Preface

Nowhere do the cooperative powers of DNA sequencing, high-resolution protein structure, biochemistry and molecular genetics shine more intensely than on Mobile DNAs. In Mobile DNA II, we knew that almost half of the human genome is comprised of retroelements. What discoveries since Mobile DNA II could surpass that claim? Very simply: everywhere DNA is dynamic, and we now meet the elegant protein machines, co-evolved DNA partners, and diverse RNA choreographers. These pages hold something for every reader, beginning with the introductory overview of mechanisms. Novices will find some of the most lucid reviews of these complex topics available anywhere. Specialists will be able to pick and choose advanced reviews of specific elements, but will be drawn in by unexpected parallels and contrasts among the elements in diverse organisms. Biomedical researchers will find documentation of recent advances in understanding immune-antigen conflict between host and pathogen. Biotechnicians will be introduced to amazing tools for *in vivo* control of designer DNAs. And long-time aficionados will simply fall in love all over again.

Questions still abound about the Transposable Elements (TE) described in this volume. Perhaps none is more profound than the basis and consequences of TE diversity even among related genomes. Active DNA TE show perhaps the most disparate distribution among organisms, being dominant in prokaryotes, and in some animals, including some insects and fish, but with the exception of certain bats, virtually absent in mammals. Plants illustrate expansion of genomes, mediated not only by increasing ploidy, but also by expansion of DNA-based TE and Long Terminal Repeat (LTR) retrotransposons. Although reverse transcriptases are found throughout all kingdoms, autonomous retroelements simply explode together with their non-autonomous partners in mammals with remarkably species-specific types. These differences in mobile DNAs define self and mate, sister species, host and pathogen.

The most striking impression from these pages must be the raw power of genetic material to refashion itself to any purpose. DNA exchange between bacteria and their environment blurs the boundaries between host, transposon, and phage, as organisms secrete and take up DNA, stash genetic material in integrons for future use, conjugate, are attacked by phage and fight back. Delving into mechanisms, we see single-stranded hairpin structures and G quartets that anchor rearrangements in multiple ways; chemically diverse nucleophilic centers–hydroxyls couched in pentose, tyrosine or serine moieties–that covalently bond or attack directly in strand-transfer reactions. Proteins act as clamps to topologically constrain DNA or act as mechanical swivels, linking and unlinking mobilizing strands. Resolution of transposition intermediates might also involve host replication or recombination machinery. More recently discovered helitrons offer unexpected opportunities for expansion of DNA-based elements by rolling-circle replication.

RNA, the primal, catalytic nucleic acid, is in evidence everywhere. In retroelements, RNA partners with reverse transcriptase to deliver on transcriptional expansion of autonomous and non-autonomous TE sequences. Group II introns in bacteria likely gave rise to eukaryotic organelle group II self-splicing, retro-homing introns, Long INterspersed Elements (LINEs), telomerase reverse transcriptases and in addition, spliceosomal introns. Phylogenetic analysis of bacterial genomes previously revealed group II introns, diversity-generating retroelements Diversity-Generating Retroelements (DGR), and retrons, but next generation sequencing now identifies a multitude of novel reverse transcriptases of unknown function.

In ciliates, *Paramecium, Tetrahymena* and *Oxytrichia,* RNA directs massive genome reduction between germ-line and somatic nuclei, mediated by ancient transposase-like enzymes. LINEs containing restriction-enzyme like- or AP-endonucleases dominate in some eukaryotic cells. Others are dominated by LTR retrotransposons and their offspring, the retroviruses; stripped down Penelope-like elements with GIY-YIG endonucleases; DIRS elements with tyrosine recombinases: and attendant non-autonomous elements.

Exceptional elements provide evidence for the interaction of domains over evolutionary time, including LTR retrotransposons encoding envelope proteins, retroviruses replicating intracellularly, and DIRS elements in which retroelement RT/RNaseH is associated with a Crypton-type DNA element tyrosine recombinase.

Nowhere is the sharp focus of structural biology and biochemistry more apparent than in studies of key retroelement enzymes reverse transcriptase and integrase motivated by the quest for inhibitors of human immunodeficiency virus (HIV) replication. Reverse transcriptase structures for multiple retroviruses, as well as now one retrotransposon, demonstrate the robustness of the palm, thumb, fingers model. However, as a caution against too much generalization, subdomains are re-arranged in monomeric, homodimeric, and heterodimeric forms in different enzymes, and catalytic activities operate in *cis* or *trans* within the complex, depending on the enzyme. The structure of full length retrovirus integrase notoriously resisted high resolution structural analysis, but now has rewarded efforts of many labs with key insights (cover of this volume). These include a surprising dimer-dimer interface where active sites are juxtaposed to a trapped, and dramatically bent and widened, major groove target. Whereas LTR retrotransposons target integration to transcriptionally-repressed regions through interactions with heterochromatic protein domains or Pol III-transcribed genes thought to repress Pol II transcription, next generation sequencing has surfaced less dramatic, but significant, retrovirus integration bias, favoring transcriptionally-active regions.

This distribution has been shown now in two cases to be mediated by interactions between integrase and epigenetic mark-associated proteins.

While it has been argued that mobile elements are "selfish DNAs", these pages are replete with examples of the positive contributions of mobile elements to host genome function. Bacterial transposons encode and mobilize selectable markers including antibiotic resistance, detoxifying enzymes, and conjugation and virulence functions. In eukaryotic cells, mobile elements contribute to chromosome structure: constituting centromeres or telomeres in some organisms and seeding heterochromatin in others. TE constitute a large fraction of transcription factor binding sites and provide an ongoing source of novel combinations which are responsive to stress signaling, MAP kinase activation and other developmental signals. Insertions of LINEs and Alu elements affect RNA processing because they encode cryptic splice sites, termination signals, and can target RNA editing.

Exapted mobile DNA coding sequences appear in novel contexts: transposases have evolved into the RAG endonuclease for V(D)J immunoglobulin gene diversification and into heterochromatic factor CENP-B; a reverse transcriptase evolved into telomerase; retrovirus envelope proteins became the trophoblast fusion protein syncytin. There are other examples of TE Open Reading Frames (ORFs) under selection, but with, as yet, unknown functions. Endogenous retroviruses forego prior allegiances and join strategies to resist new infections. For example, Fv–1, a retroviral Gag relic, thwarts incoming retroviruses of similar type. Repeated TE sequences are susceptible to DNA rearrangements via non-allelic recombination, aborted transposition, and generation of pseudo-genes–all of which might ultimately contribute to the resiliency of host genomes.

TE exploit their hosts as well. The bacterial XerCD tyrosine recombinases which function in bacteria to unlink multimeric chromosomes are exploited to integrate phage genomes or mediate invasion of the host chromosome on behalf of certain plasmid-borne mobile elements. Transposases, resolvases and integrases *in vivo* likely associate with host factors as they are joined with host genomes. TE are generally tightly controlled by host regulation so that some display opportunistic bursts of activity during specific windows of development. This is exemplified by yeast Ty transcription in response to MAP kinase signaling and activation of reverse transcription by DNA checkpoints. A common theme more generally is TE activation during stress. Diverse retroelements are derepressed during periods of germ cell development ensuring their vertical spreading in populations.

Despite these examples of cooperation, mobile DNAs are also in conflict with their hosts. RNA interference likely arose in part to combat mobilization of retroelements. Invaders of one sort or another engage in a dizzying unscored dance with their hosts. One result of this conflict is rapid evolution of genes encoding host innate immunity restriction factors, which for retroviruses include ones that prematurely uncoat incoming viruses, starve reverse transcriptases for nucleotides, and deaminate cytidines in replicating cDNA. Some of these same factors also suppress movement of endogenous retroelements.

Programmed variation is used by invaders and hosts alike for purposes of immune evasion or resistance, respectively. Examples include *Salmonella* Hin invertase flipping a promoter sequence to switch between expression of different antigenic flagellar proteins and DGR directing mutagenic reverse transcription of a template transcript coupled with directed conversion of a target expression site. *Neisseria gonorrhea*, *Borrelia burgdorferi*, *Trypanosoma brucei*; and *Plasmodium falciparum*, agents of gonorrhea, Lyme disease, sleeping sickness, and malaria, respectively, use amazingly diverse mechanisms to program variation of their antigenic surfaces for immune evasion. To counter this assault, there

is programmed variation of host immune proteins. In human immunoglobulin production, a DDE TE-derived RAG site-specific endonuclease initiates V(D)J switching, followed by transcription-activated somatic hyper-mutation (activation-induced cytidine deaminase), nuclease introduction of DSB, and final joint formation by redundant NHEJ pathways.

Next generation sequencing and development of methods for rapid TE mapping have greatly improved understanding of the distribution of TE as well as the utility of transposons for functional genomics. The bacterial Tn5 system has been exploited in particular for *in vitro* mutagenesis and next generation sequencing libraries by collapsing fragmentation and adapter ligation into a single step. P, Hermes, piggyBac, and Sleeping Beauty transposons have wide activity in eukaryotic systems and have been harnessed for genome-wide profiling, gene disruption and tagging, and genome modification. Retroviruses are additionally used in lineage tracing. The controlled, high-frequency mobilization of Mutator has made it indispensable for gene discovery in maize.

In medical research, understanding the impact of DNA mobilization is critical. In addition to individual TE, other mobile DNAs such as plasmids, Integrative Conjugative Elements (ICE) and both transposon-borne and chromosomal integrons are bacterial reservoirs of mobilizable antibiotic resistance. HIV, malaria, and sleeping sickness, and other pathogens, too numerous to mention here, remain threats to global health. Mobile element vectors transposons piggyBac, Sleeping Beauty, lenti-retroviruses and adenoviruses are being used as vectors to introduce exogenous DNAs in research, and in clinical trials. They differ with respect to targeting, excision footprints, payload size, and host activity profiles. Their mechanisms of DNA breakage and joining were among the systems first analyzed, now enabling them to be harnessed and used extensively for genome engineering including with developmentally-regulated expression, inactivation, and self-deletion strategies to enable probing essential or tissue- specific functions.

What challenges remain? One goal is to connect key findings from basic research, to clinical developments in drug resistance and genetic engineering. This volume is based on the considerable increase in understanding of molecular mechanisms of mobilization in the last decade. However, we have likely seen only the tip of the iceberg of how mobile DNAs affect the day to day biology of their hosts.

In the human genome alone, retroelements provide promoters for long non-coding and other RNAs of completely unknown function; Alu elements redirect RNA processing and delivery, and mobilization is occurring during neuronal development and in cancer with unknown consequences, just to mention a few. Finally, endogenization of a gamma retrovirus in Australian koalas is ongoing and those studies should provide insights into retroelement-host interaction. How have transposition events after separation from other great apes contributed to traits that make us human? What transposition events will provide key substrates for future evolution? And of course, perhaps the ultimate question, could we survive as a species were there no transposition?

We give our heartfelt thanks to all the authors who contributed diverse and fascinating chapters to Mobile DNA III. We express special thanks to Patti Kodeck, Administrative Assistant to Editor in Chief N. Craig, who mediated recruitment of and communications with authors and interactions between them and the publisher. Finally, our most sincere thanks to all of our supporters at ASM Press for their dedication in producing this volume, but especially to: Gregory Payne, Senior Editor; Larry Klein, Production Editor; Christine Charlip, Director of Administration; and Cathy Balogh, Administrative Assistant for Production.