) Model of Ty3 integration site selection. The Ty3 intasome (preintegration complex) consists at least of Ty3 IN and Ty3 cDNA. Note that it is not known whether RT/RNH is part of the Ty3 intasome. Ty3 selects tRNA genes by direct interaction of Ty3 IN with the TBP/Brf1 heterodimer (35,36). Ty3 IN also interacts with subunit Tfc1, which is not required for integration site selection but affects the orientation of integrated Ty3 copies relative to the target (37). (**B**) Model of target site selection by DGLT-A. After assembly of the TFIIIB Brf1–TBP complex, the τB complex is displaced by Bdp1 binding to the TPR8 region of Tfc4 (26). Both DGLT-A RNH and IN-NED target the TPR7/8 region of Tfc4, suggesting that the intasome of DGLT-A may compete with Bdp1 for binding on τA during the initiation of Pol III transcription. How the interaction of the DGLT-A intasome with the right TPR array of Tfc4 provides the integrase access to a DNA region immediately upstream of the transcription start site remains unknown.

Roles of RNH-IN interactions for retrotransposition

In the present study, we observed the robust interaction of DGLT-A RNH with IN-NED, suggesting that this interaction may be required to mediate DGLT-A retrotransposition. Several reports underscore the importance of RT/RNH-IN interactions during the amplification of retroviruses and retrotransposons. For example, HIV-1 RT interacts with IN and inhibits its catalytic activity, likely to prevent the premature integration of cDNA (51–53). During retrotransposition of the *S. pombe* retrotransposon Tf1, cross-talk between RT and IN is important for RT-mediated primer removal after reverse transcription (54). The direct interaction of Tf1 RNH with IN was observed in yeast two-hybrid assays, although in this case, IN-CCD and not IN-NTD mediated the interaction with RNH (55).

Ty3 RT/RNH forms dimers when bound to DNA (40). The DGLT-A RNH dimerization observed in our yeast two-hybrid assays is unlikely the driving force of RT/RNH dimer formation because in the solved structure of Ty3 RT/RNH (40) and in the derived model of DGLT-A RT/RNH the RNH domains are not in contact and actually face away from each other. However, the RT/RNH structure may suggest that exposed parts of RNH contribute platforms for the formation of higher order complexes with additional RT/RNH molecules or the IN protein. It has been shown that the Ty3 polyprotein is cleaved proteolytically between RT/RNH and IN (56) and it is likely that the proteins remain non-covalently bound to each other because during Ty3 retrotransposition the IN protein is required for the efficient reverse transcription and subsequent stabilization of the resulting cDNA and preintegration complex (57,58). There is no obvious similarity between Ty3 and DGLT-A around the regions reported as RT/RNH-IN cleavage sites by Kirchner and Sandmeyer (56). Thus, it remains unknown for now whether DGLT-A IN acts as a RT/RNH-IN fusion protein or as a separate protein but in contact with RT/RNH. The latter is implied by the robust binding of IN-NTD to RNH observed in the present study. These observations suggest that the interaction of RNH with IN is of significance for DGLT-A retrotransposition.

Because we determined the robust interaction of DGLT-A RNH with Tfc4, we must consider that RT/RNH is part of the DGLT-A preintegration complex. Although the IN protein bound to linear retrotransposon cDNA may be the minimal requirement to facilitate integration, the literature suggests that several other mobile element-derived and host proteins are present in preintegration complexes (intasomes). The presence of RT/RNH is particularly documented in HIV-1 intasomes (59–63). The roles of RT/RNH in post-reverse transcription processes of retroviral infection remain elusive but may involve the condensation and protection of retroviral cDNA and the stabilization of protein-DNA complexes in intasomes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [WI 1142/14-1 to T.W.]. Funding for open access charge: Deutsche Forschungsgemeinschaft (German research Foundation).
Conflict of interest statement. None declared.

REFERENCES

1. Orgel,L.E. and Crick,F.H.C. (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604–607.
2. Doolittle,W.F. and Sapienza,C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**, 601–603.
3. Sultana,T., Zamborlini,A., Cristofari,G. and Lesage,P. (2017) Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.*, **18**, 292–308.

4. Kim,J.M., Vanguri,S., Boeke,J.D., Gabriel,A. and Voytas,D.F. (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.*, **8**, 464–478.
5. Glöckner,G., Szafranski,K., Winckler,T., Dingermann,T., Quail,M., Cox,E., Eichinger,L., Noegel,A.A. and Rosenthal,A. (2001) The complex repeats of *Dictyostelium discoideum*. *Genome Res.*, **11**, 585–594.
6. Glöckner,G. and Heidel,A.J. (2009) Centromere sequence and dynamics in *Dictyostelium discoideum*. *Nucleic Acids Res.*, **37**, 1809–1816.
7. Spaller,T., Kling,E., Glöckner,G., Hillmann,F. and Winckler,T. (2016) Convergent evolution of tRNA gene targeting preferences in compact genomes. *Mob. DNA*, **7**, 17.
8. Kaller,M., Euteneuer,U. and Nellen,W. (2006) Differential effects of heterochromatin protein 1 isoforms on mitotic chromosome distribution and growth in *Dictyostelium discoideum. Eukaryot*. *Cell*, **5**, 530–543.
9. Xie,W., Gai,X., Zhu,Y., Zappulla,D.C., Sternglanz,R. and Voytas,D.F. (2001) Targeting of the yeast Ty5 retrotransposon to silent chromatin is medaited by interactions between integrase and Sir4p. *Mol. Cell. Biol.*, **21**, 6606–6614.
10. Boeke,J.D. and Devine,S.E. (1998) Yeast retrotransposons: finding a nice quiet neighborhood. *Cell*, **93**, 1087–1089.
11. Sandmeyer,S. (1998) Targeting retrotransposition: at home in the genome. *Genome Res.*, **8**, 416–418.
12. Winckler,T., Szafranski,K. and Glöckner,G. (2005) Transfer RNA gene-targeted integration: an adaptation of retrotransposable elements to survive in the compact *Dictyostelium discoideum* genome. *Cytogen. Genome Res.*, **110**, 288–298.
13. Hull,M.W., Erickson,J., Johnston,M. and Engelke,D.R. (1994) tRNA genes as transcriptional repressor elements. *Mol. Cell. Biol.*, **14**, 1266–1277.
14. Kendall,A., Hull,M.W., Bertrand,E., Good,P.D., Singer,R.H. and Engelke,D.R. (2000) A CBF5 mutation that disrupts nucleolar localization of early tRNA biosynthesis in yeast also suppresses tRNA gene-mediated transcriptional silencing. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 13108–13113.
15. Kinsey,P.T. and Sandmeyer,S.B. (1991) Adjacent pol II and pol III promoters: transcription of the yeast retrotransposon Ty3 and a target tRNA gene. *Nucleic Acids Res.*, **19**, 1317–1324.
16. Bolton,E.C. and Boeke,J.D. (2003) Transcriptional interactions between yeast tRNA genes, flanking genes and Ty elements: a genomic point of view. *Genome Res.*, **13**, 254–263.
17. Devine,S.E. and Boeke,J.D. (1996) Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev.*, **10**, 620–633.
18. Chalker,D.L. and Sandmeyer,S.B. (1992) Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev.*, **6**, 117–128.
19. Szafranski,K., Glöckner,G., Dingermann,T., Dannat,K., Noegel,A.A., Eichinger,L., Rosenthal,A. and Winckler,T. (1999) Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* transfer RNA genes. *Mol. Gen. Genet.*, **262**, 772–780.
20. Geiduschek,E. and Tocchini-Valentini,G. (1988) Transcription by RNA polymerase III. *Annu. Rev. Biochem.*, **57**, 873–914.
21. Baker,R.E., Gabrielsen,O. and Hall,B.D. (1986) Effects of $tRNA^{Tyr}$ point mutations on the binding of yeast RNA polymerase III transcription factor C. *J. Biol. Chem.*, **261**, 5275–5282.
22. Camier,S., Gabrielsen,O., Baker,R. and Sentenac,A. (1985) A split binding site for transcription factor tau on the $tRNA_3^{Glu}$ gene. *EMBO J.*, **4**, 491–500.
23. Ciliberto,G., Raugei,G., Costanzo,F., Dente,L. and Cortese,R. (1983) Common and interchangeable elements in the promoters of genes transcribed by RNA polymerase III. *Cell*, **32**, 725–733.
24. Kassavetis,G.A., Braun,B.R., Nguyen,L.H. and Geiduschek,E.P. (1990) S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. *Cell*, **60**, 235–245.
25. Moir,R.D. and Willis,I.M. (2004) Tetratricopeptide repeats of TFC4 and a limiting step in the assembly of the initiation factor TFIIIB. *Adv. Protein Chem.*, **67**, 93–121.
26. Male,G., von Appen,A., Glatt,S., Taylor,N.M., Cristovao,M., Groetsch,H., Beck,M. and Müller,C.W. (2015) Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. *Nat. Commun.*, **6**, 7387.
27. Ducrot,C., Lefebvre,O., Landrieux,E., Guirouilh-Barbat,J., Sentenac,A. and Acker,J. (2006) Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors. *J. Biol. Chem.*, **281**, 11685–11692.
28. Chaussivert,N., Conesa,C., Shaaban,S. and Sentenac,A. (1995) Complex interactions between yeast TFIIIB and TFIIIC. *J. Biol. Chem.*, **270**, 15353–13358.
29. Dumay-Odelot,H., Acker,J., Arrebola,R., Sentenac,A. and Marck,C. (2002) Multiple roles of the τ131 subunit of yeast transcription factor IIIC (TFIIIC) in TFIIIB assembly. *Mol. Cell. Biol.*, **22**, 298–308.
30. Liao,A.K., Willis,I.M. and Moir,R.D. (2003) The Brf1 and Bdp1 subunits of transcription factor TFIIIB bind to overlapping sites in the tetratricopeptide repeats of TFC4. *J. Biol. Chem.*, **275**, 44467–44474.
31. Liao,Y., Moir,R.D. and Willis,I.M. (2006) Interactions of Brf1 peptides with the tetratricopeptide repeat-containing subunit of TFIIIC inhibit and promote preinitiation complex assembly. *Mol. Cell. Biol.*, **26**, 5946–5956.
32. Moir,R.D., Puglia,K.V. and Willis,I.M. (2000) Interactions between the tetratricopeptide repeat-containing transcription factor TFIIIC131 and its ligand, TFIIIB70. *J. Biol. Chem.*, **275**, 26591–26598.
33. Ruth,J., Conesa,C., Dieci,G., Lefebvre,O., Dusterhoft,A., Ottonello,S. and Sentenac,A. (1996) A suppressor of mutations in the class III transcription system encodes a component of yeast TFIIIB. *EMBO J.*, **15**, 1941–1949.
34. Rozenfeld,S. and Thuriaux,P. (2001) Genetic interactions within TFIIIC, the promoter-binding factor of yeast RNA polymerase III. *Mol. Genet. Genomics*, **265**, 705–710.
35. Yieh,L., Kassavetis,G.A., Geiduschek,E.P. and Sandmeyer,S.B. (2000) The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific Integration of the gypsy-like element, Ty3. *J. Biol. Chem.*, **275**, 29800–29807.
36. Qi,X. and Sandmeyer,S. (2012) In vitro targeting of strand transfer by the Ty3 retroelement integrase. *J. Biol. Chem.*, **287**, 18589–18595.
37. Aye,M., Dildine,S.L., Claypool,J.A., Jourdain,S. and Sandmeyer,S.B. (2001) A truncation mutant of the 95-kilodalton subunit of transcription factor IIIC reveals asymmetry in Ty3 integration. *Mol. Cell. Biol.*, **21**, 7839–7851.
38. Dohmen,R.J., Strasser,A.W., Höner,C.B. and Hollenberg,C.P. (1991) An efficient transformation procedure enabling long-term storage of competent cells of various yeast genera. *Yeast*, **7**, 691–692.
39. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **24**, 4876–4882.
40. Nowak,E., Miller,J.T., Bona,M.K., Studnicka,J., Szczepanowski,R.H., Jurkowski,J., Le Grice,S.F.J. and Nowotny,M. (2014) Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nat. Struct. Mol. Biol.*, **21**, 389–396.
41. Li,X., Krishnan,L., Cherepanov,P. and Engelman,A. (2011) Structural biology of retroviral DNA integration. *Virology*, **411**, 194–205.
42. Moir,R.D., Puglia,K.V. and Willis,I.M. (2002) Autoinhibition of TFIIIB70 binding by the tetratricopeptide repeat-containing subunit of TFIIIC. *J. Biol. Chem.*, **277**, 694–701.
43. Marck,C., Lefebvre,O., Carles,C., Riva,M., Chaussivert,N., Ruet,A. and Sentenac,A. (1993) The TFIIIB-assembling subunit of yeast transcription factor TFIIIC has both tetratricopeptide repeats and basic helix-loop-helix motifs. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 4027–4031.
44. Maertens,G., Cherepanov,P., Pluymers,W., Busschots,K., De Clercq,E., Debyser,Z. and Engelborghs,Y. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.*, **278**, 33528–33539.
45. Cherepanov,P., Ambrosio,A.L.B., Rahman,S., Ellenberger,T. and Engelman,A. (2005) Structural basis for the recognition between

HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 17308–17313.
46. Leem,Y.E., Ripmaster,T.L., Kelly,F.D., Ebina,H., Heincelman,M.E., Zhang,K., Grewal,S.I.S., Hoffman,C.S. and Levin,H.L. (2008) Retrotransposon Tf1 is targeted to pol II promoters by transcription activators. *Mol. Cell*, **30**, 98–107.
47. Jacobs,J.Z., Rosado-Lugo,J., Cranz-Mileva,S., Ciccaglione,K.M., Tournier,V. and Zaratiegui,M. (2015) Arrested replication forks guide retrotransposon integration. *Science*, **349**, 1549–1553.
48. Bridier-Nahmias,A., Tchalikian-Cosson,A., Baller,J.A., Menouni,R., Fayol,H., Flores,A., Saib,A., Werner,M., Voytas,D.F. and Lesage,P. (2015) An RNA polymerase III subunit determines sites of retrotransposon integration. *Science*, **348**, 585–588.
49. Cheung,S., Ma,L., Chan,P.H., Hu,H.L., Mayor,T., Chen,H.T. and Measday,V. (2016) Ty1-Integrase interacts with RNA Polymerase III specific subcomplexes to promote insertion of Ty1 elements upstream of Pol III-transcribed genes. *J. Biol. Chem.*, **291**, 6396–6411.
50. Ishiguro,A., Kassavetis,G.A. and Geiduschek,E.P. (2002) Essential roles of Bdp1, a subunit of RNA polymerase III initiation factor TFIIIB, in transcription and tRNA processing. *Mol. Cell. Biol.*, **22**, 3264–3275.
51. Oz,I., Avidan,O. and Hizi,A. (2002) Inhibition of the integrase of human immunodeficiency viruses type 1 and 2 by reverse transcriptase. *Biochem. J.*, **361**, 557–566.
52. Tasara,T., Maga,G., Hottiger,M.O. and Hubscher,U. (2001) HIV-1 reverse transcriptase and integrase enzymes physically interact and inhibit each other. *FEBS Lett.*, **507**, 39–44.
53. Hehl,E.A., Joshi,P., Kalpana,G.V. and Prasad,V.R. (2004) Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins. *J. Virol.*, **78**, 5056–5067.
54. Herzig,E., Voronin,N. and Hizi,A. (2012) The removal of RNA primers from DNA synthesized by the reverse transcriptase of the retrotransposon Tf1 Is stimulated by Tf1 integrase. *J. Virol.*, **86**, 6222–6230.
55. Steele,S.S. and Levin,H.L. (1998) A map of interactions between the proteins of a retrotransposon. *J. Virol.*, **72**, 9318–9322.
56. Kirchner,J. and Sandmeyer,S. (1993) Proteolytic processing of Ty3 proteins is required for transposition. *J. Virol.*, **67**, 19–28.
57. Nymark-McMahon,M.H. and Sandmeyer,S. (1999) Mutations in nonconserved domains of Ty3 integrase affect multiple stages of the Ty3 life cycle. *J. Virol.*, **73**, 453–465.
58. Nymark-McMahon,M.H., Beliakova-Bethell,N.S., Darlix,J.L., Le Grice,S.F.J. and Sandmeyer,S. (2002) Ty3 integrase is required for initiation of reverse transcription. *J. Virol.*, **76**, 2804–2816.
59. Bukrinsky,M.I., Sharova,N., McDonald,T.L., Pushkarskaya,T., Tarpley,W.G. and Stevenson,M. (1993) Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 6125–6129.
60. Gallay,P., Swingler,S., Song,J., Bushman,F. and Trono,D. (1995) HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell*, **17**, 569–576.
61. Gallay,P., Hope,T., Chin,D. and Trono,D. (1997) HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 9825–9830.
62. Farnet,C.M. and Haseltine,W.A. (1991) Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J. Virol.*, **65**, 1910–1915.
63. Miller,M.D., Farnet,C.M. and Bushman,F.D. (1997) Human Immunodeficiency Virus Type 1 preintegration complexes: Studies of organization and composition. *J. Virol.*, **71**, 5382–5390.