

Estructura, funcionamiento y significado de los integrones bacterianos

Juan María García Lobo

Departamento de de Biología Molecular, Universidad de Cantabria.

Unidad asociada al Centro de Investigaciones Biológicas, CSIC.

E-mail: jmglobo@medi.unican.es

A lo largo del último cuarto de siglo, la atención de los que nos hemos dedicado al estudio de las bases genéticas de la resistencia a antibióticos en bacterias ha estado fijada en una serie de elementos, que al final han encajado unos dentro de otros como un juego de muñecas rusas. En los años 70 nuestro objeto de deseo eran los plásmidos, a finales de esa década y durante toda la siguiente tuvimos un nuevo sujeto de estudio, los trasposones, y en estos momentos el interés de algunos de nosotros se centra en una nueva clase de elementos genéticos, los integrones.

Definición.

Las primeras pistas sobre la existencia de esta clase de elementos se obtienen estudiando el trasposón Tn21. Analizando las funciones de recombinación de este elemento se observaron una clase de recombinantes sitio-específicos, diferentes de los producidos por la trasposasa e independientes de ésta y de los extremos invertidos del trasposón. El estudio detallado de estos recombinantes y de las secuencias requeridas para su obtención, revelaron la presencia de un elemento independiente, más tarde denominado integrón, dentro del trasposón Tn21.

Un integrón se puede definir como un elemento genético dinámico, en el que por un mecanismo de recombinación sitio específica se acumula una combinación de genes estructurales organizados como un operón. Los genes estructurales presentes en los integrones conocidos son mayoritariamente, pero no exclusivamente, genes de resistencia a antibióticos. El dinamismo de los integrones se refiere a la capacidad de los genes estructurales presentes en el integrón para escindirse en forma de círculos autónomos (no replicativos) y a la capacidad de estos círculos autónomos para integrarse en un integrón diferente.

Cada día se acumulan nuevos datos que nos indican que los integrones son ubicuos, sobre todo en plásmidos o cromosomas de bacterias Gram negativas, pero también se han descrito en Gram positivos, y quizás el hallazgo más sorprendente fue el encontrar un integrón en el cromosoma de una cepa de *Mycobacterium fortuitum*.

Estructura genética

Para la actividad de los integrones se requiere un gen que codifica una recombinasa sitio específica, la integrasa. Esta enzima cataliza la recombinación sitio-específica entre dos secuencias cortas de DNA que pueden ser de dos clases: *attI* o *attC*, que son los sitios primarios de reconocimiento de la integrasa.

En la actualidad se habla de hasta cuatro clases de integrones, que se distinguen por la integrasa que codifican. Sin embargo puesto que los más comunes son con mucho los integrones de tipo I, la mayor parte de este texto se referirá a ellos salvo referencia expresa a alguno de los otros grupos.

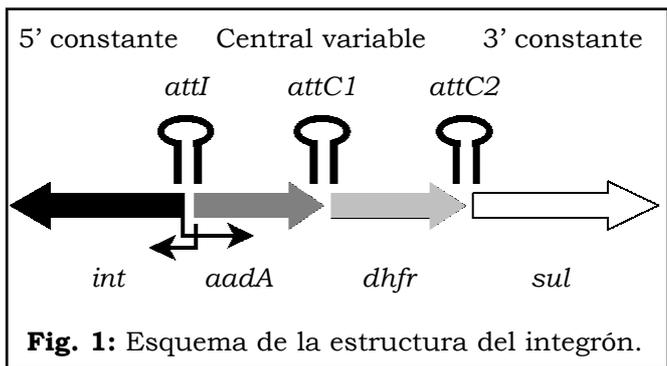


Fig. 1: Esquema de la estructura del integrón.

Estructuralmente se suele hablar de tres regiones en un integrón (Figura 1): Una **región constante 5'** que contiene básicamente el gen de la integrasa que se transcribe de derecha a izquierda. En el extremo 5' del gen de la integrasa se encuentra una región con dos promotores divergentes que controlan respectivamente la transcripción del gen de la integrasa y la de los genes estructurales situados a la derecha. A continuación nos encontramos una **region central variable** en la que se localizan los genes estructurales del integrón (casi siempre genes de resistencia a antibióticos) en número variable que suele oscilar de uno a cuatro, aunque conocemos integrones descargados (In0) que no contienen ningún gen en la región central. Las regiones constante 5' y central variable están separadas por la secuencia *attI*, uno de los sitios de reconocimiento de la integrasa. Los genes constitutivos

en la región central están separados unos de otros por secuencias *attC*. Una característica de esta región central es su compacidad genética. Las regiones no codificantes suelen ser siempre muy cortas, generalmente menos de 10 pb. De este modo es posible la organización de todos los genes en una sola unidad transcripcional controlada por el promotor presente en la región constante 5', aunque las secuencias *attC* intergénicas pueden actuar como terminadores parciales de la transcripción. Dependiendo de los genes presentes en la región central se ha propuesto una nomenclatura para los integrones In0, In1, In2... No obstante, la enorme diversidad que se está encontrando en la región central de los integrones hace este sistema poco práctico, y resulta más sencillo hacer una descripción de los integrones por los genes presentes en la región variable.

A continuación de la región central encontramos otra **región constante 3'** en la que encuentra un gen de resistencia a sulfamidas y otros dos marcos abiertos de lectura.

El integrón como elemento trasponible.

Al estudiar en los trasposones de la familia del Tn21, las secuencias colindantes con las regiones constantes del integrón que acabamos de definir se encuentran genes para una trasposasa y dos repeticiones invertidas flanqueando el conjunto que constituyen un trasposón dentro del Tn21. En algunas ocasiones este hallazgo ha llevado a definir el integrón como la región comprendida entre estas dos repeticiones invertidas. Sin embargo, es más correcto considerar esta región como un trasposón, dentro del cual se encuentra un integrón. Este trasposón es activo y de hecho, junto con el Tn21 que lo alberga, puede ser responsable de la gran diseminación de los integrones en enterobacterias.

La integrasa.

En los integrones de clase I la integrasa es una proteína de 337 aminoácidos con un tamaño calculado de 38 kDal. La comparación de secuencias ha permitido determinar que la integrasa del integrón pertenece a una amplia familia cuyo prototipo es la integrasa del fago λ . A esta familia pertenecen la mayoría de las integrasas de bacteriófagos y proteínas de resolución de genomas circulares, tanto en bacterias como en plásmidos (XerC y XerD de *Escherichia coli*, Cre del plásmido P1) y proteínas como FLP codificadas por plásmidos de levaduras. Las señas de identidad de esta familia de integrasas se localizan en el extremo carboxilo terminal de la proteína, donde encontramos cua-

tro residuos totalmente conservados: la tríada H-R-H, más una tirosina directamente implicada en la reacción de trans-esterificación. Las similitudes entre integrasas de la familia en la mitad amino-terminal son casi inexistentes. La cristalografía de rayos X ha permitido determinar la estructura tridimensional de cuatro proteínas de esta familia: las integrasas de I y del fago HP1, la proteína Cre y XerD. De estas La similitud en el extremo carboxilo permite aventurar que la estructura de la integrasa del integrón es similar en esta región a las otras de la familia, pero no sabemos nada de la estructura en la región amino terminal. La proteína se ha sobreexpresado y purificado, y es probable que se encuentre formando dímeros.

Los sitios primarios de actuación de la integrasa.

Ya he mencionado antes que la integrasa actúa sobre dos clases de sitios primarios llamados *attI* y *attC*. En trabajos anteriores los sitios de reconocimiento de la integrasa se han denominado también RHS (*Recombination Hot Spots*) y los sitios *attC* también se han llamado elementos de 59 pares de bases (59 pb-e).

Los sitios *attC* se localizan entre los genes estructurales del integrón. Se conocen muchos sitios *attC*, generalmente asociados a genes diferentes. Aunque los primeros caracterizados tenían un tamaño de alrededor de 60 pb, en la actualidad se han descrito sitios *attC* de hasta 120 pb. La descripción exacta de un sitio *attC* concreto exige cuando menos indicar el gen al que se encuentra asociado, y en caso de que esto no fuese suficiente, habría que indicar también el integrón en el que se encuentra: por ejemplo, *attCaadA1In2* sería el sitio *attC* que se halla al lado 3' del gen *aadA1* en el integrón In2. Además de ser de tamaño variable, los sitios *attC* también son muy variables en su secuencia y solamente muestran alguna constancia en los extremos. Lo que siempre sucede en todos los sitios *attC* es que su secuencia es un palíndromo imperfecto, lo que les confiere la capacidad de formar estructuras tipo horquilla cuando están en forma de DNA de cadena sencilla o de formar cruciformes cuando se localizan en DNA superenrollado. La secuencia más frecuente en los extremos de los sitios *attC* es CTAAC(50-110 pb).....GTTAG

El sitio *attI* se localiza al lado 5' del gen de la integrasa, y marca la separación entre la región conservada 5' y la región central variable. Al contrario de lo que ocurre con los sitios *attC*, el sitio *attI* es prácticamente igual en todos los integrones. A pesar de esta constancia no hay un con-

sensu sobre los límites precisos del sitio *attI*. Una de las propuestas considera al sitio *attI*, por analogía con *attC*, también como un palíndrome imperfecto de 45 pb: GCAAC...(35 pb)...GTTAA.

El sistema de ensayo de la actividad de la integrasa *in vivo*.

La manera más sencilla de probar la actividad recombinativa sitio-específica de la integrasa *in vivo* es un ensayo de conducción. Se requiere una cepa donadora *recA* que contenga simultáneamente tres plásmidos compatibles: un plásmido (A) que sobreexpresa la integrasa, un plásmido (B) conjugativo, pero sin trasposones, que contenga algún sitio *attI* o *attC*, y un tercer plásmido (C) no movilizable, también con sitios *attI* o *attC*. El ensayo consiste en estudiar la frecuencia de movilización del marcador del plásmido (C) no movilizable, lo que exige su unión física al plásmido (B). En los recombinantes se encuentra un único plásmido de tamaño igual a la suma de los tamaños de los plásmidos B y C. La naturaleza sitio-específica del proceso debe comprobarse por secuenciación de las uniones, o al menos por restricción de los recombinantes. Habitualmente se usa como fuente de la integrasa el plásmido pSU2056, que es un pUC con el gen *int* bajo el control del promotor *lac*. Estas condiciones de transcripción y número de copias son las apropiadas, ya que un mayor nivel de expresión de la integrasa hace a la bacteria inviable. El plásmido conjugativo empleado suele ser el plásmido IncW R388.

Juan María García Lobo es Licenciado y Doctor en Ciencias Biológicas por la Universidad de Bilbao (1981). Desde el comienzo de su tesis doctoral está dedicado al estudio de los mecanismos de diseminación de genes de resistencia a antibióticos en el Departamento de Biología Molecular de la Universidad de Cantabria, laboratorio que ha contribuido a la formación de muchos investigadores del área. Ha realizado estancias postdoctorales en La Universidad de California en Los Angeles, dedicada al estudio de la patogenicidad de *Yersinia*, y en la Universidad de Medicina y Estomatología de New Jersey (UMDNJ), dedicado al estudio del papel de la transcriptasa inversa bacteriana en el flujo genético. En la actualidad continúa en la Universidad de Cantabria, donde compagina la investigación en resistencia antibióticos con el estudio de los mecanismos de patogenicidad y epidemiología del género *Brucella*.



Utilizando este ensayo se encuentra que la frecuencia de recombinación entre dos sitios *attC* o entre un sitio *attI* y un sitio *attC* puede ser hasta de 10^{-2} .

Posición del sitio de entrecruzamiento.

La existencia de sitios *attC* muy diferentes ha permitido localizar con precisión el sitio de entrecruzamiento de la reacción de la integrasa en la G conservada en el extremo derecho de los sitios *attI* y *attC*:

```
attI:  GCAAC.....(35 pb).....GTTAA
attC:  CTAAC....(50-110 pb)....GTTAG
```

Esta localización muestra diferencias importantes con las otras integrasas de la familia. Éstas generalmente poseen un sitio principal (*core*) que consiste en un palíndrome con unas bases centrales. El sitio de entrecruzamiento se encuentra en la región central. La integrasa, generalmente un dímero, se une a cada brazo del palíndrome de forma simétrica y produce la transferencia de cadenas en la región central del palíndrome. Además, las dos moléculas que participan en la reacción son idénticas en esta región, lo que permite cierta migración de las ramas del intermediario de Holiday antes de la resolución. En el caso de la integrasa del integrón puede no haber ninguna homología entre los dos sitios participantes, y el sitio de entrecruzamiento no se encuentra en el centro, sino en un extremo del sitio *attC* (o *attI*).

Integración o excisión. Cassettes de resistencia.

Dependiendo de que los sitios que intervienen en la reacción estén en replicones distintos o en el mismo replicón, se producen dos reacciones diferentes. En el primer caso tenemos una reacción de integración o de fusión de replicones, como ocurre en el ensayo de conducción descrito. Si la reacción ocurre entre dos sitios en el mismo replicón se produce una delección. La reacción de delección ocurre en los integrones naturales y da lugar a la producción de círculos covalentemente cerrados, independientes pero no replicativos, que consisten en un gen de resistencia y un sitio *attC*. Estos círculos se denominan cassettes de resistencia. Los cassettes de resistencia son sustrato para la reacción de la integrasa, ya que contienen un sitio *attC*. Un cassette puede integrarse generalmente en el sitio *attI* de otro integrón. Por este procedimiento se pueden generar integrones nuevos con cualquier combinación imaginable de cassettes preexistentes.

Sitios secundarios de actuación de la integrasa.

Utilizando el ensayo de conducción descrito anteriormente, observamos la formación de recombinantes dependientes de la integrasa a frecuencias bajas (10^{-7}) cuando el plásmido no movilizable C no contenía un sitio *attI* ni *attC*. El análisis de estos recombinantes nos indicó que se trataba de fusiones simples entre el plásmido conjugativo B y el plásmido C. El estudio de los sitios implicados nos mostró que la recombinación usaba los sitios *attI* o *attC* del plásmido B. En el plásmido C se usaban sitios diferentes, pero en todos ellos se observó la secuencia GTWMW. Además de los ensayos *in vitro* que demostraban el uso de estos sitios sencillos se ha descrito la integración *in vivo* de cassettes en estas secuencias. El hecho de que la integrasa sea capaz de actuar sobre sitios sencillos (pueden aparecer cada 128 pb en el DNA) abre la posibilidad de que prácticamente cualquier gen cromosómico pueda escindirse como un cassette y llegar a formar parte de un integrón.

Además de estos pentanucleótidos sencillos, el análisis de diferentes sistemas nos ha permitido detectar sitios sobre los que la integrasa actúa con frecuencias intermedias entre las de los sitios primarios y las de los pentanucleótidos. El estudio de las secuencias de estos sitios ha mostrado que generalmente poseen combinaciones variables de pentanucleótidos consenso y cierta extensión de simetría de secuencia.

Estructura modular de los sitios primarios *attC* y *attI*.

La propuesta de una estructura modular para los sitios primarios de la integrasa deriva de dos observaciones: una, la comparación de las secuencias de los diferentes sitios *attC* conocidos, y dos, el hallazgo de sitios secundarios con actividad intermedia. La mayoría de los sitios primarios presenta una estructura palindrómica que consiste en dos pares de pentanucleótidos invertidos en cada extremo del sitio y una región central rica en G+C.

attI: GCAAC (3 pb) GTTAC... (19 pb) ...AAAAC (3 pb) GTTAA
attC: CTAAC (5 pb) GTTCA... (28 pb) ...TTAAC (6 bp) GTTAG
 pnt1 pnt2 pnt3 pnt4

De este modo se puede formar una horquilla en la que el pentanucleótido 1 se aparea con el 4 y el 2 con el 3. Sitios incompletos en los que falta alguno de los pentanucleótidos (generalmente el 1) o

con escasa simetría son los que se utilizan con frecuencias intermedias.

Resultados obtenidos con la integrasa purificada.

Al menos tres laboratorios diferentes han llevado a cabo la purificación de la integrasa y la han utilizado para experimentos *in vitro*. Con estas preparaciones se ha demostrado que la integrasa se une de forma específica a sus sitios primarios. Sin embargo, hay un matiz: mientras que la integrasa se une al sitio *attI* en forma de cadena doble, el sitio *attC* lo reconoce sólo en forma de cadena sencilla y se une sólo a una de las cadenas. La integrasa también se une al sitio *attI* en forma de cadena sencilla y también a una de las dos cadenas. Experimentos de protección con DNasaI han detectado zonas protegidas por la integrasa dentro de los sitios primarios, y aunque los datos no son determinantes, se pueden interpretar como que un pentanucleótido es el sitio mínimo de unión de la integrasa, lo que explicaría su uso *in vivo* como sitios de recombinación. En cuanto a la unión a DNA de cadena sencilla, se piensa que puede reflejar la forma en que los sitios se usan *in vivo*. Sería posible que en el DNA superenrollado los sitios primarios palindrómicos formen cruciformes, y que este paso sea necesario para la acción de la integrasa. No se ha conseguido reproducir *in vitro* ninguna actividad bioquímica relacionada con la capacidad recombinante de la integrasa. Esta carencia puede ser debida, entre otras causas, a la necesidad de proteínas accesorias para alcanzar una estructura recombinativa correcta.

Otros integrones. El integrón como sistema generador de operones.

Se ha descrito en *Vibrio cholerae* la presencia de una región genómica con estructura similar a la del integrón, en la cual los genes variables son genes relacionados con la patogenicidad en lugar de genes de resistencia a antibióticos. Como he pretendido reflejar hasta ahora, no existe ninguna razón para que la actividad de los integrones se circunscriba a la diseminación de genes de resistencia a antibióticos. Además de éstos y de los genes de patogenicidad descritos en *V. cholerae*, cualquier gen es susceptible de incorporarse en un integrón y es posible que si no se observan integrones con otro tipo de genes lo sea por que no existe la presión selectiva necesaria para ponerlos de manifiesto.

La organización de los genes en la región variable de los integrones sugiere un mecanismo gene-

ral para la formación de operones. La acción del integrón coloca genes sucesivos bajo el control de un mismo promotor separados por sitios *attC*. La maduración del nuevo operón consistirá en la pérdida o modificación de la actividad de la integrasa que podría convertirse en un gen regulador, y en la pérdida de los sitios *attC* intergénicos que pueden actuar como terminadores prematuros de la transcripción. Las secuencias repetitivas palindrómicas que son habituales en posición intergénica en muchos operones bacterianos pueden ser vestigios de integrones primitivos. Una peculiaridad adicional de los integrones es la falta de secuencias intergénicas no codificantes. A pesar de que los mecanismos de recombinación general y sitio-específicos usando sitios secundarios pueden ser suficientes para explicar el origen de nuevos genes en integrones, no se ha descartado (ni demostrado) una hipótesis que sugiere que los cassettes se puedan generar por retrotranscripción de mensajeros bacterianos. La presencia en bacterias de transcriptasas inversas, muchas veces asociada a bacteriofagos que cuentan además con una integrasa, hace plausible esta hipótesis, pero nos queda por encontrar las especies bacterianas en que este proceso pueda tener lugar.

Cuestiones abiertas para el futuro.

Es evidente que se requiere un esfuerzo en la reproducción de la reacción de la integrasa *in vitro* y en mejorar nuestro conocimiento bioquímico de la misma. La determinación de la estructura tridimensional de la integrasa será fundamental para alcanzar estos objetivos. Estos detalles nos deben clarificar los mecanismos de formación de integro-

nes y los de incorporación de nuevos genes. Los nuevos datos de secuencias genómicas nos pueden revelar estructuras tipo integrón en cromosomas bacterianos ayudándonos a entender el origen de estos elementos. En otro orden de cosas, cada día son más abundantes los datos sobre la implicación real de los integrones en la resistencia a antibióticos, especialmente en gérmenes Gram negativos, y asimismo crece el número de genes de resistencia que se encuentran asociados con estos elementos. La carencia de nuevos antibióticos y los problemas cada día más acuciantes de resistencias, pueden hacer necesario pensar en inhibidores de la integrasa como medida de control.

BIBLIOGRAFIA

- Martínez E, de la Cruz F (1990) Genetic elements involved in Tn21 site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. *EMBO J.* 9:1275-1281.
- Hall RM, Brookes DE, Stokes HW (1991) Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. *Mol. Microbiol.* 5:1941-1959.
- Francia MV, de la Cruz F, García Lobo J (1993) Secondary-sites for integration mediated by the Tn21 integrase. *Mol. Microbiol.* 10:823-828.
- Gravel A, Fournier B, Roy PH (1998) DNA complexes obtained with the integron integrase IntI1 at the *attI1* site. *Nucleic Acids Res.* 26:4347-4355.
- Mazel D, Dychinco B, Webb VA, Davies J (1998) A distinctive class of integron in the *Vibrio cholerae* genome. *Science* 280:605-608.