Introduction: My excitement to use evolutionary approaches to study biochemistry and my studies of the RAG complex led to my interest in the evolutionary origins of adaptive immunity. V(D)J recombination is the process responsible for generating the massive diversity of antigen receptors that characterizes the vertebrate immune system. RAG1 and RAG2, the protein products of *recombination activating genes 1 and 2*, cooperate to initiate V(D)J recombination in lymphoid cells by making double-stranded breaks at recombination signal sequences (RSSs)[(1)]. The recombinational DNA rearrangements catalyzed by RAG have long been biochemically likened to the cleavage reactions effected by transposases (TPs)[(1)]. In 1998, the demonstration that RAG displays transposition activity *in vitro* strongly suggested that this likeness can be explained by homology and that RAG is a descendant of an ancient transposable element[(1)]. Due to extensive sequence divergence, a close homolog of RAG within the modern diversity of TPs evaded detection until targeted PSI-BLAST searches linked the RAG1 core to the *Transib* family of TPs[(2)]. Subsequent biochemical analysis of *Hztransib*, a *Transib* transposon active in the genome of the corn earworm, revealed that, like RAG, *Hztransib* TP cleaves DNA through nicking and hairpinning steps that produce blunt transposon ends and hairpinned flanking ends[(3)] (*Fig. 1*). Additionally, insertion events create CG-rich 5-bp target-site duplications, as is typical for RAG[(3)]. These results conform to expectations of a RAG-like TP, but it is reasonable to suppose that some of RAG's properties are specific to V(D)J recombination and do not describe an ancestral TP. I propose to conduct an exhaustive biochemical analysis of the *Transib* transposon in the Yale Department of Molecular Biophysics & Biochemistry, in the laboratory of David Schatz, who discovered and biochemically characterized RAG1 and RAG2. Biochemical similarities between RAG and *Transib* can lend further support to their homology. Biochemical differences can suggest which functional aspects of RAG are evolutionarily recent recombinase-specific innovations, perhaps due to association of RAG1 with other factors (e.g. RAG2) or perhaps due to structural changes within the endonuclease itself.

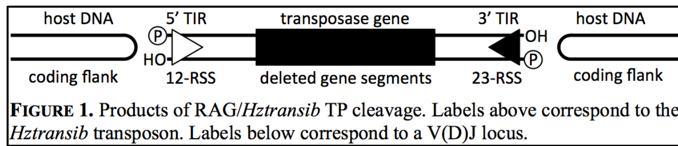

Figure 1. Products of RAG/*Hztransib* TP cleavage. Labels above correspond to the *Hztransib* transposon. Labels below correspond to a V(D)J locus.

Aim 1: Determine the substrate requirements for *Hztransib* TP activity. Each RSS comprises a conserved heptamer and nonamer separated by a nonconserved spacer of either 12 or 23 bp[(1)]. RAG's activity is governed by the 12/23 rule: cleavage can only occur if both a 12- and a 23-RSS are present[(1)]. *Hztransib* TP has already been demonstrated to cleave at paired 12/23 RSSs (unpublished data in the Schatz lab), but other RSS combinations have not been tested. To evaluate *Hztransib* TP's adherence to the 12/23 rule, I will incubate purified *Hztransib* TP protein with DNA substrates containing various combinations of 12- and 23-RSSs, and I will determine the efficiency of cleavage by visualizing and characterizing radiolabeled DNA products on a denaturing polyacrylamide gel. In this and all other described experiments, a negative control reaction will contain no endonuclease, and a positive control reaction will use RAG as the endonuclease. RAG activity is highly dependent on conservation of the first 3 bp of the heptamer (CAC), while flanking sequences have little effect on cleavage efficiency[(1)]. I will investigate the precise sequence requirements of *Hztransib* TP by quantifying cleavage of DNA substrates with various point mutations in the RSSs and their flanking DNA. The sequences with greatest cleavage efficiency will likely approximate *Hztransib*'s own terminal inverted repeats (TIRs), which resemble RSSs and begin with the same CAC sequence. Accordingly, for all described experiments, I will compare reactions that use RSS-containing substrates to reactions using TIR-containing substrates to determine the sequence dependence of any effects I observe.

Aim 2: Determine structural characteristics of *Hztransib* TP's catalytic state. RAG can nick

individual RSSs, but completion of cleavage via hairpin formation can only occur in a synaptic complex containing a 12- and a 23-RSS[1]. To assay *Hztransib* TP for nicking and hairpinning activity in the absence of synapsis, I will immobilize low concentrations of biotinylated DNA substrates containing a single 12- or 23-RSS on streptavidin agarose beads, and I will characterize products after addition of TP. I will then add free DNA substrates to the slurry to assay for cleavage activity with specific synaptic pairings. RAG's cleavage efficiency is greatly enhanced by the DNA-bending high-mobility-group protein HMGB1 because cleavage requires DNA distortion[1]. I will add HMGB1 to standard *Hztransib* TP cleavage reactions and observe its effect on cleavage efficiency. Following RAG cleavage, the four newly created DNA ends remain synapsed in a postcleavage complex[1]. To probe for an *Hztransib* postcleavage complex, I will biotinylate specific DNA ends, pull down biotinylated cleavage products with streptavidin agarose beads, and characterize any unbiotinylated DNA species that are also pulled down.

Aim 3: Determine secondary nuclease activities of *Hztransib* TP. *In vitro*, RAG exhibits various nuclease activities besides cleavage at RSSs: it cleaves single-stranded heptamers, it cuts off 5'-ended overhangs on duplex DNA, and it removes 3'-terminated single-stranded flaps[1]. By incubating *Hztransib* TP with representative radiolabeled substrates and characterizing products, I can determine whether *Hztransib* TP also exhibits these activities.

Aim 4: Suggest catalytic and regulatory roles for RAG2. While RAG1 requires RAG2 for activity[1], *Hztransib* TP bears sequence similarity only to RAG1[2] and is able to effect cleavage without supplementary protein factors[3]. To elucidate RAG2's role in V(D)J recombination, I will include in each of the previously described experiments an additional reaction containing both *Hztransib* TP and RAG2. If RAG2 enhances a RAG-like biochemical property of *Hztransib* TP, that property may have evolved due to recombinase-specific selection pressures.

Challenges: As of yet, cleavage activity in low-purity *Hztransib* TP preparations from another lab has been observed only after addition of Mn^{2+} [2], which deregulates RAG endonuclease activity when substituted for the physiological electrophile Mg^{2+} [1]. The Schatz lab has ample experience developing expression constructs and purification/reaction protocols for RAG, expertise that can now be applied to *Hztransib* TP to increase purity and, I predict, allow cleavage with Mg^{2+}. *Hztransib* may not represent the entire *Transib* family in all details; whereas several other *Transib* transposons contain V(D)J-like asymmetric TIRs[2], *Hztransib* has symmetric TIRs. I will use my background in computational sequence analysis to identify and conduct experiments with an active *Transib* transposon bearing asymmetric TIRs, allowing requirements of asymmetric synapsis to be evaluated both with RSSs and with the transposon's own TIRs.

Broader Impacts: I will recruit undergraduate mentees from my classes and from oSTEM to get involved in this work, capitalizing on the multifaceted nature of the project to teach them to approach problems from various angles. Additionally, because the USA currently falls far behind other scientifically advanced nations in popular acceptance of evolutionary theory, I will present the exciting history of this tamed transposon at high school teacher conferences to encourage DNA-level approaches to evolution pedagogy, obviating higher-order misinterpretations.

Intellectual Merit: Because V(D)J recombination is an essential step in the development of antigen-specific lymphocytes, a complete functional comparison between *Transib* TP and the RAG complex would strengthen the current model for adaptive immune system development. It would also offer clues as to how early organisms acquired pathogen defense capabilities, a significant evolutionary hurdle that, once cleared, initiated a dramatic increase in organismal complexity.

References: (1) Gellert M. 2002. *Annu Rev Biochem* 71: 101-32. (2) Kapitonov VV, Jurka J. 2005. *PloS Biol* 3: e181. (3) Hencken CG, Li X, Craig NL. 2012. *Nat Struct Mol Biol* 19: 834-6.