Integrons and the Origin of Antibiotic Resistance Gene Cassettes

Super integrons with thousands of gene cassettes may have set the stage for pathogens to develop antibiotic resistance very rapidly

Didier Mazel

Integrons, which are natural genetic engineering platforms, can incorporate open reading frames and convert them to functional genes by ensuring correct expression. Although such integrons play a major role in multidrug resistance phenomena among gram-negative species, similar platforms occur in numerous bacterial species, where they can play other roles. Meanwhile, we have evidence suggesting that they are a major source of multidrug resistance integrons and their resistance gene cassettes observed in clinical isolates.

Recruiting exogenous genes represents a rapid adaptation against antimicrobial compounds, and the integron functional platform seems perfectly suited for capturing those specific genes that enable bacterial pathogens to face challenges posed by multiple antibiotic treatment regimes. With the discovery of the superintegron and the thousands of cassettes entrapped in integrons of environmental species, we now have a better view of the immense resources embedded within this system. However, several questions of importance about this system are still without satisfactory answers, including insights about specific recombination processes, cassette genesis, and cassette exchange dynamics within complex bacterial populations.

Antibiotic Resistance-Encoding Integrons

With few exceptions, antibiotic resistance in bacterial pathogens was identified soon after particular drugs were introduced into clinical practice, illustrating the genetic flexibility of bacteria. For instance, pathogens developed resistance to sulfonamides (Su) and penicillin in the late 1930s, and in the late 1940s, mycobacteria developed resistance to streptomycin.

These single-resistance phenotypes were not entirely unforeseen. For example, laboratory studies had demonstrated penicillin-resistant point mutants. In contrast, investigators did not anticipate multidrug resistance because the co-appearence of multiple mutations conferring such phenotypes was considered to be beyond the potential of a given bacterial population. However, in 1956 in Japan, six years after physicians began prescribing—and companies began massively producing—the antibiotics streptomycin (Sm), tetracycline (Tc), and chloramphenicol (Cm), investigators were observing isolates of *Shigella dysenteriae* resistant to up to four antibiotics simultaneously (Tc, Cm, Sm, Su).

This emergence of multiple resistant strains could not be attributed to mutation alone. It was soon established that bacteria were acquiring genes that confer resistance, relying on that means to escape antimicrobial activity rather than on mutations arising in resident genes. At that time, transposons and integrons began appearing among human gram-negative pathogens as well as the commensal bacterial population, playing a key role in disseminating resistance genes while hitchhiking on conjugative plasmids and interspecies transfer.

Nevertheless, integrons were not formally identified as agents of resistance gene recruitment until the late 1980s, following the observation that loci in transposons and resistance (R)-plasmids expressing different antibiotic resistance spectra share the same genetic backbone and differ only in the resistance genes that...
they harbor. Regardless of this lag in awareness, however, those integrons were certainly part of the first multidrug resistance outbreaks in the 1950s. One proof is that Tn21, an integron-containing transposon, was involved in the resistance phenotype propagated by plasmid NR1 (R100) in the very first events in Japan.

In light of the six-year time scale for the emergence of multidrug-resistant Shigella strains in Japan in 1956, it was hardly disputable that bacteria were already equipped with appropriate genetic tools for meeting the challenge of multidrug assaults. Indeed, molecular studies during the last 15 years show how powerful the bacterial integron recombination machinery is, despite its relative functional simplicity.

Thus, all known integrons are composed of three essential elements for procuring exogenous genes: (i) a gene coding for an integrase (intI), (ii) a primary recombination site (attI), and (iii) a strong promoter (Pc). Integron integrases recombine, in a recA-independent manner, discrete units of circularized DNA known as gene cassettes downstream of the resident Pc promoter at the proximal attI site, permitting expression of their encoded proteins (Fig. 1). Even considering only those that differ in nucleotide sequence by more than 5%, more than 70 different resistance cassettes have been described, and they confer resistance to all beta-lactams, all aminoglycosides, chloramphenicol, trimethoprim, streptomycin, rifampin, erythromycin, and antiseptics of the quaternary ammonium compound family.

Moreover, all these integron-inserted cassettes share several specific structural characteristics. For instance, the boundaries of each integrated cassette are defined by two GTTRRRY (core-site) sequences in the same orientation, which are the targets of the recombination process. The integrated cassettes also generally include a single gene and an imperfect inverted repeat located at the 3’ end of the gene called an attC site (or “59-base element”), a diverse family of sequences that function as recognition sites for the site-specific integrase. The attC sites vary from 57 bp to 141 bp in length, and their nucleotide sequence similarities are primarily restricted to the inverse core-site and the core-site (Fig. 2).

Three classes of integrons (RI) involved in multidrug resistance expression were defined on the basis of homology of their integrase genes. Each class appears capable of sharing and acquiring the same gene cassettes. Several cassettes appear to belong in two different classes of integrons, and the class 1 integrase can recombine several structurally diverse attC sites. The IntI integrases belong to the catalytic family of the tyrosine (Y) recombinases that are involved in the movement of numerous phages through site-specific recombination (such as the lambda phage integrase, λInt) or in fundamental cellular processes such as chromosome dimer resolution in cell division (XerC/D). In spite of these rela-
tionships, the integron integrases have a unique characteristic among the Y-recombinases, as they recombine sequences that are only remotely related.

The RI platforms are defective for self-transposition, but this defect is often complemented through association with IS, transposons, and/or conjugative plasmids which can serve as vehicles for the intra- and interspecies transmission of genetic material. With this system, bacteria are capable of stockpiling exogenous genetic loci to establish an appreciable antimicrobial armamentarium. Some bacteria harbor up to eight different resistance cassettes in a single integron.

Another Type of Integron: the Chromosomal Superintegrons

A pivotal question is, what are the origins of the RI and their cassettes? The degree of homology between the three RI integrases (45–58%) suggests that their evolutionary divergence extends longer than the 50 years since antibiotics came into use. In the late 1990s, studies examining the relationship between RI gene cassette arrays and a cluster of repeated sequences identified in the Vibrio cholerae genome, called VCRs (V. cholerae repeats), led to discovery of another type of integron, a superintegron (SI) (Fig. 2). This distinct type of integron is now known to be an integral component of many g-proteobacterial genomes.

The integron discovered in chromosome 2 of V. cholerae possesses a specific integrase, IntIA, that is related to the RI integrases and also is responsible for inserting coding sequences (ORFs) into a unique chromosomal attI site. However, this SI has two structural characteristics that distinguish it from other RIs—namely, the large number of cassettes that it gathers and the high homology observed among the attC sites of these cassettes, which in the case of V. cholerae are known as VCRs.

These key features plus the sedentary nature of the functional platform (IntIA + attI site) define a superintegron. Such SI structures also are found among the Vibrionaceae and their close relatives, the xanthomonads, and a branch of the pseudomonads. They share the same general characteristics, such as their large number of more than 20 cassettes and a high homology between their endogenous cassette attC sites. More importantly, they predate the antibiotic era, and have been identified in isolates collected from the 19th century.
The SI carried in three *Vibrio* species whose genomes are sequenced all show a large number of cassettes, from 72 in *V. parahaemolyticus* to more than 200 in *V. vulnificus*. In the case of the *V. cholerae* strain El Tor N16961, the SI gathers at least 216 mostly unidentified genes in an array of 179 cassettes that starts from the VchiIntIA gene and occupies about 3% of the genome. The high level of identity shared by the *attC* sites carried by the majority of these cassettes suggests that they are assembled in the SI-carrying species through the physical association of an *attC* site with an incoming DNA fragment. However, the mechanics of this process are not known.

Integrons that do not share all the characteristics of the SI have been found in the genome of two *Shewanella* species. These integrons gather only a handful of cassettes with structurally heterogeneous *attC* sites, a situation reminiscent of the RI cassette arrays. Furthermore, integron integrase-like genes have also been identified in the genomes of other proteobacteria, including *Nitrosomonas europaea*, *Microbulbifer degradans*, *Geobacter sulfurreducens*, and *Treponema denticola*. Apart from the one carried in *N. europaea*, they have not been characterized.

Using PCR primers directed against conserved regions of the integron-integrase genes and *attC* sites, Harold Stokes of Macquarie University, Sydney, Australia, and Steven Schmidt of the University of Colorado, Boulder, and their respective collaborators amplified 19 new integron inte-grases from markedly different environmental DNA samples, including from national parks or metal-contaminated soils. Unfortunately, in most cases their protocol does not permit determining the source of these integrons, be it the endogenous SI of a soil bacterium or an integron located on a mobile structure. However, their findings support the hypothesis that integrons are widespread among bacterial populations either as components of mobile DNA elements or the chromosome, and that they are not confined to pathogenic or multidrug-resistant bacteria.

**Phylogeny Indicates Integrons Are Ancient, Super Integrons Possibly Ancestral**

According to comparative analyses, the integron-integrases form a specific clade within the Y-recombinase family. Moreover, the integron platform appears to be ancient based on the species-specific clustering of the respective SI integrase genes in a pattern that mostly adheres to the line of descent among the bacterial species in which they are found. Thus, the establishment of SIs likely predates speciation within respective genera, suggesting that integrons are ancient structures that steered the evolution of bacterial genomes for hundreds of millions of years. However, it is possible that transfer of either a part or all of a SI occurred long ago, perhaps explaining some of the discrepancies observed between the SI-integrase and 16S rRNA gene trees, such as for *V. fischeri*.

Because the RIs and SIs have a common structural organization, the antiquity of SIs suggests that they are the ancestors of RIs. Hence, we proposed that RIs evolved from SIs through entrapment of *intI* genes and their cognate *attI* sites by mobile structures, such as easily and randomly assembled compound transposons. Furthermore, we noticed that 12 different resistance cassettes carry an *attC* site almost identical to those specifically found in and characteristic of *Xanthomonas* and *Vibrio* SIs (Fig. 3). Once mobile, one can imagine that the subsequent harvesting of cassettes from various SI sources led to contemporary RIs, including the great diversity of *attC* sites associated with these gene cassettes.

Two recent observations support this account of how RIs could have arisen from SIs. The first came from Henning Sørum of the Norwegian School of Veterinary Science in Oslo and his collaborators. While analyzing a plasmid associated with trimethoprim resistance in *V. salmonicida*, they found that resistance is due to a *dfrA1* cassette carried in an integron (Genbank #AJ277063) that contains a “classical” *dfrA1* cassette in second position and is followed by six cassettes. This new integron has two features that corroborate our model: (i) its integrase originated from a *Vibrio* species, or one from a closely related genus, as attested to by its 74% similarity to the *V. cholerae* and *V. mimicus IntI*; and (ii) the *attC* sites of three of the seven cassettes are structurally homogeneous, a characteristic until now only found in the SI cassette arrays. Because this integron is surrounded by IS sequences, we suspect that it corresponds to an intermediate of what could possibly became a true RI by harvesting other resistance cassettes,
using the mobility of the assembled compound transposon and multiple transfers.

The second observation from our group involves SI cassettes that could be recruited by a RI, and, moreover, confer a resistance phenotype. The majority of the examined SI cassettes appear unique to the host species, and a majority of their encoded genes have no counterparts in the database or are the sole homologs of unassigned ORFs. Nevertheless, the reservoir of adaptive functions residing within SIs includes cassettes with significant homology to known antibiotic resistance genes. We have demonstrated that class 1 integrons can randomly recruit any gene cassettes harbored within a SI. After we applied selective pressure for antibiotic resistance, we discovered a chloramphenicol acetyltransferase gene cassette, catB9, in the V. cholerae SI.

Meanwhile, another resistance cassette, coding for a novel carbenicilllinase, has been identified in the SI of another V. cholerae isolate. This cassette, like the catB9 cassette, is structurally identical to the other V. cholerae SI cassettes, as its attC site is a canonical VCR. Along with results demonstrating that SI cassettes are substrates for the integrase of class 1 integrons when present on a high-copy-number plasmid, these results suggest that environmental conditions, such as the presence of antibiotics, dictate which of the randomly recruited cassettes are retained within the RIs of clinical isolates.

It is likely that the assembly of complex RIs having more than two resistance cassettes occurred through recombination between different RI, rather than through successive direct recruitment from the SI cassette arrays of environmental species. Niches exist which could fa-
vor the exchange of cassettes between RI on a high scale. The recent characterization of a remarkable collection of cassettes, integrons, and plasmids circulating in a single wastewater treatment plant, as well as the unpredicted demonstration that aerobic, gram-positive *Corynebacteria* can be a reservoir of class 1 RI in poultry litter, indicates that environments other than obvious clinical settings exist in which cassette exchanges among RI may be common.

**Trying To Trace Sources of Resistance Genes in Integron Cassettes**

As for antibiotic resistance genes that are not carried in cassettes, key questions about their origins and how they disseminate remain unanswered. Nonetheless, evidence suggests that horizontal gene transfer among prokaryotes is a perpetual activity. So ancient is this process that determining a definitive source for any particular resistance marker will surely prove difficult—unless a recent event can be identified.

In 1973, Raoul Benveniste and Julian Davies proposed that intrinsically resistant or antibiotic-producing organisms routinely supply resistance genes to clinical isolates, and some recent findings substantiate their hypothesis. Their proposal was based in part on the structural relationship between the aminoglycoside phosphotransferase (APH), which protects the actinomycetes that produce aminoglycosides, and the APH encoded in transposons such as Tn5. The best example is certainly the strong similarity in sequence and genetic organization observed by Gerard Wright of McMaster University, Hamilton, Ontario, Canada, between the vancomycin resistance gene cluster found in enterococci and those of the glycopeptide-producing actinomycetes. While no other direct evidence is presently available, many antibiotic-producing strains other than those used industrially exist in nature and could be a source of resistance genes.

The recruitment of general housekeeping genes provides an alternate route by which resistance determinants might evolve. For instance, several examples of chromosomally encoded homologs whose functions are not related to antibiotic resistance have been identified in the genomes of *Serratia marcescens*, *Providencia stuartii*, *Streptomyces* spp., and *Mycobacterium* spp.

The marked differences in codon usage among cassettes within the same integrons indicate that these several antibiotic resistance determinants are of diverse origins. However, evolutionary rates and the passing of genes through intermediates undoubtedly created descendants that are appreciably divergent from the original loci, meaning the original source of a given resistance gene will remain elusive.

**ACKNOWLEDGMENTS**

I thank all the members, past and present, of my group. I especially thank D. Rowe-Magnus and M.-C. Ploy for their critical reading of the manuscript. This work was supported by the Pasteur Institute and the CNRS.

**SUGGESTED READING**


