Ty1 integrase is composed of an active N-terminal domain and a large disordered C-terminal module dispensable for its activity *invitro*

Phong Quoc Nguyen, Christine Conesa, Elise Rabut, Gabriel Bragagnolo, Célia Gouzerh, Carlos Fernández-Tornero, Pascale Lesage, Juan Reguera, Joël Acker

JBC RESEARCH ARTICLE

Ty1 integrase is composed of an active N-terminal domain and a large disordered C-terminal module dispensable for its activity *in vitro*

Received for publication, June 4, 2021, and in revised form, August 5, 2021 Published, Papers in Press, August 17, 2021,
https://doi.org/10.1016/j.jbc.2021.101093

Phong Quoc Nguyen[1,‡], Christine Conesa[2,‡], Elise Rabut[2], Gabriel Bragagnolo[1], Célia Gouzerh[2], Carlos Fernández-Tornero[3], Pascale Lesage[4], Juan Reguera[1,5,*], and Joël Acker[2,*]

From the [1]Aix-Marseille Université, CNRS, AFMB UMR 7257, Marseille, France; [2]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France; [3]Centro de Investigaciones Biológicas Margarita Salas, CSIC, Madrid, Spain; [4]INSERM U944, CNRS UMR 7212, Genomes and Cell Biology of Disease Unit, Institut de Recherche Saint-Louis, Université de Paris, Hôpital Saint-Louis, Paris, France; [5]INSERM, AFMB UMR7257, Marseille, France

Edited by Craig Cameron

Long-terminal repeat (LTR) retrotransposons are genetic elements that, like retroviruses, replicate by reverse transcription of an RNA intermediate into a complementary DNA (cDNA) that is next integrated into the host genome by their own integrase. The Ty1 LTR retrotransposon has proven to be a reliable working model to investigate retroelement integration site preference. However, the low yield of recombinant Ty1 integrase production reported so far has been a major obstacle for structural studies. Here we analyze the biophysical and biochemical properties of a stable and functional recombinant Ty1 integrase highly expressed in *E.coli*. The recombinant protein is monomeric and has an elongated shape harboring the three-domain structure common to all retroviral integrases at the N-terminal half, an extra folded region, and a large intrinsically disordered region at the C-terminal half. Recombinant Ty1 integrase efficiently catalyzes concerted integration *in vitro*, and the N-terminal domain displays similar activity. These studies that will facilitate structural analyses may allow elucidating the molecular mechanisms governing Ty1 specific integration into safe places in the genome.

Ty LTR retrotransposons, the only transposable elements present in the yeast *Saccharomyces cerevisiae* (1, 2), are structurally and functionally related to retroviruses. Among the five distinct families, Ty1, Ty2, Ty3, Ty4, and Ty5, identified in the reference strain S288 C, Ty1 is the most abundant and is still active (3–6). Following Ty1 transcription by RNA polymerase II (Pol II), the Ty1 mRNA is translated into Gag and the Gag-Pol polyprotein that includes sequentially the capsid protein (Gag), the protease (PR), the integrase (IN), and the reverse transcriptase (RT). Gag and Gag-Pol polyprotein pack the Ty1 mRNA into cytoplasmic virus-like particles (VLPs), where occurs a PR-dependent maturation of the different proteins. The Ty1 mRNA is then reverse transcribed by RT into a linear double-stranded DNA copy (cDNA). IN binds to the cDNA to form the preintegration complex, which migrates into the nucleus where the cDNA is integrated into the host genome (6), preferentially upstream of Pol III-transcribed genes (7–10) through a direct interaction between IN and Pol III (11, 12). Previous studies suggest that, like retroviruses, Ty1 integration process is carried out by complexes composed of multiple copies of IN assembled on the LTR DNA ends, commonly called intasomes (13, 14).

Ty1 IN has the same three-domain organization as that of other LTR-retrotransposons and retroviruses (15): (i) the N-terminal domain (NTD), which is involved in integrase multimerization and is characterized by an HHCC zinc finger motif; (ii) the catalytic core domain (CCD), which contains the active site of the enzyme with the invariant D,D(35)E motif; and (iii) the less conserved C-terminal domain (CTD). The CTD, which is particularly extended in Ty1 compared with retroviral INs, plays distinct roles during Ty1 replication cycle. The last 115 C-terminal residues are important for Ty1 reverse transcription *in vivo* (16, 17), and their presence is critical to detect recombinant RT activity *in vitro* (18). In addition, a nuclear localization signal and the targeting domain of Ty1 responsible for IN recruitment to Pol III-transcribed genes are located at the extreme C-terminus of the protein (11, 13, 19).

In recent years, substantial structural and functional information on retroviral intasomes has provided unprecedented detail on the molecular mechanism of retroviral DNA integration (20–26). By contrast, there are no structural data on the Ty1 integration process and the interaction between IN and Pol III. A major obstacle to obtain structural information on the Ty1 integration mechanisms is the purification of high yields of homogeneous and active IN protein. Our current knowledge on Ty1 IN catalytic activity *in vitro* stems from a poorly soluble partially purified ectopic integrase, which was active *in vitro* in integration, disintegration, and concerted integration assays and exhibited biochemical properties similar to those detected with Ty1 VLPs (27–30). Unlike retrovirus (31) and Ty3 (32), Ty1 LTR DNA ends with the conserved

‡ These authors contributed equally to this work.
* For correspondence: Joël Acker, joel.acker@cea.fr; Juan Reguera, juan.reguera@inserm.fr.

inverted dinucleotides 5′-TG…CA-3′ and thus no further 3′-end processing activity is required before integration. Further studies have determined the sequence specificity of IN integration activity to be limited to the presence of A:T ends on either strand at the blunt end of the cDNA (28).

Here, we present an improved procedure for high yield production of homogeneous and functional Ty1 IN expressed in *Escherichia coli*. Through biophysical and enzymatic methods, we demonstrate the monomeric nature of recombinant Ty1 IN in the apo form and provide structural insights into the domain distribution of this unusually large integrase. We identify the precise boundary between a folded region at the N-terminal half, which has an extended CTD, and an intrinsically disordered region at the C-terminal half, which includes the protein fragment required for RT activity and the Ty1 targeting domain that directs the interaction of IN with Pol III. State-of-the-art fluorescence-based activity assays show that this recombinant Ty1 IN is fully functional *in vitro* and suggest an unexpected role for Mn^{2+} in Ty1 integration *in vivo*. These results pave the way for further structural and functional analyses aimed to decipher the molecular mechanisms allowing Ty1 specific integration upstream of Pol III-transcribed genes.

Results

Ty1 integrase expression and purification

A low expression of endogenous Ty1 IN ORF in bacteria was observed in preliminary tests (Fig. 1*A*, left, IN_{wt}). Therefore, we engineered an optimized coding DNA for expression in *E. coli* (Fig. S1), and two constructs were designed that both expressed IN as N-terminal fusion proteins with six histidine residues but differed with the second one, which consisted of the Fh8 tag, a small domain known to be a good solubility enhancer (33) or the Sso7d tag, a nonspecific DNA binding domain that was shown to help to produce a highly soluble and hyperactive HIV integrase (34). Both constructs allowed the production of high level of IN protein (Fig. 1*A*, left, 5–10 mg/l of *E. coli* culture) that could only be solubilized under high salt conditions (1 M NaCl) as previously described for recombinant Ty1 IN expressed in yeast (27). We observed that the apparent molecular weight determined on SDS-PAGE was particularly higher than the calculated molecular weight (82 kDa). Both proteins were first purified by nickel-affinity chromatography followed by gel filtration, a procedure that produced a low level of homogenous IN and larger peaks that might correspond to Ty1 IN multimerization or aggregation (Fig. 1*A*, right panel). Additional analysis performed using a Superose 6 column showed that these large peaks eluted with an apparent molecular weight higher than 5000 kDa (Fig. S2*A*) and thus probably corresponded to aggregates and not to multimers of Sso7d-IN. Since integrases are known to bind nonspecifically to DNA, we assessed whether the addition of benzonase that degrades all form of DNA and RNA might help to improve the obtention of homogenous nonaggregated Ty1 IN. We also assayed the addition of NP40, a good protein solubility enhancer found to prevent multimerization of HIV IN (35). Extracts were prepared according to the procedure shown in Figure 1*B*. Compared with untreated extracts, the addition of NP40 (0.05%) enhanced the total protein solubility (Fig. 1*C*, left). The presence of both benzonase and NP40 increased the yield of homogeneous nonaggregated Ty1 IN, which migrated in size exclusion chromatography (SEC) with an apparent molecular weight of ∼200 kDa relative to protein standards (Fig. S2*B*). During the course of our experiments, we also noted that once extracted at high salt concentration, both Ty1 IN fusion proteins could remain soluble at lower salt conditions (Fig. 1*D*, left). However, attempts to perform the whole purification at 300 mM NaCl mainly led to the production of aggregates (Fig. 1*D*, right). In conclusion, we provide an optimized protocol for the expression and purification of high yields of pure and homogeneous Ty1 integrase for biochemical and structural studies.

Architectural characterization of Ty1 integrase

Previous intragenic complementation experiments suggested that Ty1 IN functions *in vivo* as a multimer like all retroviral integrases (13, 14). We thus wondered whether our recombinant Ty1 IN formed multimers *in vitro*. SEC coupled to a multi-angle light scattering (36) revealed that the molecular weight of the protein is 82 kDa (Fig. 2*A*), which corresponds to the monomeric form of the integrase plus the tags (71.5 kDa and 10.5 kDa, respectively). The higher apparent molecular mass detected in both gel filtration (Fig. 1*C* and Fig. S2*B*) and SDS-PAGE (Fig. 1*A*) is a feature characteristic of intrinsic disordered proteins (36). We thus used remote structural homology search with HHpred (37) and protein disorder predictors, PrDOS (38) and Globprot 2 (39) to better understand the disordered nature of Ty1 IN (Fig. S3). HHpred revealed that the N-terminal half of Ty1 IN, residues 9 to 311, folds like retroviral integrases detecting PFV, HIV, and Rouse Sarcoma Virus with probabilities above 99% and E values below 7×10^{-8} (Fig. S3*A*). The alignment defines the conserved three-domain organization of retroviral integrases, with the N-terminal (NTD), the central catalytic (CCD), and the C-terminal (CTD) domains. PrDOS predicted the N-terminal half region to be folded and Globprot-2 identified three domains in this region (Fig. S3*B*). Both servers predicted a disordered C-terminal half region (Fig. S3*B*). To experimentally assess the disordered nature of the C-terminal half, purified Sso7d-IN was subjected to limited proteolysis experiments using papain protease (Fig. 2*B*). One main protease resistant fragment was obtained with a molecular weight of 52.9 kDa as determined by mass spectrometry (Fig. S4, *A* and *B*). Our attempt to determine the N-terminal sequence of this fragment by Edman method was unsuccessful, suggesting that this peptide was the very N-terminus of Ty1 IN and was blocked by posttranslational modifications. To bypass this limitation, additional proteolysis experiments were performed using a SUMO tagged Ty1 IN purified after the N-terminal tag removal by the TEV protease (Fig. S4*C*). In this case, a papain resistant fragment of 43.1 kDa was obtained. N-terminal sequencing revealed that the fragment started with the two last

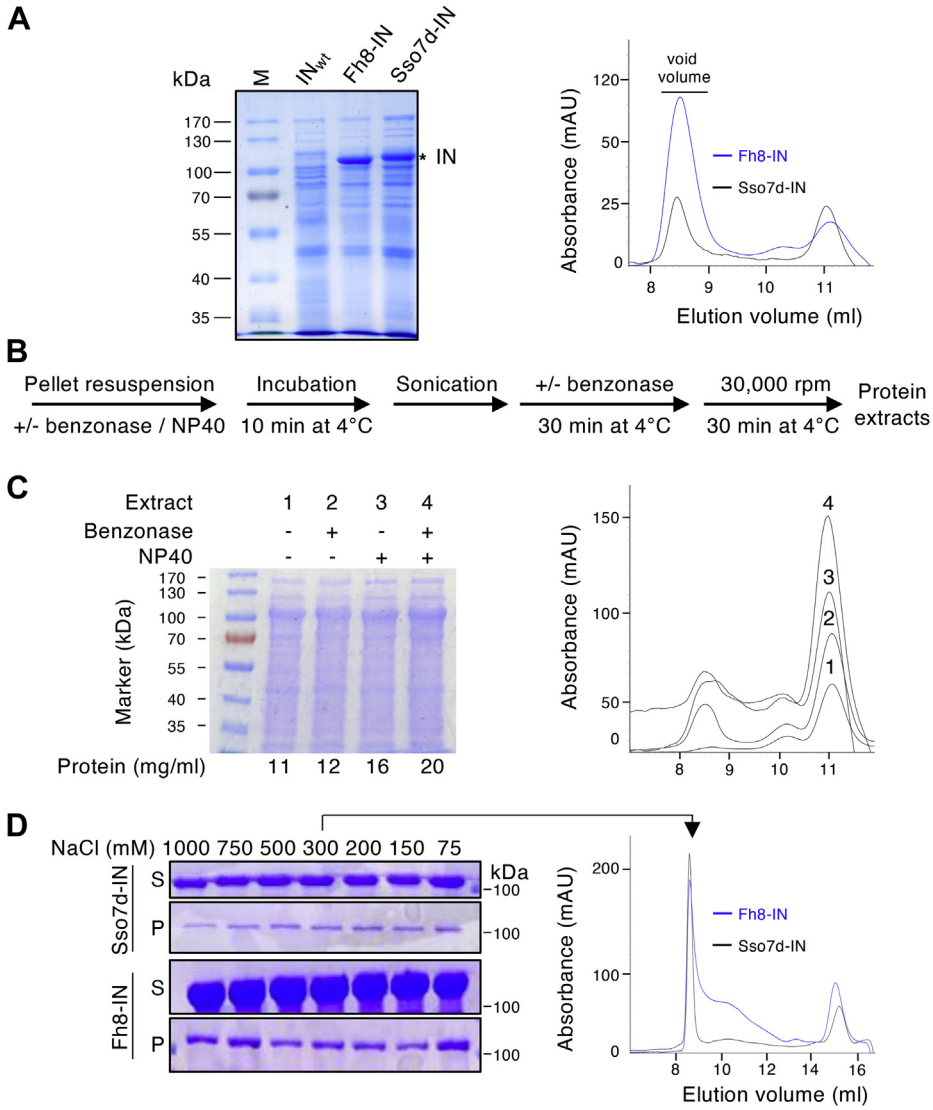

Figure 1. Production of soluble homogeneous Ty1 IN in *E. coli*. *A*, expression of Ty1 IN in *E. coli*. *Left panel*, soluble extracts (20 μg) obtained from *E. coli* expressing the indicated constructs were analyzed by SDS-PAGE and stained with Coomassie blue. IN_{wt} corresponds to the endogenous Ty1 IN cDNA expressed as an N-terminal fusion with six histidine residues (6H). Ty1 IN cDNA was codon optimized and expressed as an N-terminal fusion with 6H and Fh8 (Fh8-IN) or 6H and Sso7d (Sso7d-IN). The protein sequences of IN_{wt}, Fh8-IN and Sso7d-IN are shown in Fig. S1. Expected sizes are 75 kDa for IN_{wt} and 82 kDa for Fh8-IN or Sso7d-IN. *Right panel*, extracts prepared with the initial solubilization buffer (see Experimental procedures) were purified by nickel affinity chromatography, then analyzed by gel filtration. Elution profiles on a Superdex 200 GL 10/30 column of Fh8-IN (*blue line*) or Sso7d-IN (*black line*) extracts are shown. The large peaks obtained for both Fh8-IN and Sso7d-IN extracts were excluded from this column and might be mainly composed of aggregated proteins. *B*, schematic representation of the different steps of the purification optimized to prepare large amounts of homogeneous Ty1 IN soluble extracts (for details, see Experimental procedures). *C*, optimization of homogeneous Ty1 IN protein extraction. *Left panel*, the same pellet of *E. coli* cells was extracted with solubilization buffer in the presence (+) or not (–) of NP40 (0.05%) and benzonase (25 U per ml of extract) using the procedure shown in panel *B*. The four resulting extracts (1–4 as indicated) were analyzed by SDS-PAGE and Coomassie blue staining. The protein concentration of the soluble extracts and the sizes of the bands present in a protein ladder (M) are indicated. *Right panel*, the elution profiles of the four (1–4 as indicated) extracts analyzed by gel filtration on a Superdex 200 increase 10/300 GL column are shown (See Fig. S2*B* for the elution profiles of molecular weight markers). *D*, solubility of recombinant Ty1 IN. *Left panel*, Fh8-IN or Sso7d-IN extracts number 4 (panel *C*) prepared with the optimized solubilization buffer in the presence of 1 M NaCl were diluted to the indicated NaCl concentration and incubated for 30 min at room temperature. After centrifugation, pellets (P) and supernatants (S) were analyzed by SDS-PAGE and Coomassie blue staining. *Right panel*, Fh8-IN or Sso7d-IN extracts prepared with the optimized solubilization buffer in the presence of 1 M NaCl were purified by nickel affinity chromatography, dialyzed to decrease salt concentration to 300 mM NaCl, and analyzed by gel filtration on a Superose 6 increase 10/300 G column. The elution profiles of Fh8-IN (*blue line*) or Sso7d-IN (*black line*) are shown. The large peaks obtained for both Fh8-IN and Sso7d-IN extracts were excluded from this column and might be mainly composed of aggregated proteins.

amino acids from the TEV protease cleavage site (GA) followed by the N-terminal residues of Ty1 IN (NVHTS-). According to the size of the peptide derived from Sso7d-IN obtained after papain cleavage (52.9 kDa) and of the Sso7d tag (9.9 kDa), we can conclude that the papain resistant fragment from both constructs ends with the same C-terminus (residue K375 from Ty1 IN). These data indicate that (i) the C-terminal region encompassing the last 260 residues of IN is disordered in good agreement with the *in silico* prediction (Fig. S3*B*); (ii) an extended CTD region (eCTD), which is papain-resistant is present in the C-terminal half predicted disordered region (residues I306 to K375) and may represent an enlargement of the domain or a domain that is independent of the CTD homologous to retroviral INs (Fig. 2*D*).

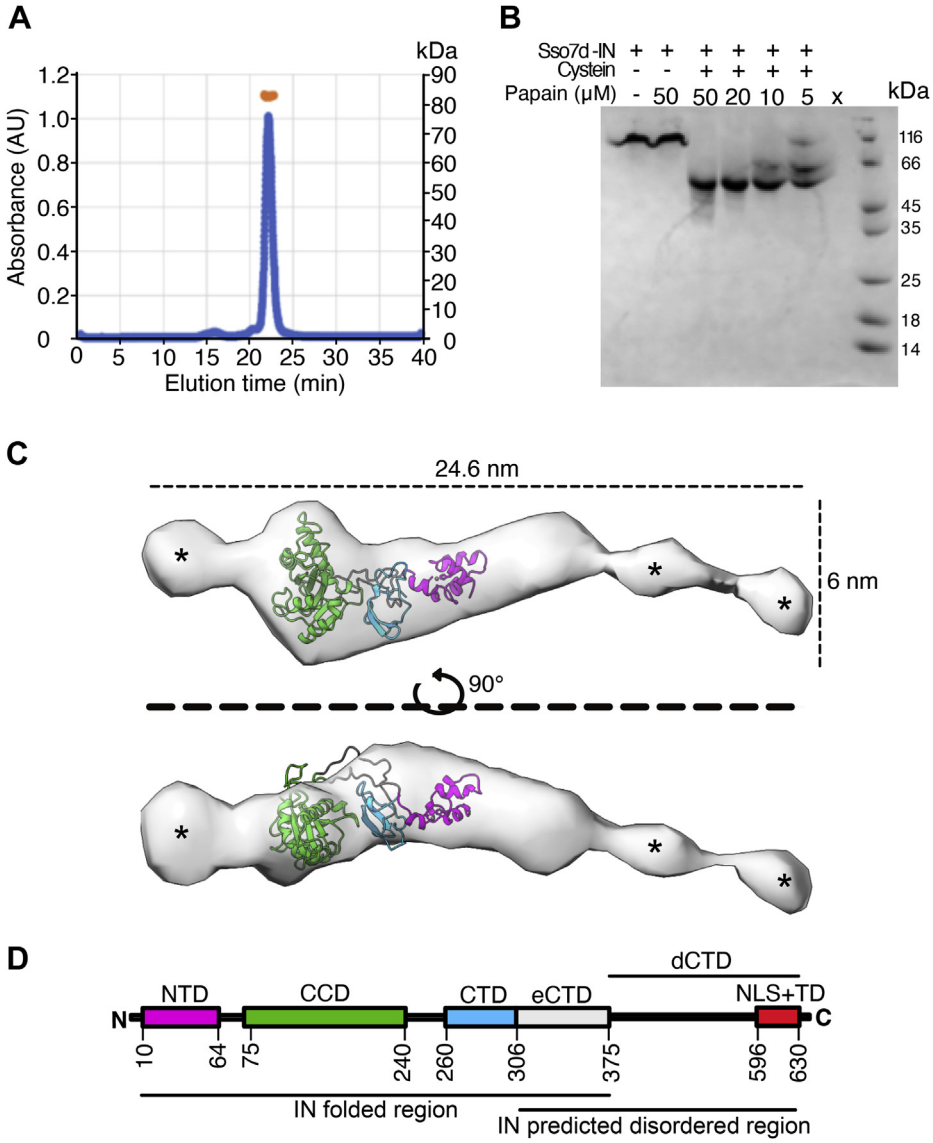

Figure 2. Biophysical properties of Ty1 IN. *A*, recombinant Ty1 IN is monomeric. About 50 μg of purified Sso7d-IN was subjected to size-exclusion chromatography coupled to MALS/RI/UV detectors as described in Experimental procedures. The molecular mass (*orange line*), derived from refractive index measurements, and the absorption at 280 nm (*blue line*) were plotted as a function of the elution time. *B*, limited proteolysis of Ty1 IN. Analysis by SDS-PAGE and Coomassie blue staining of papain limited digestion of Sso7d-IN at different concentrations of papain with and without the cysteine protease cofactor. A protease-resistant fragment appears between the 45 kDa and 66 kDa molecular weight markers (*right lane*). *C*, Ty1 IN reconstructed model. Orthogonal views of the SAXS calculated volume of Sso7d-IN are represented as a *gray transparent surface*. The Ty1 IN model calculated by SWISS-MODEL is superposed to the SAXS volume and is represented in *cartoons* with the CCD colored in *green*, the NTD in *magenta*, and the CTD in *cyan*. *D*, schematic representation of the Ty1 IN structure. The NTD, CCD, and CTD boundaries are defined by HHpred and SWISS-MODEL sequence alignments. The extended CTD region (eCTD) and the disordered region (dCTD) are defined by the size of the papain resistant fragment. The Nuclear Localization Signal + Targeting Domain (NLS+TD) region are defined in Asif-Laidin *et al.*, 2020 (11). *Asterisks* correspond to the three additional globular domains.

In order to have further information on Ty1 IN architecture, small angle X-ray scattering (SAXS) experiments were carried out with Sso7d-IN (see Experimental procedures and Fig. S4*D*). The log I(s) *versus* s plot represents the primary SAXS data. The P(r) *versus* r profile shows that the smooth approaches zero at r = 0 and r_{max} = 266 Å (Fig. S4*D*). The shape of P(r) *versus* r profile and d_{max} = 248.6 Å suggested that the protein is partially folded. The volume calculated is shown as a gray envelope in Figure 2*C* with 0.00654 contour level (represented by Chimera X). The volume appears elongated with globular enlargements and measures 24.6 nm long by 6 nm broad in the largest region. Interestingly, the Ty1 IN atomic model produced by SWISS-MODEL (40) using the homologous PFV integrase (PDB: 30YM-A, Fig. S5) assembled in the intasome fits in size and shape with the calculated volume. The CCD domain of Ty1 IN can be unambiguously located into the largest globular stretch. This fitting strongly suggests that the NTD and CTD regions could reside in an annexed elongated tubular area that could also harbor the extra folded residues determined by limited proteolysis (Fig. S4*C*). In addition to the central volume corresponding to the integrase CCD, three additional globular volumes were detected. The larger one is located next to the CCD domain, whereas the two others, smaller, are connected to the NTD

region. These regions likely correspond to the disordered C-terminus of Ty1 IN and to the N-terminal Sso7d tag, which is a folded globular domain of 29 Å in diameter, equivalent to the dimensions of the additional globular volume. In conclusion, the apo form of Ty1 IN is monomeric and has an elongated shape harboring the integrase conserved domains at the N-terminal half, an extra folded region (eCTD) after the retroviral-like CTD up to residue K375, and an intrinsically disordered region of 260 amino acids at the C-terminal half (dCTD, Fig. 2*D*).

Ty1 integrase catalytic activity

To detect recombinant Ty1 IN activity, we tested its ability to mediate strand transfer reaction using an oligonucleotide integrase assay with short blunt-end double-stranded DNA probes of 30 bp mimicking the end of the U3 or U5 regions of Ty1 LTR (Fig. S6*A*, (27)). Strand transfer reaction is a concerted cleavage and ligation reaction. *In vitro*, integrases exhibit little DNA specificity. Thus, the cleavage of the donor DNA generates multiple nonspecific (NS) products smaller than the substrate, whereas the ligation of the cleaved DNA yields various strand transfer (ST) products longer than the substrate (Fig. S6). Both NS and ST products are detected when Ty1 U3 or U5 probes are incubated with an active integrase. To monitor IN activity, Ty1 U3 and U5 probes were labeled with a fluorescent dye that can be attached to either end of both complementary oligonucleotides designed to assemble the double-stranded DNA probe. To determine the optimal probes for the detection of integration activity *in vitro*, we assayed all eight possible combinations for each probe. A single dye attached to the probe was sufficient to detect IN activity and no significant improvement was observed in the presence of two dyes (for details, see Fig. S6). The experiments also showed that, as expected, none or very weak activity could be detected when the dye is attached close to the highly conserved CA dinucleotide present at the end of the U3 or U5 domains. According to these data, subsequent oligonucleotide integration assays were performed with Ty1 U3 or U5 substrates labeled with a single dye attached to the 5′ end of the oligonucleotide harboring the conserved CA dinucleotide at the 3′ end (Fig. S6*B*, substrate a).

In order to identify the best tag for biochemical and structural studies, we next compared the activity of Sso7d-IN and Fh8-IN using the Ty1 U3 probe as a substrate. Both recombinant INs exhibited ST activity as shown by the presence of labeled products above the unreacted substrate, only when the reaction mixtures contained a divalent cation, as expected (27) (Fig. 3*A*). Sso7d-IN appeared to be more active since similar fluorescent signals could be detected with five times less protein than for Fh8-IN. A time course experiment showed that the catalytic activity of Sso7d-IN was more stable (Fig. 3*B*). First, the increase in ST product levels stopped very rapidly for Fh8-IN (10–15 min) as compared with Sso7d-IN (∼30 min). Second, an incubation of Fh8-IN at 30 °C in reaction mixtures without the substrate resulted very rapidly in a complete loss of activity, whereas Sso7d-IN remained active (Fig. S7*A*). These data suggest that Fh8-IN is less stable as compared with Sso7d-IN and in consequence less active *in vitro*. The DNA-binding activity of Sso7d may explain the enhanced activity of this specific construct. However, in the case of the HIV hyperactive Sso7d-IN, a mutation that greatly reduced Sso7d DNA binding did not result in a significant decrease of the integrase activity (34) suggesting that this DNA binding activity is not involved in the enhanced activity of Ty1 IN. For further optimization of the integration reaction, we analyzed the salt requirement for Sso7d-IN activity, performing assays with increasing amounts of KCl (Fig. 3*C*). An optimal signal was observed between 50 and 125 mM KCl that rapidly decreased with higher salt concentrations. No activity was detected at 300 mM, in agreement with what was observed for Ty1 IN purified from VLPs (30). We also examined the requirement for Mn^{2+} or Mg^{2+} using a constant amount of Sso7d-IN or Fh8-IN and increasing amounts of each divalent cation. Using Ty1 U3 as a substrate (Fig. 3*D*, left), a similar efficiency of integration was observed for both cations as assessed by the levels of the predominant ST product, indicating no preference for Mn^{2+} over Mg^{2+}. However, the concentrations needed to reach optimal signals were different (less than 0.1 mM for Mn^{2+} and 2.5 mM for Mg^{2+}). On the other hand, the global patterns of both ST and NS reaction products appeared to be distinct. In the presence of Mg^{2+}, one major band was detected, whereas several bands of weak intensity were observed with Mn^{2+} suggesting less integration specificity with Mn^{2+}. The same experiment performed with Ty1 U5 as a substrate gave different results, with a clear preference for Mn^{2+} over Mg^{2+}, evidenced by weaker intensity bands in the presence of Mg^{2+}, as previously described (27) (Fig. 3*D*, left panel). In addition, a higher $MgCl_2$ concentration (20 mM) than for Ty1 U3 was necessary to detect maximal IN activity. In these assays, the ST and NS signals obtained with Ty1 U3 as a substrate were stronger and more homogeneous than with Ty1 U5, in accordance with the results of *in vitro* integration assays using Ty1 IN prepared from VLPs (29).

A body of evidence suggests that Mg^{2+} is a more physiologically relevant cofactor of IN than $Mn^{2+}$and that IN activity is less specific in the presence of Mn^{2+} (31). In the case of Ty1 IN, additional ST products were observed in the presence of Mn^{2+} compared with Mg^{2+} (Fig. 3*D*, left panel). However, the cellular level of Mn^{2+} is much lower than the conditions typically used *in vitro*. A study showed that in yeast, the cellular concentration of manganese (1.5 μM) is 1000-fold lower than that of magnesium (1.5 mM) (41). We thus performed additional experiments to determine the minimum concentration of Mn^{2+} required for Ty1 IN activity (Fig. 3*D*, right panel and Fig. S7*B*). Interestingly, a very limited amount of Mn^{2+} was sufficient to observe activity (<0.1 μM for Ty1 U3; 2.5 μM for Ty1 U5) and an optimal signal was obtained for Ty1 U3 in the presence of physiological concentration (1.5 μM). Upon these experimental conditions, the global patterns were similar for both cofactors with mainly one major ST product suggesting that lower amounts of Mn^{2+} resulted in higher integration specificity (Fig. 3*D*, right panel). Experiments performed with increasing amounts of Mn^{2+} in the

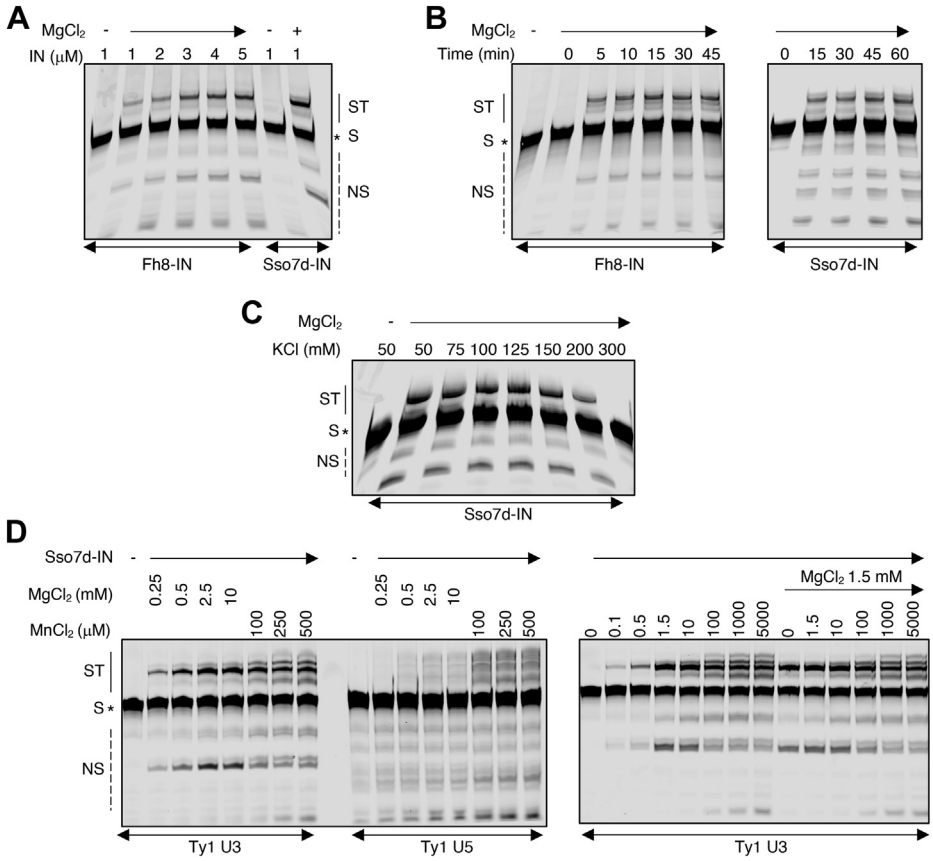

Figure 3. Biochemical characterization of recombinant Ty1 IN activity. Analysis of strand transfer (ST) and nonspecific endonuclease (NS) activities of recombinant Fh8-IN or Sso7d-IN were performed in an oligonucleotide integrase assay as described in Experimental procedures using 30 bp fluorescent double-stranded DNA probes mimicking Ty1 LTR-U3 or Ty1 LTR-U5 ends as substrates (Ty1 U3 or Ty1 U5, for details see Fig. S6). Samples were loaded on a 20% denaturing polyacrylamide gel that was then analyzed using an Odyssey CLx imaging system. *Dark lines* indicate the positions of ST products, *dashed lines* the NS products (and the shortmers), and *asterisks* the substrates (S). *A*, Sso7d-IN is an hyperactive IN. Increasing amounts of Fh8-IN from 1 to 5 μM as indicated or 1 μM Sso7d-IN were assayed at 30 °C for 60 min using a Ty1 U3 probe in reaction mixtures containing 100 mM KCl and 5 mM $MgCl_2$. Samples without $MgCl_2$ (–) were used as negative controls. *B*, strand transfer time course experiment. Reaction mixtures containing Fh8-IN (5 μM) or Sso7d-IN (1 μM) were incubated at 30 °C for the indicated times with a Ty1 U3 probe in the presence of 100 mM KCl and 5 mM $MgCl_2$. *C*, optimal salt conditions. Assays were performed at 30 °C for 30 min using 2 μM Sso7d-IN and a Ty1 U3 probe in the presence of 5 mM $MgCl_2$ and increasing amounts of KCl as indicated. *D*, metal cofactor requirement. *Left panel*, assays were performed at 30 °C for 15 min using a Ty1 U3 or Ty1 U5 probe in reaction mixtures containing Sso7d-IN (1 μM) in the presence of 100 mM KCl and increasing amounts of $MgCl_2$ or $MnCl_2$ as indicated. *Right panel*, a μM amount of Mn^{2+} is sufficient for Ty1 IN activity. Assays were performed as in the *left panel* but with a larger scale of $MnCl_2$ concentrations in the presence or not of 1.5 mM $MgCl_2$ as indicated.

presence of 1.5 mM of Mg^{2+} showed that the integration pattern was similar to that of Mn^{2+} alone, suggesting a clear preference for this metallic cofactor.

In conclusion, a recombinant Ty1 IN produced in *E. coli* with an Sso7d tag is stable and exhibits cation-dependent strand transfer activity with behavior identical to that reported for yeast-derived IN. In addition, the physiological concentration of Mn^{2+} is sufficient for an optimal integration activity of Ty1 IN.

Ty1 integrase substrate specificity

We extended our study to investigate the sequence specificity of Ty1 IN *in vitro*. No activity was detected for the Sso7d-IN $D_{154}A$ mutant confirming that a single mutation in the D,D(35)E triad motif of the catalytic domain was sufficient to completely abolish the enzymatic activity (Fig. 4*A*, INd). On the other hand, an N-terminal fragment of Sso7d-IN generated upon papain treatment and harboring the catalytic domain carried out strand transfer but less efficiently than full-length IN (Fig. S7*C*). Our data also confirmed that the activity of Ty1 IN was specific for Ty1 LTR substrates since neither NS or ST activity was detected when Sso7d-IN was incubated with a 30 pb random probe (Fig. 4*A*). To determine the minimal length of DNA that could be used in such assays, distinct Ty1 U5 probes with decreasing lengths from 30 to 10 bp have been designed (Fig. 4*B*). Upon incubation with Sso7d-IN, clear signals of integration were observed with the 25 pb and 15 pb substrates, whereas weaker signals could be detected with the 10 pb probe suggesting that this is the minimal size of the probe that can be used to assay Ty1 IN activity *in vitro*.

Previous studies reported low DNA sequence specificity for Ty1 IN and that alterations of the 3′ end of the probe interfered with IN activity (28). Our data indicated some preference for Ty1 U3 *versus* Ty1 U5 and for Mn^{2+} *versus* Mg^{2+}. In the case of HIV IN, the presence of large amounts of Mn^{2+} was shown to reduce specificity for cDNA termini (42). We thus tackled Ty1 IN substrate specificity in the presence of each

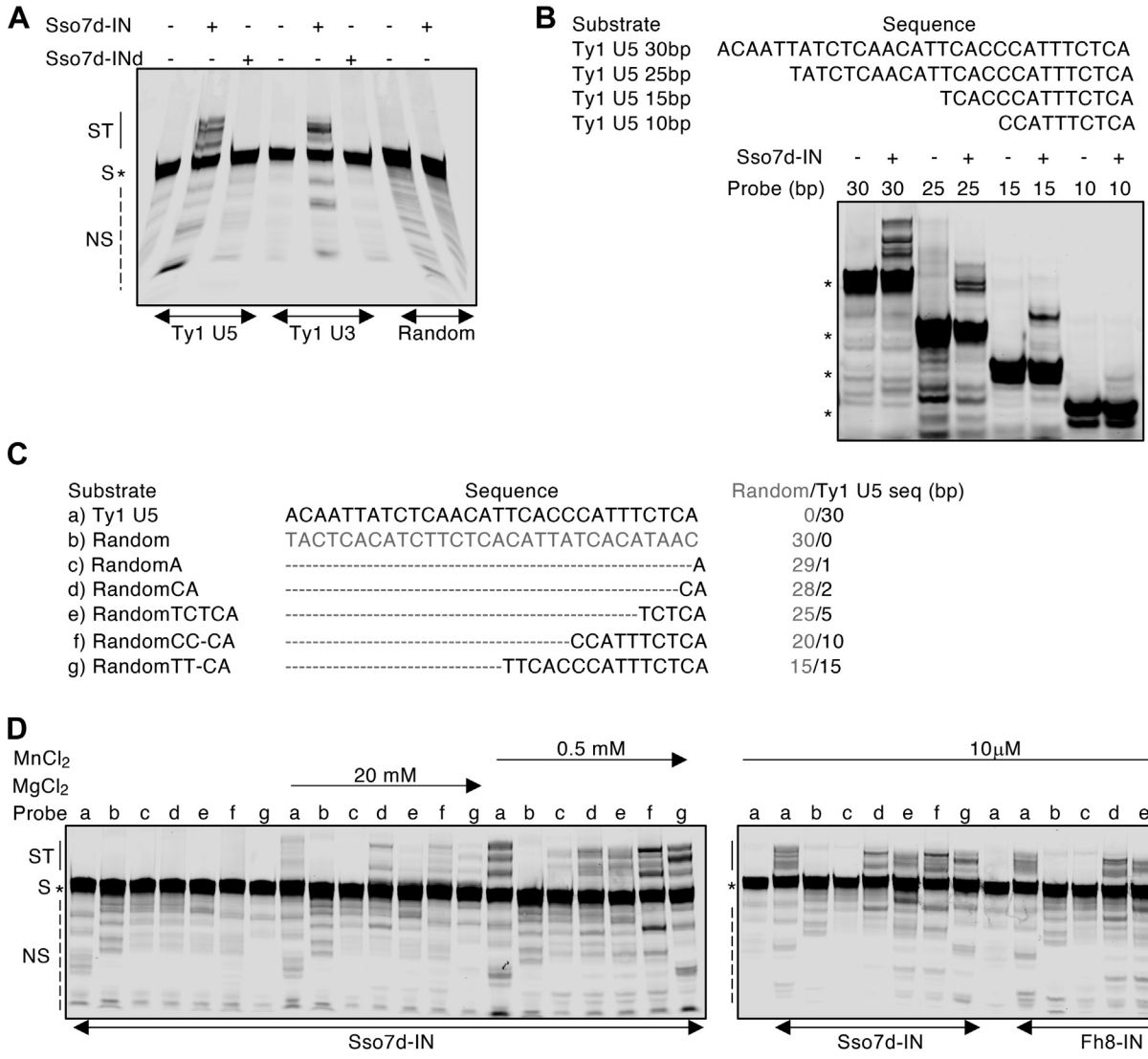

Figure 4. Ty1 IN substrate specificity. *A*, Ty1 IN activity is specific to DNA substrates mimicking Ty1 LTR. Strand transfer activity of Sso7d-IN (2 μM) was assayed using Ty1 U3, Ty1 U5 or a 30 bp random DNA sequence as substrates. A catalytically inactive mutant of Sso7d-IN (D_{154}A, INd, 2 μM) was used as a control. Strand transfer products (ST), substrate (S), and nonspecific endonuclease products (NS) are indicated. *B*, effect of DNA length on Ty1 IN activity. *Upper panel*, the sequences of the positive strand (represented from 5′ to 3′) of the three probes of 25, 15, or 10 pb derived from the 30 pb Ty1 U5 probe are shown. *Lower panel*, assays were performed using Sso7d-IN (2 μM) and Ty1 U5 probes with decreasing lengths as substrates. The positions of the different substrates are indicated by *asterisks*. Strand transfer products are located above the substrates. *C*, the sequences of the positive strand of different random sequences harboring at their 3′ end from 0 to 15 bases specific to the 3′ end of the Ty1 U5 probe are represented from 5′ to 3′. Only the sequences different from the 30 pb random sequence are shown. The ratio of random to specific bases is indicated for each probe. *D*, analysis of Ty1 IN substrate specificity at the 3′ end of Ty1 U5. Assays with the probes shown in panel *C* were performed using Sso7d-IN (2 μM) in the presence of 20 mM $MgCl_2$ or 0.5 mM $MnCl_2$ (*left panel*) or with Sso7d-IN (2 μM) or Fh8-IN (8 μM) in the presence of 10 μM $MnCl_2$ (*right panel*) as indicated.

metallic cofactor. First, we compared the sequences of the ends of the LTRs of all Ty1 elements present in *S. cerevisiae* genome (Fig. S8*A*). Among the 32 copies (43) found in *Saccharomyces* genome database (https://www.yeastgenome.org), 27 have two distinct LTRs, two share a common LTR, two harbor an uncomplete LTR, and 1 has only one defined LTR, resulting in 60 complete LTRs. Sequence alignment of the last 30 bp of each Ty1 U3 and U5 3′ ends for these 60 LTRs showed that all Ty1 LTRs end with the CA dinucleotide that is highly conserved in retrovirus and retrotransposon LTRs (Fig. S8, *B* and *C*). The 3′ end of U3 is much more conserved than that of U5 (14 out of 15 bases and nine out of 15 bases strictly conserved, respectively). To investigate the potential role of the sequence composition of the 3′ end of Ty1 U5 in substrate specificity, we performed oligonucleotide integration assays with various probes that harbored an increasing number of specific bases at the 3′ end (Fig. 4*C*) according to the consensus sequence present in all Ty1 U5 LTRs (Fig. S8*C*). The reactions were set up using standard concentration of metallic cofactors (Fig. 4*D*, left panel). Upon these conditions, the importance of the CA dinucleotide was confirmed since its presence at the end of the substrate was sufficient to detect a strand transfer activity independently of the metal cofactor used (Fig. 4*D*, left panel, probe d). With additional Ty1 U5 specific bases, the integration profile was modified and appeared to be similar to the complete Ty1 U5 probe in the

presence of 15 specific bases (Fig. 4*D*, left panel, compare probes e–g with probe a). These results show that the CA dinucleotide is sufficient for productive integration and that the substrate specificity is dependent of the nucleotide sequence of the acceptor DNA. Note that the fluorescent signals were different in the presence of Mn^{2+} or Mg^{2+}. In particular, some ST activity could be detected when the CA dinucleotide was substituted for an AA dinucleotide in the presence of Mn^{2+} (Fig. 4*D*, left panel probe c) as previously described (28). On the other hand, when the assay was performed in the presence of micromolar concentration of Mn^{2+}, no integration signal was observed with the AA dinucleotide, and the CA dinucleotide at the end of the probe was the minimal requirement for integration activity (Fig. 4*D*, right panel). Similar results were obtained with both Sso7d-IN and Fh8-IN suggesting that the N-terminal tag of the recombinant integrases did not interfere with substrate specificity (Fig. 4*D*, right panel).

All together, our data indicate that *in vitro*, Ty1 IN integration activity is essentially identical whether Mn^{2+} or Mg^{2+} is present in physiological concentration.

Ty1 integrase disintegration activity

In vitro, integrases are able to catalyze a disintegration activity that corresponds to a reverse reaction of the strand transfer (31). To test this activity, we used a single oligonucleotide previously designed to assay yeast Ty1 IN disintegration activity (28) that, upon correct annealing, could form a three-armed dumbbell, one of which corresponds to the terminal 6 bp of Ty1 U3 (Fig. 5*A*) or U5 (Fig. S9) LTR, and where two Msp I restriction sites are present in the target DNA. First, the reconstituted substrate was submitted to Msp I digestion to control the correct formation of the DNA structure (Fig. 5*B*). Although the Msp I digestion of the DNA fragment was not complete, two fragments of respectively 28 and 17 bases were detected indicating that most of the substrate was correctly folded. Unsurprisingly, the less accessible Msp I site generated much less 17-base fragments that the other one. In this assay, the strand cleavage should occur precisely at the junction between the LTR arm and the target DNA sequence, releasing the LTR end DNA, whereas the nick present in the target DNA fragment should be covalently joined. In the presence of Sso7d-IN, disintegration of the substrate was observed with the appearance of short fluorescent DNA fragments of the expected

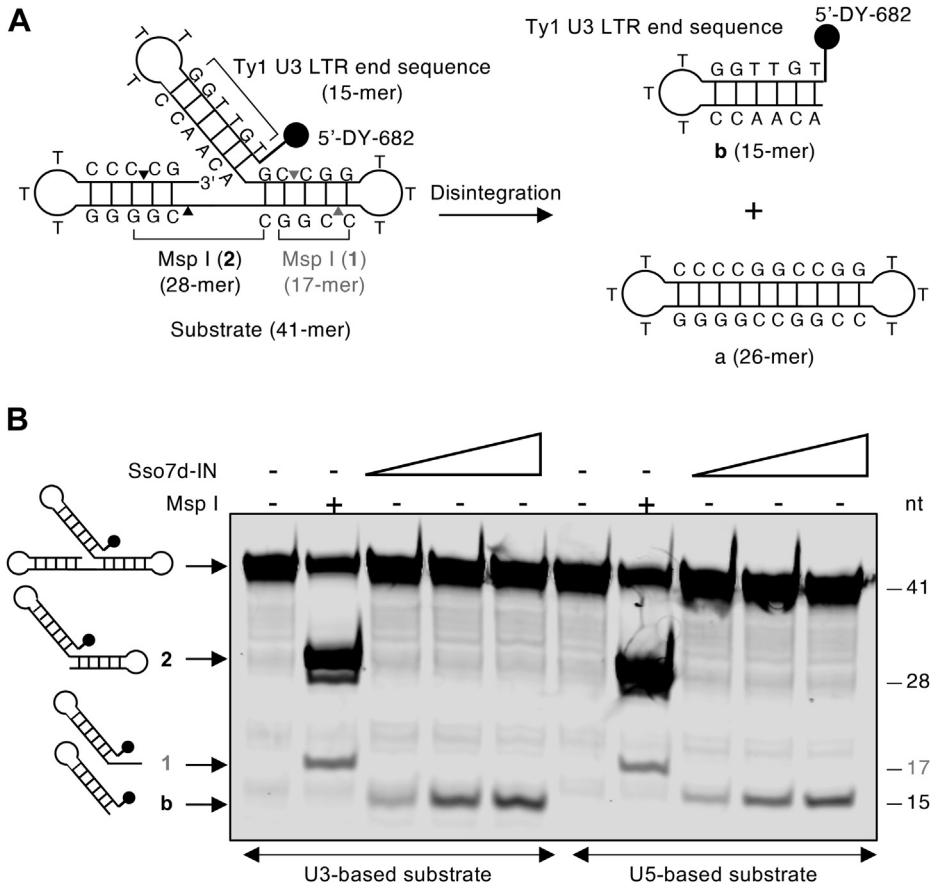

Figure 5. Ty1 IN disintegration activity. *A*, scheme of the disintegration assay using a Ty1 U3-based dumbbell-shaped substrate. The correct folding of the 41-mer labeled substrate results in the presence of two Msp I restriction sites. Upon Msp I digestion, the most accessible restriction site (*black arrows*) generates a 28-mer labeled DNA fragment, whereas a 17-mer labeled DNA fragment is produced from the other one (*gray arrows*). Disintegration activity of Ty1 IN results in the production of two DNA fragments, a and b as indicated. Only the substrate and the b fragment, labeled with the fluorescent dye, could be detected by fluorescent imaging. As shown in Fig. S9, disintegration assay and Msp I digestion using a 41-mer Ty1 U5-based dumbbell-shaped substrate generate labeled DNA fragments of the same sizes. *B*, increasing amounts (0.6 μM, 1.3 μM, and 1.9 μM) of Sso7d-IN were incubated at 30 °C for 30 min with a Ty1 U3- or U5-based dumbbell-shaped substrate as indicated. To control the correct folding of both substrates, Msp I digestion was performed in the absence of IN. Fluorescent DNA fragments were detected after separation on a 20% denaturing polyacrylamide gel. Predicted structures and sizes of the substrates and of the products of Msp I digestion and disintegration assay are indicated.

size (Fig. 5*B*, 15-base). More 15-base products were detected with increasing amounts of integrase and the activity was similar using the U3- or U5-based substrates.

Ty1 integrase concerted integration assay

Since Sso7d-IN was able to catalyze both strand transfer and disintegration activities, we next assayed its ability to mediate concerted integration. In this assay, as depicted in Figure 6*A*, a short 22 bp labeled double-stranded DNA mimicking Ty1 U5 LTR end is used as donor DNA (dDNA) to be integrated into a target DNA plasmid (pUC19). The insertion of a single dDNA fragment into one strand of the plasmid generates a circular product referred to as half-site concerted integration (hs.i) that comigrates with the open circular plasmid. On the other hand, the integration of one dDNA to each strand of the plasmid forms a linear product referred to as concerted integration (c.i) that migrates as the linearized form of the plasmid. In the presence of Sso7d-IN, both hs.i and c.i were easily detected, whereas no activity was observed in the presence of the catalytically defective IN_d mutant (Fig. 6*B*). The activity was very efficient since the supercoiled form of the plasmid completely disappeared. Interestingly, the full-length IN and INt, obtained after papain treatment, had similar activity (Fig. 6*B*), indicating that the C-terminal disordered domain of Ty1 IN was not directly involved in concerted integration. Further analysis indicated that the activity was maximal with a protein/dDNA ratio of about 8:1 (Fig. 6*C*). However, upon these experimental conditions, the half-site integration was clearly favored. Thus, we conducted additional assays to optimize the conditions resulting in more efficient concerted (full-site) integration (Fig. S10, for details, see Experimental procedures). The use of a longer dDNA (30 pb) in the presence of 5% glycerol but without PEG clearly favored the full-site integration. We also addressed the influence of the metallic cofactor in the concerted integration assay using increasing amounts of each divalent cation with either the Ty1 U3 or U5 substrate and confirmed the results obtained in the oligonucleotide integration assays (Fig. 3*D*).

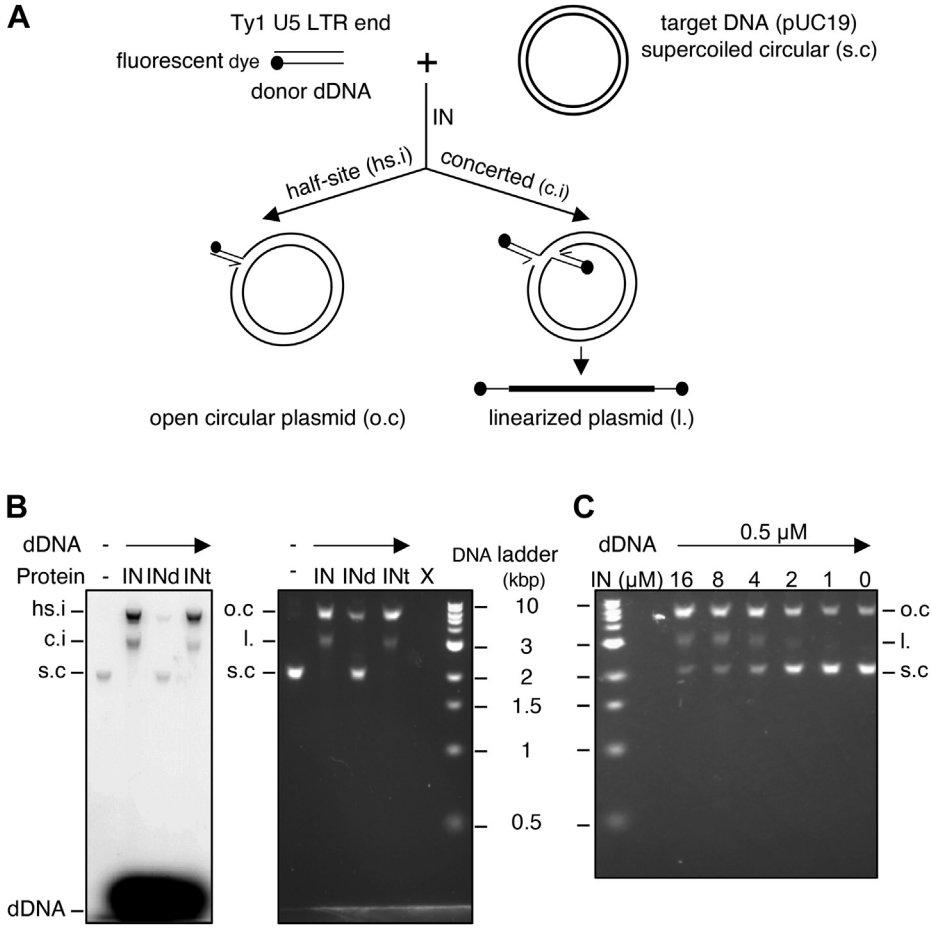

Figure 6. Concerted integration assay. *A*, schematic of the concerted integration assay. In the presence of a circular supercoiled (s.c) plasmid target DNA (pUC19), a double-stranded labeled donor DNA (dDNA) mimicking the end of Ty1 U5 LTR and an active Ty1 IN, two integration products are generated. The integration of a single dDNA in the target DNA corresponds to half-site integration (hs.i). A nick is introduced in the target DNA that migrates as an open circular plasmid (o.c). Concerted integration (c.i) corresponds to the insertion of two dDNA in the target DNA that migrates as a linearized plasmid (l.). *B*, Ty1 IN efficiently catalyzes concerted integration. Sso7d-IN and the N-terminal protein purified after papain treatment (INt, Fig. S4*A*) were assayed using a 6-FAM labeled dDNA. The samples were loaded on an agarose gel (1.5%) that was visualized by either SYBR Safe staining (*right panel*) or a Typhoon fluorescence scanner (*left panel*). Open circular (o.c), linear (l.) and supercoiled (s.c) forms of the plasmid and integration products (hs.i and c.i) are indicated. Samples without IN or in the presence of IN-D_{154}A (INd), a catalytically inactive mutant, were used as negative controls. Positions of a DNA ladder bands are indicated. *C*, characterization of the optimal ratio of Sso7d-IN/dDNA for concerted integration. Assays were performed as in panel *B* with 0.5 µM dDNA and increasing amounts of integrase as indicated. The gel was visualized by SYBR Safe staining.

First, a clear preference for Ty1 U3 over Ty1 U5 substrate was detected independently of the metallic cofactor used (Fig. 7*A*). Second, 1.5 μM Mn^{2+} was sufficient to detect concerted integration products and increasing its concentration strongly stimulated the reaction (Fig. 7*A*, left panel). Millimolar concentrations of Mg^{2+} were required to detect similar activity (Fig. 7*A*, right panel). In the presence of physiological concentration of metallic cofactor (1.5 μM of Mn^{2+} or 1.5 mM of Mg^{2+}), concerted integration was very efficient, whereas single site integration was hardly detected.

Finally, three distinct random probes ending with AC, AA, or CA dinucleotide were assayed to determine the substrate specificity (Fig. 7*B*). In the presence of physiological concentrations of Mn^{2+} or Mg^{2+}, concerted integration was detected only with the random probe harboring the conserved CA dinucleotide (Fig. 7*B*, compare lane 3 with lanes 1 and 2). The reaction was strongly enhanced and much less specific with millimolar amounts of Mn^{2+} since the integration products were also detected with the random probe ending with a dinucleotide AA (Fig. 7*B*, compare lanes 6, 7 and 8). These data correlate with a previous *in vivo* study showing that upon overexpression of Ty1 IN, DNA fragments harboring the terminal dinucleotides 5′-TG…CA-3′ could be used as substrates for integration in the yeast genome (44).

Altogether, our data indicate that *in vitro* (i) recombinant Ty1 IN catalyzes concerted integration, (ii) the C-terminal disordered region is not required for Ty1 IN activity, (iii) Ty1 IN integration activity is detected with a similar profile, efficiency, and specificity in the presence of physiological amounts of Mn^{2+} or Mg^{2+}.

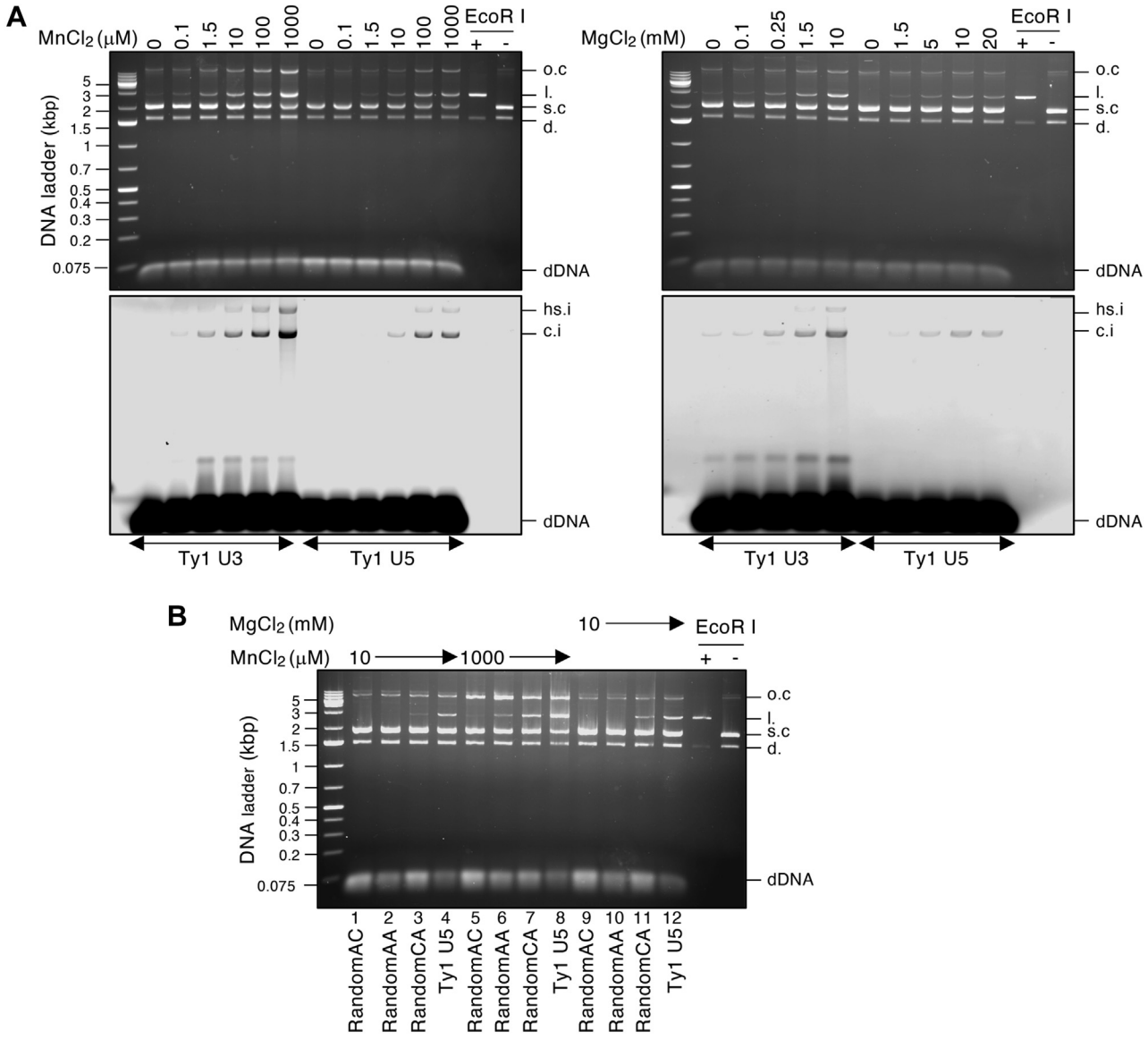

Figure 7. Limited amounts of $MnCl_2$ are sufficient for concerted integration. *A*, concerted integration assays were carried out as in Figure 6 with increasing amounts of $MnCl_2$ (*left panel*) or $MgCl_2$ (*right panel*) except that Ty1 U3 or U5 dDNA was labeled with the DY-682 fluorescent dye. The gel was visualized by ethidium bromide staining (*upper panel*) or with an Odyssey CLx imaging system (*lower panel*). The target pUC19 DNA was digested with EcoR I as indicated to visualize the supercoiled circular (s.c) and linear (l.) forms of the plasmid. An additional denatured form (d.) of the plasmid (65) was observed. *B*, specificity of the concerted integration. Assays were carried out as in panel *A* in the presence of Ty1 U5 dDNA or random probes harboring AA, AC, or CA dinucleotide (as in Fig. 4*C*) at their 3′ end as indicated. The gel was visualized by ethidium bromide staining.

Discussion

In this study, we have produced an active highly purified homogenous Ty1 IN. Biophysical analyses reveal an N-terminal half folded region including the three domains common to all retroviral integrases together with an extra folded region after the CTD followed by a large intrinsically disordered domain. Biochemical analyses demonstrate that bacterial recombinant Ty1 IN is functionally similar to yeast ectopic Ty1 IN and suggest that the metallic cofactor Mn^{2+} may play a role in Ty1 integration *in vivo*.

The in-depth understanding of retrotransposon integration mechanisms through structural and biochemical studies has been burdened by the absence of comprehensive protocols for the production of high yields of active protein samples of high purity and homogeneity. We show that this is achievable for Ty1 IN with the potent and versatile *E.coli* expression system by including an Sso7d solubility tag and optimizing the purification protocol to prevent protein aggregation and nucleic-acids contaminations. The highly purified Ty1 Sso7d-IN is a monomer as observed for HIV Sso7d-IN (34). However, in yeast cells, Ty1 intasome is assembled in the VLPs and thus should not exist as a monomer. The structural analysis of HIV intasomes assembled with Sso7d-IN corresponds to the conserved intasome core present in all intasome structures determined to date (45), suggesting that this tag does not affect the multimerization that occurs during the assembly of intasomes. We can speculate that it should be the same for the multimerization of Ty1 IN. Further structural studies of Ty1 intasome will certainly allow to clarify this point.

Sso7d-IN shows a behavior in terms of integration and disintegration activities similar to that reported for IN produced in yeast and for the first time, a concerted integration reaction could be directly detected without a PCR amplification step (46, 47) or southern blotting (14) from a yeast retroelement integrase. These results pave the way for further biochemical and structural studies aiming to better understand Ty1 retrotransposon integration mechanisms.

Although Mg^{2+}, more abundant in the cells, is considered to be the physiological metallic cofactor of integrases, we showed that the low concentration of Mn^{2+} present in yeast cells (41) was sufficient for Ty1 IN activity *in vitro* suggesting that Mn^{2+} might play a role in Ty1 integration *in vivo*. The three residues that compose the highly conserved D,D(35)E motif form the basis of the active site of the nucleotidyltransferase superfamily (48) that could be bound by Mg^{2+} or Mn^{2+} as previously described (49, 50). Our data suggest that Ty1 IN active site has a higher apparent affinity for Mn^{2+} compared with Mg^{2+} since 1000-fold less manganese is sufficient for Ty1 IN activity. On the other hand, the strong differences in Ty1 IN activity detected in the presence of low or high Mn^{2+} concentration indicate that this metallic cofactor could bind to the two metal-binding sites but with different affinity. At low concentration of Mn^{2+}, one site could be favored, whereas in the presence of a large amount of Mn^{2+}, the occupation of both sites might interfere with Ty1 IN activity. Due to the 1000-fold excess of free Mg^{2+} over Mn^{2+} in the cell, we can speculate that *in vivo* most metal-binding sites may be occupied by Mg^{2+} in normal growth conditions.

Mn^{2+} is present at much lower concentration than other metals (for a review, (51)) but a previous analysis revealed that significant changes in the level of this metallic cofactor could take place *in vivo* (41). A set of genes is involved in the regulation of Mn^{2+} abundance in yeast (52) and the inactivation of several of them impacts the cellular level of Mn^{2+} indicating that conditions that could modulate the use of Mn^{2+} by Ty1 IN exist *in vivo*. Among these genes, *PMR1*, a transporter that balances Ca^{2+} and Mn^{2+} between the Golgi and the cytosol, was found to play a central role in Mn^{2+} homeostasis (52). An accumulation of Mn^{2+} was observed in *pmr1Δ* cells (53). Interestingly, *pmr1Δ* cells also presented a severe defect in Ty1 retrotransposition due to a strong inhibition of Ty1 reverse transcriptase (54). Our data suggest that although Mg^{2+} could be considered as the physiological cofactor upon normal growth conditions, a possible role for Mn^{2+} in the control of Ty1 retrotransposition *in vivo* should be investigated. In particular, it would be interesting to assay Ty1 retrotransposition in conditions that balance Mn^{2+} or Mg^{2+} concentrations in the cell.

Transposable elements (TEs) are a source of genetic variability in yeasts and potentially instrumental in the evolution of their genome and the adaptation to the environment (55). Some TEs have maximized their chance of propagation because of an accurate control of integration events specifically targeting non-deleterious regions of the genome. This is the case for Ty1, which targets its integration upstream of Pol III-transcribed genes through direct interactions between IN and Pol III (11, 12). The architecture that we reveal with our SAXS model and limited proteolysis suggests that, to achieve this specificity, Ty1 IN has evolved maintaining the three conserved integrase domains in the N-terminus and adding a large intrinsically disordered region at the C-terminus that includes the Ty1 targeting domain. Although Ty1 IN is supposed to be in a multimeric form *in vivo* (13, 14), highly purified recombinant Ty1 IN is monomeric. The SAXS calculated volume of this construct reveals that the integrase central globular shape matches the size and shape of the PFV integrase structure (20), which underscores the highest homology with Ty1 IN. This observation suggests that the monomeric and oligomeric forms of IN maintain a similar spatial arrangement of the three conserved domains, despite the flexibility of the inter-domain linkers. In addition, we observed additional globular volumes at both sides of the central region, which could reflect possible preferential conformations of the disordered C-terminus and of the Sso7d tag.

Interestingly, the folded N-terminal half and the intrinsically disordered C-terminal regions of Ty1 IN appear to act independently. We showed that the structure of the first half of the protein is similar to that of retroviral integrases and is sufficient to carry out integration. On the other hand, a single mutation in the targeting domain located in the C-terminus of IN that disrupts the interaction with Pol III redirects Ty1 insertion in the subtelomeric regions (10). Moreover, the presence of the Ty1 IN C-terminus region in a related retroelement is sufficient to redirect its integration to Pol III-transcribed genes (11). All these data support that the engineering of the Ty1 IN C-terminus region is a good strategy for the design of retroviral vectors with tuneable integration site specificity.

Experimental procedures

Plasmids and primers

All plasmids and primers used in this study are reported in Supporting information (Tables S1 and S2).

Plasmids construction

A codon-optimized DNA fragment (Fig. S1) was synthesized and subcloned in pET17b by GenScript to generate the pET17b-6H-Fh8-Sso7d-IN-Ctag plasmid. The two primers FW6HTEVKas and RW6HTEVKas were phosphorylated and then annealed before to be subcloned in the pET17b-6H-Fh8-Sso7d-IN-Ctag plasmid digested by Nde I and Kas I restriction enzymes to generate the pET17b-6H-IN-Ctag plasmid. The plasmids pET17b-6H-Fh8-IN-Ctag and pET17b-6H-Sso7d-IN-Ctag were generated using the same strategy with the FWNheTEV and RWNheTEV or the FWNde6H and RWNde6H primers, respectively. Ty1 IN was cloned in a pET9 6H-SUMO-vector (56) using the In-fusion cloning technique and recloned with the tag in a pFastBac expression vector using In-fusion cloning technique again.

Protein expression and purification of recombinant Ty1 IN

The plasmids encoding Ty1 IN were transformed into T7 express *E.coli* competent cells (New England Biolabs) and streaked on LB (Luria–Bertani) broth agar plates supplemented with ampicillin, chloramphenicol, and 1% glucose. Transformants were inoculated in the same medium overnight at 30 °C, diluted to 0.05 A_{600} in fresh 1000 ml of TB terrific broth medium with ampicillin and chloramphenicol without glucose, and grown at 37 °C to an A_{600} = 0.6. Cultures were transferred at 4 °C for 1 h, then at 16 °C. After 30 min, protein expression was induced by adding 50 μM isopropyl 1-thio-β-D-galactopyranoside (IPTG), and the cells were grown overnight at 16 °C. Cells were harvested at 1100*g* for 15 min at room temperature, washed with PBS 1×, and weighed. The pellets were resuspended in two volumes (2 ml/g of pellet) of cold solubilization buffer A (10% glycerol, 20 mM Tris-HCl pH7.8, 1 M NaCl, 0.05% Nonidet P40 (NP40), 20 mM imidazole) complemented with 5 mM mercapto-2-ethanol, 0.1 mM PhenylMethylSulfonyl Fluoride (PMSF), 1 μg/ml of lysozyme, benzonase (1 μl at 250,000 U/ml per 10 ml of extract), and complete protease inhibitor cocktail, before being incubated for 10 min at 4 °C on a wheel. The extracts were then lysed by sonication (Q700 sonicator, Qsonica) on ice for five cycles (output amplitude 10%, 5 s ON, 40 s OFF). Additional benzonase was then added to the extracts (1 μl at 250,000 U/ml per 10 ml of extract), which were incubated on a wheel for 30 min at 4 °C before to be ultracentrifuged at 118,000*g* at 4 °C for 30 min. The supernatants were collected on separated tubes and the lysate concentration was determined using a Bradford assay (Biorad). The solubilized extracts were loaded onto a Ni^{2+}-charged HisTrap column (Cytiva) previously equilibrated with buffer A. Proteins were eluted with buffer B (buffer A complemented to 250 mM imidazole) using the AKTA purifier system (Cytiva). The fractions containing IN identified by Coomassie blue staining after SDS-PAGE were pooled and dialyzed against buffer C (buffer A without imidazole). Aggregated proteins were removed by gel filtration on a Superdex 200 increase 10/300 GL equilibrated with buffer C. If necessary, the samples were loaded on an Amicon centrifugal concentrator (EMD Millipore) to obtain a final protein concentration of 2 to 5 μg/μl, flash-frozen in liquid nitrogen, and stored at −80 °C. The purified recombinant Ty1 IN was stored in high-salt buffer (1 M NaCl). Note that in the preliminary test of expression (Fig. 1*A*), neither NP40 nor benzonase was added to the buffers and the ultracentrifugation was performed right after the sonication.

Ty1 IN was also expressed in High Five cells (Thermo Fisher) using baculoviruses. The construct includes an N-terminal Histidine tag followed by a SUMO tag and a TEV protease-cleaving site. Recombinant bacmids and viruses were produced in YFP-DH10 Bac cells (57) and Sf21 cells (Thermo Fisher) respectively, according to the manufacturer's protocols. For protein expression, High Five cells in suspension at a density of 0.5 to 1.0 × 10^6 ml^{-1} were infected with baculoviruses at 2% volume of the culture. Cells were harvested at 2 to 3 days post cell arrest, when Ty1 IN expression was ascertained to be the highest by Western blotting.

Assays for Ty1 IN activity

All the oligonucleotides labeled with a DY-682 fluorescent dye were purchased from Eurofins Genomics. The 6-FAM labeled primer was purchased from Thermo Fisher Scientific. To prepare double-stranded DNA substrates, two complementary oligonucleotides were annealed by boiling on a thermomixer at 95 °C for 5 min in the presence of 1 M NaCl and 20 mM Tris-HCl pH8 before to stop the heat and let the temperature decrease to reach room temperature. For oligonucleotide integration assays, the indicated amounts of Fh8-IN or Sso7d-IN were incubated with 20 pmol of fluorescent DNA substrates on ice in integration buffer (10 mM Tris-HCl pH7.5, 0.8% glycerol, 5% PEG 8000, 5 mM $MgCl_2$, 0.1 mM DTT) in the presence of the indicated concentrations of KCl, $MnCl_2$, or $MgCl_2$ in a 20 μl reaction volume. Standard assays contained 125 mM KCl (taking into account the 1 M NaCl concentration of Ty1 IN protein fractions), 0.5 mM $MnCl_2$ or 5 mM $MgCl_2$. After 15 to 30 min at 30 °C, the reaction was stopped by the addition of 20 μl of loading buffer (20 mM Tris-HCl pH7.5, 8 M urea, 0.5% SDS, 2× Tris-Borate-EDTA). The sample was boiled 5 min at 95 °C and loaded on a 20% polyacrylamide/8 M urea gel. After electrophoresis, the DNA was visualized using an Odyssey CLx fluorescence near-infrared imaging system at 700 nm.

Disintegration assays were performed under the same conditions but with a specific Ty1 U3 or U5 -based dumbbell-shape substrate. Optimized concerted integration assays were performed by incubating the indicated amounts of Fh8-IN or Sso7d-IN with 20 pmol of fluorescent DNA substrates in the concerted integration buffer (10 mM Tris-HCl pH7.5, 5% glycerol, 4 μM $ZnCl_2$, 0.1 mM DTT) in the presence of the indicated KCl, $MnCl_2$, or $MgCl_2$ concentrations in a 20 μl reaction volume. After 30 min on ice, 1 μl of the target plasmid DNA (pUC19) at 0.3 μg/μl was added and the reaction was transferred to 30 °C for 1 h.

The reaction was stopped by the addition of 5 µl of the stop reaction buffer (0.25 M EDTA, 1% SDS) and 1 µl of pronase (10 mg/ml, Sigma) and transferred at 37 °C for 1 h. The samples were subjected to electrophoresis in an agarose gel (1.5%) supplemented with ethidium bromide in a 1× TBE buffer to allow a direct visualization of the DNA using a transilluminator. The DY-682 fluorescent dye was next detected using an Odyssey CLx fluorescence near-infrared imaging system at 700 nm.

Size-exclusion chromatography (SEC) coupled with small-angle X-ray scattering (SEC-SAXS)

SEC-SAXS data for Ty1 IN sample were collected at the SWING beamline, Synchrotron SOLEIL. In line SEC-SAXS was performed using an HPLC equipped with the BioSec3-300 column and 30 µl of Sso7d-IN (16 mg/ml) was loaded at a flow rate of 300 µl/min before data collection. Scattering images with an exposure of 1 s were collected every 0.01 s using an Eiger 4M (Dectris) detector at a photon energy of 12 keV (λ = 1.033 nm) and sample to detector distance of 2 m. The experiment was performed at 10 °C. Beamline SWING-specific software was used for subtracting background buffer signal from protein signal, with a previous averaged data. Data were analyzed using the ATSAS 3.0.2 (58) and RAW 2.0.2 (59). *Ab initio* reconstructions were generated with the program DAMMIF (60). Twenty independent DAMMIF runs were averaged using the program DAMAVER (61) and clustered by the program DAMCLUST (62). The resulting model was refined using DAMMIN (63).

Size-exclusion chromatography coupled with multiangle light scattering (SEC-MALS)

SEC was performed on a HPLC system (1200 pump, Agilent Technologies) using a precalibrated Superdex 200 increase 10/300 GL column (Cytiva) run in 20 mM Tris-HCl pH8, 1 M NaCl, 10% glycerol, and 5 mM DTT at 0.5 ml/min. MALS, UV spectrophotometry, QELS, and RI were achieved with a DAWN EOS detector (Wyatt Technology), a SpectraSYSTEM UV2000 UV/VIS detector (Thermo Finnigan), a WYATT QELS module (Wyatt Technology), and an Optilab rEX Refractive Index detector (Wyatt Technology), respectively. Mass and hydrodynamic radius calculation was done with ASTRA 5.3 software (Wyatt Technology) using a dn/dc value of 0.175 ml/g, calculated as described (64).

Papain digestion

Purified recombinant Sso7d-IN (5 µM) was incubated with 5, 10, 20, or 50 nM of papain (from *Carica papaya* latex, Sigma-Aldrich) respectively, for 2 h at 4 °C in buffer containing 20 mM Tris-HCl pH8, 1 M NaCl, 2 mM tris(2-carboxyethyl) phosphine (TCEP), 10% glycerol, and 5 mM L-cysteine (Sigma-Aldrich). The reaction was stopped by adding E-64 (Sigma-Aldrich) to reach a final concentration of 10 µM. After 5 min incubation on ice, the mixture was subjected to SDS-PAGE. To purify INt, sample that has a 1:250 ratio of papain to Sso7d-IN was loaded to a Superdex 200 increase 10/300 GL gel filtration column equilibrated with 20 mM Tris-HCl pH8, 1 M NaCl, 2 mM TCEP, and 10% glycerol. The flow rate was 0.3 ml/min and the fraction size was 0.5 ml. Fractions were then subjected to SDS-PAGE. Fraction C7 was sent to mass spectrometry analysis and used for INt integration assays.

MALDI-TOF-TOF mass spectrometry

Protein masses were determined on solution samples. Ten microliter of protein at 0.5 mg/ml was desalted on ZipTip C8 and eluted with 2 µl of a saturated solution of sinapinic acid in 0.1% TFA/CH3CN (50:50 v/v). One microliter was then spotted on the target and analyzed by MALDI-TOF-TOF on a Ultraflex III spectrometer (Bruker Daltonics) controlled by the Flexcontrol 3.0 package (Build 51) and operated in the linear mode, using a maximum accelerating potential of 25 kV and a 20,000 to 80,000 m/z range (LP_66 kDa Method). The laser frequency was fixed to 100 Hz and ca. 1000 shots per sample were cumulated. Four external standards (Protein Calibration Standard II, Bruker Daltonics) were used to calibrate each spectrum to a mass accuracy within 100 ppm. Peak picking was performed using the FlexAnalysis 3.0 software with an adapted analysis method. Parameters used were: centroid peak detection algorithm, S/N threshold fixed to 5, and a quality factor threshold of 30.

Protein identification by fingerprinting

All SDS-PAGE bands were cut and digested by trypsin. Mass spectrometry analyses were carried out by LC-MSMS using a LTQ-Velos-Orbitrap online with a nanoLC Ultimate 3000 chromatography system. The peaks obtained have been compared with the theoretical digestion peak list of the expected protein added to a database containing 20,150 human sequences, 4306 *E.coli* sequences.

N-terminal Edman sequencing

Twenty-five microliter (122 pmol) of a 0.259 mg/ml solution (fraction C7, Fig. S4*A*) was titrated with TFA to reach pH=1. The protein was transferred on a polyvinylidene fluoride membrane using the ProSorb Applied system. To remove the salt, the membrane was washed with 100 µl of 0.1% TFA five times. After that, dried membrane was placed in the cartridge of the Shimadzu PPSQ 31B sequencer. Seven Edman degradation cycles were performed.

Data availability

All data are contained within the manuscript and accompanying supporting information.

Supporting information—This article contains supporting information (66).

Acknowledgments—We thank Dr Patrick Fourquet (CRCM, Marseille Proteomics (MaP)) for mass spectrometry analyses. The MaP proteomic facility (marseille-proteomique.univ-amu.fr) is supported by « Infrastructures Biologie Santé et Agronomie » (IBiSA); « Plateformes Technologiques Aix-Marseille »; « Cancéropôle PACA »; « Région Sud Provence-Alpes-Côte d'Azur »; « Institut Paoli-Calmettes »; « Centre de Recherche en Cancérologie de Marseille »

(CRCM); « Fonds Européen de Développement Régional » (FEDER) and « Plan Cancer ».
This work was supported by grants from the Agence Nationale de la Recherche through the generic call project ANR-17-CE11-0025. This work has been supported by the Bettencourt Shueller Foundation (FBS: AGDI-531281) and an ATIP-avenir grant (CNRS/INSERM).

Author contributions—C. C., J. R., and J. A. conceptualization; P. Q. N., C. C., E. R., G. B., C. G., J. R., and J. A. data curation; P. Q. N., C. C., E. R., G. B., C. G., J. R., and J. A. formal analysis; C. F.-T., P. L., J. R., and J. A. funding acquisition; P. Q. N., C. C., E. R., C. G., J. R., and J. A. investigation; P. Q. N., C. C., E. R., C. G., C. F.-T., and J. A. methodology; P. L. project administration; J. R. and J. A. supervision; G. B., C. F.-T., P. L., J. R., and J. A. validation; C. C., J. R., and J. A. writing—original draft; C. C., J. R., and J. A. writing—review and editing.

Funding and additional information—P. Q. N. was supported by the ARC Foundation with a fourth-year PhD fellowship.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CCD, catalytic core domain; CTD, carboxy terminal domain; dDNA, donor DNA; IN, integrase; LTR, long terminal repeat; NTD, N-terminal domain; NS, nonspecific nuclease activity; PFV, Prototype foamy virus; Pol II, RNA polymerase II; Pol III, RNA polymerase III; RT, Reverse transcriptase; SAXS, small-angle X-ray scattering; ST, strand transfer activity; TEV, tobacco Etch virus; VLP, virus-like particle.

References

1. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., *et al.* (1996) Life with 6000 genes. *Science* **274**, 546, 563-567
2. Engel, S. R., Dietrich, F. S., Fisk, D. G., Binkley, G., Balakrishnan, R., Costanzo, M. C., Dwight, S. S., Hitz, B. C., Karra, K., Nash, R. S., Weng, S., Wong, E. D., Lloyd, P., Skrzypek, M. S., Miyasato, S. R., *et al.* (2014) The reference genome sequence of Saccharomyces cerevisiae: Then and now. *G3 (Bethesda)* **4**, 389–398
3. 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
4. Carr, M., Bensasson, D., and Bergman, C. M. (2012) Evolutionary genomics of transposable elements in Saccharomyces cerevisiae. *PLoS One* **7**, e50978
5. Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985) Ty elements transpose through an RNA intermediate. *Cell* **40**, 491–500
6. Curcio, M. J., Lutz, S., and Lesage, P. (2015) The Ty1 LTR-retrotransposon of budding yeast, Saccharomyces cerevisiae. *Microbiol. Spectr.* **3**. MDNA3-0053-2014
7. Ji, H., Moore, D. P., Blomberg, M. A., Braiterman, L. T., Voytas, D. F., Natsoulis, G., and Boeke, J. D. (1993) Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell* **73**, 1007–1018
8. 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
9. Mularoni, L., Zhou, Y., Bowen, T., Gangadharan, S., Wheelan, S. J., and Boeke, J. D. (2012) Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res.* **22**, 693–703
10. Baller, J. A., Gao, J., Stamenova, R., Curcio, M. J., and Voytas, D. F. (2012) A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. *Genome Res.* **22**, 704–713
11. Asif-Laidin, A., Conesa, C., Bonnet, A., Grison, C., Adhya, I., Menouni, R., Fayol, H., Palmic, N., Acker, J., and Lesage, P. (2020) A small targeting domain in Ty1 integrase is sufficient to direct retrotransposon integration upstream of tRNA genes. *EMBO J.* **39**, e104337
12. 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) Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposon integration. *Science* **348**, 585–588
13. Moore, S. P., Rinckel, L. A., and Garfinkel, D. J. (1998) A Ty1 integrase nuclear localization signal required for retrotransposition. *Mol. Cell. Biol.* **18**, 1105–1114
14. Moore, S. P., and Garfinkel, D. J. (2009) Functional analysis of N-terminal residues of ty1 integrase. *J. Virol.* **83**, 9502–9511
15. Wilhelm, F. X., Wilhelm, M., and Gabriel, A. (2005) Reverse transcriptase and integrase of the Saccharomyces cerevisiae Ty1 element. *Cytogenet. Genome Res.* **110**, 269–287
16. Wilhelm, M., and Wilhelm, F. X. (2005) Role of integrase in reverse transcription of the Saccharomyces cerevisiae retrotransposon Ty1. *Eukaryot. Cell* **4**, 1057–1065
17. Wilhelm, M., and Wilhelm, F. X. (2006) Cooperation between reverse transcriptase and integrase during reverse transcription and formation of the preintegrative complex of Ty1. *Eukaryot. Cell* **5**, 1760–1769
18. Wilhelm, M., Boutabout, M., and Wilhelm, F. X. (2000) Expression of an active form of recombinant Ty1 reverse transcriptase in Escherichia coli: A fusion protein containing the C-terminal region of the Ty1 integrase linked to the reverse transcriptase-RNase H domain exhibits polymerase and RNase H activities. *Biochem. J.* **348 Pt 2**, 337–342
19. Kenna, M. A., Brachmann, C. B., Devine, S. E., and Boeke, J. D. (1998) Invading the yeast nucleus: A nuclear localization signal at the C terminus of Ty1 integrase is required for transposition *in vivo*. *Mol. Cell. Biol.* **18**, 1115–1124
20. Hare, S., Gupta, S. S., Valkov, E., Engelman, A., and Cherepanov, P. (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236
21. Ballandras-Colas, A., Brown, M., Cook, N. J., Dewdney, T. G., Demeler, B., Cherepanov, P., Lyumkis, D., and Engelman, A. N. (2016) Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358–361
22. Yin, Z., Shi, K., Banerjee, S., Pandey, K. K., Bera, S., Grandgenett, D. P., and Aihara, H. (2016) Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362–366
23. Passos, D. O., Li, M., Yang, R., Rebensburg, S. V., Ghirlando, R., Jeon, Y., Shkriabai, N., Kvaratskhelia, M., Craigie, R., and Lyumkis, D. (2017) Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355**, 89–92
24. Bhatt, V., Shi, K., Salamango, D. J., Moeller, N. H., Pandey, K. K., Bera, S., Bohl, T. E., Kurniawan, F., Orellana, K., Zhang, W., Grandgenett, D. P., Harris, R. S., Sundborger-Lunna, A. C., and Aihara, H. (2020) Structural basis of host protein hijacking in human T-cell leukemia virus integration. *Nat. Commun.* **11**, 3121
25. Barski, M. S., Minnell, J. J., Hodakova, Z., Pye, V. E., Nans, A., Cherepanov, P., and Maertens, G. N. (2020) Cryo-EM structure of the deltaretroviral intasome in complex with the PP2A regulatory subunit B56gamma. *Nat. Commun.* **11**, 5043
26. Pandey, K. K., Bera, S., Shi, K., Rau, M. J., Oleru, A. V., Fitzpatrick, J. A. J., Engelman, A. N., Aihara, H., and Grandgenett, D. P. (2021) Cryo-EM structure of the Rous sarcoma virus octameric cleaved synaptic complex intasome. *Commun. Biol.* **4**, 330
27. Moore, S. P., and Garfinkel, D. J. (1994) Expression and partial purification of enzymatically active recombinant Ty1 integrase in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 1843–1847
28. Moore, S. P., Powers, M., and Garfinkel, D. J. (1995) Substrate specificity of Ty1 integrase. *J. Virol.* **69**, 4683–4692
29. Braiterman, L. T., and Boeke, J. D. (1994) Ty1 *in vitro* integration: Effects of mutations in cis and in trans. *Mol. Cell. Biol.* **14**, 5731–5740

30. Braiterman, L. T., and Boeke, J. D. (1994) *In vitro* integration of retrotransposon Ty1: A direct physical assay. *Mol. Cell. Biol.* **14**, 5719–5730
31. Delelis, O., Carayon, K., Saib, A., Deprez, E., and Mouscadet, J. F. (2008) Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology* **5**, 114
32. Kirchner, J., and Sandmeyer, S. B. (1996) Ty3 integrase mutants defective in reverse transcription or 3'-end processing of extrachromosomal Ty3 DNA. *J. Virol.* **70**, 4737–4747
33. Costa, S. J., Coelho, E., Franco, L., Almeida, A., Castro, A., and Domingues, L. (2013) The Fh8 tag: A fusion partner for simple and cost-effective protein purification in Escherichia coli. *Protein Expr. Purif.* **92**, 163–170
34. Li, M., Jurado, K. A., Lin, S., Engelman, A., and Craigie, R. (2014) Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS One* **9**, e105078
35. Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., and Brochon, J. C. (2000) Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* **39**, 9275–9284
36. Schramm, A., Bignon, C., Brocca, S., Grandori, R., Santambrogio, C., and Longhi, S. (2019) An arsenal of methods for the experimental characterization of intrinsically disordered proteins - how to choose and combine them? *Arch. Biochem. Biophys.* **676**, 108055
37. Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Soding, J., Lupas, A. N., and Alva, V. (2020) Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinformatics* **72**, e108
38. Ishida, T., and Kinoshita, K. (2007) PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460–W464
39. Linding, R., Russell, R. B., Neduva, V., and Gibson, T. J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**, 3701–3708
40. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258
41. Eide, D. J., Clark, S., Nair, T. M., Gehl, M., Gribskov, M., Guerinot, M. L., and Harper, J. F. (2005) Characterization of the yeast ionome: A genome-wide analysis of nutrient mineral and trace element homeostasis in Saccharomyces cerevisiae. *Genome Biol.* **6**, R77
42. Esposito, D., and Craigie, R. (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* **17**, 5832–5843
43. Lesage, P., and Todeschini, A. L. (2005) Happy together: The life and times of Ty retrotransposons and their hosts. *Cytogenet. Genome Res.* **110**, 70–90
44. Friedl, A. A., Kiechle, M., Maxeiner, H. G., Schiestl, R. H., and Eckardt-Schupp, F. (2010) Ty1 integrase overexpression leads to integration of non-Ty1 DNA fragments into the genome of Saccharomyces cerevisiae. *Mol. Genet. Genomics* **284**, 231–242
45. Engelman, A. N., and Cherepanov, P. (2017) Retroviral intasomes arising. *Curr. Opin. Struct. Biol.* **47**, 23–29
46. Moore, S. P., and Garfinkel, D. J. (2000) Correct integration of model substrates by Ty1 integrase. *J. Virol.* **74**, 11522–11530
47. Qi, X., and Sandmeyer, S. (2012) *In vitro* targeting of strand transfer by the Ty3 retroelement integrase. *J. Biol. Chem.* **287**, 18589–18595
48. Nowotny, M. (2009) Retroviral integrase superfamily: The structural perspective. *EMBO Rep.* **10**, 144–151
49. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1996) The catalytic domain of avian sarcoma virus integrase: Conformation of the active-site residues in the presence of divalent cations. *Structure* **4**, 89–96
50. Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D., and Cherepanov, P. (2010) Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20057–20062
51. Cyert, M. S., and Philpott, C. C. (2013) Regulation of cation balance in Saccharomyces cerevisiae. *Genetics* **193**, 677–713
52. Yu, D., Danku, J. M., Baxter, I., Kim, S., Vatamaniuk, O. K., Vitek, O., Ouzzani, M., and Salt, D. E. (2012) High-resolution genome-wide scan of genes, gene-networks and cellular systems impacting the yeast ionome. *BMC Genomics* **13**, 623
53. Lapinskas, P. J., Cunningham, K. W., Liu, X. F., Fink, G. R., and Culotta, V. C. (1995) Mutations in PMR1 suppress oxidative damage in yeast cells lacking superoxide dismutase. *Mol. Cell. Biol.* **15**, 1382–1388
54. Bolton, E. C., Mildvan, A. S., and Boeke, J. D. (2002) Inhibition of reverse transcription *in vivo* by elevated manganese ion concentration. *Mol. Cell* **9**, 879–889
55. Chuong, E. B., Elde, N. C., and Feschotte, C. (2017) Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86
56. Reguera, J., Gerlach, P., Rosenthal, M., Gaudon, S., Coscia, F., Gunther, S., and Cusack, S. (2016) Comparative structural and functional analysis of bunyavirus and arenavirus cap-snatching endonucleases. *PLoS Pathog.* **12**, e1005636
57. Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., and Berger, I. (2010) New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol.* **172**, 45–54
58. Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., and Franke, D. (2021) ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355
59. Hopkins, J. B., Gillilan, R. E., and Skou, S. (2017) BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **50**, 1545–1553
60. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346
61. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864
62. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350
63. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886
64. Amartely, H., Avraham, O., Friedler, A., Livnah, O., and Lebendiker, M. (2018) Coupling multi angle light scattering to ion exchange chromatography (IEX-MALS) for protein characterization. *Sci. Rep.* **8**, 6907
65. Sayers, J. R., Evans, D., and Thomson, J. B. (1996) Identification and eradication of a denatured DNA isolated during alkaline lysis-based plasmid purification procedures. *Anal. Biochem.* **241**, 186–189
66. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641