Published in final edited form as:

Proteins. 2019 April ; 87(4): 313–325. doi:10.1002/prot.25653.

Rigidity and flexibility characteristics of DD[E/D]-transposases Mos1 and Sleeping Beauty.

Christopher M. Singer¹, Diana Joy¹, Donald J. Jacobs^{1,2}, and Irina V. Nesmelova^{1,2,*}

¹Department of Physics and Optical Science, University of North Carolina, Charlotte, NC 28223, USA.

²Department of Center for Biomedical Engineering, University of North Carolina, Charlotte, NC 28223, USA.

Abstract

DD[E/D]-transposases catalyze the multistep reaction of cut-and-paste DNA transposition. Structurally, several DD[E/D]-transposases have been characterized, revealing a multi-domain structure with the catalytic domain possessing the RNase H-like structural motif that brings three catalytic residues (D, D, and E or D) into close proximity for the catalysis. However, the dynamic behavior of DD[E/D]-transposases during transposition remains poorly understood. Here, we analyze the rigidity and flexibility characteristics of two representative DD[E/D]-transposases Mos1 and Sleeping Beauty (SB) using the minimal Distance Constraint Model (mDCM). We find that the catalytic domain of both transposases is globally rigid, with the notable exception of the clamp loop being flexible in the DNA-unbound form. Within this globally rigid structure, the central β -sheet of the RNase H-like motif is much less rigid in comparison to its surrounding α -helices, forming a cage-like structure. The comparison of the original SB transposase to its hyperactive version SB100X reveals the region where the change in flexibility/rigidity correlates with increased activity. This region is found to be within the RNase H-like structural motif, comprising the loop leading from beta-strand B3 to helix H1, helices H1 and H2, which are located on the same side of the central beta-sheet, and the loop between helix H3 and beta-strand B5. We further identify the RKEN214-217DAVQ mutations of the set of hyperactive mutations within the catalytic domain of SB transposase to be the driving factor that induces change in residue-pair rigidity correlations within SB transposase. Given that a signature RNase H-like structural motif is found in DD[E/D]-transposases and, more broadly, in a large superfamily of polynucleotidyl transferases, our results are relevant to these proteins as well.

Keywords

transposase; dynamics; flexibility and rigidity; fluorescence; Distance Constraint Model; DNA transposon; protein-DNA complex

*To whom correspondence should be addressed: Irina V. Nesmelova, Grigg Hall 306, Department of Physics and Optical Sciences, University of North Carolina, 9201 University City Blvd., Charlotte, NC 28223. Tele: 704-687-8145; Fax: 704-687-8197; Irina.Nesmelova@uncc.edu.

Introduction

DNA transposons are mobile DNA elements that can move (transpose) from one location in a genome to another or from one DNA molecule to another. Autonomous DNA transposons encode a transposase enzyme that is needed for their activity. Transposase enzymes are divided into several families based on the mechanism utilized during transposition¹. DD[E/D]-transposases use the cut-and-paste mechanism of DNA transposition, where they cleave DNA to excise and reinsert the transposon DNA into the host genome. Importantly, DD[E/D]-transposases have several advantages over existing viral and nonviral gene delivery technologies²⁻⁴. In particular, the Sleeping Beauty (SB) transposase has gained much attention because it can mediate stable integration and sustained expression of therapeutic transgenes in the genomes of primary human somatic cells⁵⁻⁸. It has been used in a first human clinical trial for the treatment of hematologic cancer⁹ and is under evaluation for another clinical trial to treat age-related macular degeneration (project TargetAMD). Thus, understanding the molecular mechanism for the SB transposase, and DD[E/D]-transposases more generally, has practical medical importance.

Members of the DD[E/D]-transposase family possess a characteristic triad of catalytic residues, two aspartic acids and one either glutamic or a third aspartic acid, brought into close proximity for catalysis by a common RNase H-like structural motif¹⁰⁻¹⁴. Furthermore, the RNase H structural motif is a signature motif of the members of a widespread and diverse RNase H-like (RNHL) superfamily of proteins that includes enzymes involved in replication, recombination, DNA repair, splicing, transposition, RNA interference (RNAi), and CRISPR-Cas immunity^{15,16}. The RNase H-like structural motif is schematically shown in Figure 1. It consists of a central β -sheet comprised of five α -strands (ordered 32145 with the second α -strand B2 antiparallel to other strands) and three surrounding α -helices. Two of the three α -helices (H1 and H2) are inserted in the amino acid sequence between β -strands B3 and B4 and B4 and B5, respectively, and are located on one side of the β -sheet. The third α -helix (H3) is located on the opposite side of the β -sheet and either immediately follows β -strand B5 or is separated from it by inserted amino acid sequence of various length. The most abundant proteins in the RNHL superfamily are DD[E/D]-transposases¹⁵. Although there are significant differences in sequence and structure across the RNHL superfamily, the RNase H architecture represents a highly conserved core of the catalytic domain of RNHL proteins, suggesting its functional role across the family^{17,18}.

The rationale for this study comes from the fact that cut-and-paste DNA transposition requires a two-step hydrolysis reaction during which all reaction steps are carried out by the same catalytic site, suggesting that DD[E/D]-transposases must undergo conformational rearrangements in order to bind the transposon DNA and to supply a catalytic site during each step of the reaction^{18,19}. Furthermore, the reaction is carried out within a nucleoprotein complex, a transpososome, which includes two ends of the transposon DNA and at least two transposase enzymes. The transpososome provides a precise architecture for the progression of transposition by adopting different conformational states in the course of the reaction, and the transposase enzyme must enable such changes. Collectively, these observations suggest the role for flexibility and dynamics in DNA transposition mediated by DD[E/D]-transposases. However, the information on the flexibility and dynamics of DD[E/D]-

transposases is limited. Nevertheless, it is reasonable to expect that the flexibility of DD[E/D]-transposases originates from the multi-domain structure where the catalytic and DNA-binding domains are connected by long flexible linkers^{10–12,20}. In addition to flexible linkers, computational studies of one of the most studied members of the RNHL family, the HIV integrase, revealed that the loop (residues 140–153) containing a catalytic residue E152 is dynamic^{21–24}, and replacing inherently flexible glycine residues in the catalytic loop with alanine significantly reduces integrase activity²⁵. Experimental NMR relaxation data show that the catalytic loop of HIV integrase is moderately dynamic on the picosecond to nanosecond time scale²⁶. Besides the catalytic loop, the crystal structures of DD[E/D]-transposase Mos1 and SB^{27,28} reveal a clamp loop, which is more than 30 residues long, between the two beta-strands of the RNase H-like structural motif²⁹. The clamp loop becomes constrained upon the interactions with the transposon DNA and participates in the protein-protein interactions within the transpososome^{27,30}. It has been further proposed that the conformational change in the clamp loop is necessary for the DNA-binding signal to propagate across the complex^{30,31}.

In this paper, we investigate the role of rigidity and flexibility within the catalytic domain of the DD[E/D]-transposases, and of the RNase H structural motif. First, we compare rigidity and flexibility properties between two closely related Mos1 and SB DD[E/D]-transposases to reveal a common trend. This comparison is facilitated by the availability of high-resolution structures^{27,28,32} and is done in the absence or presence of the transposon DNA. Next, we compare the original SB transposase to its hyperactive version SB100X (approximately 100-times activity increase)³³ to infer how rigidity/flexibility relate to changes in activity. Finally, we use extensive mutagenesis data available for SB transposase^{33,34} to define how single-point mutations with different effect on transposase activity alter its flexibility and rigidity.

Materials and Methods

Protein expression, purification and sample preparation

DNA plasmids encoding the catalytic domain containing amino acid residues 111–340 of the original SB transposase sequence³⁵ or its hyperactive SB100X version³³ were ordered from GenScript USA (Piscataway, NJ). Plasmids were transformed into chemically competent BL21-A1 E. coli cells. Cells were grown in LB medium at 30°C in the presence of ampicillin at the concentration of 0.01 mg/mL until OD₆₀₀ reached 0.6–0.8. Protein expression was induced by adding 1 mM IPTG and 0.2% L-arabinose for 4 hrs. Cells were collected by centrifugation and lysed in 50 mM TRIS and 500 mM NaCl lysis buffer at pH 8.0 by sonication. Soluble extract containing catalytic domain of SB transposase was prepared by centrifugation of cell lysate at 20,000 g for 1.5 hrs followed by overnight resuspension of pellet, and another centrifugation of cell lysate at 15,000g for 1hr. The protein was purified using Ni-NTA affinity resin (ClonTech, Mountain View, CA) and refolded by gradual decrease of urea concentration. The final sample was prepared in 25 mM TRIS buffer at pH 8.0. The presence and purity of proteins were monitored by SDS-page gel electrophoresis.

Fluorescence Spectroscopy

Fluorescence emission spectra were recorded on a PTI (Photon Technology International) QuantaMaster fluorescence spectrofluorometer using a quartz cell with the path length of 1 cm and with a cylindrical insert to minimize effects of evaporation at higher temperatures. The temperature of the jacketed cell holder was maintained by circulating water and monitored by a Hanna Instruments 93530 K thermocouple thermometer. The temperature range of 5 to 85°C for thermally-induced unfolding experiments was used.

ANS (1-anilinonaphthalene-8-sulfonic acid) fluorophore was used as an extrinsic probe of transposase unfolding. For ANS fluorescence measurements, protein samples were prepared in 25 mM aqueous TRIS buffer at pH 8.0. The concentration of both the original SB and the hyperactive SB100X transposase was 0.8 μ M. Stock solutions of ANS was prepared in ethanol at 160 and 16 μ M concentrations. ANS fluorophore was added to protein solution at increasing concentrations, and the intensity of ANS fluorescence was monitored by recording spectra at 25 °C. The optimal concentration of ANS (no further changes of fluorescence intensity were observed) was 20 μ M, and this ANS concentration was used for thermal unfolding experiments. Prior to recording emission spectra, ANS was mixed with protein samples for 1 hr. ANS emission spectra were recorded in the range of 420 – 575 nm with the excitation at 380 nm. Blank spectra of ANS alone were subtracted from the corresponding spectra recorded with the protein present to yield the spectra used for the final analysis.

Intrinsic fluorescence measurements were done using protein samples prepared in 25 mM aqueous TRIS buffer at pH 8.0 at the concentration of 3 μ M. The excitation wavelength was 286 nm, and the fluorescence emission was collected from 300 to 400 nm. Urea titration experiment was performed at 25 °C, while changing urea concentrations from 0 M to 6 M in steps of 0.25 M. To find the melting temperature T_m , measurements were done at urea concentrations ranging from 1 M to 4 M.

Structural Models

For input structures, we used crystallographic structures of the catalytic domain of SB100X transposase²⁸ (pdb code 5CR4) and of the Mos1 transposase²⁹ (pdb code 2F7T). We focus on the catalytic domain because it is separated from the DNA-binding domain by a long flexible linker, making the rigidity signal propagation unlikely. Although Mos1 and SB transposases crystallize as dimers, we focus on an individual monomer because there is a large separation between each catalytic domain in a dimer in the presence of DNA³⁴. The presence of DNA was modeled by introducing external torsional constraints on amino acid residues that make contact with DNA. For Mos1, the DNA contacts were identified from the DNA bound structure of Mos1 (pdb code: 4U7B)³². For SB, we used the reported DNA contacts²⁸.

For comparative analysis to the original SB transposase, we consecutively introduced amino acid residue mutations (RKEN214–217DAVQ, M243H, T314N) into the SB100X structure to obtain the structure of the original SB and five variants containing different combinations of the six point-mutations. Each structure was developed by mutagenesis in PyMol³⁶ and

then subsequently minimized using GROMACS^{37,38} using AMBER99SB-LN force field and explicit TIP3 water model³⁹. The structure for Mos1 was prepared by minimization using GROMACS in an identical manner. The SB100X crystal structure was further used to construct 10 structures with single amino acid mutations selected from a set of 80 mutations that has available functional assay data³⁴. Our selection criteria included (1) the location in proximity to helix H1 and the RKEN214–217DAVQ mutations, which were identified as the key determinant factor that distinguishes between SB and SB100X transposases, (2) the location in the interior of the protein to avoid the contribution of solubility to transposase activity, and (3) the broad range of effects on transposition activity: increase, decrease, no change.

Comparative Stability and Flexibility Analysis

A workflow is summarized in Supplementary Figure S1. To determine the flexibility of Mos1 and SB transposases, we used the minimal Distance Constraint Model (mDCM)^{40–43}. The input to the mDCM is a three-dimensional structure of the native state of a protein, from which an ensemble of graphs is generated where vertices represent atoms and edges represent interatomic interactions that include covalent bonds, H-bonds, and atomic packing. Salt bridges are considered as special types of H-bonds. While covalent bonds are fixed in the graph, native H-bonds are allowed to break and reform. An interatomic interaction is modeled by a distance constraint with assigned energy and entropy contribution. Atomic packing is modeled by fluctuating distance constraints that fix rotatable torsion angles into good or poor packing environments. A change from poor to good atomic packing lowers both energy and entropy. Accordingly, a protein with N_{rb} rotatable bonds and N_{hb} native H-bonds defines an ensemble of $2^{N_{rb} + N_{hb}}$ distinct graphs, each representing a constraint network with total energy defined by the interactions present. Unlike energy, entropy is non-additive among the interactions due to mechanical correlations created by the distance constraints. As such, the entropy of a graph is reduced only when flexibility is reduced by the addition of a constraint. Consequently, the mechanism of molecular cooperativity is captured because collective formation of a group of favorable interactions will reduce the entropy of a protein less than each interaction independently.

A macrostate of a protein is specified by the number of H-bonds N_{hb} and the number of good packing constraints N_{gp} within the protein. Energy and entropy contributions are calculated for each protein macrostate, and the free energy of a macrostate is given as:

$$G(N_{hb}, N_{gp}, T) = U(N_{hb}) - u(N_{hb}) + v(N_{gp}) - T[S_c(N_{hb}, N_{gp}|\delta) + S_m(N_{hb}, N_{gp})], \quad (1)$$

where $U(N_{hb})$ is the energy of the intramolecular H-bond, u is the average energy of H-bond between solvent and protein, v is the energy of a good packing constraint, and δ is the entropy of a torsion constraint representing good packing. A poor packing constraint corresponds to a random coil conformational state, and it is assigned an entropy that is considered to be transferable⁴¹. Therefore, the difference in entropy between a well packed residue and one that resembles a random coil can be adjusted by adjusting the δ parameter. The pair of parameters v and δ directly controls enthalpy-entropy compensation associated

with protein packing. Compared to an unfolded state, a folded state has a lower energy (i.e., $v < 0$), and a greater entropy drop occurs as δ decreases. The S_c term is the total conformational entropy as determined by network rigidity. S_c depends on all interaction types that are modeled as constraints (e.g. covalent bonds, H-bonds, torsions). Total conformational entropy and energy depend on parameters δ , u , and v , because the rigidity network is modified when these parameters are varied. S_m is the mixing entropy associated with the ensemble of graphs representing the same microstate, which is defined by the number of H-bonds and good packing torsion constraints. Parameter u is negative, indicating that when the number of intramolecular H-bonds increases, protein-solvent H-bond energy increases due to a commensurate loss in protein-solvent H-bonds.

Initially, the mDCM calculates the free energy landscape (FEL) with (N_{hb}, N_{gp}) as order parameters and the heat capacity $C_p(T|u, v, \delta)$ for a range of temperatures between 27 and 127°C for a trial set of parameters (u, v, δ) . This calculation requires random sampling of a few hundred constraint networks per macrostate, providing detailed flexibility information, including the total number of degrees of freedom (DOF) in the protein. The flexibility measures are averaged using statistical weights obtained from the FEL over all constraint networks per macrostate. The FEL of two order parameters (N_{hb}, N_{gp}) is collapsed onto a plot of $G(\theta, T)$ versus θ to obtain the free energy $G(\theta, T)$ as a function of global flexibility, where θ is the number of DOF in the protein per residue. For all the protein structures studied here, $G(\theta, T)$ had the signature shape of a “W” indicating a two state folding process (Supplementary Figure S2). After the initial calculation with a trial set of parameters (u, v, δ) , a grid search is performed, during which the u , v , and δ parameters are continually adjusted. The u and v parameters were changed in increments of $u = v = 0.01$ kcal/mol for each discretized value of parameter δ to perform the rigidity calculation. The sweep over δ is slow because rigidity calculations are performed for each different δ to calculate entropy, whereas the sweeps over u and v are fast because the entropy terms are fixed. Potential solutions were selected from this large dataset based on two state behavior (“W” shape) of $G(\theta, T)$, the peak in the predicted heat capacity matching experimental melting temperature (T_m values for Mos 1 and SB transposases), and $C_p(T)$ with a well-defined peak maximum values within expected ranges. The reported results in this work were averaged over all predictions based on 67 different parameter sets. The variation in predictions from different mDCM parameters were used for error analysis. The overall workflow for the grid search methodology is shown in Supplementary Figure S3.

Rigidity analysis of a constraint network determines which constraints are redundant or independent, where DOF are located, which regions are flexible, and which regions form rigid substructures⁴⁴. The Flexibility Index (FI) and Cooperativity Correlation (CC) are two measures that are used to characterize protein flexibility. The FI is defined as the ratio of the number of DOF in a given protein region divided by the number of rotatable covalent bonds in that region and represents the density of excess DOF within a flexible region reported as a positive number or redundant constraints within a rigid region reported as a negative number. A protein region is isostatic when it has no redundant constraints. In this case the FI is zero. The CC measure is a two-dimensional matrix that is built using FI values and has pairs of negative numbers for excess constraints, zeros for pairs of residues that are isostatic, and pairs of numbers that have positive numbers for excess DOF. The averaging of the FI

and CC over many constraint networks provides the information on where the protein is mostly flexible or rigid.

Following previous work⁴⁵, we also applied a signal beyond noise ratio (SBNR) to CC matrices to quantify statistically significant results. We calculated SBNR by first calculating the standard deviation, σ , associated with all CC_{jk} matrix elements. When $|CC_{jk}| < \sigma$ then $SBNR(CC_{jk}) = 0$. If $|CC_{jk}| > \sigma$ then $SBNR(CC_{jk}) = (CC_{jk} - \sigma)/\sigma$ for $CC_{ij} > 0$, and $SBNR(CC_{jk}) = (CC_{jk} + \sigma)/\sigma$ when $CC_{ij} < 0$. As such, when the magnitude of SBNR was greater than 1 the results were statistically significant, while variations within one standard deviation were not considered. When many CC matrices were compared, the standard deviation was calculated over all matrix elements across all the CC matrices.

Results

Experimental determination of T_m

The melting temperature T_m is needed to provide experimental constraints for parameterizing the mDCM. The T_m for Mos1 transposase has been reported previously⁴⁶. In this work, we determine the melting temperature for the original SB and its hyperactive version SB100X by extrinsic (ANS) and intrinsic (Tyr and Trp) fluorescence. The fluorescence properties of ANS strongly depend on the polarity of the environment⁴⁷. Accordingly, ANS fluorophore is commonly used to detect protein unfolding, typically accompanied by the exposure of hydrophobic surfaces. Figure 2A exemplifies ANS fluorescence spectra of SB and SB100X transposase at several temperatures. For both proteins, the significant increase of ANS fluorescence intensity and a blue shift of the emission maximum are observed (519 $\rightarrow$ 511 nm for SB, 514 $\rightarrow$ 507 nm for SB100X) as the temperature increases. Such changes occur when the ANS fluorophore transfers from a polar aqueous to the essentially non-polar environment⁴⁸. Control experiments with ANS alone showed no dependence of ANS fluorescence on temperature. Therefore, the observed changes of ANS fluorescence in the presence of transposase indicate that non-polar regions become sterically accessible to bind the dye, likely due to the change of tertiary structure. Figure 2B shows the temperature dependence of ANS fluorescence, quantified as the total area under the curve. Non-linear Boltzmann sigmoidal fits⁴⁹ of experimental data (solid lines) yielded similar T_m values of 65 ± 1.0 and 66 ± 1.5 °C for SB and SB100X transposases, respectively.

To supplement extrinsic fluorescence data and to verify that the observed ANS fluorescence is not affected by protein aggregation, we also performed intrinsic fluorescence experiments in the presence of urea. SB and SB100X transposases contain nine tryptophan and eight tyrosine residues distributed throughout the protein. While the emission wavelength of tyrosine does not change when its environment changes, the tryptophan fluorescence maximum (λ_{max}) and intensity are highly influenced by the polarity of the micro-environment, hydrogen bonding, and other non-covalent interactions, displaying a red shift when the polarity of Trp surrounding increases⁵⁰. All tryptophan residues and five tyrosine residues are located in the catalytic domain of the transposase. Five of the nine tryptophan residues and one tyrosine residue are contained within the RNase H-like fold. The

sidechains of remaining tryptophan and tyrosine residues, with the exception of Y167 and W169 located on the clamp loop and exposed to solvent, are oriented towards the interior of the protein. Thus, the measurement of changes in the intrinsic Tyr and Trp fluorescence as a function of increasing temperature should reflect the conformational changes throughout the whole protein and reveal the value of T_m .

Both the SB and SB100 transposases give intense fluorescence spectra shown in Figure 2C. The fluorescence emission maximum (λ_{max}) of SB transposase occurs at 342 nm at 25 °C, using an excitation wavelength of 286 nm, whereas it is shifted to 346 nm for SB100X transposase. Both proteins show red-shifted emission spectra, indicating that the fluorophores are located in sufficiently polar environment consistent with the high content of charged and polar residues in SB or SB100X transposase^{33,35}. Furthermore, upon the denaturation with urea that effectively suppresses protein aggregation and induces an extended conformation with nearly fully solvent accessible tryptophan residues, the position of emission maximum has a pronounced red shift to 351 nm and 353 nm for SB and SB100X transposase respectively, as shown in Figure 2C. We monitored thermally-induced unfolding of SB and SB100X transposases in the presence of urea at different concentrations ranging between 1 M and 3M. The value of T_m decreases as the concentration of urea increases as expected. The melting temperatures of SB and SB100X were obtained by extrapolating T_m values measured in the presence of urea to zero concentration of denaturant. These values were 62 ± 1.0 for SB and 63 ± 1.0 °C for SB100X in good agreement with ANS fluorescence data.

Flexibility and rigidity of DD[D/E]-transposases Mos1 and Sleeping Beauty

We compare the flexibility and rigidity characteristics of Mos1 and SB as two representative DD[D/E]-transposases. We examine the flexibility and rigidity of the RNase H-like fold, which is a common motif of DD[D/E]-transposases and RNHL proteins, as well as protein regions outside the RNase H-like fold. Although, these regions show a great structural diversity between the members of the RNHL family, they are similar in Mos1 and SB transposases.

The RNase H-like structural motif, with five beta-strands colored green and three alpha-helices colored purple, is shown in Figure 3A on the three-dimensional crystallographic structure of Mos1 transposase (pdb code 2F7T)²⁹. Figures 3 B-D show the flexibility index (FI) of Mos1 and SB catalytic domains, as a function of the residue number (B), which is then mapped onto their three-dimensional structures (C-D). The blue and red colors correspond to negative and positive values of FI, indicating that on average over the ensemble of rigidity networks a given residue is rigid or flexible, respectively. White color indicates that the magnitude of FI is zero, corresponding to the isostatic case where there are no excess constraints or DOF. The backbone flexibility profiles are similar between Mos1 and SB transposases, indicative of common flexibility and stability trends of these structurally homologous proteins. We find that Mos1 and SB are globally rigid with a differentiating degree of rigidity throughout the protein. Several isostatic regions are observed in loops, and a high degree of flexibility is confined almost exclusively within the clamp loop that extends out from the catalytic domain, consistent with this loop found to be

disordered in the x-ray structures of Mos1 and SB catalytic domains solved without DNA^{28,29}. It is striking that the central β -sheet of the RNase H-like motif is much less rigid in comparison to its surrounding α -helices, more so in SB transposase, forming a cage-like structure which could assist the rearrangements in the catalytic site and its vicinity to accommodate the DNA molecule. There is no significant difference in rigidity between the helices in or outside the RNase H-like fold.

A differentiation in rigidity extends to the three catalytic residues. In comparison to the first catalytic residue (D156 in Mos1 and D153 in SB transposase) located at the end of strand B3, the second catalytic residue (D249 in Mos1 and D244 in SB transposase) located on the loop joining strand B4 to helix H2, is much less rigid. The third catalytic residue (D284 in Mos1 and E279 in SB transposase) located at the edge of helix H3 is the most rigid of the three catalytic residues. Based on the definition of FI, residues will have different degrees of rigidity when they are members of distinct substructures with different levels of redundant constraint densities. This implies that there is a hierarchical pattern in how rigidity extends throughout these proteins, suggesting that certain critical hydrogen bonds (H-bonds) need to break to create collective flexibility within the transposase structure. While the backbone FI profile is similar in Mos1 and SB transposases, the beta-sheet and helix H3 of the RNase H-like motif are more flexible in SB transposase (Figure 3B). This trend extends to the second and third catalytic residues, indicating that the constraint network of Mos1 is more densely packed in the core of the protein, while SB has a more distributed network of H-bonds, which is noted in the H-bond contact map (Supplemental Figure S4).

Effect of DNA binding on the flexibility and rigidity of Mos1 and Sleeping Beauty transposases

Next, we investigated changes in backbone rigidity of Mos1 and SB transposases upon DNA binding. The DNA binding was modeled by placing external torsion constraints in the vicinity of the DNA binding site based on DNA contacts outlined in the literature^{27,28}. Despite added torsion constraints to the Mos1 and SB structures, locally increasing rigidity along the backbone, the difference in backbone rigidity between DNA-bound and DNA-unbound states of Mos1 and SB transposases demonstrates a rich distribution of regions with increased and decreased rigidity (Figure 4). This happens because of the enthalpy/entropy compensation due to global changes in flexibility/rigidity as was previously observed as a general effect along the lines of a Le Chatelier principle⁵¹. As expected, we find that there is a dramatic loss of flexibility in the clamp loop that forms multiple protein-protein and protein-DNA contacts²⁷. Additionally, flexible loops joining secondary structure elements of the RNase H-like fold, B3 and H1 or B4 and H2 or B5 and the following helix (not part of the RNase H-like motif) gain rigidity. The central β -sheet of the RNase H-like motif becomes more rigid. The most significant increase in rigidity at the catalytic site is observed for the second catalytic residue D249 or D244 in Mos1 or SB transposase, respectively. In contrast, the helices of the RNase H-like motif and other α -helices flanking the β -sheet lose rigidity with the exception of the N-terminal helix. These regions are less rigid or more flexible to compensate for the increase in rigidity at points of perturbation for the protein to maintain stability. Despite the proximity to the DNA binding site, no appreciable change is observed for the first catalytic residue D156 or D153. Interestingly, the third catalytic

residue shows different trends between Mos1 and SB, where it becomes more rigid in Mos1 and less rigid in SB transposase. Although the overall trends in how rigidity changes are the same for Mos1 and SB, the magnitude of the change differs. For example, upon DNA binding, the beta sheet rigidifies more in SB than in Mos1 transposase, and the loss of rigidity in helices H1 and H2 is greater in Mos1 than in SB transposase. Note that Mos1 without DNA is more rigid overall than SB. Consequently, DNA binding tends to enhance similarities.

Link between the flexibility/rigidity and transposition activity

We now address whether there is a link between the flexibility/rigidity of DD[D/E]-transposases and activity. Here, we focus on SB transposase, for which a large set of mutagenesis and functional data has been reported^{33,34}, including single amino acid residue substitutions and their combinations. In particular, using the molecular evolution approach, the transposase version with approximately 100-times increase in activity over the original transposase, the hyperactive SB100X transposase, was created³³. The SB100X differs from the original SB transposase by nine mutations, of which three mutations (K14R, K33A, and R115) are located outside the catalytic domain. Within the catalytic domain, SB100X differs from the original SB transposase by four sequential mutations RKEN214–217DAVQ, located at the beginning of helix H1 of the RNase H-like motif, mutation M243H located on strand B3 next to the catalytic residue D244, and mutation T314N located on the surface of SB100X that falls outside the RNase H-like motif (Figure 5A). In this work, we limit our focus to the catalytic domain because the DNA-binding domain is joined to the catalytic domain by a long and flexible linker that does not permit flexibility/rigidity changes to propagate between the two domains. We compare the original SB to SB100X transposase by computationally performing all the hyperactive mutations (RKEN214–217DAVQ, M243H, and T314N) within the catalytic domain.

Figure 5B shows the comparison of SB100X to the original SB transposase. Qualitatively, SB and SB100X transposases have similar FI profile, although SB100X exhibits somewhat less variation in the region comprising beta-strands B1 and B2, helices H1 and H2, and corresponding connecting loops. Quantitative differences of FI values, color-coded on the SB100X transposase structure, reveal a few marked changes in the flexibility/rigidity introduced by the six hyperactive mutations (Figure 5C). The central β -sheet is more rigid in SB100X than in the original SB transposase. In contrast, the surrounding α -helices become less rigid with one exception found at the C-terminus end of helix H3. In addition, there is a substantial increase in rigidity within the loop connecting strand B1 to helix H3, corresponding to the catalytic loop in HIV integrase that was found to be dynamic on the picosecond to nanosecond time scale, but did not show the characteristics often seen for flexible linker regions in proteins²⁶. Interestingly, the least rigid of the three catalytic residues, D244, does not exhibit an appreciable change in rigidity, whereas the other two catalytic residues, D153 and E279 decrease in rigidity. In particular, the catalytic residue E279 along with helix H2 and the C terminal helix show the strongest increase of FI. The crystallographic structure of SB100X revealed that the new H-bond, formed between E216 and D210 after RKEN214–217DAVQ mutations, played an important stabilizing role for SB100X transposase²⁸. The mDCM confirms that this H-bond (5.2 kcal/mol) is a major

contributing factor for the increase in rigidity and, additionally, finds several other weaker H-bonds that form between D216–N217 (3.2 kcal/mol), and between R214–E216, R214–K215, and R214–I212 with energies less than 1.0 kcal/mol.

Another level of protein dynamics description can be gained from the analysis of pairwise residue-to-residue mechanical couplings, visualized by a cooperativity correlation (CC) plot. CC plot is a symmetrical matrix that indicates how frequent (in the ensemble of rigidity networks characterizing a given macrostate) each residue pair is likely to be within the same rigid cluster or flexible region, outlining pathways of rigidity and flexibility propagation through the protein. We note that in rigidity theory, flexibility is distinct from mobility. An analogy can be drawn to a swinging pendulum, where a stationary pivot is highly flexible but not mobile, whereas the end of the pendulum is rigid and highly mobile⁵². Nevertheless, there is a strong correlation between flexibility and mobility in globular proteins⁴². Qualitatively, an increase in rigidity or flexibility, respectively, correlates to a decrease or increase in dynamics (mobility) in a sense that when two residues belong to the same rigid cluster their motion is highly correlated, whereas for a pair of co-flexible residues, the motion of either residue affects the other residue. Furthermore, if there is no mechanical coupling between the two residues, they will not show correlation regardless of whether they are rigid or flexible, meaning that their motion is independent of other residues.

Figure 6 shows CC plot differences, calculated by subtraction of CC plot of the original SB transposase from CC plots of hyperactive mutants (M243H, T314N, RKEN214–217DAVQ) separately (Figures 6A-C) or combined (Figures 6D-F), where Figure 6F represents SB100X transposase. Only statistically significant changes, e.g., variations over one standard deviation, are captured by applying the SBNR measure. Shades of red correspond to decreased correlation in rigidity or increased correlation in flexibility between a pair of residues. In contrast, shades of blue correspond to increased correlation in rigidity or a decreased correlation in flexibility. White color indicates protein regions where no change in correlation occurs. There is no significant difference in CC properties between the original SB and the T314N mutant. Previously, it was suggested that this mutation increases activity through solubility effects²⁸. Our results indicate that this mutation does not affect intra-molecular couplings related to network rigidity, indirectly corroborating the suggestion that solubility (protein-protein interactions) is the sole driver for T314N substitution effects. Figure 6B represents the M243H mutation, which is adjacent to the DDE catalytic triad of residues. Despite the importance of the position of M243H mutation and the potential formation of pi-bonding between histidine residues as a result of this mutation²⁸, the changes in the constraint network are insignificant. Figures 6C-F show RKEN214–217DAVQ substitution alone (C) or in combination with T314N and M243H mutations. Qualitatively similar characteristics of these plots suggest that the RKEN214–217DAVQ substitution has the greatest effect on the mechanical framework of the protein. The RKEN214–217DAVQ substitution leads to a much greater correlated rigidity of the loop joining beta-strand B3 to helix H1 and increases rigidity correlation in helix H1, albeit to a lesser degree. Moreover, a few residues following the beta-strand B1 gain in correlated rigidity, while there is an increase in correlated flexibility in helix H2 positioned right next to helix H1. Helix H1 contains the RKEN214–217 residues, the substitution of which causes an additional H-bond to form between the helix and the adjacent loop (209–215) connecting

residues several positions apart and rigidifying the entire region. This change results in rearrangement of H-bonds in the helix to better connect other regions of the helix. The M243H substitution in combination with RKEN214–217DAVQ leads to a minor rigidification of helix H2 as demonstrated by less intense red color and to the increased correlated flexibility of the loop connecting beta-strand B5 to helix H3. The latter loop corresponds to the catalytic loop of HIV integrase, the dynamics of which has been linked to functional activity²⁵. Overall, the analysis of CC plot differences identifies the correlated rigidity/flexibility of the loop between beta-strand B3 and helix H1, helices H1 and H2, and the loop between helix H3 and beta-strand B5, all from the RNase H-like structural motif, as functionally important.

In addition to the hyperactive SB100X, more than 80 mutations along with their activities have been reported³⁴. Out of these reported mutations, we selected an additional set of ten single-point mutations (C197A, F198A, T203V, A205P, A205K, H207V, D210E, E216K, N217H, N217K) based on the following criteria: (1) proximity to helix H1 so that the potential effect on function could be related to the perturbation of the residue constraint network in its vicinity; (2) opposite effects (increase or decrease of the transposition activity); (3) different substitutions of the same residue with adverse functional effects³⁴. These point mutations are shown in Figure 7A. The activity of these mutations is also indicated as the percent of the original SB transposase activity, where 100% corresponds to no change of activity (light cyan), whereas values smaller and greater than 100% correspond to lesser (blue) and higher activity (red), respectively. We note that the reported maximum activity change due to single-point mutations is considered small (700 %) when compared to SB100X (approximately 10,000 %).

Despite the location of the majority of mutations on beta-strand B1 facing helix H1, a greater change for most mutations is experienced by helix H2, highlighting a long-range propagation of the signal (Figure 7B). Interestingly, as with SB100X transposase, helix H2 becomes moderately more flexibly correlated to the rest of the molecule for all activity-increasing mutations (T203V, H207V, D210E, and N217H). Unlike hyperactive mutations of the SB100X transposase, none of the ten single-point mutations selected for additional analysis show the trend of increased correlated rigidity in helix H1 and the loop that connects the beta-strand B3 to helix H1. In effort to further identify the mechanism behind this effect, we found a hydrogen bond between two histidine residues H207 and H225. In the original SB transposase, the H207-H225 H-bond has the energy of 2.1 kcal/mol. In SB100X, the energy of H207-H225 H-bond decreases to 1.4 kcal/mol. Seven of the ten single-point mutations disrupt the local environment enough to reduce the H207-H225 H-bond energy to below 1.0 kcal/mol, and in the case of H207V mutation the bond does not form. Without the H207-H225 H-bond, the backbone flexibility of residue H225 dramatically increases, leading to local rearrangement of the intra-helical H-bonds in helix H1. This specific H-bond is critical, because upon its introduction, part of the H2 helix from H207 to the RKEN residues locally rigidifies according to flexibility index, and the local rigidity in this region correlates with activity level. Additionally, the mDCM reveals the H-bond formed between residues Y218-H243 with a binding energy of 4.1 kcal/mol. M243H is one of the mutations that separates SB from SB100X, and H243 was suggested by crystallographers to potentially participate in pi bonding with H249. However, removing the Y218-H243 H-bond, or

disrupting the H243-H249 interaction does not alter the QSFR characteristics, suggesting that these bonds are redundant constraints.

While small changes are observed, the FI profile is similar in all ten mutants. To reveal the correlation between the change in the correlated flexibility/rigidity of the protein due to mutation and the transposition activity, we applied the clustering approach adapted by the software Clustergram^{53–56}. To use Clustergram, we coarse-grained the FI and CC differences between the mutant and the original transposase by selecting five segments within the transposase structures that showed changes upon mutation (beta-strands B1, B2, B3 and alpha-helices H1 and H2) and averaging FI and CC difference values over all residues for each of these segments. The results of clustering analysis present a phylogenetic distance as a function of difference in flexibility index of the defined regions (Figure 7C). The FI based clustering matched functional assays of the ten single-point mutations at 300K. This supports our assumptions that both local and long-range rigidity are the key for catalytic enhancement.

Discussion

We characterized the flexibility and rigidity properties of two representative transposases from the DD[E/D]-transposase family, Mos1 and SB transposases. The catalytic domain of these transposases contains a signature structural motif, an RNase H-like fold, present in a diverse number of enzymes that belong to the RNHL superfamily¹⁵. Thus, general trends in flexibility/rigidity found in the RNase H-like structural motif of Mos1 and SB transposases may be expected for other members of the RNHL family. Our data show that the catalytic domain in Mos1 and SB transposases has a highly constrained topology with globally conserved flexibility/rigidity profiles along their backbones. Other than the flexible clamp loop and several loops that are close to be isostatic, we find that the catalytic domain is globally rigid, which can be traced back to the hydrogen bond (H-bond) network and atomic packing, dictated by the architecture of the fold. Finding a rigid core is somewhat surprising, because the catalytic domain must catalyze both the DNA breaking and joining reactions and, hence, possess the ability of conformational adjustment during the reaction, typically associated with a considerable degree of flexibility. Instead, there is a differentiating degree of rigidity throughout the protein. We note that within the overall constrained topology, the central β -sheet of the RNase H structural motif is less rigid than the surrounding α -helices, forming a cage-like structure. This trend reverses upon DNA binding where the central beta-sheet becomes comparatively more rigid than the surrounding helices. Additionally, the second catalytic residue (D249 in Mos1 and D244 in SB transposase) is considerably more flexible than the other two catalytic residues in the absence of DNA. Upon binding DNA, this catalytic residue rigidifies while the other two catalytic residues gain flexibility. Although the overall structure is seen to be rigid, the long clamp loop shows a high degree of flexibility. However, upon DNA binding, the clamp loop directly interacts with DNA and subsequently rigidifies.

We find that the SB structure without DNA is more rigid than Mos1 structure without DNA within the helix bundle surrounding the core and preceding the RKEN214–217 residues of helix H1. One likely reason why this occurs is because the short loop in Mos1 transposase

has three more residues than SB transposase. The short loop preceding the H1 helix in both Mos1 and SB is found to be near isostatic, although it is slightly more flexible in Mos1. Therefore, this short loop enhances rigidity found elsewhere in the catalytic domain, which is counter intuitive if one thinks that loops are always flexible. Our results suggest that high levels of rigidity in the core beta-sheet, coupled with its surrounding alpha helices have a marked similarity with the mean field treatment of rigidity within proteins that leads to a sharp binary switching effect (e.g., two-state cooperativity) as described in a general context on the mechanism for how rigidity induces cooperativity⁵⁷. To make this connection quantitative, the free energy landscapes show well-defined two-state behavior (Supplementary Figure S2) in both transposases. Furthermore, the global increase in residue-residue rigidity coupling also increases cooperativity upon DNA binding to DNA as reflected by the increase of barrier height in the free energy landscapes. Consequently, we can conclude that the catalytic domain in DD[E/D]-transposases interacts with the transposon DNA cooperatively, tracking the clamp loop motion.

The hyperactive SB100X transposase provides a unique opportunity to investigate the link between protein mechanics/dynamics and function. We focused on the comparison of the catalytic domains of SB and SB100X transposases. From the comparison, we identified the region of the protein where the flexibility/rigidity differs. Interestingly, this region is located within the RNase H-like structural motif, comprising the loop leading from beta-strand B3 to helix H1, helices H1 and H2, which are located on the same side of the central beta-sheet, and the loop between helix H3 and beta-strand B5. While in the hyperactive SB100X transposase helix H1 and the loop between beta-strand B3 and helix H1 become more rigid, helix H2 and the loop between helix H3 and beta-strand B5 gain flexibility. We note that our analysis captures the loop between helix H3 and beta-strand B5, which corresponds to the catalytic loop in the HIV integrase. The dynamics of the HIV catalytic loop has been extensively studied computationally and experimentally^{21,22,25,26}. When HIV integrase is not bound to DNA, the catalytic loop is moderately dynamic on the picosecond to nanosecond time scale as it samples different conformations²⁶, also exemplified by crystallographic studies^{58–60}. Our data indicating the gain in flexibility by the loop between helix H3 and beta-strand B5 in a hyperactive SB100X transposase are compatible with this observation. In addition, our analysis identified new segments in the RNase-like structural motif (the loop leading from beta-strand B3 to helix H1 and helices H1 and H2), the dynamics of which is distinct between the original SB and SB100X transposases, suggesting its role in functional activity. From six hyperactive mutations (RKEN214–217DAVQ, M243H, and T314N), we find that the RKEN214–217DAVQ sequential mutations drive long range rigidity alterations, whereas M243H mutation provides a smaller adjustment when combined with RKEN214–217DAVQ mutations. The local rigidity of the DAVQ stretch is controlled by strong hydrogen bonds between residues H207 and H225, E216 and D210, and D216 and N217. The disruption of one of these bonds causes the rearrangement in the H-bond network leading to the rigidification of the entire helix H1.

In addition to SB100X transposase hyperactive mutations (approximately 10,000% of the original transposase activity), we studied ten single-point mutations within the catalytic domain with different impacts on the transposition activity (approximately 0–700% of the original transposase activity). We selected these mutations in the vicinity of helix H1,

however the majority of the mutations induced the flexibility/rigidity change in helix H2, highlighting a long-range propagation of the signal. While all activity-increasing mutations (T203V, H207V, D210E, and N217H) lead to increased flexibility of helix H2, none of the ten single-point mutations show the trend of increased correlated rigidity in helix H1 and the loop that connects the beta-strand B3 to helix H1 as the SB100X mutations, highlighting two important points. First, while a single mutation can have a significant effect on the transposition activity, a combination of mutations can modify the global network of correlated motions differently and in a more optimal way for function. Second, the flexibility of helix H2 is a consistently significant contribution to SB transposase activity.

Previously, the analysis of sectors in SB transposase revealed that mutations within the catalytic domain with a strong effect on free energy of folding significantly reduce the transposition rate. However, most hyperactive mutants were found on the protein surface, suggesting that SB hyperactivity results from the modification of protein interactions, rather than the stabilization of protein fold³⁴. Here, we identified the region within the RNase H-like structural motif, where mutations alter the flexibility/rigidity network of the protein leading to increased transposition activity. From practical standpoint, mutations in the loop leading from beta-strand B3 to helix H1, helices H1 and H2, and the loop between helix H3 and beta-strand B5 of the RNase H-like fold should be explored first to affect the transposase function through altered dynamics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the research grant the National Institute of Health grant R15GM112076 and by the University of North Carolina funds to IVN and by NIH NIGMS 1R01 GM 073082 to DJJ. Key to the distance constraint model is the use of graph-rigidity algorithms, claimed in U.S. Patent 6,014,449, which has been assigned to the Board of Trustees of the Michigan State University, and was used with permission.

References:

1. Curcio MJ, Derbyshire KM. The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* 2003;4(11):865–877. [PubMed: 14682279]
2. Kebriaei P, Izsvak Z, Narayanavari SA, Singh H, Ivics Z. Gene Therapy with the Sleeping Beauty Transposon System. *Trends Genet* 2017;33(11):852–870. [PubMed: 28964527]
3. Munoz-Lopez M, Garcia-Perez JL. DNA transposons: nature and applications in genomics. *Curr Genomics* 2010;11(2):115–128. [PubMed: 20885819]
4. VandenDriessche T, Ivics Z, Izsvak Z, Chuah MK. Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. *Blood* 2009;114(8):1461–1468. [PubMed: 19471016]
5. Izsvak Z, Hackett PB, Cooper LJ, Ivics Z. Translating Sleeping Beauty transposition into cellular therapies: victories and challenges. *Bioessays* 2010;32(9):756–767. [PubMed: 20652893]
6. Hackett PB, Largaespada DA, Cooper LJ. A transposon and transposase system for human application. *Mol Ther* 2010;18(4):674–683. [PubMed: 20104209]
7. Hudecek M, Izsvak Z, Johnen S, Renner M, Thumann G, Ivics Z. Going non-viral: the Sleeping Beauty transposon system breaks on through to the clinical side. *Crit Rev Biochem Mol Biol* 2017;52(4):355–380. [PubMed: 28402189]

8. Hodge R, Narayanavari SA, Izsvak Z, Ivics Z. Wide Awake and Ready to Move: 20 Years of Non-Viral Therapeutic Genome Engineering with the Sleeping Beauty Transposon System. *Hum Gene Ther* 2017;28(10):842–855. [PubMed: 28870121]
9. Kebriaei P, Singh H, Huls MH, Figliola MJ, Bassett R, Olivares S, Jena B, Dawson MJ, Kumaresan PR, Su S, Maiti S, Dai J, Moriarity B, Forget MA, Senyukov V, Orozco A, Liu T, McCarty J, Jackson RN, Moyes JS, Rondon G, Qazilbash M, Ciurea S, Alousi A, Nieto Y, Rezvani K, Marin D, Popat U, Hosing C, Shpall EJ, Kantarjian H, Keating M, Wierda W, Do KA, Largaespada DA, Lee DA, Hackett PB, Champlin RE, Cooper LJ. Phase I trials using Sleeping Beauty to generate CD19-specific CAR T cells. *J Clin Invest* 2016;126(9):3363–3376. [PubMed: 27482888]
10. Dyda F, Chandler M, Hickman AB. The emerging diversity of transpososome architectures. *Q Rev Biophys* 2012;45(4):493–521. [PubMed: 23217365]
11. Montano SP, Rice PA. Moving DNA around: DNA transposition and retroviral integration. *Curr Opin Struct Biol* 2011;21(3):370–378. [PubMed: 21439812]
12. Nesmelova IV, Hackett PB. DDE transposases: Structural similarity and diversity. *Adv Drug Deliv Rev* 2010;62(12):1187–1195. [PubMed: 20615441]
13. Nowotny M Retroviral integrase superfamily: the structural perspective. *EMBO Rep* 2009;10(2):144–151. [PubMed: 19165139]
14. Jaskolski M, Alexandratos JN, Bujacz G, Wlodawer A. Piecing together the structure of retroviral integrase, an important target in AIDS therapy. *FEBS J* 2009;276(11):2926–2946. [PubMed: 19490099]
15. Majorek KA, Dunin-Horkawicz S, Steczkiewicz K, Muszewska A, Nowotny M, Ginalski K, Bujnicki JM. The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res* 2014;42(7):4160–4179. [PubMed: 24464998]
16. Moelling K, Broecker F, Russo G, Sunagawa S. RNase H As Gene Modifier, Driver of Evolution and Antiviral Defense. *Front Microbiol* 2017;8:1745. [PubMed: 28959243]
17. Craig NL. Unity in transposition reactions. *Science* 1995;270(5234):253–254. [PubMed: 7569973]
18. Mizuuchi K Polynucleotidyl transfer reactions in transpositional DNA recombination. *J Biol Chem* 1992;267(30):21273–21276. [PubMed: 1383220]
19. Mizuuchi M, Baker TA, Mizuuchi K. Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. *Cell* 1992;70(2):303–311. [PubMed: 1322248]
20. Rice PA, Baker TA. Comparative architecture of transposase and integrase complexes. *Nat Struct Biol* 2001;8(5):302–307.
21. Lee MC, Deng J, Briggs JM, Duan Y. Large-scale conformational dynamics of the HIV-1 integrase core domain and its catalytic loop mutants. *Biophys J* 2005;88(5):3133–3146. [PubMed: 15731379]
22. Lins RD, Briggs JM, Straatsma TP, Carlson HA, Greenwald J, Choe S, McCammon JA. Molecular dynamics studies on the HIV-1 integrase catalytic domain. *Biophys J* 1999;76(6):2999–3011. [PubMed: 10354426]
23. Perryman AL, Forli S, Morris GM, Burt C, Cheng Y, Palmer MJ, Whitby K, McCammon JA, Phillips C, Olson AJ. A dynamic model of HIV integrase inhibition and drug resistance. *J Mol Biol* 2010;397(2):600–615. [PubMed: 20096702]
24. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a novel binding trench in HIV integrase. *J Med Chem* 2004;47(8):1879–1881. [PubMed: 15055986]
25. Greenwald J, Le V, Butler SL, Bushman FD, Choe S. The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry* 1999;38(28):8892–8898. [PubMed: 10413462]
26. Fitzkee NC, Masse JE, Shen Y, Davies DR, Bax A. Solution conformation and dynamics of the HIV-1 integrase core domain. *J Biol Chem* 2010;285(23):18072–18084. [PubMed: 20363759]
27. Richardson JM, Colloms SD, Finnegan DJ, Walkinshaw MD. Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* 2009;138(6):1096–1108. [PubMed: 19766564]

28. Voigt F, Wiedemann L, Zuliani C, Querques I, Sebe A, Mates L, Izsvak Z, Ivics Z, Barabas O. Sleeping Beauty transposase structure allows rational design of hyperactive variants for genetic engineering. *Nat Commun* 2016;7:11126. [PubMed: 27025571]
29. Richardson JM, Dawson A, O'Hagan N, Taylor P, Finnegan DJ, Walkinshaw MD. Mechanism of Mos1 transposition: insights from structural analysis. *EMBO J* 2006;25(6):1324–1334. [PubMed: 16511570]
30. Bouuaert CC, Walker N, Liu DX, Chalmers R. Crosstalk between transposase subunits during cleavage of the mariner transposon. *Nucleic Acids Research* 2014;42(9):5799–5808. [PubMed: 24623810]
31. Liu DX, Chalmers R. Hyperactive mariner transposons are created by mutations that disrupt allosterism and increase the rate of transposon end synapsis. *Nucleic Acids Research* 2014;42(4):2637–2645. [PubMed: 24319144]
32. Dornan J, Grey H, Richardson JM. Structural role of the flanking DNA in mariner transposon excision. *Nucleic Acids Res* 2015;43(4):2424–2432. [PubMed: 25662605]
33. Mates L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K, Matrai J, Ma L, Samara-Kuko E, Gysemans C, Pryputniewicz D, Miskey C, Fletcher B, VandenDriessche T, Ivics Z, Izsvak Z. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet* 2009;41(6):753–761. [PubMed: 19412179]
34. Abrusan G, Yant SR, Szilagyi A, Marsh JA, Mates L, Izsvak Z, Barabas O, Ivics Z. Structural Determinants of Sleeping Beauty Transposase Activity. *Mol Ther* 2016;24(8):1369–1377. [PubMed: 27401040]
35. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 1997;91(4):501–510. [PubMed: 9390559]
36. Schrödinger L The PyMOL Molecular Graphics System. 2.0; 2015.
37. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* 2008;4(3):435–447. [PubMed: 26620784]
38. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* 2005;26(16):1701–1718. [PubMed: 16211538]
39. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013;29(7):845–854. [PubMed: 23407358]
40. Jacobs DJ. Ensemble-based methods for describing protein dynamics. *Curr Opin Pharmacol* 2010;10(6):760–769. [PubMed: 20965786]
41. Jacobs DJ, Dallakyan S. Elucidating protein thermodynamics from the three-dimensional structure of the native state using network rigidity. *Biophys J* 2005;88(2):903–915. [PubMed: 15542549]
42. Livesay DR, Dallakyan S, Wood GG, Jacobs DJ. A flexible approach for understanding protein stability. *FEBS Lett* 2004;576(3):468–476. [PubMed: 15498582]
43. Livesay DR, Jacobs DJ. Conserved quantitative stability/flexibility relationships (QSFR) in an orthologous RNase H pair. *Proteins* 2006;62(1):130–143. [PubMed: 16287093]
44. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. *Proteins* 2001;44(2):150–165. [PubMed: 11391777]
45. Herring CA, Singer CM, Ermakova EA, Khairutdinov BI, Zuev YF, Jacobs DJ, Nesmelova IV. Dynamics and thermodynamic properties of CXCL7 chemokine. *Proteins* 2015;83(11):1987–2007. [PubMed: 26297927]
46. Trubitsyna M, Morris ER, Finnegan DJ, Richardson JM. Biochemical characterization and comparison of two closely related active mariner transposases. *Biochemistry* 2014;53(4):682–689. [PubMed: 24404958]
47. Hawe A, Sutter M, Jiskoot W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* 2008;25(7):1487–1499. [PubMed: 18172579]
48. Lubkowski J, Yang F, Alexandratos J, Wlodawer A, Zhao H, Burke TR, Jr, Neamati N, Pommier Y, Merkel G, Skalka AM. Structure of the catalytic domain of avian sarcoma virus integrase with a

- bound HIV-1 integrase-targeted inhibitor. *Proc Natl Acad Sci U S A* 1998;95(9):4831–4836. [PubMed: 9560188]
49. Huynh K, Partch CL. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* 2015;79:28 29 21–14.
50. Lakowicz JR. Principles of fluorescence spectroscopy New York: Springer; 2006 xxvi, 954 p. p.
51. Li T, Tracka MB, Uddin S, Casas-Finet J, Jacobs DJ, Livesay DR. Redistribution of flexibility in stabilizing antibody fragment mutants follows Le Chatelier's principle. *PLoS One* 2014;9(3):e92870. [PubMed: 24671209]
52. Verma D, Jacobs DJ, Livesay DR. Changes in Lysozyme Flexibility upon Mutation Are Frequent, Large and Long-Ranged. *PLoS Comput Biol* 2012;8(3):e1002409. [PubMed: 22396637]
53. Bar-Joseph Z, Gifford DK, Jaakkola TS. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* 2001;17 Suppl 1:S22–29. [PubMed: 11472989]
54. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95(25):14863–14868. [PubMed: 9843981]
55. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997;278(5338):680–686. [PubMed: 9381177]
56. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286(5439):531–537. [PubMed: 10521349]
57. Vorov OK, Livesay DR, Jacobs DJ. Nonadditivity in conformational entropy upon molecular rigidification reveals a universal mechanism affecting folding cooperativity. *Biophys J* 2011;100(4):1129–1138. [PubMed: 21320459]
58. Bujacz G, Alexandratos J, Qing ZL, Clement-Mella C, Wlodawer A. The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Lett* 1996;398(2–3):175–178. [PubMed: 8977101]
59. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 1994;266(5193):1981–1986. [PubMed: 7801124]
60. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci U S A* 1998;95(16):9150–9154. [PubMed: 9689049]

Figure 1. Schematic representation of the RNase H-like fold.

Circles and arrows represent α -helices and β -strands, respectively. Arrow orientation shows the orientation of the β -strands. Black lines show connecting loops. Red spheres represent the catalytic residues D, D, and E or D.

Figure 2. Determination of T_m by extrinsic and intrinsic fluorescence.

(A) ANS fluorescence spectra are shown at temperatures 15, 55, and 80 °C for original SB (solid circles) and SB100X (open circles) transposases. (B) Temperature-induced change of ANS fluorescence measured at λ_{max} for original SB (solid circles) and SB100X (open circles) transposases. Solid lines represent non-linear Boltzmann sigmoidal fits of experimental data. Determined T_m values are 65 ± 1.0 and 66 ± 1.5 °C for SB and SB100X transposases, respectively. (C) Intrinsic fluorescence spectra collected for original SB (solid symbols) and SB100X (open symbols) transposases at 25 °C and urea concentrations of 0 M (circles) and 5 M (diamonds). Red spectral shift in the presence of urea indicates protein unfolding. (D) T_m values determined for original SB (solid circles) and SB100X (open circles) transposases at different urea concentrations. T_m values determined by extrapolation to 0 M urea concentration are 62 ± 1.0 and 63 ± 1.0 °C for SB and SB100X transposases, respectively.

Figure 3. Flexibility index (FI) of Mos1 and SB transposases.

(A) Cartoon representation of the catalytic domain of Mos1 transposase (pdb code 2F7T²⁹) showing the RNase H-like structural motif consisting of a five-stranded beta-sheet shown in green and three alpha-helices shown in purple. Catalytic residues are shown using spheres, and metal ions are shown by cyan spheres. (B) Backbone FI is shown versus the residue number for Mos1 (orange) and SB (black) transposases, where their amino acid sequences are aligned according to secondary structure elements. Residue numbering corresponds to Mos1. Positive values indicate flexibility while negative values indicate rigidity. Secondary structure elements are indicated by rectangles, where grey color represents alpha-helices outside the RNase H-like motif, and green and purple colors represent beta-strands and alpha-helices of the RNase H-like motif, respectively. (C-D) FI is mapped onto the cartoon structure of Mos1 (C) and SB (D) transposases. Blue color corresponds to negative FI values

(rigid), red color corresponds to positive FI values (flexible), and white color corresponds to $FI = 0$ (isostatic). The intensity of color reflects the magnitude of FI.

Figure 4. The effect of DNA binding on backbone flexibility of Mos1 and SB transposases. The difference of FI values between DNA bound and unbound states is shown for Mos1 (A) and SB (B) transposases. Blue regions indicate increased rigidity and red regions indicate decreased rigidity caused by the binding of the transposon DNA.

Figure 5. Flexibility index (FI) of the original SB and SB100X transposases.

(A) Cartoon representation of SB100X transposase with hyperactive mutations shown in orange color using spheres. The RNase H-like structural motif is shown by green (beta-sheet) and purple (alpha-helices) colors. Catalytic residues are shown in red color using spheres. (B) The difference of FI values between the original SB and SB100X transposases is mapped onto the cartoon structure of SB transposase. (C) Backbone flexibility index is shown versus the residue number for SB (blue) and SB100X (red) transposases. Positive values indicate flexibility, while negative values indicate rigidity. Dashed line corresponds to zero (isostatic case) and is shown as guide to the eye. Secondary structure elements are indicated above the graph and colored as in the structure shown in (A).

Figure 6. Change in correlated cooperativity (CC) in SB transposase upon introducing hyperactive mutations.

CC differences are calculated by subtracting the CC of the original transposase from the CC of respective mutant. Only changes over one standard deviation, calculated using the SBNR⁴⁵, are shown. (A-C) CC differences of T314N, M243H, and RKEN214–217DAVQ mutants, respectively. (D-E) CC differences of RKEN214–217DAVQ mutant combined with T314N and M243H mutations, respectively. (F) The CC difference between the original SB³⁵ and SB100X³³ transposases. Red color shows regions of decreased correlation in rigidity or increased correlation in flexibility between a pair of residues. Blue color shows regions of increased correlation in rigidity or a decreased correlation in flexibility. White color indicates protein regions with no change in correlation. Secondary structure elements are indicated above plots with the RNase H-like fold colored in purple (helices) and green (beta sheets) in accordance with Figure 2A. The elements of the RNase H-like structural motif are also labeled.

Figure 7. Single-point mutant analysis.

(A) Ten point mutations selected for the analysis are shown using spheres on cartoon representation of SB transposase (pdb code 5cr4²⁸). Residues colored red indicate mutations that enhance transposition activity, blue – decrease transposition activity, and light cyan – do not significantly affect the transposition activity of SB transposase. Transposition activity is also shown as percent of the original transposase activity, where 100% corresponds to the activity of the original SB transposase. (B) Correlated cooperativity plots showing changes of flexibly and rigidly correlated regions in SB transposase upon introducing a single-point mutation. Red color shows regions of decreased correlation in rigidity or increased correlation in flexibility between a pair of residues. Blue color shows regions of increased correlation in rigidity or a decreased correlation in flexibility. White color indicates protein regions with no change in correlation. (C) Clustering of FI difference between the original

SB transposase and ten mutants based on values corresponding to the average over all residues in beta-strands B1, B2, B3 and alpha-helices H1, H2.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript