

Genética molecular I

- Estructura de ácidos nucleicos.

Estructura primaria.

El estudio se puede enfocar desde el punto de vista de los monómeros que constituyen los ácidos nucleicos, es decir, sus propiedades, relaciones y constituyentes, y también desde el punto de vista de la secuencia.

Los ácidos nucleicos, a semejanza de las proteínas, son heteropolímeros constituidos por un esqueleto de fosfato y azúcar que tiene una direccionalidad, que se representa desde 5' a 3', o también desde el extremo fosfato hasta el extremo OH. Este esqueleto se asemeja a la secuencia N-C-C de proteínas. Del esqueleto salen una serie de bases nitrogenadas o nucleobases que son las que, según el orden en que estén, determinan la identidad de cada molécula individual de ácido nucleico, de modo parecido a los grupos R de las proteínas. Este orden se denomina secuencia.

Cada monómero individual se denomina nucleótido, compuesto por el fosfato, el azúcar y una base. La unión de estos dos últimos componentes se llama nucleósido, con lo que los nucleótidos también se pueden llamar nucleósidos-monofosfato. La unión de la base con el azúcar se realiza mediante un enlace N-glucosídico, con lo que los nucleósidos son N-glucósidos. En esta nueva molécula se añade una prima a los carbonos del azúcar, para no confundirlos con los átomos de las bases. Los nucleótidos se forman al unir un fosfato en 5'.

Los constituyentes de los nucleótidos son intermedios de un metabolismo mucho más complejo, intrincado con el resto de rutas metabólicas de la célula. De hecho los nucleósidos que son intermediarios en la síntesis de ARN se utilizan habitualmente para formar moléculas activadas (ADP ribosilación, UDP-azúcares), transferir energía (ATP, GTP), segundos mensajeros (AMPc, GMPc) o sirven de cofactores enzimáticos. Las rutas de degradación son principalmente de reciclaje, pero también se degradan hasta azúcares y ácido úrico.

Bases nitrogenadas.

Las hay de dos tipos. En ambos casos son heterociclos con nitrógeno y con carácter hidrófobo y aromático. Son moléculas planas. Hay dos familias, según deriven de la purina (Pu) o la pirimidina (Py).

- Púricas:
 - **Adenina** (A): 6-aminopurina.
 - **Guanina** (G): 2-amino-6-oxopurina.
- Pirimidínicas:
 - **Citosina** (C): 2-oxo-4-aminopirimidina.
 - **Uracilo** (U): 2-oxo-4-oxopirimidina. Sólo aparece en ARN
 - **Timina** (T): 2-oxo-4-oxo-5-metilpirimidina. Sólo aparece en ADN.

*Regla mnemotécnica: **AGriCULTor**: las dos primeras púricas y las tres últimas pirimidínicas.

Su carácter aromático los hace muy hidrófobos, pero los átomos de nitrógeno permiten el establecimiento de enlaces de hidrógeno, con lo que esta estructura determina todas las propiedades de las bases y de las moléculas que forman, ya que tenderán a huir del agua formando enlaces de hidrógeno entre bases de la misma cadena de ácido nucleico o también con otras moléculas distintas.

Bases modificadas.

Bases metiladas.

Son muy frecuentes en el ADN, y se modifican después de sintetizado el polímero:

- N6-metiladenina.
- 5-metilcitosina.

Hay enzimas que reconocen específicamente una secuencia corta de unos seis nucleótidos y metilan alguna de estas dos bases. Esto genera un patrón de metilación en el ADN que regula la capacidad que tiene esa zona del ADN para transcribirse y, además, modifica o condiciona la estructura secundaria y posteriores.

Bases del ARNr.

Como en todos los ARN hay muchas más bases modificadas que en el ADN, pero también abunda la 5-metilcitosina.

Bases del ARNt.

El caso de estas moléculas es especial, porque su función depende de su estructura secundaria, y para conseguir las interacciones deseadas y el plegamiento correcto se recurre en un porcentaje muy alto a la modificación de bases. Entre otras, aparecen:

- Inosina (I): La base nitrogenada es la hipoxantina (6-oxopurina).
- Metilinosina.
- Dihidrouracilo.
- Dimetilguanina.
- Timina. En este caso unida al azúcar propio del ARN, y no al del ADN.
- Pseudouridina (δ). Caracteriza uno de los elementos estructurales del ARNt.

Otras bases.

Xantina y ácido úrico nunca están presentes en los ácidos nucleicos, pero sí son parte integrante de su metabolismo.

Propiedades.

Tautomería.

Todas las bases presentan varios tautoisómeros posibles que por comodidad se estudian a pares. Esto se debe al carácter aromático de los anillos, en los que los electrones tienen cierta capacidad de movimiento y, además, en los anillos hay átomos muy electronegativos ricos en electrones, como nitrógeno y oxígeno, que van a competir por estar protonados en determinadas circunstancias. Se conocen las constantes de protonación de los nitrógenos del anillo. Con estas consideraciones se pueden determinar pares de tautomería para cada base:

- **Adenina y citosina** pueden presentarse en forma amino o en forma imino, según cómo se encuentre este radical en posición 6 (A) o en posición 4 (C).
- **Guanina y timina** pueden encontrarse en forma ceto o en forma enol, según el grupo oxo en 6 (G) o en 2 (T).
- **Uracilo** puede estar en forma lactama o lactima, según tenga los grupos oxo en 2 y 4 como ceto o como enol, respectivamente.

La preferencia de una u otra forma depende de los valores de los pK de protonación de los nitrógenos del anillo. A valores de pH fisiológico las formas predominantes son las amino y ceto, que se deduce del cuadro de pK que se repartió en clase. Las formas tautoméricas son esenciales para la configuración correcta de enlaces de hidrógeno y, por tanto, para la mejor función a través del reconocimiento de cada base por su complementaria. Esto se debe a que los distintos tautómeros cambian el carácter aceptor o dador de átomos de hidrógeno de determinados grupos esenciales en la interacción, y es este cambio el que determina las formas de estructura secundaria de ADN: B (normal), A y Z (tautómeros).

Absorción UV.

Todas las bases, debido a su aromaticidad, absorben radiación UV con un máximo alrededor de 260 nm. Este hecho permite que se puedan cuantificar mediante técnicas de espectrofotometría. Además, se puede relacionar la variación en el espectro de absorción con el grado de protonación de los nitrógenos de las bases. En este apartado se observa una mayor variación en las bases pirimidínicas que en la púricas, debido al apantallamiento que ejerce la estructura de doble anillo de purina.

La absorción y, por tanto, la cuantificación, se puede hacer extensiva a los nucleósidos, nucleótidos y a todos los ácidos nucleicos en general, por lo que es de una gran importancia.

Azúcares.

Los únicos glúcidos que pueden entrar a formar parte de los ácidos nucleicos son la ribosa y la 2'-desoxirribosa, y son lo que diferencia los dos tipos de ácidos nucleicos que existen: ribonucleico (ARN) y desoxirribonucleico (ADN). Ambos azúcares se presentan siempre ciclados como δ -ribofuranosa o δ -2'-desoxiribofuranosa. Las bases se unen mediante un enlace N-glucosídico con el N1 de las pirimidinas o con el N9 de las purinas para formar los nucleósidos. Éstos son estables en medio básico, pero no en medio ácido para el enlace con purinas.

Los nucleósidos recién formados se nombran de la siguiente forma:

- **Adenosina** con adenina.
- **Guanosina** con guanina.
- **Citidina** con citosina.
- **Uridina** con uracilo.
- **Timidina** con timina.

La unión de una molécula hidrófila como es el azúcar a la base nitrogenada, que es altamente apolar, provoca el aumento de la solubilidad de éstas, permitiendo un mejor transporte.

La estructura de la pentosa no es plana, sino que tiene estructura tipo E (recordar la bioquímica estructural), en la que se admiten cuatro conformaciones distintas:

- C2' endo.
- C3' endo.
- C2' exo.
- C3' exo.

Endo y exo significan la posición del átomo que se sale del plano como dirigido hacia el C5' o bien hacia el otro lado, respectivamente. De estas cuatro las más abundantes son las dos primeras, pero todas son más o menos posibles, lo que unido a la capacidad de rotación alrededor de numerosos enlaces del esqueleto de azúcar-fosfato permite muchos grados de flexibilidad conformacional a la molécula de ácido nucleico.

Por otro lado, el enlace glucosídico tiene posibilidad de giro (ángulo δ), con dos posiciones más energéticamente favorecidas denominadas syn y anti, según si la base se dispone sobre el cuadro del azúcar (como si le hiciera sombra) o bien si ocurre exactamente lo contrario. En el caso de las purinas puede ocurrir cualquiera de las dos, aunque está más favorecida la anti, pero para las pirimidinas sólo está permitida la anti, porque el grupo ceto en C2 interacciona con el esqueleto del polímero a través del fosfato. Aun así, esta flexibilidad conformacional unida a lo anteriormente descrito para el azúcar, hace de los nucleótidos y, por tanto, de los ácidos nucleicos unas estructuras muy flexibles.

Nucleótidos.

Los nucleótidos se forman al unir un resto de ácido fosfórico al carbono 5' de un nucleósido mediante un enlace fosfoéster. Se nombran según el nucleósido del que provengan:

Nucleósido	Nucleótido		
<i>Adenosina</i>	Ác. adenílico	Adenosina-5'-monofosfato	AMP
		Adenosina-5'-difosfato	ADP
		Adenosina-5'-trifosfato	ATP
<i>Guanosina</i>	Ác. guanílico	Guanosina-5'-monofosfato	GMP
		Guanosina-5'-difosfato	GDP
		Guanosina-5'-trifosfato	GTP
<i>Citidina</i>	Ác. citidílico	Citidina-5'-monofosfato	CMP
		Citidina-5'-difosfato	CDP
		Citidina-5'-trifosfato	CTP
<i>Uridina</i>	Ác. uridílico	Uridina-5'-monofosfato	UMP
		Uridina-5'-difosfato	UDP
		Uridina-5'-trifosfato	UTP
<i>Timidina</i>	Ác. timidílico	Timidina-5'-monofosfato	TMP
		Timidina-5'-difosfato	TDP
		Timidina-5'-trifosfato	TTP

Todos los nucleótidos son ácidos porque el fosfato está ionizado a pH fisiológico, y su carga se encuentra contrarrestada o apantallada por iones divalentes, generalmente Mg^{2+} . Los tres posibles grupos fosfato se nombran δ , β y α , empezando por el más próximo al azúcar y, además, los nucleótidos con desoxirribosa se escriben con una de minúscula delante en la forma abreviada (dCMP) o con el prefijo desoxi.

Es importante destacar el papel que juegan los nucleótidos con más de un grupo fosfato en el metabolismo energético, porque la unión por deshidratación de varias moléculas de ácido fosfórico origina un ácido anhídrido (enlace anhídrido) capaz de liberar una elevada energía de hidrólisis, sin embargo, en los ácidos nucleicos solo se encuentran nucleósidos monofosfato (NMP: N = cualquier nucleósido). De esto se desprende que para marcar ácidos nucleicos con fósforo el átomo que se tiene que marcar es el δ .

Propiedades de los nucleótidos.

Como los nucleósidos, presentan máximos de absorción a 260 nm. La absorción de un ácido nucleico depende de la proporción de cada uno de sus nucleótidos, y se puede fabricar una tabla con los coeficientes de extinción de los ácidos nucleicos de varias especies y tipos (hebra sencilla, ss, o doble hebra, ds; ADN o ARN). Mediante la relación entre la absorción a 260 nm y la absorción a 280 nm se puede determinar la pureza del aislado de ácido nucleico. Esto se debe a que las impurezas que acompañan a los ácidos nucleicos suelen ser proteínas y fenol, utilizado en las preparaciones, ambos con absorciones a 280 nm, con lo que una relación entre 1,8 y 2,1 se considera buena para poder decir que el ácido nucleico está puro.

Funciones biológicas de los nucleótidos.

- Mensajeros:
- Intercelulares: Adenosina (cerebro).
- Intracelulares: AMPc, GMPc.
- Intercambio de energía:
- ATP: Moneda de cambio universal.
- GTP: Imprescindible en la síntesis de proteínas.
- CTP: Imprescindible en la síntesis de fosfolípidos.
- UTP: Imprescindible en la síntesis y degradación de azúcares.
- NTP: Base de la síntesis de ácidos nucleicos.
- Coenzimas de oxidorreducción: NAD⁺, FAD⁺.

Polinucleótidos.

La formación de polímeros, independientemente de su naturaleza, siempre requiere un aporte de energía externo. En el caso de la unión de NMP para formar un dinucleótido, la reacción es muy endergónica, con un $\Delta G' = 25$ KJ/mol. Este problema se soluciona en medio fisiológico mediante el acoplamiento a la hidrólisis del enlace anhídrido entre α y β de un NTP, que cursa con $\Delta G' = -31$ KJ/mol. Sumando a esta variación de energía libre la causada por la hidrólisis del pirofosfato ($\Delta G' = -33$ KJ/mol) y la anterior de unión de dos NMP, resulta que se obtiene $\Delta G' = -39$ KJ/mol.

La unión de nucleótidos siempre se realiza mediante un enlace éster entre el hidroxilo en 3' de uno y el fosfato en 5' de otro. De este modo se genera un polímero polar, en el que el primer nucleótido es el que tiene el fosfato en 5' libre, y el último es el que tiene el OH en 3' libre. Estos extremos se llaman 5'-P y 3'-OH, y siempre se escriben de izquierda a derecha. Igual que en las proteínas resultaba un esqueleto N-C-C, en los ácidos nucleicos el esqueleto es 5' P-azúcar-...-P-azúcar-OH 3'. La unión de dos nucleótidos a través de dos ésteres con un grupo fosfato se llama enlace fosfodiéster.

En el nuevo polímero hay una gran cantidad de enlaces susceptibles de rotación, como ya se ha adelantado, con lo que una molécula de ácido nucleico no tiene conformación definida ni fija a no ser que establezca interacciones con otros ácidos nucleicos o con proteínas (ocurre como con el colágeno). Los enlaces capaces de rotar son los del fósforo con el oxígeno unido al 5' (O5'), el de éste con 5', el de éste con 4', una vez más con 3' (define las conformaciones endo y exo), éste con el oxígeno unido a él (O3') y el de éste con el fósforo del nucleótido siguiente. Además está el enlace glucosídico.

En las representaciones y nomenclaturas hay que tener en cuenta que los extremos pueden estar fosforilados o no, con lo que hay que nombrarlo explícitamente.

Diferencias químicas entre ADN y ARN.

Sustitución de uracilo por timina.

Esta diferencia se debe principalmente a que la citosina sufre espontáneamente y a una frecuencia

muy baja una desaminación que la convierte en uracilo. Esta mutación se podría transferir a la descendencia si no hubiera un mecanismo de control fiable. La ausencia de uracilo en el ADN se ha seleccionado positivamente para permitir la detección y eliminación de estas mutaciones, utilizando para ello la enzima propia al efecto, la uracil ADN N-glucosidasa. Esta enzima lee permanentemente el ADN e hidroliza el enlace glucosídico cuando la base es uracilo, dejando un hueco que otro conjunto de enzimas rellena con la citosina correcta.

Realmente la timidina es un uracilo metilado, y realiza la misma función que el uracilo en el ARN, así que se podría llamar el uracilo del ADN, capaz de hacer que se reconozca el ADN como lo que es para no confundirlo con el ARN.

Ribosa frente a desoxirribosa.

El ADN tiene desoxirribosa porque le permite ser totalmente estable en medio alcalino. Los enlaces fosfodiéster del ADN son especiales, porque se une a dos átomos de carbono, con lo que recibe el nombre especial de enlaces dialquil fosfato. Estos enlaces son extremadamente resistentes a hidrólisis. Sin embargo, en medio básico los radicales hidroxilo podrían provocar un ataque nucleofílico al fosfato, pero hay impedimentos estéricos que impiden que se acerque lo suficiente. En el caso del ARN los hidroxilos del álcali tiran del hidrógeno en 2' y dejan al hidroxilo en 2' con un exceso de carga negativa, que provoca un ataque nucleofílico al fosfato y obliga a la hidrólisis de la molécula al romper el enlace entre el fosfato y el 5'. El producto inmediato es un nucleótido cíclico en 2' y 3' que se rompe a partes iguales generando nucleótidos 2'-fosfato y 3'-fosfato.

Para este proceso de catálisis intramolecular se necesita un OH en 2', cosa que no hay en el ADN. Por esto se ha seleccionado la desoxirribosa en lugar de la ribosa. Además, esta diferencia se utiliza para eliminar ARN de una mezcla de ácidos nucleicos.

Hidrólisis de polinucleótidos.

Los ácidos nucleicos se hidrolizan por los métodos químicos que ya se han tratado: hidrólisis del enlace N-glucosídico de las purinas en medio ácido (ADN o ARN apurínico) e hidrólisis total en monómeros del ARN en medio ligeramente básico. Además, hay una gran diversidad de métodos enzimáticos de rotura más o menos específica de los ácidos nucleicos. Este conjunto de técnicas y métodos de rotura es una herramienta esencial para el estudio de los ácidos nucleicos, ya que siempre se parte de una molécula extremadamente larga del orden de 1 Mb, y hay que ir cortándola en trocitos menores para poder analizarla en profundidad, principalmente para determinar la secuencia.

A la hora de utilizar un método enzimático de hidrólisis, llamado generalmente digestión, hay que tener muy en cuenta por dónde se va a cortar el polímero, es decir, a qué lado del fosfato, si son endo- o exonucleasas (cortan desde dentro o por un extremo) y en este último caso por qué extremo empiezan, si su especificidad es de ADN o de ARN, y dentro de cada clase, si cortan cadenas sencillas, dobles o moléculas híbridas ADN-ARN. Si sólo cortan ácidos nucleicos de doble cadena, se debe conocer cuál de ellas es la que se rompe, o si por el contrario son las dos (si el corte es en una sola hebra se suele denominar *nick*: corte o muesca). Además, la especificidad puede depender del tipo de bases e incluso de la secuencia concreta. Así, hay enzimas glucosidasas específicas de tipo de base (purinas o pirimidinas) o incluso de una sola base (la uracil ADN N-glucosidasa). También la especificidad de secuencia o de bases se observa en las enzimas que cortan el enlace fosfodiéster, particularmente en las enzimas de restricción, cuya diana es una secuencia doblemente simétrica en una molécula de ADNds (palíndromo) de entre 4 y 8 nt (últimamente se han visto dianas de más nucleótidos). Estas enzimas suelen estar fabricadas por microorganismos para defenderse de la invasión por ADN de virus bacterianos (fagos), y la secuencia diana está protegida en el propio genoma de la

bacteria.

La importancia de saber la mecánica de corte de cada enzima es que generan productos de digestión totalmente distintos en cada caso, y se debería poder escoger lo que se desea obtener en cada momento.

Con las enzimas de restricción se empieza a cortar específicamente el material genético de partida, que suele ser muy grande. Se abordan varias aproximaciones con varias enzimas distintas, que se sabe que cortan frecuentemente el ADN, y los productos de la digestión de cada una, y de varias juntas, se corren en un gel y se obtiene una cantidad de fragmentos de diferente longitud que se debe poder encajar como un puzzle y asignar sobre un esquema del fragmento de partida los sitios de corte de cada una. Una vez conseguida esta distribución de las dianas de cada enzima se dice que se ha elaborado un mapa de restricción del fragmento de partida. Con esta información se utilizan estos fragmentos ya conocidos (tanto en número de bases como en orden) para proceder al estudio posterior que se considere, generalmente para poder secuenciar el fragmento grande partiendo de sus productos de digestión.

Secuenciación de ADN.

Después de tratar el material de la manera descrita anteriormente, es decir, manejando varios fragmentos de unos cientos de bases cada uno, se puede proceder a su secuenciado, teniendo cuidado de escoger fragmentos que solapen parcialmente al utilizar varias enzimas distintas.

El principal problema a la hora de abordar una secuenciación es que se parte de muy poco material de partida (ng), por lo que se hace necesario aumentar su cantidad mediante su amplificación en plásmidos (utilizando su origen y promotores, tanto naturales como fabricados) o utilizando una PCR. Además, se necesitan unas técnicas analíticas de una extremada sensibilidad, como son la fluorografía y, últimamente, la quimioluminiscencia.

Para obtener una secuencia hace falta separar fragmentos que difieran en un solo nucleótido, con lo que se necesita separarlos en geles de poliacrilamida, cuyo poro es mucho menor que el de la agarosa. Sin embargo, para afinar más la técnica y obtener una secuencia fiable hay que secuenciar ambas cadenas de una molécula de ADNds.

Hay dos métodos de secuenciación modernos: el primero se basa en el análisis del ADN modificado químicamente y el segundo y más utilizado se basa en la interrupción de la síntesis en un determinado momento.

Método de terminación de cadena.

Desarrollado por Sanger, parte de la imposibilidad de continuar la polimerización al incorporar un nucleótido didesoxi (ddNTP), es decir, que no tiene un 3'OH libre y, por tanto, no se le puede unir nada. Por esto también se llama método didesoxi. Para todo esto se necesita en cuatro tubos distintos una ADN polimerasa, que se compra a unos precios desorbitados, el fragmento que se quiere secuenciar, el cebador correspondiente que también se compra o se fabrica (de unos 20 nt) y una colección de dNTP marcados. En cada uno de los tubos se añade una cantidad de uno de los ddNTP en la proporción correcta para que no se incorpore siempre en lugar del dNTP correspondiente y se puedan generar fragmentos de distinta longitud. También se puede incorporar la radiactividad con el cebador, en lugar de en los nucleótidos que se incorporan. Yo creo que es lo más recomendable y manejable, pero el profesor no ha dicho nada al respecto ni aclaró mucho cuando se lo comentamos.

El mecanismo del método consiste en que cuando se comienza la polimerización, a la temperatura correcta, de cada vez que tiene que incorporarse un dNTP, existe una probabilidad de incorporar un ddNTP en su lugar, con lo que esa cadena ya no se puede seguir sintetizando. Esa proporción depende de la relación entre dNTP y

ddNTP en cada tubo, además de la concentración absoluta de ambos. El resultado es una colección de fragmentos de distinta longitud que, en cada tubo, ha acabado en el momento en que se ha incorporado un ddNTP. Para analizar el resultado se fabrica un gel de poliacrilamida en el que se pone en cuatro pocillos una muestra de cada uno de los tubos, y se marcan con la letra del ddNTP. Estos geles se leen de abajo arriba, ya que los fragmentos menores son los que incorporaron un ddNTP antes y, por tanto, son los que están primero en la secuencia. Después de obtener la secuencia que se lee en el gel en sentido 5' a 3', hay que tener en cuenta que se trata de la secuencia complementaria a la del ADN que se quiere secuenciar, por lo que hay que fabricar su complementaria y leerla nuevamente en el sentido 5' a 3', lo que implica darle la vuelta.

A la vista de los geles que se hacen en los laboratorios, de una sola vez sólo se pueden leer una pequeña longitud de cada gel porque en el resto las bandas no se separan bien o están distorsionadas, con lo que limita el número de bases que se pueden diferenciar de cada vez a unas cien (como mucho 300). Esta es la principal limitación de esta técnica, porque se necesita realizar muchos geles y que sean buenos, y amplificar muchos segmentos que tienen que solapar bastante entre sí para luego poder unirlos todos uno detrás de otro. Además, dado que es una técnica muy reiterativa y rutinaria, se corre el riesgo de cometer muchas veces un fallo determinado (que se rompa el gel, que no tenga la resolución adecuada...).

Los problemas que presenta esta técnica se pueden solucionar con un secuenciador automático, en el que el cebador de cada tubo se marca fluorescentemente con un color distinto. Luego se hace una mezcla con una muestra de cada tubo y se somete a una electroforesis, es decir, que no separa en cuatro pocillos, sino que se juntan todos en uno. Esto permite a un haz de láser iluminar la electroforesis según se va corriendo y leerla al tiempo, porque los fragmentos tienen distinta fluorescencia según en qué tubo se hayan sintetizado. Es un método más limpio, más eficaz, más rápido, más capaz (10 Kb/día), más seguro (no utiliza radiactividad) y extremadamente caro (el aparato cuesta unos diez millones). La carestía de este método hace surgir un negocio paralelo en el que se crean empresas que ofrecen este servicio a precio de oro.

Método de rotura de cadena.

No se hace una copia del ADN que se quiere secuenciar, sino que se parte del propio material al que se le ha añadido una marca radiactiva en alguno de los dos extremos a través de un grupo fosfato. Este método se basa en reacciones específicas de retirada de bases, seguidas de determinadas reacciones químicas que provocan la rotura del ADN en ese punto. Estas reacciones son:

- Eliminación de G.
- Eliminación de purinas (A y G).
- Eliminación de C.
- Eliminación de pirimidinas (C y T).

Todas las reacciones se realizan en dos pasos: eliminación/degradación y rotura. Esta última es igual para cada una de las cuatro mencionadas anteriormente, y consiste en la unión de piperidina en medio básico al carbono anomérico (desplazando una base degradada, si la hay) mediante un doble enlace con el nitrógeno. Se provoca la aparición de un catión y un efecto de inducción electrónica que obliga al azúcar a romper los enlaces con los dos fosfatos que lo flanquean, liberando dos fragmentos de ADN, uno con 3'-fosfato y otros con 5'-fosfato. Este proceso se llama δ -eliminación.

La eliminación selectiva de guanina pasa por una metilación suave con sulfato de dimetilo, que añade un grupo $-\text{CH}_3$ en posición N7, pero se tiene que hacer de tal modo que no se metilen todas las bases, sino sólo las necesarias para no romper todas las moléculas por el mismo sitio, ya que sino, no se podría leer ninguna secuencia. El producto de la metilación es un ion amonio que se estabiliza por resonancia con el N9, pero al pasar a medio básico se rompe el anillo y la base se degrada y se hace susceptible de una sustitución por piperidina, que ya se incluye en ese medio alcalino.

La eliminación de purinas consiste en un sencillo tratamiento en medio ácido suave, que hidroliza el enlace N-glucosídico (no todos, de nuevo), como ya se ha visto antes. Esto deja al descubierto el carbono anomérico para su posterior tratamiento con piperidina.

La eliminación de pirimidinas pasa por la degradación del anillo, como el caso de la guanina, pero esta vez con hidracina, que ataca los carbonos 4 y 6 de los anillos de pirimidina y cambiando totalmente la estructura de la molécula. Este cambio la hace susceptible de ser sustituida por piperidina en medio básico. En valores de fuerza iónica cercanas a las fisiológicas, esta reacción se produce con cualquiera de las dos pirimidinas, pero a altas concentraciones de sal (NaCl 2M) se hace específica de citosina.

Después de obtener la colección de fragmentos y someterlos a electroforesis y revelado, hay que hacer notar que no hay carriles para cada base, sino para G, purinas, pirimidinas y C, con lo que no basta con leer una banda detrás de la otra, sino que los sitios donde aparecen dos a la misma altura son bien de G o bien de C, y las bandas que sólo aparecen en los carriles de purinas o pirimidinas se corresponden con A y T, respectivamente. Además, hay que tener en cuenta qué extremo es el que está marcado, porque de ello depende el sentido de la lectura. Si se marca el extremo 5' hay que leer como en el método didesoxi, pero no fabricar luego la complementaria, porque la electroforesis es ya del ADN, no de su complementario, si por el contrario es el extremo 3' el que está marcado, el gel se lee de arriba abajo.

Comparación de métodos.

El método didesoxi es mucha más fácil porque no hay que partir del ADN puro para posteriormente modificarlo, sino que basta con tenerlo. Además, en el método didesoxi se incorpora más radiactividad, con lo que es unas 200 veces más sensible, es más rápido y barato, y con una calidad similar a la hora de secuenciar.

Actualmente se utiliza el método didesoxi unido a la PCR para secuenciar grandes trozos de ADN, porque cada día se fabrican polimerasas mejores mediante ingeniería genética y, además, como se realiza a elevadas temperaturas se evita la formación de estructuras secundarias (pares, islas o bucles GC) que interrumpen la transcripción.

Sin embargo, el método de rotura de cadena se utiliza mucho para realizar estudios de huella o *footprinting*, para determinar la secuencia que protege un determinado factor de transcripción o proteína de unión a ADN (promotores, reguladores, intensificadores).

Secuenciación de ARN.

Se suelen seguir técnicas de rotura de cadena, con lo que de nuevo obliga a purificar ARN a homogeneidad, lo que a veces es imposible porque el material de partida es muy heterogéneo y los ARN se parecen todos mucho entre sí en sus características físicas. Además, un problema de los ARN es que se desintegran en álcali incluso suave, por lo que hay que buscar otras estrategias para romper la cadena en distintos puntos. Si se sigue un abordaje químico, las reacciones son bastante parecidas a las del método para el ADN:

- G con sulfato de dimetilo.
- Principalmente A, aunque G en proporción baja con pirocarbonato de dimetilo.
- U principalmente, aunque también C con hidracina.
- C con hidracina en alta sal.

El abordaje enzimático con endonucleasas específicas también permite una rotura selectiva de la cadena de ARN, aunque por puntos distintos. Hay muchas diferentes, generalmente derivadas de microorganismos, que se pueden utilizar como mejor convenga. Un ejemplo es éste.

- ARNasa T1: pGp*N. Libera G unida al fosfato del siguiente nucleótido.

- ARNasa U2: pAp*N. Libera A unida el fosfato del nucleótido siguiente.
- ARNasa PhvM: pAp*N y pUp*N. Corta detrás de A y U y se llevan el fosfato del nucleótido siguiente.
- ARNasa BC: pUp*N y pCp*N. Corta detrás de U y C y se llevan el fosfato del siguiente nucleótido.

Todos los fragmentos obtenidos por el método anterior tienen extremos 3'-P y 5'-OH. En el gel de electroforesis se carga cada uno de estos carriles y determinarán la situación de las G, A, A + U y U + C. Lo que no se encuentre en el carril A, pero sí en el A + U será una U, y de modo parecido lo que no se encuentre en el carril A + U, pero sí en el U + C será una C. La resolución de este método, comparándolo con la del ADN, es bastante menor y se permite identificar muchas menos bases en un solo experimento.

El principal problema de este método, ya mencionado anteriormente, es que los ARN son muy difícilmente purificables, excepto los muy abundantes, como genomas de virus ARN o los mayoritarios (ARNr y ARNt). El conjunto de ARNm se podría separar con suerte porque tienen una cola de poli-A, pero hay unos 25000 tipos distintos en una célula eucariota, y pescar el que se quiere en una gran mancha de gel es extremadamente difícil. Para pescar ARNm se utiliza la transcripción inversa, que forma una molécula de ADNc más fácilmente secuenciable. Sin embargo, para esto es necesario tener una sonda, es decir, conocer la secuencia que interesa. Ésta se obtiene a partir de la secuencia de aminoácidos de una proteína conocida de la que interesa obtener el gen, que puede traducirse nucleótidos a través del código genético (no olvidar que hay que dejar una base floja en el tercero de cada codón para permitir el tambaleo), y fabricar un cebador con esa secuencia. Otra estrategia es el clonaje por homología, en el que se buscan similitudes con genes o secuencias conocidas, lo que permite clonar una pequeña cantidad de genes con un cierto grado de identidad.

La secuenciación de ARN mediante ADNc permite también caracterizar genes y determinar dónde se producen las transiciones intrón-exón mediante su comparación con la secuencia de ADNg (genómico).

La principal limitación es que la transcriptasa reversa puede introducir errores y depende mucho de las condiciones de transcripción la longitud de la cadena sintetizada. Además, el principal problema es que no se puede contrastar el resultado de la secuenciación de una sola hebra de ARN con la complementaria porque, sencillamente, no la hay, por lo que si se introduce un solo fallo, éste persiste hasta el final y no se corrige hasta que se identifica el gen en el ADN genómico.

Estructura secundaria de ADN.

El modelo de Watson y Crick.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

El estudio de la estructura secundaria del ADN arranca desde el modelo de Watson y Crick de doble hélice de ADN. Ellos propusieron su modelo en una coyuntura científica en la que se sabían los siguientes datos:

- Avery y otros investigadores ya habían determinado irrefutablemente que el material genético es el ADN.
- Por esta cualidad debería tener una serie de propiedades:
- Ser perpetuable.
- Llevar información gracias a una complicada estructura.
- Esto permitiría fabricar proteínas tan complejas como las que se conocían.
- La información debía poder ser modulada a algún nivel.
- El modelo debía poder explicar o ser coherente con los fenómenos de mitosis y meiosis.
- También debía poder explicar las mutaciones y la evolución genómica.
- Se conocían las reglas de Chargaff:
- Igual número de A que de T.
- Igual número de C que de G.
- Igual número de purinas que de pirimidinas.
- Por tanto, no existe el tetranucleótido.
- El modelo debería explicar o tener en cuenta los tautómeros de las bases que entran a formar parte del ADN.
- Debería considerar que las bases son hidrófobas y huyen del agua.

- Años antes Pauling descubrió la hélice α en proteínas, y propuso la primera estructura helicoidal de moléculas biológicas.
- Se conocían datos de cristalografía de ADN.
- Se tenían que considerar los impedimentos estéricos.
- Por estudios de centrifugación se conocía que el ADN era de doble cadena.
- Se conocía la existencia de un esqueleto de azúcar–fosfato, que debería estar expuesto al agua por ser hidrófilo.
- El modelo debería poder sugerir un método para la replicación de ADN.

Datos estructurales.

Los datos de cristalografía de los que disponían Watson y Crick se determinaron a partir de aislado de ADN nuclear en solución y no muy puro. Estos datos se determinan mediante el estudio de la difracción de un haz de rayos X al atravesar una muestra convenientemente tratada y en estado cristalino. Como los rayos X tienen una longitud de onda comparable a la distancia internuclear, la desviación que se provoque y la posterior impresión en una película será representativa de la estructura del material que se traspassa. Si ese material tiene estructura cristalina, es decir, repetitiva, se generará una difracción coherente. En el caso del ADN, las moléculas se pueden estirar y disponerse todas paralelamente de modo que los rayos X incidan perpendicularmente. Los expertos de cristalografía consiguieron determinar algunas de las periodicidades del ADN: cada 3,4 y cada 34 Å, además de una estructura helicoidal. Posteriormente se ha descubierto que estos datos son resultado de un promedio, porque el ADN presenta muchas estructuras distintas, aunque parecidas.

Con los datos disponibles, Watson y Crick determinaron un modelo de ADN que ha sido plenamente demostrado y que, además, se conoce como el tipo mayoritario de ADN en la célula. Éste se estructura como una doble hélice antiparalela (cada cadena va en un sentido) y plectonémico (no se pueden separar estirando solamente de una de las dos cadenas) arrollada a derechas. En ella las bases están en el interior y el esqueleto de azúcar–fosfato en el exterior. En su modelo las bases aparean A con T y C con G, es decir, una purina con una pirimidina y en un plano perpendicular al eje de la hélice, pero que luego se ha visto que tiene un pequeño ángulo de -6° . El apareamiento descrito es el único posible compatible con el modelo de ADN, aunque luego se han determinado otros apareamientos diferentes merced a una configuración distinta de puentes de hidrógeno y tautómeros. El paso de la hélice es de 35,7 Å (cercano a los 34 que se observaban en cristalografía), y la altura de cada residuo es de 3,4 Å, exactamente lo que se observaba en cristalografía. Cada par de bases provoca un giro de 34,3 Å en la hélice, lo que lleva a que cada vuelta tenga 10,5 pb por vuelta. El diámetro de la doble hélice es de 20 Å y es independiente de la secuencia, dado el apareamiento únicamente de purinas con pirimidinas, ya que de otro modo sería variable y haría la hélice inestable. Esto permite almacenar información siempre del mismo modo pero independientemente de la secuencia sin alterar la estructura. El eje pasa entre las bases de cada par, justo por las zonas de interacción de puentes de hidrógeno.

La doble hélice determina dos surcos de distinta anchura: el mayor y el menor. Esto se debe a las características antes citadas, principalmente por tres causas:

- La base no está centrada en la estructura del nucleótido.
- Las cadenas corren antiparalelas.
- Las bases tienen que aparear, con lo que se determina bastante la estructura.

Esto provoca que las dos hélices no discurren exactamente una frente a la otra, sino ligeramente desplazadas, de ahí los surcos de distinta anchura. Teniendo en cuenta la tautomería de las bases, los puentes de hidrógeno sólo pueden ser los que describe el modelo, y como el diámetro es constante, la distancia entre C1' también es constante, exactamente de 10,85 Å.

Ya se ha mencionado que pueden aparecer en determinadas estructuras distintas de la del modelo de Watson y Crick puentes de hidrógeno diferentes entre tautómeros diferentes de cada base. Así se generan apareamientos

inversos de Watson y Crick, o también de Hoogsten, en los que cambia tanto la configuración de los puentes de hidrógeno como los átomos implicados en su formación.

Consecuencias del modelo.

El modelo que describieron Watson y Crick es muy bueno porque aporta la idea del apareamiento de bases como algo básico en Biología molecular, permite e incluso sugiere un método de replicación, propone el antiparalelismo como uno de los elementos característicos de los ácidos nucleicos. Además, introduce el concepto de empaquetamiento de bases y excluye la posibilidad de la inclusión de moléculas de agua entre ellas. Los modelos anteriores partían de la idea de una escalera en la que las bases fueran los peldaños, pero esto deja 6,8 Å de distancia, que es suficiente para que entren moléculas de agua y la hagan inestable. El hecho de enrollar la escalera permite el acercamiento de los pares de bases hasta los 3,4 Å, que es justamente la distancia óptima de interacción mediante fuerzas de Van der Waals e interacciones hidrofóbicas, además de excluir físicamente el agua. Todo esto redundando en una mayor estabilidad del ADN.

Como consecuencia del modelo de ADN de Watson y Crick se permite la creación de teorías sobre la replicación y la mutación del ADN. El hecho de tener dos cadenas, cada una complementaria de la otra, provoca automáticamente la aparición de dos moldes, cada uno de los cuales puede dar lugar a una copia, cada una para cada célula hija. La duplicación de la información provoca automáticamente el aumento de la seguridad y la protección del material genético, además de porque la estructura es muy cerrada y las bases están protegidas en el centro frente a agresiones químicas del exterior.

Antes del modelo de Watson y Crick se desarrollaron tres posibles hipótesis de replicación del ADN: la conservativa, en que la doble cadena hace de molde de otra doble cadena igual; la semiconservativa (propuesta en el modelo), en la que cada hebra sirve de molde para fabricar la cadena complementaria, con lo que se obtienen dos dobles hélices; y la dispersiva, en la que la replicación se hace a trozos dispersos más o menos al azar que luego se unen y forman dos dobles hélices con secuencias de distintos orígenes. Meselson y Stahl elaboraron, sobre la base del modelo de Watson y Crick, experimentos de centrifugación diferencial de ADN en gradientes de CsCl según tuviera el isótopo ^{15}N o el normal ^{14}N . Los resultados de los ensayos sólo son coherentes con el modelo semiconservativo y, además, descartan cualquier otro propuesto hasta la fecha.

Las dos hebras de una molécula de ADN se pueden separar de diversas formas, ya sea por temperatura o por fuerza iónica, y esa cualidad se utiliza para su estudio. Esto significa que la estabilidad de la estructura helicoidal depende de unos factores tanto externos como internos, entre los que se encuentran:

- Puentes de hidrógeno internos entre las bases.
- Fuerzas de Van der Waals e interacciones hidrofóbicas entre los pares de bases.
- Puentes de hidrógeno externos entre el esqueleto y el agua.
- Estabilización catiónica de los fosfatos del esqueleto mediante iones divalentes metálicos como Mg^{2+} y, en menor medida Ca^{2+} , y también con policationes orgánicas como espermidina y putrescina.

Las bases exponen cada uno de sus lados a los surcos mayor y menor. De estos dos, sólo el mayor es representativo de la secuencia, y es el que utilizan las enzimas para leer el ADN, aparte de que el surco menor es ya de por sí muy pequeño para que quepa nada. Se podría considerar que las bases exponen su lado con información en el surco mayor, porque todas son distintas por ese lado. Esto es de extrema importancia para la detección de secuencias específicas por parte de los factores de transcripción y de todas las demás proteínas que interaccionan específicamente con ADN.

Otras estructuras en ADN.

El modelo de Watson y Crick representa la estructura de ADN más común en la célula, llamada ADN-B. Además de esta estructura, también se han encontrado otras variantes en determinados sitios del genoma.

Estos cambios se producen en secuencias muy cortas y suelen tener implicaciones importantes en la regulación de la expresión génica porque representan cambios reconocibles en la superficie de la molécula de ADN. Los cambios estructurales responden a cambios en la conformación de los enlaces de los nucleótidos: conformación endo o exo, anti o syn, y todos los demás enlaces que implican al fosfato y los carbonos del esqueleto externo.

ADN-A.

Es una hélice dextrógira de 11 residuos por vuelta, cada uno de los cuales sube 2,55 Å, con lo que queda una hélice más ancha y corta que la del ADN-B de Watson y Crick (23 Å de diámetro frente a 20 y 28 Å de paso frente a 35,7). Esto origina una rotación menor de cada par de bases respecto al anterior, que pasa de 34,3 a 33 Å, con lo que se permite una mayor compresión de las bases que genera una inclinación mayor (20 Å frente a -6) y que todos los pares de bases se sitúen en la periferia interna de la hélice. Por esta característica el ADN-A tiene un hueco en su eje y los puentes de hidrógeno ya no están centrados en él.

Todos estos cambios que permiten un mayor apilamiento de las bases están originados por un único cambio en la conformación de la pentosa, que pasa de tener C2'-endo a C3'-endo, lo que obliga al esqueleto a, literalmente, contraerse porque se reduce la distancia entre átomos consecutivos de fósforo de 7 Å en el ADN-B a 5,9 Å en este ADN-A. Esto permite recorrer más residuos por vuelta, lo que deriva en todos los cambios que se han descrito.

La importancia del cambio de ADN-B a ADN-A cambia la conformación de la superficie de la molécula, es decir, la anchura y profundidad de los surcos. Esto tiene una especial importancia para la regulación de una expresión génica, dado que las regiones reguladoras con sitios de unión de factores de transcripción pueden sufrir estas transiciones estructurales.

Algo que el profesor no ha mencionado, pero que yo he notado y que viene en las fotocopias, es que el surco mayor pasa a ser el menor, y viceversa, con lo que el lado accesible de las bases ya no es el informativo, sino el que no se puede leer, y esto podría tener implicaciones como las descritas antes.

ADN-Z.

Es una hélice más estilizada que las dos anteriores, pero la principal característica del ADN-Z es que la doble hélice es levógira. Evidentemente esto implica dos cosas: la estructura sufre un cambio radical, y la hélice sólo se puede conseguir dándole la vuelta a las bases, para mantener los puentes de hidrógeno descritos por Watson y Crick. Esta estructura de ADN sólo se puede dar en regiones cuya secuencia es PuPyn, y se ha estudiado *in vitro* mediante oligonucleótidos GCn. En este tipo de secuencia las guaninas (purina) se unen al azúcar en conformación syn, en lugar del anti normal en ADN-A y B, y cambian de C2' endo en ADN-B a C3' endo en ADN-Z. Las citosinas podrían hacer lo mismo, pero como la conformación syn está estéricamente impedida en las pirimidinas, por el oxígeno en 2', en este caso es todo el nucleósido (la citosina y la desoxirribosa) el que se da la vuelta, gracias a la flexibilidad de los enlaces que flanquean cada átomo de oxígeno (O5' y O3') [**corregir esto en la anterior entrega**]. Como consecuencia del giro del nucleósido entero, el esqueleto de azúcar-fosfato aparece retorcido en el ADN-Z, de modo que las distancias entre átomos de fósforo ya no son constantes y es mayor entre los que flanquean una G (7 Å) que entre los que flanquean una C (5,9 Å), produciendo una línea quebrada. Todas estas características hacen que en el ADN-Z la unidad estructural básica sea el dinucleótido GC, en lugar de los nucleótidos individuales.

Las zonas de ADN-Z se intercalan como pares PuPy en regiones puntuales de ADN-B. Este cambio provoca una distorsión enorme, además de generar un punto de inestabilidad, ya que en la interfase estructural las bases no están apiladas correctamente unas encima de las otras y se pierde la interacción, y se permiten más giros o curvamientos bruscos que en otros tipos de estructura de ADN.

Localización de cada tipo de estructura en la célula.

El ADN-B es el tipo mayoritario de ADN, y se encuentra normalmente en condiciones fisiológicas de humedad y salinidad. El ADN-A se forma al bajar la humedad en el laboratorio hasta el 75%, pero también se ha comprobado su existencia *in vivo* en algunas esporas (que están muy deshidratadas) y en determinadas secuencias cortas polidC·polidG, y en todos los segmentos de doble cadena de ARN (en ARNt y ARNr) y en los híbridos de ADN·ARN debido al hidroxilo en C2'. En este caso se dice que el ARN tiene estructura ARN-A o ARN-11, porque caben once residuos por vuelta como en el ADN-A. Hay muchas más estructuras de ADN:

- ADN-C: Reduciendo más la hidratación.
- ADN-D: 8,5 pb/vuelta. Se presenta en regiones polidA·polidT.
- ADN-T: En el genoma de algunos tipos de fagos.

El ADN-Z aparece en plásmidos para relajar tensiones debidas a superenrollamiento negativo, y en eucariotas aparecen secuencias que potencialmente adquieren esta estructura en regiones esenciales para el metabolismo de ácidos nucleicos y la expresión génica, porque se asocia a regiones de comienzo de la transcripción (caja GACA o secuencia -35).

Variaciones en la estructura.

La estructura global del ADN, independientemente de cuál se trate, es en realidad un promedio de pequeñas variaciones sobre el total, cada una de ellas generadas por cada par de bases. Es decir, la secuencia determina las variaciones puntuales en la estructura global, pero no cambia esta última. Esta característica de conservar una estructura superior y promediada a base de acumular, o a pesar de que existan pequeñas variaciones incorporadas o producidas por cada parte de ella, se llama metaestabilidad. El ADN es un inmenso repertorio de estructuras distintas y cambiantes, incluidas cada una de ellas por cada par de bases distinto, pero que sumadas todas las variaciones resulta la estructura promedio que antes se ha descrito, por esto se dice que el ADN es metaestable.

Sin embargo, dado que el ADN no es una molécula rígida, la estructura se puede distorsionar muy fácilmente por, entre otras cosas, los agentes intercalantes. Éstos son moléculas pequeñas, apolares y generalmente aromáticas o planas, de tal modo que pueden introducirse entre dos pares de bases consecutivos y forzar una distorsión en la hélice. Estos compuestos se utilizan en el laboratorio para detectar ácidos nucleicos, aprovechando que la mayoría de estas sustancias también son fluorescentes cuando se exponen a luz ultravioleta (bromuro de etidio, naranja de acridina). Estas sustancias son capaces de introducirse entre las bases porque la estructura del ADN no es inmutable, sino que existen determinados momentos en los que la vibración propia de toda molécula provoca un estiramiento puntual de la hélice y, consecuentemente, la separación de los pares de bases adyacentes. Ese momento es el que permite al agente intercalante introducirse entre los dos pares de bases, donde se quedará porque establece interacciones hidrofóbicas con ellas.

ADN triple.

Se pueden producir asociaciones de tres cadenas de ADN si tienen secuencias poliPu o poliPy. Se puede considerar como una doble hélice poliPu·poliPy a la que se enrolla otra de cualquiera de los dos tipos anteriores sobre su surco mayor, con orientación distinta según tenga sólo Pu o sólo Py. Esta asociación ocurre en determinados sitios dependiendo de la secuencia, y generalmente forma parte de alguna estructura terciaria sobre una sola hélice doble en la que una hebra vuelve a enrollarse sobre sí misma. En esta triple hélice se observan apareamientos entre bases que no son los de Watson y Crick, sino que aparecen mezclados con otros raros, como los de Hoogsteen, pero la estructura es bastante menos estable que un hélice doble convencional.

Estabilidad del ADN.

La estabilidad de una doble hélice de ADN depende de muchos factores, tanto internos como externos, de los que todos se producen a la vez y cooperativamente:

- Grado de polaridad del solvente. El ADN es más estable cuanto más polar sea el medio.
- Tipo y concentración de iones. La estabilidad aumenta cuanto mayor sea la concentración de cationes divalentes en el medio.
- pH. Cuanto menor es el grado de pH, más estable es la doble hélice, porque al aumentar la basicidad pueden llegar a desprotonarse las bases e impedir la formación de puentes de hidrógeno.
- El porcentaje de pares G·C frente a A·T. La unión de G con C se realiza a través de tres puentes de hidrógeno, mientras que la unión A·T sólo tiene dos. Cuanto mayor sea la proporción de G + C, mayor será la estabilidad. Este apareamiento se produce espontáneamente entre bases sueltas.
- Apilamiento de bases. Este fenómeno es también espontáneo en una solución de bases sueltas.

Siempre se tiende a pensar que el ADN se mantiene en su configuración de doble hélice gracias a los puentes de hidrógeno, pero en la mayoría de los casos no es así, sino que lo que mantiene unida la estructura son las interacciones hidrofóbicas y las fuerzas de Van der Waals que se dan entre pares de bases contiguos. La importancia de los puentes de hidrógeno es, más bien, la especificidad que otorgan a la molécula y a todos los procesos en los que se ve implicado algún tipo de ácido nucleico, siendo su contribución a la estabilidad de la doble hélice algo secundario aunque nada desdeñable.

Desnaturalización y renaturalización del ADN.

Desnaturalización.

La desnaturalización de una molécula de ADN de doble cadena implica la separación de las dos hebras. Esto se consigue aumentando el pH hasta niveles en los que todas las bases se encuentren desprotonadas y, por tanto, no se establezcan puentes de hidrógeno, o también elevando la temperatura para separar las hebras mediante el aumento de la vibración térmica. Cualquiera de los métodos anteriores, aunque se utiliza más el segundo, permite obtener dos moléculas de ADN de cadena sencilla sin conformación alguna en la que no hay puentes de hidrógeno ni apilamiento de bases, partiendo de una doble hélice.

La separación en dos hebras lleva aparejados una serie de cambios en las variables macroscópicas de una disolución de ADN:

- Se pierde la viscosidad característica de las disoluciones de ADN de doble cadena. Esto se debe a la pérdida de la conformación helicoidal, relativamente rígida y que forma hilos largos, y la sustitución por una estructura sin conformación definida, capaz de formar ovillos que presentan mucha menos fricción intermolecular y, por tanto, menos viscosidad.
- Se produce un aumento de la absorción a 280 nm. Este es el llamado efecto hipercrómico, que se produce porque la pérdida de las interacciones hidrofóbicas y las fuerzas de Van der Waals entre las bases libera a los electrones que estaban anteriormente implicados, con lo que ahora contribuyen a la absorción de radiación. Este aumento puede llegar a ser de un 40% entre las formas de cadena doble y sencilla, pero la forma del espectro no cambia.

El efecto hipercrómico se utiliza para cuantificar la cantidad de ADN que se tiene en forma de doble hebra o de cadena sencilla. Al representar la temperatura frente a la absorción a 280 nm se observa una curva sigmoidea, con lo que se deduce que las fuerzas que estabilizan la estructura helicoidal son cooperativas, es decir, cuesta mucho romper las primeras interacciones estabilizantes, pero una vez conseguido el resto se rompe muy fácilmente. Hay un umbral mínimo de energía a partir del cual el proceso de desnaturalización es muy rápido.

La temperatura a la cual se ha conseguido separar el 50% de las hebras se llama temperatura de fusión TM (de *melting*: fusión en inglés), y da un valor representativo de la estabilidad de una secuencia determinada de ADN. Se ha comprobado que este valor depende de todos los factores que dan estabilidad al ADN, pero sobre todo depende de la cantidad de pares G·C que haya, es decir, de la cantidad de puentes de hidrógeno que se establezcan. Esto pone de relevancia que estos enlaces también contribuyen de manera importante a la estabilidad de la doble hélice o, más bien, que todos los procesos o características que estabilizan la estructura actúan conjunta y cooperativamente. Se ha caracterizado una relación lineal entre el porcentaje de G + C y la TM, de modo que ésta aumenta 0,4 °C cada 1% más de G + C. En el hombre y los mamíferos el porcentaje medio de G + C es de un 40%, con lo que la temperatura de fusión ronda los 87 °C, pero hay zonas del genoma, concretamente los promotores, en los que la proporción de G + C sube hasta un 60%, con lo que son zonas del ADN especialmente estables en las que la TM tiene un valor de hasta 95 °C.

También se ha encontrado una relación directa de la temperatura de fusión con la concentración de cationes en el medio, que era otro de los factores estabilizantes a los que se aludía antes. Concretamente los cationes monovalentes estabilizan la estructura inespecíficamente (Na⁺, Li⁺, K⁺), y los divalentes son más específicos y efectivos y son los que se utilizan en experimentos donde es necesaria una regulación fina de la estabilidad del ADN (Mg²⁺, Ca²⁺ aunque menos utilizado). Algunos de estos procesos incluyen la amplificación en PCR, donde la concentración de sal es uno de los parámetros a optimizar, ya que esta técnica se basa en la regulación de la estabilidad de las moléculas de ADN para un apareamiento selectivo de una que interesa (el cebador). Tan previsible es la TM que hay fórmulas para hallarla, entre las cuales ésta es un ejemplo:

$$TM = 0,411(\% \text{ de G + C}) + 16,6[\text{Na}^+] + 81,5$$

Renaturalización.

La renaturalización de moléculas de ADN es el proceso contrario al de la desnaturalización, es decir, se parte de dos hebras sencillas para volver a formar una doble hélice. Éste es también un proceso cooperativo en el que hace falta un primer apareamiento estable que haga de semilla para después seguir a lo largo de las hebras a modo de cremallera. La formación de esas primeras interacciones se llama nucleación, y necesita una mínima cantidad de tiempo para que se den los contactos necesarios entre las bases complementarias. Sin embargo, a diferencia de la desnaturalización, para la correcta renaturalización del ADN hace falta que se baje la temperatura lentamente y no muy por debajo de la TM (10 ó 15 °C menos) para mantener el equilibrio desnaturalización-renaturalización (que las moléculas puedan moverse y encontrarse, pero que las uniones no correctas no sean tan estables ni permanentes) y evitar que aparezcan secuencias ligeramente homólogas donde no deben. Si se bajara la temperatura bruscamente se formaría un gel de moléculas de hebra sencilla parcialmente apareadas en segmentos cortos con homología secuencial, incluso dentro de una misma hebra. Esta característica se utiliza para guardar muestras de ADN que se quieran conservar como hebras sencillas.

Hibridación de ADN.

Las técnicas de hibridación de ADN se basan en la detección de secuencias específicas, ya sea sobre tejido (hibridación *in situ*; para ver la localización celular del ácido nucleico que se busque) o sobre ácidos nucleicos purificados (sobre ADN técnica de Southern; sobre ARN técnica de Northern), mediante el empleo de sondas. Éstas son fragmentos cortos de ácido nucleico sintetizado en el laboratorio, cuya secuencia es la complementaria de aquélla que se busca y están marcados de algún modo para poder luego detectarlos con facilidad. Existen varios tipos de sondas:

- ADNss: Se utiliza para detectar ARNm. Generalmente sólo puede tener una marca radiactiva en un extremo.
- ADNds: Primero se desnaturaliza y se enfría rápidamente para obtener ADNss, que es el que se usa. Sin embargo, este método provoca la competencia entre las mismas moléculas de la sonda, con lo que se pierde efectividad. La ventaja es que suelen ser más largos que los ADNss y, al ser de cadena

doble, se puede introducir más radiactividad y a la postre el resultado puede ser más sensible.

- ADNc (ds): Son copias complementarias a un ARNm, utilizados como las sondas anteriores.
- ARNc (ss): A veces es la técnica más sensible. Para esto se parte del ADNc de la proteína que se quiere estudiar, al que se le añade un promotor y un sitio de corte al final. Sobre esto se añade la polimerasa correspondiente a ese promotor y una colección de nucleótidos, que pueden estar todos marcados. El ARN que se forma es el ARNc, con secuencia complementaria a la del ARNm. Si no se conoce la orientación del ARNm, entonces se colocan dos promotores y dos sitios de restricción a sendos lados del ADNc, y se obtiene una colección de ARNc, cada uno complementario de una hebra, y la mitad de ellos complementarios del ARNm. Como se incorpora una gran cantidad de marca, después se trata con una ARNasa que elimina los restos de ARNss para resaltar el contraste, y ésta es la gran ventaja frente a las sondas de ADN.

Los experimentos de hibridación se basan en la complementariedad de bases, pero también puede haber secuencias cortas o incluso componentes celulares que hibriden o se asocien inespecíficamente a la sonda. Para minimizar este problema se juega con las condiciones de hibridación (renaturalización: T^a , concentraciones) y de lavado (desnaturalización), sobre todo con estas últimas. Se puede hibridar a temperaturas más altas, para que la unión sea más específica, pero sobre todo se procede a un lavado riguroso después de hibridar a unas condiciones poco severas, aumentando la temperatura y bajando la concentración de iones para que sólo queden apareadas las sondas con sus secuencias complementarias.

Estructura terciaria.

Estructuras en horquilla.

Aparece frecuentemente en ácidos nucleicos de cadena sencilla que tengan regiones complementarias en su propia secuencia. La estructura básica es una doble hélice en cuyo extremo hay un bucle de unas cuantas bases desapareadas. Normalmente también hay bucles internos dentro de la hélice en los que no hay complementariedad de bases, e incluso nucleótidos que no encajan entre dos pares de bases y bien salen hacia fuera o bien se intercalan entre ellos, pero de todos modos se encuentra en una situación energéticamente desfavorable. Esta estructura es muy importante para determinadas proteínas, porque determina la función (como ocurría con la forma de las proteínas), sobre todo de ARNt, ARNr y los ARNsn (fragmentos U).

Hay programas de ordenador capaces de determinar la estructura de una molécula de ácido nucleico teniendo en cuenta el balance energético entre la estabilización por cada base apareada y la desestabilización por cada base no desapareada, según esté en un bucle interno, terminal, o sin pareja en una doble hélice. Mediante estas técnicas se puede inferir la estructura más estable.

Esta estructura se encuentra también en algunos virus de ADN, en los que ayuda a su propia replicación.

Repeticiones invertidas.

En los ácidos nucleicos pueden aparecer secuencias de unos 10 a 15 nt (aunque pueden ser más largas) que tienen una repetición complementaria y en sentido contrario al que están definidas y que, además, pueden estar separadas o no por nucleótidos no simétricos.

En el caso de ADNds, las repeticiones invertidas están próximas entre sí y, además, también existen en la hebra complementaria, con lo que se forma una estructura doblemente simétrica (o también con un eje de simetría) llamada palíndromo. En este caso la formación de bucles en cada hebra por el método descrito antes implica la separación previa del ADN en sus dos hebras partiendo del eje de simetría y la posterior nucleación y reapareamiento sobre la propia hebra. Al final se obtiene un doble bucle que por su forma recuerda una cruz y, por tanto, se llama cruciforme. La energía necesaria para separar la doble hélice en sus dos hebras se recupera luego en la formación del cruciforme.

En el caso del ARNss, cuando se copia el palíndromo de ADN en ARNm, se permite el establecimiento de apareamientos intracatenarios, lo que generaría algo parecido a una horquilla, pero a menor escala y en un rango mucho menor de nucleótidos. Este mecanismo es el que se utiliza en la regulación de la expresión génica en bacterias mediante atenuación, que consiste en la formación de un bucle de unos 15 pb en distintos sitios de la molécula de ARNm según las disponibilidades de elementos para la síntesis de las proteínas del operón correspondiente (recordar el operón Trp).

La importancia de los cruciformes en el ADN es la capacidad de intercambio con otras estructuras como superenrollamientos negativos, lo que es muy importante en la topología y regulación metabólica del ADN.

Topología.

Se llama topología del ADN a la forma tridimensional que adquiere esta molécula en el espacio. Los elementos anteriores considerados como estructuras terciarias son, realmente, una extensión tridimensional de la estructura secundaria. Hay que decir a este respecto que el límite entre estructura secundaria y terciaria, e incluso otras estructuras de orden superior e inferior está muy difuso en los ácidos nucleicos, no como en las proteínas que los límites son claros. De todos modos, la verdadera estructura terciaria tridimensional es lo que se conoce como topología de ADN. Para estudiarla hace falta tener en cuenta una serie de conceptos, y recurrir insistentemente a ellos:

- La doble hélice de ADN es una estructura flexible y que se puede doblar, a veces muy bruscamente.
- Los cambios en la topología son esenciales para la función y regulación del metabolismo del ADN.
- El número de bases por vuelta es flexible y ajustable según las circunstancias y las necesidades de distintos procesos o secuencias concretas del ADN.
- La flexibilidad de una determinada región en la doble hélice provoca un cambio opuesto y complementario en las zonas adyacentes. Es decir, que si la hélice se relaja en un punto determinado, las regiones que flanquean la zona relajada sufren una sobretensión equivalente.
- El curvamiento del ADN responde a condiciones fisiológicas y también puede ser provocado por unión a distintas estructuras, generalmente proteicas y relacionadas con la transcripción o la replicación.
- El ADN se tiene que separar en sus dos hebras para poder realizar los procesos citados anteriormente. Paralelamente, los organismos cuyo material genético es ADNss deben adquirir estructura de doble hélice para replicarse.
- Por tanto, la estructura del ADN es extremadamente dinámica.

Curvamiento del ADN.

El curvamiento de las dobles hélices puede deberse a la propia secuencia de pares de bases, a la aparición de interacciones con proteínas, o a ambas cosas a la vez. Si nos limitamos a una doble hélice sin más elementos externos o internos que distorsionen la estructura, existe una longitud mínima, llamada de persistencia, en la que el ADN no se puede doblar y es de unas 200 pb. Sin embargo, existe un microorganismo patógeno cuyo ADN mitocondrial (ADNmit) tiene zonas de 200 pb capaces, no sólo de curvarse, sino de formar círculos, observación corroborada por los datos de movilidad electroforética, en los que se aprecia una menor distancia recorrida que para un fragmento lineal del mismo peso molecular. Esto indica un fuerte curvamiento que hace que la molécula de enganche más a menudo en el gel de electroforesis.

La hipótesis para explicar este curvamiento anómalo es la existencia de una secuencia P que específicamente hace que el ADN se curve. Para determinar la posición de la secuencia se obtuvo un dímero del ADNmit uniendo dos moléculas cabeza con cola, y tratando luego con distintas enzimas de restricción para obtener una colección de moléculas circularmente permutadas. Cada una de estas moléculas tiene el mismo peso molecular, que es igual al de la molécula de origen, pero su secuencia está desplazada ligeramente entre cada tipo de molécula modificada. Esto permitió predecir los resultados del ensayo en base a la hipótesis

formulada. La mayor o menor movilidad electroforética de cada especie deberá indicar la mayor o menor proximidad de la secuencia P al extremo, ya que si se aleja del centro de la molécula el efecto de curvatura del ADN será menor y, por tanto, la movilidad mayor.

El ensayo permitió determinar una secuencia característica alrededor del nt 140, en la que se repetían grupos de 5 ó 6 A cada 10 nt, es decir, cada vuelta. Los segmentos de poli-A tienen un ángulo de rodada característico que, junto a todos los demás, forman lo que se llama un sitio de curvamiento (*bending locus*), y su situación estratégica cada 10 nt hace que estén todos en el mismo lado de la hélice, con lo que el efecto de curvatura se suma; se dice que están en fase. Si, por el contrario, la fase es cada 16 nt (vuelta y media), entonces se disponen las secuencias alternativamente a cada lado de la hélice, con lo que se consigue una molécula en zigzag. Experimentalmente se demostró que grupos de 3 ó 4 A espaciados 10 nt hacen que el ADN se curve, y no cualquier otro espaciamiento distinto.

Aparte de secuencias con curvatura, también hay secuencias susceptibles de ser curvadas en determinados momentos, sobre todo cuando se produce la unión con proteínas. Puede darse cualquier combinación, desde ADN curvo que sirve para unir proteínas hasta proteínas que obligan al ADN a curvarse, pasando por todos los estados intermedios. Un ejemplo es la unión de la proteína CAP (*catabolyte activator protein*: proteína activadora de catabolito) a la región operadora del operón lactosa. Esta región está curvada, y la unión es específica para esta proteína. La demostración de la curvatura del ADN se realizó mediante un experimento de características similares las del de arriba.

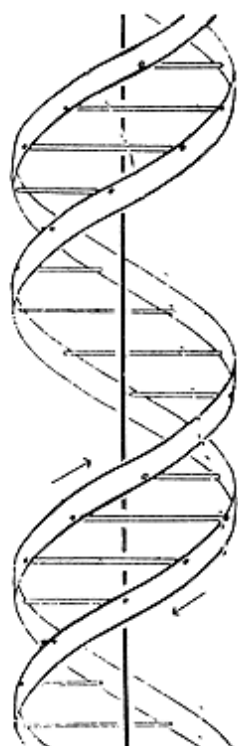
La curvatura del ADN también se encuentra en el sitio de origen de la replicación. Se ha estudiado en detalle el virus SV40, en el que existe una secuencia de poli-A que aporta una gran flexibilidad a la zona. Cuando el virus entra en la célula se produce una proteína llamada antígeno T que, dimerizado, se une específicamente a la zona del origen de la transcripción para permitir la apertura de la hélice y que comience la replicación del virus. En otros orígenes de replicación también se encuentran grupos de poli-A en fase que aportan la flexibilidad necesaria para poder abrir la hélice.

Una manera de provocar curvaturas en el ADN es mediante la introducción de fotodímeros de T debido a la excitación con luz ultravioleta. La dimerización impide el giro de una base respecto a la otra, por lo que se provoca un curvamiento en la hélice. El sistema de reparación específico de esta distorsión no reconoce el dímero de T hasta que no encuentra previamente la distorsión de la doble hélice.

En eucariontes el inicio de la transcripción depende siempre de la unión de factores proteicos específicos. Siempre hace falta la unión de un primer factor a la zona del promotor, exactamente a la caja TATA. Este factor (TBP: *TATA binding protein*; proteína de unión a TATA) provoca un curvamiento muy brusco, casi se diría que es una doblez, en la doble hélice. Esta distorsión tan enorme es la que realmente permite la unión posterior de otros factores de transcripción y, finalmente, la ARN polimerasa. Sin la curvatura inducida por TBP no sería posible la transcripción.

1

32



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis