4. Discussion

4.1. The DNA-bending protein HMGB1 is a cellular cofactor of *Sleeping Beauty* transposition

Many DNA recombinational processes in prokaryotic systems require the association of sequence-specific recombinase proteins and some other proteins that alter DNA conformation (Echols, 1986). Integration host factor (IHF) and HU are closely related histone-like DNA bending proteins that are widespread in prokaryotes. IHF binds to a defined consensus DNA sequence, while HU binds to DNA non-specifically (Gossen and Van de Putte, 1995). In the cell, they serve as architectural factors in many cellular processes, such as transcription, replication, and site-specific recombination helping to assemble many different nucleoprotein complexes bringing distant DNA sites closer to each other by virtue of their ability to introduce sharp bends into the DNA. For example, IHF was found to enhance assembly of Tn10 synaptic complexes (Sakai et al., 1995). HU protein is absolutely necessary for the Mu nucleoprotein complex assembly. It brings the binding sites of the Mu transposase together by its ability to bend the Mu transposon DNA (Lavoie et al., 1990). Hin recombinase-mediated recombination and bacteriophage lambda integration are strongly stimulated by HU (Haykinson and Johnson, 1993) and integration host factor (IHF) (Goodman and Nash, 1989), respectively. The eukaryotic high mobility group (HMG) proteins can functionally replace HU and IHF in some recombination reactions, indicating some level of exchangeability between these DNA-bending proteins (Segall et al., 1994). All of these DNA-bending proteins are widely involved in assisting many recombinational mechanisms by facilitating the formation of active recombinase-DNA complexes (Paull et al., 1993; Bustin, 1999).

In this thesis, evidence is presented that HMGB proteins are cellular cofactors of *Sleeping Beauty* transposition. In HMGB1 knockout cells, transpositional activity was found to be marginal (Fig. 5B). This residual activity can probably be accounted for by the presence of HMGB2 in these cells. HMGB2 was found to partially or fully complement the absence of HMGB1 in some reactions (Bustin, 1999), so these two proteins are interchangeable to a certain degree. Indeed, transient overexpression of HMGB2 partially complemented the HMGB1-deficiency in our transposition assays (Fig. 5B). Mammalian cells contain significant amounts of HMGBs; there might be one molecule of HMGB1 for every 2 kb of the human genome (Bianchi and Beltrame, 2000). Therefore, our finding that transient

overexpression of HMGBs in wild-type mouse cells enhances transposition (Fig. 5B) was unexpected. However, this phenomenon is not without precedent: transient overexpression of HMGB1 by transfection enhances the activity of certain HMGB1 interactors, such as RAG1/2 (Van Gent et al., 1997), some Hox proteins (Zappavigna et al., 1996), and nuclear hormone receptors (Boonyaratanakornkit et al., 1998). These findings suggest that HMGB1 can be a limiting factor of *SB* transposition, and that different cellular levels of HMGB1 might modulate the efficiency of transposition in different tissues or species.

4.1.1. Possible roles of HMGB1 in transposition

In prokaryotes, the DNA-bending proteins HU and IHF bind directly to DNA, and no protein-protein interactions are required for their targeting (Lavoie and Chaconas, 1990). In contrast, HMGs have low affinity to standard, B-form DNA, and interactor proteins usually guide them to certain sites. It has been shown that the Sleeping Beauty transposase is an HMG interactor (Figs 9 and 10). The interaction was detectable in the absence of DNA, suggesting that SB might actively recruit HMGB1 to sites of transposition. At which step is HMGB1 required for transposition? This could be explained by considering the following, not mutually exclusive, possibilities: 1) HMGB1 induces a structural change in transposon DNA, which is required for efficient transposition; 2) HMGB1 enhances binding of the transposase to the transposon inverted repeats; 3) HMGB1 induces a conformational change of the transposase that makes the transposase more active; 4) HMGB1 plays a role in transposon integration by making contacts with chromatin components and/or by DNA-bending at target sites (Vigdal et al., 2002). In this thesis, evidence is provided that HMGB1 promotes circle formation of transposon DNA (Figs. 6 and 7), that it significantly enhances specific transposase binding to the transposon inverted repeats (Fig. 8), and that it can form a ternary complex with the transposase and transposon DNA (Fig. 10). Thus, although a role of HMGB1 in transposon integration cannot be ruled out, my results are consistent with a role of HMGB1 in the early steps of transposition, prior to catalysis.

Considering the significant drop of transposition activity in HMGB1-deficient cells (Fig. 5B), the role of HMGB1 in transposition has to be a critical one. *SB* has four transposase binding sites, directly repeated at the ends of the terminal inverted repeats. It has been previously shown that 1) presence of the four transposase binding sites is absolutely required for transposition (Izsvak et al., 2000) and 2) SB transposase forms tetramers in complex with

transposase binding sites (Izsvak et al., 2002). These observations are consistent with an interaction between the IR/DR structure and a transposase tetramer during transposition. A proposed role of HMGB1 protein is to bring the two binding sites closer to each other during synaptic complex formation (Fig. 17). HMGB1 might promote communication between DNA motifs that are otherwise distant to each other, including the DRs, the transpositional enhancer and the two IRs (Fig. 17). Similar to *SB* transposition, a DNA-bending protein, HU, is involved in looping out the linker DNA between transposase binding sites during Mu-transpososome assembly (Lavoie and Chaconas, 1990; Chaconas, 1999). If the only role of HMGB1 is to extrude the spacer region between the DRs, thereby bringing them close to each other in space, than deleting the spacer would rescue transposition activity even in the absence of HMGB1. Towards that end, transposons in which the DRs were 10, 20 and 50 bps from each other have been constructed. None of these transposons had any detectable activity (data not shown), indicating that physical proximity of the DRs is not sufficient for transposition, and that the correct geometrical configuration of the inverted repeats and the binding sites is crucial.

These observations indicate that a highly specific configuration of functional DNA elements within the inverted repeats has a critical importance in *SB* transposition (Fig. 17). This complex needs to be very precisely assembled, and probably includes the four DRs, the HDR enhancer motif, four transposase molecules (Izsvak et al., 2002) and HMGB1 (Fig. 17). Because transposase has higher affinity to the internal binding sites within the transposon inverted repeats (Fig. 8B), it appears that the order of events that take place during the very early steps of transposition is binding of transposase molecules first to the inner sites, and then to the outer sites. The pronounced effect of HMGB1 on binding of the transposase to the inner sites suggests that HMGB1 enforces ordered assembly of a catalytically active synaptic complex. If any of these molecular requirements is not fulfilled properly, the transposition reaction is hampered or does not proceed at all. Indeed, replacement of the outer transposase binding sites with the inner sites, i.e. increasing binding at the outer sites, abolishes *SB* transposition (Cui et al., 2002). An assembly pathway similar to the one we propose for *Sleeping Beauty* has been described for bacteriophage lambda. The integrase protein, together with IHF, first assembles on a high affinity attachment site (attP) on the phage genome (Richet et al., 1986), and then captures another, low affinity site on the bacterial chromosome (attB) (Patsey et al., 1995). In this system, the order of assembly is determined by the difference in affinity of the integrase for the attP and attB sites.

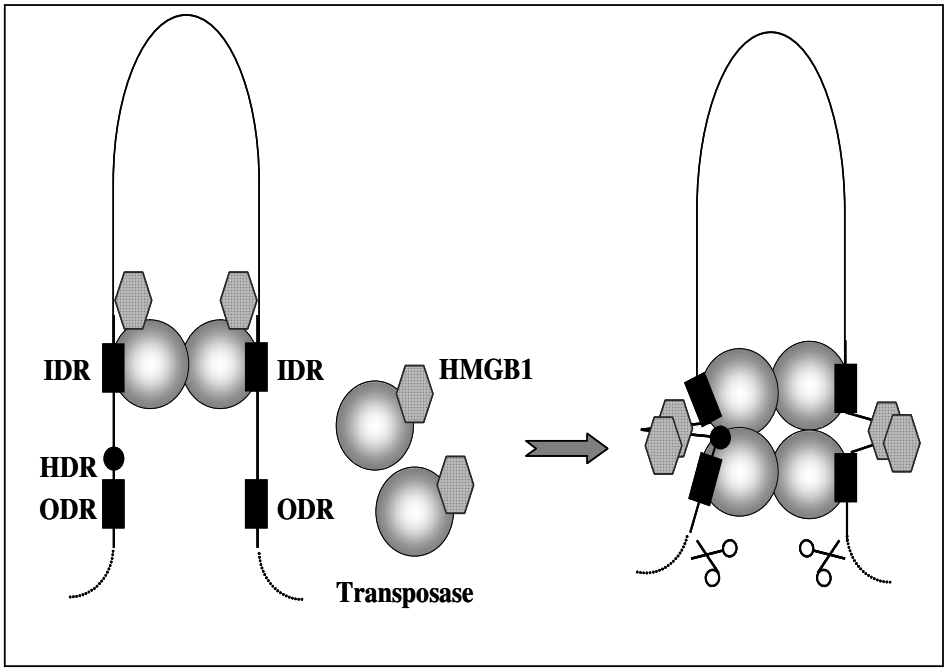

Fig. 17. A proposed model for the role of HMGB1 in *Sleeping Beauty* synaptic complex formation. Sleeping Beauty transposase (gray spheres) recruits HMGB1 (dotted hexagons) to the transposon inverted repeats. First, HMGB1 stimulates specific binding of the transposase to the inner binding sites (IDRs). Once in contact with DNA, HMGB1 bends the spacer regions between the DRs, thereby assuring correct positioning of the outer sites (ODRs) for binding by the transposase. Cleavage (scissors) proceeds only if complex formation is complete. The complex includes the four binding sites (black boxes), the HDR enhancer sequence (black circle) and a tetramer of the transposase.

4.1.2 HMGB1 has overlapping, but distinct roles in V(D)J recombination and *SB* transposition.

In V(D)J recombination, the RAG1/2 complex specifically binds to the nonamer and heptamer motifs of the RSSs (Fig. 2C), that are separated by 12 or 23 base pair spacer regions (12/23-RSS). V(D)J recombination preferentially takes place between a 12-RSS and a 23-RSS, which is termed the 12/23 rule (Van Gent et al., 1997; Aidinis et al., 1999; Hiom et al., 1998). HMGB1/2 significantly stimulates the binding of both RSSs, but this stimulation is especially pronounced at the 23-RSS (Van Gent et al., 1997). This selective enhancement of

binding is thought to enforce the specificity of the subsequent cleavage step (Van Gent et al., 1997; Aidinis et al., 1999). Recent results indicate that the RAG1/2 complex first assembles on a single RSS, and that the partner RSS is later incorporated into the complex as naked DNA (Jones et al., 2002). Initial binding of RAG1/2 to the 12-RSS results in more faithful adherence to the 12/23 rule. Because there is no substantial difference in the binding affinity of RAG1/2 for naked 12- and 23-RSSs in the presence of HMGB proteins, it has been suggested that chromatin structure may influence whether RAG1/2 binds first to a 12 RSS or a 23 RSS *in vivo* (Jones et al., 2002).

The transposase-binding sites of *Sleeping Beauty* resemble the RSSs in their sequence (Fig. 2C). Similarly to the RSSs, the spacing between the nonamer and heptamer-like motifs within the transposase-binding sites is different, 12 and 14 bps, in the internal and external DRs, respectively. It was found that SB transposase preferentially binds the inner DR (12DR) (Fig. 8B). The two-base-pair difference in spacer length between 12DR and 14DR might not be sufficient for HMGB1 to assert its DNA-bending activity to promote transposase binding. More likely, the helical phasing of the heptamer- and nonamer-like sequences in 14DR might be less favourable for transposase binding. In contrast to V(D)J recombination, the original preference of the SB transposase for binding to the 12DR is not altered, but even further emphasised in the presence of HMGB1 (Fig. 8B). In conclusion, HMGB1 seems to have overlapping, but distinct roles in transposition and in V(D)J recombination.

The IR/DR-type organization of inverted repeats is an evolutionarily conserved feature of many transposons in the Tc1 family (Plasterk et al., 1999), but its function in transposition has been enigmatic. The presented results suggest that the IR/DR introduces a higher level regulation into the transposition process: the repeated transposase binding sites, their dissimilar affinity for the transposase, and the effect of HMGB1 to differentially enhance transposase binding to the inner sites are all important for a geometrically and timely orchestrated formation of synaptic complexes, which is a strict requirement for the subsequent catalytic steps of transposition.

4.2. Modification of both of the transposon and transposase increases the functional activity of *SB* transposon system

One way to obtain proteins with novel functions and properties is to improve upon currently known natural proteins by artificial selection to phenotypes that are unlikely to have been selected in nature. One promising strategy is the change of the gene structure either by specifically replacing some of the non-essential amino acids or by the introduction of random changes in the gene structure and selection for the best characteristics. Increasing of the activity of some proteins and enzymes have been reported using directed evolution, like in the case of the ampicillin resistance gene (Yano and kagamiyama, 2001) and the aspartate aminotransferase enzyme (Yano et al., 1998). Mutants of green fluorescent protein (GFP) generated by either random mutagenesis by error-prone PCR (Heim et al., 1994) or by DNA shuffling (Crameri et al., 1996) resulted in a variety of fluorescence properties that significantly extend the usefulness of GFP for molecular and cell biology applications.

Mutations are defined as changes in DNA molecules. They can change the genetic code for amino acid sequence in proteins. Though mutations are associated with negative and deleterious consequences to the host and can cause many diseases, there are examples for positive contribution of some mutations. Mutations have two broad categories: 1) induced mutations and 2) spontaneous mutations. Induced mutations are introduced by some agents like a chemical and radiation. Spontaneous mutations include point mutations, deletions, insertions, and inversion. Transposons can spontaneously cause mutations by virtue of their randomness of integration.

Hyperactive transposase versions could be obtained by several ways, for instance mutations in the Tn5 bacterial transposase led to hyperactive versions either due to the reduction of the self-inhibitory activity of intact Tn5 transposase (Wiegand et al., 1992), or the effect on the affinity of the interaction of co-translated inhibitor protein to the transposase (Weinreich et al., 1994), or an increase in the binding affinity of the Tn5 transposase to the outside ends of its own specific binding sites in the ITR DNA (Zhou and Reznikoff, 1997). The combination of these three hyperactive mutants gave an additive effect, leading to an extraordinarily active Tn5 transposase which is very efficient in transposition (Goryshin et al., 1998). Change from acidic amino acids to basic amino acids led to hyperactive transposase mutants. Some of these hyperactive mutants were due to the change from glutamic acid (E) to

lysine (K) in the case of Tn5 (Zhou and Reznikoff, 1997) as well in Himar 1 transposase (Lampe et al., 1999). This shift from acidic to basic amino acid could make a more favorable interaction between the transposase and the negatively charged DNA backbone.

In this section I discuss development of a more active *SB* transposition system (Ivics et al., 1997), by taking two experimental approaches: 1) specifically change some of the non-conserved acids in Tc1/*mariner* family from acidic to basic amino acids in the transposase 2) altering the substrate transposon structure by increasing the numbers of transposase binding site mimicking a naturally occurring *Paris* element. Secondary structure was a prime concern not to be changed during the induction of the mutation in SB transposase, the non-conserved acidic amino acids in Tc1/*mariner* family were selected and predicted secondary structures of the mutant transposase versions were checked with the PredictProtein computer program available on the internet (Rost and Sander, 1993). Amino acid changes were spanning all SB transposase domains from the N-terminal to the C-terminal (Fig. 11A). The rationale behind the change to basic amino acids of all nonconserved acidic amino acid residues is that such changes might eliminate (or at least reduce) the unfavourable charge-charge interaction between the acidic amino acid residues and the negatively charged phosphate backbone of the transposon (or target) DNA. The assumption was true in the case of the D260K mutant which reproducibly was more active than the wild-type transposase by about 30% (Fig 11B). Although I can not confirm that the hyperactive phenotype of D260K is due to more efficient protein-DNA interactions (since I did not test its binding activity *in vitro*), it is a likely explanation for it. In case of the other mutations including E6K, D10K, D17K, D68K, D86K, E92K, E93K, E158K, D164K, E174k, E216K; E321R (Fig. 11B), the transpositional activities are either abolished or are close to zero. These results suggest that the amino acids in these positions are important for the integrity of the transposase and play a crucial role in SB's activity.

The aspartic acid (D) in the position 260 is either lysine (K) or arginine (R) in other transposases including Tc1/mariner transposases (Fig. 14), which leads us to conclude that the reason of the slight enhancement could be that lysine can better function in that sequence context. It is possible, that a particular version of fish Tc1-like transposases did contain K at position 260, but this amino acid got replaced at some point during transposase evolution, because it is functionally non-essential for the transposase. Therefore, subsequent replacement of D260 with a positively charged amino acid does not change the overall secondary structure of the transposase, but has the capacity to increase its binding activity.

Expression of an N-terminally truncated inhibitor of the Tn5 bacterial transposon inhibits transposition both in *cis* and in *trans* (Johnson et al., 1982). There are indications, that N-terminal derivatives of the SB transposase can be produced *in vivo*, and these peptides can efficiently inhibit transposition (Izsvak et al., unpublished). The change from R115 to H115 can result in hyperactivity (Fig. 11B) either because it interferes with the production of inhibitor polypeptides or because the histidine residue in position 115 results in slight changes in transposase structure and/or functional interactions. Histidine can be uncharged or positively charged, depending on its local environment. Histidine is often found in the active sites of enzymes where its imidazole ring can readily switch between the uncharged and the positively charged states to catalyze the making and breaking of bonds. During the molecular reconstruction of the SB transposase gene (Ivics et al., 1997), there was a version containing a single amino acid substitution (R143C) compared to the consensus transposase sequence. The R143C mutation slightly increases transposition activity.

Upon the combination of the individual hyperactive mutations, a synergistic effect was seen, which was not dramatic, but nevertheless indicates a sign of cooperativity. The R115H/D260K and R115H/R143C double mutants exhibited about 3.7- and 3.5-fold increase in transposition activity over wild-type, respectively (Fig. 12). It seems that the R115H mutation is important in the double mutants since the R143C/D260K mutant showed only about 2.6-fold increase in activity, and the triple mutant was nearly 2.5-fold more active than the wild-type transposase. These observations argue that the synergistic effect of R115H is relatively more pronounced upon the combination with either D260K and R143C (Fig 12).

Proline residue is a secondary structure breaker in proteins, and by virtue of its nature; it has been widely used to modify the conformation of protein structure. In case of the Tn5 transposase, enhanced binding of the transposase to the transposon DNA was seen by introducing a proline residue which is thought to interrupt the interference between the N-terminal and the C-terminal regions of the transposase (Davies et al., 2000). That led to the discovery of a hyperactive version of Tn5. In the lack of structural knowledge about the SB transposase, we can only infer the presence of certain structural motifs and predict functions for them. By looking at the predicted secondary structure of the SB transposase, there is a helix spanning the region between the N-terminal DNA-binding domain and the catalytic domain (Fig. 11A). This helix is conserved in the Tc1 family (Fig. 13), and is not part of the

DNA-binding domain (Ivics et al., 1997). However, when present in recombinant transposase polypeptides, the helix interferes with efficient DNA-binding (Izsvak et al., unpublished), supporting the hypothesis that the helix promotes an unfavorable conformation of the transposase. The idea was to change every second amino acid (only one at a time) in the predicted helix to proline. A change of one of the amino acids in this helix to proline could enhance the binding affinity of the transposase to its substrate and subsequently elevate the activity of SB transposase. Unfortunately, the mutants that I generated all showed severely impaired transposition (Fig. 11B), indicating the functional importance of this helix in *SB* transposition. Future work should be directed to a random mutagenesis of the helix in the hope that less drastic changes could result in hyperactive phenotypes.

The structure of the sandwich transposon is somewhat similar to that of the Tn5 bacterial transposon. Tn5 is thought to have originated by insertions of two insertion sequences on both sides of an immobile segment encoding resistances to kanamycin and streptomycin (Berg et al., 1984). This situation can also arise in other transposition systems, resulting in new, composite, mobile elements. Indeed, a pair of *Paris* elements that flank a nonrepetitive sequence (longer than 10kb) in an inverted orientation was shown to be able to transpose in *Drosophila virilis* (Petrov et al., 1995).

Why does the SA vector transpose long transgenes better than the wild-type *SB* transposon? In general, long elements tend to transpose less efficiently than short ones, likely because the ends of long elements cannot pair easily during synaptic complex formation. We suggest that an increase in the number of transposase binding sites (from four to eight) can partially rescue synaptic complex formation of long elements, presumably due to the more pronounced action of transposase-transposase interactions and HMGB1 at the transposon inverted repeats (Zayed et al. 2003). Example for that is the large Mu transposon that needs a tetramer of the Mu transposase, multiple transposase binding sites within the Mu genome and the HU protein, the prokaryotic counterpart of HMGB1. These components are absolutely needed for the nucleoprotein synaptic complex assembly for Mu transposition (Lavoie et al., 1990).