



UNIVERSIDAD NACIONAL  
AVENIDA DE  
MEXICO

*Genética de insectos  
Ecolología y evolución  
mutaciones*

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

21

FACULTAD DE CIENCIAS

00361

zej

CONSIDERACIONES BIOFISICAS SOBRE  
LAS MUTACIONES PUNTUALES

TESIS QUE PARA OBTENER EL GRADO DE  
MAESTRO EN CIENCIAS  
(BIOLOGIA)

PRESENTA

LUIS MEDRANO GONZALEZ.

MAYO DE 1989.

TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## INDICE

PREFACIO .....	1
RESUMEN .....	2
INTRODUCCION .....	3
REVISION GENERAL SOBRE LA MUTAGENESIS .....	7
§ Bosquejo histórico. ....	7
§ Mecanismos de las mutaciones. ....	9
§ Aspectos evolutivos del proceso de mutación. ....	15
ANTECEDENTES Y PLANTEAMIENTO DEL PROBLEMA DE ESTUDIO .....	21
§ Caracter no aleatorio de la secuencia de bases en el DNA. ....	21
§ La evolución molecular en los pseudogenes. ....	25
§ La regionalización en la composición de bases en el genoma de los vertebrados. ....	26
§ Factores físicos en la evolución molecular. ....	28
PROCEDIMIENTOS DE ANALISIS .....	31
§ Generalidades. ....	31
§ Análisis de las globinas y del hombre y el gorila y del gene <i>lacI</i> de <i>Escherichia coli</i> . ....	32
§ Distribución estadística de las energías de amarre en distintas secuencias de DNA. ....	33
§ Clasificación y agrupamiento de digramas. ....	34
§ Distribución de energías locales en secuencias específicas. ....	37
DISCUSION .....	40
§ Nociones sobre la relación de la energía local en el DNA y la mutagénesis. ....	40
§ Estructura energética de diferentes tipos de secuencias. ....	45
CONCLUSIONES .....	51
ABREVIATURAS .....	53
REFERENCIAS .....	54

## PREFACIO

Esta tesis se realizó en el Instituto de Física de la UNAM bajo la dirección de los Dres. Germinal Cocho y José Luis Rius. El trabajo es solamente una parte de un proyecto interdisciplinario en el que también participa el Físico Pedro Miramontes a quien agradezco su ayuda en la resolución de varios problemas de análisis y de cómputo. Estoy en deuda por la valiosa ayuda de los Dres. Victor Manuel Valdés, Daniel Piñero, Antonio Lazcano, Jesús Manuel León Cázares y George Dreyfus, sinodales de esta tesis así como con el Dr. Anelio Aguayo, tutor de mis estudios de maestría. También agradezco las atenciones del equipo de cómputo del Instituto de Física. Debo reconocer la ayuda y la confianza que me dieron muchas otras personas cuyos nombres son numerosos para citarlos aquí pero que no olvido. Aunque este trabajo resulta de un esfuerzo colectivo, los errores de todo tipo que subsistan son responsabilidad mía.

En algunas secciones del trabajo me he tomado la libertad de especular demasiado y de expresar opiniones personales. Considero que los trabajos de tesis pueden ser materiales valiosos de consulta porque, independientemente de sus resultados y de su calidad, deben reflejar la experiencia de quien trabaja en un tema específico. Por ello me he permitido ciertos excesos si bien, se definen los alcances y las limitaciones reales de la investigación. La tesis no tiene como objetivo hacer una exposición sistemática y detallada sobre el proceso de mutagénesis, sino plantear un enfoque de estudio que tiene la pretensión de integrar los diversos aspectos del fenómeno en un discurso global, del cual, este trabajo es un primer reporte. Por esta razón, el capítulo dedicado a la revisión sobre las mutaciones, solamente proporciona un panorama general así como algunos elementos que son útiles en la discusión posterior.

Esta tesis está dedicada a la memoria de María Elena Sandoval y Rolando Lara.

Luis Medrano G.

## RESUMEN

Se investigó la distribución de energías de amarre en varias secuencias de genes de globinas e inmunoglobulinas. Por otra parte, se desarrolló un análisis del espectro de mutaciones en cepas de *Escherichia coli* carentes de sistemas de reparación y de las mutaciones que se detectan en la comparación de secuencias alineadas de las globinas y del hombre y el gorila. Este análisis consistió en relacionar la frecuencia de mutaciones en un determinado sitio de la secuencia, con la entalpía local de amarre.

A partir de agrupar los diez diferentes enlaces en la doble hélice en tres clases de energía (D1, D2 y D3), se hizo un esquema de todas las vías posibles de transformación entre unos y otros por medio de una sustitución al cual se denominó red mutacional.

Los resultados sobre la relación de la frecuencia de mutaciones con la energía de amarre y la distribución de esta energía en diversas secuencias, indican que las mutaciones ocurren preferentemente en los sitios con menor entalpía. Esto provoca que, en ausencia de restricciones por la selección natural, a largo plazo se acumulen los enlaces fuertes GG/CC, GC/CG y CG/GC. Cuando hay metilación de la citosina, estos tres dígramas se vuelven sitios con alta incidencia de la transición 5-mC  $\rightarrow$  T y la tendencia en la red mutacional a formar enlaces fuertes, lleva a la acumulación de los dígramas TA/AT, AT/TA y AA/TT. Ya que la secuencia en los exones está integrada por la secuencia de codones, se estudió la distribución de la energía en las tres posiciones del codón. Se observó que el enlace entre la segunda y tercera base del codón, por ser el menos restringido selectivamente, refleja las restricciones físicas en los eventos de mutación. El enlace entre la primera y segunda base del codón, muestra en general valores de energía independientes de la energía total en las secuencias exónicas. Esta diferenciación origina una estructura energética característica de cada tipo de secuencia. En los intrones, la distribución de la energía es uniforme en las tres posiciones.

Se concluye que no sólo las restricciones selectivas definen la secuencia de nucleótidos en el DNA sino que existe una dinámica mutacional subyacente no aleatoria, que depende de la estabilidad de la molécula de DNA. Este proceso físico es más evidente en ausencia de restricciones selectivas.

## INTRODUCCION

En alguna ocasión Beethoven declaró que en la música todo debe ser sorprendente y esperado a la vez. Si una serie de notas es monótona, hecha con un arreglo sencillo, el resultado es el aburrimiento. Pero, si tal sucesión se da carente de algún orden, tampoco es posible apreciar algo que podamos identificar como música.

La forma en los seres vivos, y por extensión la belleza que en ellos apreciamos, también refleja la integración del orden con elementos de novedad. Entre la multitud de formas en los organismos podemos distinguir en todas ellas que la distribución de sus partes no es aleatoria sino que éstas guardan ciertas relaciones por las que se puede reconocer una estructura ordenada, sin embargo, la forma en los organismos no es resultado de alguna especie de impulso artístico metafísico. Casi siempre, es posible vincular una morfología específica con las funciones que el organismo realiza. Aquí, el término función no se interpreta en un sentido teleológico. Se puede considerar válido el uso de esta palabra pues en campos como la Matemática y la Fisiología tiene un significado específico distinto al de finalidad o propósito.<sup>(81)</sup>

Conviene recordar uno de los ejemplos clásicos de la relación entre la función y las posibilidades de la morfología. En las siete clases de vertebrados actuales es posible encontrar distintas especies que viven parcial o totalmente en el agua. A pesar de la gran variedad de formas que se pueden encontrar, se observa en todos estos animales, la existencia de elementos comunes en la anatomía externa relacionados con la natación tales como la forma fusiforme del cuerpo, apéndices en forma de aletas y comúnmente, una aleta de propulsión.

Quizás, al igual que en la música y que en todas las formas de expresión estética, la combinación del orden y la incertidumbre es una causa por la cual las formas vivas son bellas para nuestra percepción. Pero, la admiración por esta belleza no se agota con la contemplación de estructuras sino que abarca el acto que de ellas resulta. La existencia de una morfología específica en los organismos se relaciona estrechamente con los procesos que éstos

realizan. En tanto que el organismo es un sistema que tiende a mantener la continuidad de su propio orden interno, los procesos orgánicos se conjugan de modo que éste supera las limitaciones físicas que impiden dicha continuidad, lo cual implica que el organismo asimila, de alguna forma, energía libre de su entorno. Las transformaciones energéticas involucradas en los procesos biológicos llaman la atención además porque su eficiencia termodinámica es mayor que la de procesos análogos realizados por los sistemas más perfeccionados hechos por el hombre.

Lo anterior no significa que para explicar la existencia de la vida y la aparente racionalidad en sus procesos, tengamos que recurrir a entidades metafísicas o a principios teleológicos sino que, como lo plantea Oparin,<sup>(80)</sup> para comprender estos sistemas debemos estudiar el origen de la vida a partir de la materia inanimada y su evolución en la tierra.

En este contexto, debemos hacer algunas consideraciones a nuestro juicio importantes. Primero, que los seres vivos son sistemas físicos que interaccionan con un ambiente que también lo es por lo que, al concebir la Física como la ciencia que estudia la estructura y las propiedades de la materia, debemos incluir en este estudio a los organismos. Esto conduce a interpretar que, como sostiene Volkenshtein,<sup>(100)</sup> la Biofísica no es solamente una rama auxiliar de la Biología sino que es el instrumento teórico de fundamentación científica del conocimiento biológico, lo cual, no quiere decir que la Biología deba reducirse a la Física.

Los seres vivos son entes físicos en los que a diferencia de los sistemas no biológicos, existe una integridad relativa de su orden interno que se logra a partir de la asimilación de alguna forma de energía libre, la cual, en última instancia proviene del sol. A largo plazo, la continuidad de los sistemas vivos involucra la reproducción de ellos mismos. Aunque no pueda darse una definición de lo que es la vida, sabemos que, en general, podemos distinguir a los seres vivos a partir de estas dos características: el metabolismo y la reproducción.

El organismo como tal es resultado de las características particulares que en él tiene el metabolismo el que, a su vez, se origina con base en la información contenida en la molécula del

ácido desoxirribonucleico (DNA) y de ahí, entonces, que sea el DNA la molécula portadora de la información transferida en los organismos durante la reproducción. Esa información genética se encuentra en la secuencia de las bases nitrogenadas que constituyen la molécula de DNA. Dichas bases son: la adenina (A), la guanina (G), la timina (T) y la citosina (C). En algunos virus el genoma está constituido por el ácido ribonucleico (RNA). Se piensa que el RNA fue el portador de la información genética en las primeras células que existieron en nuestro planeta. (61)

La información almacenada en el DNA no se conserva íntegra ya que ocasionalmente ocurren cambios en la molécula, los cuales, son de naturaleza diversa y en conjunto se les conoce como mutaciones. Como explica la teoría de la selección natural, es por las mutaciones por lo que en los organismos puede haber un proceso de adaptación al medio ambiente pues son la fuente de cambio en la evolución de éstos. No debemos olvidar que los fenómenos de recombinación son un factor determinante en la formación de variabilidad sobre la cual opera la selección natural.

De acuerdo a Dobzhansky et al., (26) podemos agrupar las mutaciones en dos clases: las mutaciones génicas y las cromosómicas de las que, en particular, nos interesan las primeras por razones que serán evidentes más adelante. A su vez, las mutaciones génicas comprenden fenómenos como las inserciones, deleciones y sustituciones de segmentos cortos de DNA. Cuando el cambio involucre sólo un nucleótido, lo cual ocurre sobre todo en las sustituciones, la mutación se denomina puntual. Este no es el sentido que originalmente tuvo el término al ser propuesto por Morgan (74) quien no conocía la naturaleza del material de la herencia.

Si bien la mutación es un proceso que existe en el ámbito de lo biológico, es también un fenómeno físico y, por lo tanto, sujeto a los principios de la Física. En el contexto de la teoría de la evolución, pensamos habitualmente que las mutaciones ocurren de modo aleatorio, pero formalmente, la afirmación se refiere a que las mutaciones ocurren al azar en relación a la adecuación del organismo. Partimos de que, al ser el origen del cambio evolutivo, las mutaciones no obedecen a la determinación de voluntades



metafísicas pero eso no significa que, siendo un fenómeno de la realidad, se conciben como indeterminadas y fuera de las leyes naturales establecidas por la Física. Por lo tanto, es posible plantear la hipótesis de que existen restricciones físicas en el proceso de mutación que pueden reflejarse en la secuencia de nucleótidos del DNA. Veremos más adelante que la secuencia de nucleótidos no es aleatoria. Sin embargo, existen diferentes tipos de fenómenos involucrados en la determinación de la secuencia de bases entre los cuales podemos distinguir distintos factores físicos.

El objetivo de este trabajo ha sido el estudiar algunos aspectos físicos de los procesos de mutación y analizar sus implicaciones biológicas. Merced de que se trata de un tema complejo, en el presente trabajo nos hemos avocado, inicialmente, a estudiar algunas cuestiones globales relacionadas con las sustituciones de nucleótidos que son la forma más común de mutaciones. Las sustituciones, al ser más fáciles de analizar, nos brindan nociones que se pueden aplicar al tratar con procesos mutacionales más complicados. Pensamos que con ese enfoque se tiene un buen modelo para estudiar el proceso de la evolución desde el punto de vista físico.

## REVISION GENERAL SOBRE LA MUTAGENESIS

### § Bosquejo histórico.

Desde el siglo XIX, algunos naturalistas habían denominado mutaciones al surgimiento de variaciones hereditarias discontinuas de una generación a otra de organismos. A finales del siglo pasado y principios del presente, Hugo de Vries reconoció la existencia de mutaciones en la planta *Oenothera lamarckiana* y propuso que éstas conforman el cambio de los seres vivos en la evolución, lo que dio lugar a la teoría que en aquellos años se llamó mutacionismo.

El origen de las mutaciones fue un problema sujeto a muchas especulaciones. Muchos naturalistas pensaban que eran espontáneas. Muller<sup>(75)</sup> descubrió que las mutaciones se pueden inducir con rayos X, y que surgen con una cierta regularidad que se puede medir. Además de los rayos X, se descubrió que otras radiaciones de alta energía y algunas sustancias pueden provocar mutaciones, lo cual, hacía evidente la naturaleza física del proceso y, por lo tanto, la posibilidad de ser investigado. Muller encontró que muchos de los mutantes obtenidos por irradiación eran idénticos a los que aparecen espontáneamente y propuso que las mutaciones naturales probablemente son el resultado de los efectos de las radiaciones que hay en la naturaleza.

Aunque las ideas de Muller son correctas actualmente, es a Morgan<sup>(74)</sup> a quien se atribuye el concepto moderno de mutación, el cual, desarrolló con sus estudios en la mosca *Drosophila melanogaster*. Morgan descubrió moscas mutantes cuyos cromosomas, a diferencia de los mutantes de De Vries, no presentaban alteraciones morfológicas evidentes. Morgan determinó, con base en la segregación mendeliana, que estas mutaciones involucraban sólo un gen y comunmente aparecían en forma recesiva. A estas mutaciones las denominó puntuales. De acuerdo a Morgan, definimos la mutación como un cambio en los genes, y por ende heredable, que se puede manifestar en la morfología, el comportamiento, la reproducción, o cualquier otra característica del organismo, y puede tener efectos drásticos o sutiles.

Esta noción sobre las mutaciones adelantó con el descubrimiento de las mutaciones auxotróficas en el hongo *Neurospora*, i.e., se había encontrado que algunos microorganismos mutantes no pueden crecer en un medio si no se adiciona un nutriente específico. Beadle y Tatum<sup>(4)</sup> encontraron que esta dependencia se debe a que los mutantes auxotróficos carecen de una enzima necesaria en la síntesis del nutriente en cuestión. De esta forma, se encontró que los genes determinaban la estructura de una enzima y, en general, de una proteína, y que una mutación en ellos conducía a una alteración en la función de dicha enzima.

Sin embargo, el origen de las mutaciones permanecía oscuro porque no se conocía el sustrato material de la herencia, lo que fue resuelto por Avery y sus colaboradores en 1944<sup>(2)</sup> y por Hershey y Chase en 1952.<sup>(47)</sup> El descubrimiento de la estructura del DNA por Watson y Crick<sup>(103)</sup> puso de manifiesto que en la secuencia de nucleótidos en la molécula, se encuentra la información hereditaria. Esto hizo pensar que, en las mutaciones puntuales, se altera dicha secuencia. Watson y Crick<sup>(104)</sup> propusieron que la mutación puntual ocurre como un proceso en el que las bases nitrogenadas ocasionalmente adquieren una de sus formas tautómeras raras y ello provoca que éstas hagan apareamientos de manera distinta a la normal de modo que en la síntesis de DNA puede incorporarse una base equivocada.

El desarrollo de la investigación en los microorganismos particularmente *Escherichia coli* y sus fagos, llevó a un conocimiento detallado de los procesos genéticos en estos sistemas. La sencillez relativa de los procesos de la herencia en estos organismos, ha permitido la realización de una cantidad innumerable de experimentos tendientes a explicar los fenómenos de mutagénesis, aunque un problema no trivial es el de en qué medida los resultados obtenidos pueden aplicarse a otras formas de vida. Inicialmente, las mutaciones se estudiaron de modo indirecto, en forma de alteraciones fenotípicas. En los últimos años, las técnicas de clonación y secuenciación de DNA han favorecido la caracterización directa de las mutaciones.

Existen diferentes formas de mutaciones y, por lo tanto, no es posible referirse a un mecanismo único de mutación. Se mencionó que las mutaciones en general pueden ser a nivel cromosómico, o del genóforo en el caso de los procariontes, y a nivel únicamente de los genes. En lo sucesivo, al aludir a las mutaciones, nos referiremos a las últimas.

Las mutaciones consisten en inserciones y deleciones de segmentos con pocas bases o de una sola base y sustituciones en un nucleótido. Si bien, todos estos fenómenos se pueden agrupar como mutaciones puntuales,<sup>(26)</sup> en este trabajo hemos preferido restringir el término al caso de mutaciones en las que solamente hay cambio en una base. En este caso, se puede tratar de sustituciones, inserciones o deleciones y estas dos últimas provocan lo que se conoce como cambio de pauta o fase de lectura, especialmente si ocurren en segmentos de DNA que codifican cadenas polipeptídicas. Las sustituciones pueden ser cambios de purina a pirimidina y viceversa, a los que se denominan transversiones, o cambios de purina a purina o pirimidina a pirimidina, a los que se llaman transiciones.

Los mecanismos por los cuales ocurren las mutaciones puntuales, especialmente las sustituciones, se pueden agrupar en dos clases de acuerdo a Drake y Baltz<sup>(30)</sup>: mutaciones por apareamiento incorrecto o por deficiencias en la reparación del DNA. Utilizamos el término de apareamiento incorrecto como traducción al vocablo 'mispairing'. En este caso, se trata de mutaciones en las cuales una base incorrecta se incorpora a la cadena de DNA directamente como resultado de un determinado proceso mutagénico. La incorporación puede ocurrir en la fase de doble hélice o durante la replicación e involucra la formación de un par de bases que no son normales de acuerdo al modelo de Watson y Crick, esto es, la formación de pares que no son G-C o A-T. Mencionamos que Watson y Crick propusieron que los apareamientos incorrectos ocurren por la formación de tautómeros de las bases nitrogenadas. Sin embargo, Topal y Fresco<sup>(96)</sup> observaron que la frecuencia de mutaciones por base ( $10^{-9}$  a  $10^{-12}$ ) no coincide con

la frecuencia de formación de tautómeros ( $10^{-4}$  a  $10^{-5}$ ). A partir de esta observación, estos autores proponen que puede haber apareamientos de bases no complementarias, en el sentido del modelo de Watson y Crick, que son estables y pueden acomodarse en la hélice del DNA. El mecanismo de síntesis, de acuerdo a Topal y Fresco, explica cómo a partir de la tasa de formación de tautómeros se obtiene la tasa de mutación. Además, se proponen dos tipos de apareamientos incorrectos: el apareamiento incorrecto de transición formado por una purina y una pirimidina, y el apareamiento incorrecto de transversión formado por dos purinas. El apareamiento pirimidina-pirimidina, de acuerdo a estos autores, tiene una estructura que no se acomoda en la doble hélice y por ello no lo toman en cuenta.

Drake y Baltz<sup>(30)</sup> consideran que, no necesariamente, los apareamientos incorrectos involucrados en la mutación deben acomodarse bien en la doble hélice. Casi todos los horrores moleculares, como denominan a los apareamientos incorrectos, pueden ocurrir espontáneamente, y eso concuerda con la observación de que todas las sustituciones posibles pueden darse espontáneamente. Los apareamientos incorrectos se pueden inducir por varios procesos como la tautomerización, la ionización y la sustitución de los radicales de las bases nitrogenadas, entre otros. De hecho, el efecto mutagénico de varias sustancias se puede explicar a través de la formación de apareamientos incorrectos. No obstante, no siempre es posible determinar qué apareamiento incorrecto es el involucrado en una mutación específica.

El ácido nitroso, por ejemplo, provoca la desaminación oxidativa de la citosina para formar uracilo, o de la 5-metilcitosina (5-mC) para formar timina. Estas bases pueden aparearse con la adenina. Este mutágeno promueve también la desaminación de la adenina para formar hipoxantina, la cual, puede aparearse con la citosina. El efecto mutagénico de la 2-aminopurina (2-AP), al parecer, también consiste en la desaminación oxidativa de la 5-mC.<sup>(21,30)</sup> aunque también actúa directamente sobre el anillo de la citosina.<sup>(31)</sup> Se ha demostrado que la desaminación espontánea de la adenina para formar

hipoxantina es el mecanismo por el cual ocurre la transición A-T  $\longrightarrow$  G-C que se da en la posición +6 del gene *lacZ* de *E. coli*.<sup>(24,87)</sup>

Los efectos de sustituyentes como el 5-bromouracilo involucran procesos de ionización y alteraciones en el empaquetamiento de pares de bases vecinos.<sup>(30)</sup> Ciertos mutágenos como el diamino, dicloro, platino (DDP) pueden intercalarse en la doble hélice y desestabilizarla.<sup>(1)</sup> Como veremos más adelante, la estabilidad local en pares vecinos de bases es un factor del proceso de mutagénesis que tiene implicaciones importantes a nivel evolutivo.

Otro grupo importante de mutágenos son los reactivos alquilantes entre los cuales podemos citar al metil, metanosulfonato (MMS) y al etil, metanosulfonato (EMS). La mutagenicidad de estas sustancias se basa, por un lado, en apareamientos incorrectos a su vez formados por defectos en la incorporación de nucleótidos. Además, las sustancias alquilantes causan mutaciones por defectos en los mecanismos de reparación.<sup>(30)</sup>

Por último, citaremos el carácter mutagénico del calor. Se estima que en ausencia de mutágenos, el calor, a largo plazo, provoca más mutaciones que los errores de replicación. La mutagenicidad del calor se debe a la desaminación de la citosina o de la 5-mC.<sup>(29,30,85)</sup> Las mutaciones por efecto de la temperatura pueden ocurrir en la etapa de doble hélice o durante la replicación. La tasa de error de copiado en la síntesis del DNA, por lo tanto, depende de la temperatura.<sup>(50,85)</sup> Los efectos del calor en relación a las mutaciones se deben tener en cuenta al tratar lo que, en general, podemos llamar la integridad del genoma. La estabilidad genética ante los efectos de la temperatura se plantea como una restricción natural para las formas de vida que existen a temperaturas relativamente altas como los mamíferos. En la práctica humana, existen problemas como el almacenamiento de genomas y la estabilidad de cepas mutantes de microorganismos importantes en la industria y la investigación.

La formación de apareamientos incorrectos no es el único factor involucrado en la mutagénesis. En las células existen

varios sistemas enzimáticos que reconocen sitios en la molécula de DNA en los cuales no hay un apareamiento o, en general una estructura, normal, y pueden corregirlos. Cuando hay alguna alteración en el funcionamiento de estos sistemas, se incrementa la tasa de mutación. De ello se infiere que los sistemas de reparación en las células evitan la acumulación de las mutaciones. Además, dichos sistemas intervienen en diversos procesos de recombinación genética como la conversión de genes, y la integración de elementos genéticos móviles como las secuencias virales. (30,73,83)

La existencia de sistemas de reparación se dedujo a partir de estudios en mutantes producidos por radiación ultravioleta. (30) Debemos aclarar que los efectos mutagénicos por luz ultravioleta están mediados por alteraciones en los sistemas de reparación. La confirmación plena de los sistemas de reparación, se logró a partir de una serie de estudios de transfección viral los cuales también proporcionaron un modelo de experimentación. (73) Actualmente, se conocen varias formas de reparación del DNA que se clasifican, de acuerdo a Friedberg, (38) como mecanismos de reversión de lesión ('reversal of damage'), excisión de lesión ('excision of damage') y tolerancia al daño en el DNA ('DNA damage tolerance').

En lo que respecta a los apareamientos incorrectos, una cuestión importante acerca del funcionamiento de los sistemas de reparación es plantear de qué manera, al reconocer un apareamiento incorrecto, éstos operan sobre la cadena en la cual se ha generado la base equivocada. Un modelo inicial que describe el funcionamiento del sistema de reparación, fue propuesto por Wagner y Meselson en 1976. (102) Estos autores proponen que la reparación se relaciona con el sistema de replicación. El modelo de Wagner y Meselson postula que la discriminación necesaria para la reparación depende de la identificación de la cadena recién sintetizada. Esta identificación se basa en que la cadena matriz está metilada en la adenina de la secuencia (...GATC...). Cuando el sistema de reparación reconoce un apareamiento incorrecto, a partir de la secuencia (...GATC...), éste corta un segmento de aproximadamente 3,000 bases en la dirección 3' de la cadena no

metilada. La reparación subsecuente del hueco termina el proceso de reparación. La operación de esta forma de reparación se basa en que la cadena recién sintetizada no se metila inmediatamente, sino después de un cierto tiempo. Si se modifica el proceso de metilación, al inhibir o aumentar el grado de metilación en las dos cadenas, se incrementan las mutaciones. Esto demuestra que la metilación es el factor determinante en la discriminación. Este sistema es el que mejor se conoce y se le denomina sistema de reparación dirigido por metilación ('methyl directed repair system').<sup>(73,102)</sup> Este sistema funciona mejor para eliminar transiciones que transversiones y es poco efectivo en la reparación de inserciones o deleciones. La eficiencia de este sistema de reparación no se correlaciona con la estabilidad de los diversos apareamientos incorrectos aislados. Existe, sin embargo, una cierta relación con la estabilidad del empaquetamiento de bases aledañas, esto es, apareamientos incorrectos que originan energías de estabilidad local similares a las de los apareamientos normales, se corrigen mejor que aquellos apareamientos incorrectos que generan estabilidades locales menores o que inducen conformaciones extrahelicoidales, como es el caso del apareamiento pirimidina-pirimidina.<sup>(73,83)</sup>

Este sistema de reparación se ha estudiado en procariontes y existen evidencias de que un sistema similar existe en los eucariontes.<sup>(73)</sup> Además de éste, se ha identificado un mecanismo de reparación de segmentos cortos ('very short patch repair system') que funciona en la reparación de los pares G-T formados por la desaminación espontánea de la 5-metil,citosina. La desaminación espontánea de la citosina forma uracilo. Este apareamiento es reconocido por la enzima uracil-N-glucosidasa que corta el enlace de la base con la desoxirribosa.<sup>(83)</sup> La desaminación de la 5-metil,citosina conduce a la formación de timina la cual no es reconocida por la uracil-N-glucosidasa.<sup>(21,31)</sup> De este modo, los sitios en los que hay 5-metil,citosina constituyen puntos de mutación con incidencia alta.<sup>(21)</sup> El sistema de reparación de segmentos cortos comparte algunos elementos del sistema dirigido por metilación, sin embargo, no es muy eficiente.<sup>(73)</sup>



En lo que respecta a las inserciones y deleciones, se considera que los mecanismos de estas formas de mutación son esencialmente los mismos. Uno de los mecanismos propuestos para explicar este tipo de mutaciones consiste en la formación de discontinuidades en una de las cadenas de la doble hélice. Como consecuencia de ello, los nucleótidos en los extremos de la discontinuidad, pueden tomar configuraciones no helicoidales. La reparación subsecuente del hueco puede llevar a la formación de una cadena con una base más que la cadena complementaria. Otro proceso ulterior de reparación puede retirar la base adicional o complementarla. Algunos mutágenos, como las acridinas, pueden estabilizar la estructura extrahelicoidal de los nucleótidos, lo que favorece su formación.<sup>(30)</sup> Otro modelo que explica la deleción e inserción de bases, sobre todo si se trata de segmentos con pocas de ellas, se basa en el hecho de que alrededor de las inserciones y deleciones, generalmente existen secuencias repetidas en diversas formas. Ya que estas secuencias son complementarias si se doblan, pueden formar fácilmente rizados en la molécula de DNA. Asociadas a la formación de rizados, existen discontinuidades cuya reparación puede originar deleciones o inserciones. Además de las discontinuidades, en los rizados o en la hélice principal, pueden existir nucleótidos que adquieran estructuras no helicoidales como consecuencia de la formación del rizo. La reparación posterior de estos nucleótidos extrahelicoidales provoca inserciones o deleciones.<sup>(39,87)</sup>

La preservación de la información genética se realiza, además de los sistemas de reparación, por la polimerasa del DNA. Este complejo enzimático, aparte de su actividad como polimerasa  $5' \rightarrow 3'$ , tiene actividad de exonucleasa  $3' \rightarrow 5'$  en *E. coli*. La conjunción de estas dos actividades implica una lectura de prueba ('Kinetic proofreading'). Hopfield<sup>(50)</sup> propone, con base en el mecanismo de polimerización, que hay un paso en el que se incrementa la energía de los intermediarios. Cuando hay un intermediario equivocado, esta energía es menor, de modo que se favorece la incorporación del precursor correcto. Con la lectura de prueba, la preferencia por el precursor correcto, y por lo tanto la fidelidad de la síntesis de DNA, se incrementa por una

elevación a la segunda potencia.

Antes del trabajo de Hopfield, se sabía que la tasa de mutación depende de la proporción entre la actividad de exonucleasa y polimerasa. Un incremento en la actividad relativa de exonucleasa incrementa el factor de lectura de prueba y, con ello, se tiene una disminución en la tasa de mutación.<sup>(76)</sup> En el complejo de la polimerasa de DNA en los mamíferos, no existe la actividad de exonucleasa, sin embargo, la fidelidad en la polimerización es alta.<sup>(38)</sup> El que la tasa de mutación pueda modificarse por factores biológicos, como las actividades de corrección y reparación, implica que dicha tasa tiene un valor selectivo. Una tasa de mutación muy alta puede alterar la continuidad genética de una especie mientras que una tasa muy baja impide que la especie pueda adaptarse a un ambiente cambiante.

### § Aspectos evolutivos del proceso de mutación.

El postulado de que las mutaciones son la base del cambio en la evolución, tampoco fue fácil de estudiar en virtud de un conocimiento relativamente escaso en lo concerniente a las relaciones de los seres vivos con el medio ambiente. El mutacionismo, que en tiempos de Morgan defendía la tesis de que las mutaciones constituyen el cambio evolutivo, se orientaba principalmente a la genética. El surgimiento de la teoría sintética que integró el principio de la selección natural con la genética mendeliana, constituyó el avance decisivo para reconocer a las mutaciones como el origen primario de la variación en las poblaciones naturales.

Aunque la teoría sintética es la que más se ha difundido y la que se acepta comúnmente, desde su origen ha sido fuertemente criticada. Primero, no se admite del todo que la acumulación gradual de las mutaciones dé lugar a la formación de categorías taxonómicas por encima del nivel de especie. Goldschmidt declara en 1948,<sup>(40)</sup> que el neodarwinismo es reduccionista al intentar explicar la evolución por encima del nivel de especie con los principios que se encuentran en la evolución de las poblaciones. Para Goldschmidt, esa extrapolación le imprime un carácter

dogmático a la teoría neodarwinista. Este debate en la actualidad continua con el desarrollo de la teoría del equilibrio puntuado, que ha surgido sobre todo a partir del análisis del registro fósil. (33)

Se encuentra que en el mismo estrato de roca, los fósiles de una cierta especie casi no varían. En general, las especies diferentes entre sí y que guardan relaciones de descendencia se encuentran en estratos diferentes. El surgimiento de la mayoría de los grupos taxonómicos ocurre en periodos tan cortos que, éste, difícilmente se puede atribuir al gradualismo. Consecuentemente, se ha propuesto que la microevolución y la macroevolución están desacopladas y que esta última ocurre no por un proceso de selección natural de individuos sino por un proceso análogo de selección a nivel de especies. (33;95) La existencia de cambios evolutivos profundos en tiempos relativamente pequeños puede explicarse por distintos modelos de regulación génica como el de Britten y Davidson (17) que postula la existencia de genes cuyos productos intervienen en la regulación genética. Así, mutaciones en los genes reguladores pueden originar cambios amplios en las poblaciones naturales. No se ha encontrado evidencia experimental en favor de este modelo pero se han desarrollado otras hipótesis que explican cómo ocurren cambios fenotípicos profundos sin alteraciones grandes en el genoma. (108)

Otro de los aspectos de la teoría de la evolución sujeto a fuertes debates es el de cómo es la expansión, en el seno de las poblaciones, de los alelos mutantes. Por una parte, el neodarwinismo clásico postula que, en las poblaciones naturales sujetas a presiones de selección, se fijan las variantes genotípicas que conlleven la mejor adecuación al medio ambiente. Durante la década de los sesentas, se desarrolló la teoría neutralista de la evolución molecular. Kimura, (54) uno de sus exponentes más importantes, considera, a partir del tamaño del genoma y la proporción de guanina y citosina que, por lo menos en los mamíferos, la sustitución de nucleótidos es la forma más común en los eventos de mutación subyacentes al proceso evolutivo. Kimura estima la tasa de evolución en términos de las sustituciones de nucleótidos y obtiene un valor tan alto que

deduce que la mayoría de las mutaciones deben ser neutrales y, por consiguiente, su fijación en la población debe ser aleatoria.

Se desarrolló un esquema estocástico de la dinámica genética en las poblaciones, y se obtuvo teóricamente que la tasa de evolución molecular, esto es, en términos de la sustitución de nucleótidos, es igual a la tasa de mutación y no depende de ningún parámetro poblacional lo que también apoya la idea de la fijación azarosa de los alelos en las poblaciones. La genética de poblaciones neodarwinista considera un factor de adecuación que la teoría neutralista incorpora como una corrección que se refiere a la tasa de aparición de mutantes deletéreos. Esto significa que, de acuerdo a la teoría neutralista, la mayoría de los mutantes son neutrales y su fijación en la población es un proceso aleatorio. Existe una determinada proporción de mutantes deletéreos que son retirados de la población por selección negativa, es decir, por un proceso en el cual no se selecciona el genoma mejor adaptado sino que se elimina aquel con una adecuación menor.

Desde que se conocieron las primeras secuencias de proteínas, se hicieron comparaciones de éstas y se descubrió que la tasa de sustitución de nucleótidos era constante para un tipo determinado de proteína, aunque, había una serie de complicaciones en cuanto a los cálculos del tiempo de divergencia de las especies, las medidas de diferencia de las especies y las correcciones relacionadas con la estimación de la sustitución de nucleótidos inferida a partir de la secuencia de aminoácidos.<sup>(108)</sup> La inexactitud en la medición de los tiempos de divergencia de las especies basada en el registro fósil puede explicar porqué algunos autores han obtenido evidencia en favor de cambios en las tasas de sustitución, argumento que se ha utilizado en contra de la teoría neutralista.<sup>(41)</sup> Recientemente, Britten<sup>(16)</sup> ha reportado que la tasa de mutación si varía entre distintos grupos taxonómicos, y es menor, por ejemplo, en los primates. Britten postula que las diferencias en las tasas de mutación entre los grupos taxonómicos se deben a diferencias en la eficiencia de los sistemas de reparación y que la mayor parte de los cambios en la secuencia de DNA es por la fijación aleatoria de mutantes neutrales.

Otro resultado de las comparaciones en las secuencias de

proteínas, fue que la mayor parte de las sustituciones casi no modifican la estructura del péptido.<sup>(58,108)</sup> La constancia en las tasas de mutación y la localización de diferencias en las secuencias de proteínas, inducen a pensar que la evolución transcurre por eliminación de mutantes deletéreos y fijación de los mutantes neutrales.

La teoría neutralista planteó el análisis de los sitios en que aparecen las mutaciones y de los cambios que implican éstas. Kimura y Ohta en 1974,<sup>(57)</sup> propusieron cinco principios por los cuales se rige la evolución molecular. El primero es que la tasa de sustitución de aminoácidos es constante para un cierto tipo de proteína siempre y cuando no se modifique mucho la estructura terciaria. El segundo principio postula que la tasa de sustitución para diferentes tipos de proteínas depende de las posibilidades de que una proteína mantenga su función cuando hay cambios en la estructura primaria. Para proteínas cuya función puede alterarse mucho ante una mutación, la tasa de sustitución es menor. El tercer principio, consecuencia del anterior, dice que las sustituciones que afectan menos la estructura de una proteína ocurren con más frecuencia. En el cuarto postulado se afirma que la duplicación de un gen antecede a la aparición de un gen con una nueva función. Finalmente, el quinto apartado recapitula los postulados esenciales de la teoría neutralista, esto es, que la fijación por azar o la eliminación de mutantes deletéreos ocurren más frecuentemente que la selección positiva darwiniana.

La teoría neutralista ha sido criticada en el sentido de que la fijación aleatoria de mutantes neutrales no explica o no es relevante en los eventos de adaptación.<sup>(26)</sup> Recientemente, Kimura<sup>(55)</sup> ha propuesto una solución al debate entre el neutralismo y el seleccionismo basado en los datos aportados por las secuencias de DNA acumuladas especialmente a partir de 1979. De acuerdo con Kimura, la evolución a nivel molecular no es un proceso que se rige por los mismos principios que la evolución a nivel fenotípico, lo que conduce a la existencia de un desacoplamiento entre ambas. Una de las cuestiones por las que la discusión entre seleccionistas y neutralistas ha sido difícil es que los neutralistas se refieren a la evolución molecular mientras

que en el seleccionismo, se está en el contexto de la adaptación de las especies. De esta manera, el problema que se plantearía es el del acoplamiento de lo que Kimura distingue como evolución molecular y evolución fenotípica. (53,55)

Desde las primeras comparaciones en la estructura de las proteínas, se había observado que la tasa de evolución a nivel molecular era constante para líneas filogenéticas distintas como los mamíferos y los anfibios. (109) En el caso del orden Anura aunque la tasa de evolución fenotípica evidentemente es pequeña en comparación a los mamíferos, la tasa de sustitución de nucleótidos es aproximadamente la misma para ambos grupos. De ahí entonces que debemos pensar en fenómenos distintos a la sustitución de nucleótidos para explicar la incongruencia aparente, entre la tasa de evolución molecular y la tasa de evolución fenotípica.. Entre las posibles alternativas, se puede aludir a las mutaciones cromosómicas y de hecho se ha encontrado que el cariotipo evoluciona más rápidamente en los mamíferos que en los anuros. Esto se debe a que en la estructura poblacional de los mamíferos se tienen tamaños efectivos de población menores en comparación a los anuros. (106,107)

No solamente los cambios genéticos amplios pueden ocasionar cambios fenotípicos igualmente amplios. Hemos visto que las mutaciones en los genes reguladores pueden tener efectos drásticos en el fenotipo. Incluso, es posible que una mutación provoque sólo un cambio en los niveles celulares de una enzima, lo cual, puede ser mediado por un gen regulador o por una alteración en las regiones reguladoras del gen. Se ha estimado que los cambios en los niveles de proteínas pueden dar lugar a cambios fenotípicos amplios. (108) Independientemente de esto, es claro que en la evolución molecular tratamos con fenómenos distintos a los de la evolución a nivel fenotípico. (108) El planteamiento de este problema, curiosamente reitera la cuestión de las relaciones entre la microevolución y la macroevolución si se piensa en las implicaciones evolutivas de mutaciones puntuales que tienen efectos fenotípicos profundos.

Kimura (55) postula que la evolución de mutantes neutrales ocurre a nivel fenotípico en el fenómeno de la evolución

estabilizadora. Este término se refiere a que las poblaciones tienen una cierta variación fenotípica cuya moda está muy cercana al punto máximo de adecuación. Las variaciones fenotípicas que se alejen mucho de este valor, se eliminarán de la población, mientras que las variaciones alrededor de la moda poblacional se preservarán.

Por lo expuesto, podemos pensar que el proceso de fijación de mutantes en las poblaciones naturales está determinado por la selección natural. Qué factores definen los eventos de mutaciones específicas en el genoma y qué implicaciones evolutivas tienen en la estructura de éste es el problema general que nos atañe. Existen varias ideas al respecto y nuestro propósito, al desarrollar este trabajo es el de aportar otro elemento más de consideración, llámese el efecto de algunos factores físicos.

## ANTECEDENTES Y PLANTEAMIENTO DEL PROBLEMA DE ESTUDIO

La teoría de la evolución se abstrae de los mecanismos de las mutaciones y parte de que estas aparecen en un lugar definido del DNA aleatoriamente con respecto a la adecuación del organismo. El evento de mutación aunado a la expansión de una variante en una población natural, ya sea en el marco del seleccionismo clásico o de la teoría neutralista, se concibe como el proceso que constituye la evolución a nivel molecular. Se reconoce que la selección natural, a través de una serie de restricciones específicas, determina los cambios en el genoma. Consideramos necesario estudiar algunos aspectos de la mutagénesis que, en nuestro concepto, establecen restricciones al proceso de la evolución molecular adicionales a la selección natural concebida en su forma más simple.

El que las mutaciones ocurran al azar con respecto a la adecuación no significa que éstas aparezcan con la misma probabilidad en todo el genoma. En 1961, Benzer<sup>(4)</sup> estudió el patrón de mutaciones en la región rII del fago T4 y encontró que no todos los sitios en el DNA son igualmente mutables, y que estas diferencias no se pueden atribuir a variaciones estadísticas. Sin haber podido establecer la índole de estas mutaciones, Benzer concluyó que la mutabilidad de un sitio definido en el DNA quizá depende de ciertas características locales en éste. Como hemos visto, varios estudios posteriores, demostraron que la mutabilidad en un nucleótido específico depende de varios factores como la metilación y el funcionamiento de los sistemas de reparación.

§ Caracter no aleatorio de la secuencia de bases en el DNA.

Con el desarrollo de las técnicas de secuenciación y clonación fue posible acumular una gran cantidad de datos acerca de la secuencia de nucleótidos. Los primeros análisis mostraron que la composición de bases no es aleatoria y se definieron algunas reglas referentes a la composición de nucleótidos. Estas reglas, variaban entre eucariontes y procariontes.<sup>(72)</sup>

También se observó que algunos digramas, esto es,



nucleótidos contiguos sobre la misma cadena, se distribuyen con una cierta periodicidad. Algunos autores, interpretaron sus resultados en el sentido de que ciertos pares distribuidos regularmente contribuyen al enrollamiento ordenado de la doble cadena en la formación del genóforo de los procariontes y, extensivamente, en lo que se refiere al enrollamiento de la cromatina en los eucariontes.<sup>(90)</sup> En el caso del virus SV40, la periodicidad de la secuencia de bases en la región del promotor se ha asociado a la formación de una estructura que puede servir como punto de reconocimiento a proteínas involucradas en el control de la transcripción.<sup>(31)</sup> Otros, consideraron que la periodicidad de los digramas pirimidina-purina era remanente de la síntesis prebiótica de los ácidos nucleicos.<sup>(83,84)</sup> Algunos piensan que en la tierra primitiva, los ácidos nucleicos se sintetizaron por la acción catalítica de las superficies de arcillas.<sup>(1)</sup> La periodicidad de los pares pirimidina-purina quizá podría reflejar la periodicidad cristalina de dichas arcillas.

La comparación de secuencias alineadas confirmó la existencia de restricciones por selección natural de las secuencias de aminoácidos en las diferentes posiciones de los nucleótidos en los codones. La segunda base del codón es la menos variable, esto es, la más restringida en virtud de que este sitio determina, en general, la naturaleza polar o no polar del aminoácido. Usualmente, si la segunda base del codón es una purina, se codifica un aminoácido polar. Si por el contrario, se trata de una pirimidina, se codifica un aminoácido no polar.<sup>(1)</sup> De la polaridad del aminoácido, como sabemos, depende en gran medida la estructura terciaria de las proteínas. El código genético es redundante por lo que la tercera base del codón es la menos sujeta a restricciones y, por ende, puede ser más variable. Se esperaría, que en una secuencia específica, se encontrase una distribución uniforme de los diferentes codones que codifican para el mismo aminoácido. Sin embargo, esto no es así, los codones aparecen en diferentes regiones del genoma con una frecuencia fija distinta a la que se espera por azar. Para un genoma específico, la distribución de frecuencias de los codones es característica. Este fenómeno es al que se ha denominado como uso no aleatorio de los

codones de acuerdo a una traducción literal de la terminología en la lengua inglesa. Preferimos nombrar el concepto anterior como distribución no aleatoria de frecuencias de los codones.<sup>(32,37)</sup> Este fenómeno muestra que existen restricciones, diferentes a la secuencia de aminoácidos, que determinan la secuencia de nucleótidos en el DNA.

Se han desarrollado hipótesis diversas para explicar de modo general o en casos particulares la distribución no aleatoria de los codones. En primer lugar Grantham et al.,<sup>(37)</sup> postulan que cada gen tiende a conformarse a la distribución de codones propia de la especie, de lo cual se infiere que es el genoma y no el gen individual, la unidad de selección natural. Estos autores llaman a esto la hipótesis del genoma. Este grupo de trabajo descubrió que la distribución de frecuencias de los codones depende de la magnitud con la que un gen definido se expresa. En los genes que se expresan mucho, los codones cuya tercera base implique una energía de unión débil con el anticodón aparecen con mayor frecuencia que los codones en los que la tercera base conlleve a una unión fuerte con el anticodón. En este caso, en los genes más activos en términos de transcripción, en la tercera posición aparece con más frecuencia citosina que timina.<sup>(38)</sup>

Grosjean y Fiers<sup>(32,39)</sup> proporcionan una hipótesis de alcance más general. Estos autores consideran que en el fago MS2 y en *E. coli*, el factor que determina la distribución no aleatoria de las frecuencias de los codones es la optimización en la energía de unión entre el RNA mensajero (mRNA) y el de transferencia (tRNA). Las energías de unión muy fuertes provocan que las proteínas se sintetizen con lentitud, pero, las energías de interacción muy débiles inducen errores en la traducción. Se encuentra, así, que en la tercera posición del codón en los genes que se expresan mucho, el factor de rapidez en la traducción es el más importante y, por ello, aparecen con más frecuencia bases que provocan una energía de unión menor con el anticodón, mientras que en los genes que se expresan poco, las bases de la tercera posición llevan a una interacción más fuerte entre el mRNA y el tRNA.

Se ha encontrado en *E. coli*, que la distribución no

aleatoria de frecuencias de los codones se correlaciona bien con los niveles intracelulares de los tRNA correspondientes. Esto se interpreta en el sentido de que la disponibilidad del tRNA en la célula determina la frecuencia de los codones especialmente si la actividad de síntesis de proteínas es intensa.<sup>(46)</sup> Otro ejemplo particular, es el del fago MS2 cuyo genoma es de RNA. Hasegawa et al.,<sup>(41)</sup> proponen que la restricción selectiva que define la distribución no aleatoria de las bases en la tercera posición es la estabilización de la estructura secundaria de la molécula de RNA, para la cual se ha propuesto un modelo en forma de flor similar al modelo del tRNA. Otra idea al respecto de la distribución de frecuencias no aleatoria de los codones, consiste en que, en las globinas, los codones se distribuyen de modo tal que se minimizan las posibilidades de mutaciones con efectos drásticos. En las globinas  $\alpha$  y  $\beta$ , por ejemplo, casi no existen codones que den lugar a un codón de terminación por una sola sustitución.<sup>(65)</sup>

A partir de todas las hipótesis que hemos expuesto, se puede generalizar que la explicación de la distribución no aleatoria de las frecuencias de los codones se ha orientado a la demostración de la existencia de restricciones selectivas específicas.

En 1983 Lipman y Wilbur<sup>(57)</sup> publicaron un trabajo en el que se aplica la teoría de la información al análisis de diversas secuencias de DNA agrupadas como secuencias de mitocondrias, procariontes y eucariontes. En su investigación estos autores encontraron que la hipótesis original del genoma de Grantham et al.,<sup>(37)</sup> en las mitocondrias, puede explicar la secuencia de bases en el DNA. En el caso de los procariontes, aún al considerar la magnitud en la expresión de los genes, la hipótesis del genoma no puede explicar las características de la secuencia. En el caso de los eucariontes, se deben considerar una gran cantidad de restricciones para explicar la secuencia de las bases. Lipman y Wilbur concluyen que en los procariontes, la identidad de la base de la tercera posición en el codón se relaciona con la segunda posición mientras que en los eucariontes, la identidad en la tercera posición depende de las dos bases vecinas.

### § La evolución molecular en los pseudogenes.

Se han descubierto secuencias cuya estructura es muy parecida a la de genes conocidos pero que no se expresan porque, generalmente en la estructura primaria hay una alteración que las diferencia de un gen ortodoxo y les impide funcionar como tales. Estas secuencias se han denominado pseudogenes y han abierto perspectivas inesperadas en el estudio de la evolución molecular. Mencionamos que este proceso consiste en el evento de mutación seguido del proceso de selección natural. Si nos abstraemos del proceso de mutagénesis, como generalmente ocurre, el estudio de la evolución molecular se encamina a explicar los diversos aspectos de este fenómeno en términos de la selección natural. Los pseudogenes, al no expresarse, no están sujetos a presiones selectivas y, como una primera aproximación, ofrecen una oportunidad para estudiar la evolución molecular exenta de la selección natural en los productos peptídicos. Miyata y Hayashida, (64) encontraron que los pseudogenes evolucionan a una tasa mayor a la de los genes normales y esta, es mayor en las regiones homólogas a los exones. Incluso, esta es la tasa de sustitución más alta registrada. La primera conclusión que se obtiene a partir de este resultado es la confirmación de la teoría neutralista. Si la evolución se diera por selección de los mutantes más ventajosos, se hubiera esperado que la tasa de evolución en los pseudogenes fuera menor a la de los genes normales. El que en realidad la tasa de sustitución en los pseudogenes sea mayor, como interpreta Kimura, (47,50) refleja las restricciones impuestas por la selección negativa en virtud de que, de acuerdo al neutralismo, en ausencia de restricciones, la tasa de sustitución se iguala a su valor máximo posible dado por la tasa de mutación. Un resultado adicional del trabajo de Miyata y Hayashida es que en los pseudogenes la distribución de la frecuencia de los codones alternativos es uniforme, lo que confirma que las presiones selectivas determinan esta distribución no aleatoria.

Asimismo, el estudio de los pseudogenes reveló que los diversos tipos de sustitución de nucleótidos no ocurren todos por igual. Como se ha mencionado, las transiciones son más abundantes

de lo que se espera por azar. Igualmente, los pares con guanina y citosina son más remplazados que los pares con adenina y timina. Esto es resultado de la transición 5-mC  $\longrightarrow$  T que ya hemos estudiado. Esta transición ocurre si la cantidad de guanina y citosina es relativamente alta. (55)

§ La regionalización en la composición de bases en el genoma de los vertebrados.

Las transiciones entre los pares complementarios A-T y G-C, implican cambios importantes en el DNA ya que de la cantidad de guanina y citosina depende directamente la estabilidad de esta molécula. Merced de que el par G-C se mantiene unido por tres puentes de hidrógeno mientras que el par A-T lo hace por dos de estos enlaces, es que la estabilidad térmica de la doble hélice se incrementa proporcionalmente al aumento en la fracción del par G-C. (11)

En el DNA de los vertebrados se han descubierto, por centrifugación en gradientes de CsCl, varios fragmentos con una composición de guanina y citosina característica de cada uno de ellos. A estas fracciones se les ha denominado isocoras, y se trata de segmentos de aproximadamente 200 a 300 kilobases (Kb) de largo con una cantidad de pares G-C homogénea. Quizá las isocoras sean los fragmentos que definen las bandas en los cromosomas a alta resolución, cuyo tamaño se estima en 1,250 Kb. En ese sentido, se dice que el genoma está compartimentalizado en cuanto a la composición de bases. Por su densidad, las isocoras se han clasificado en ligeras (L) y pesadas (H) las últimas de las cuales, poseen una proporción mayor de guanina y citosina. En las diferentes especies de vertebrados, el perfil de composición de los compartimientos es distinto. En los vertebrados de sangre caliente, es decir, las aves y los mamíferos, la proporción de isocoras pesadas es mayor que en los vertebrados de sangre fría. (7)

De acuerdo a Bernardi et al., (6) el aumento en la proporción de guanina y citosina en los vertebrados de sangre caliente es una adaptación que evita los efectos de desestabilización en el DNA provocados por el aumento en la temperatura corporal. En apoyo a

esto, se ha encontrado que algunas especies de peces del género *Tilapia* que viven en aguas termales, muestran una proporción mayor en la fracción de isocoras pesadas con respecto a las especies del mismo género que viven en aguas templadas.

Este mismo grupo de trabajo descubrió que la distribución de guanina y citosina no es la misma en las distintas posiciones de los codones. El aumento en la proporción de guanina y citosina, con respecto al incremento global, es mayor en la tercera posición del codón, y menor en la segunda posición.<sup>(5,6)</sup> Así, estos autores proponen la existencia de restricciones de composición en el DNA que tienen influencia en regiones que codifican proteínas y en las que no lo hacen, lo cual indica que estas regiones también tienen una cierta función. Según estos autores, las restricciones de composición están dadas por factores externos como la temperatura. En el caso de la proporción del par G-C, por un lado, el incremento en la fracción de G-C en relación a la temperatura, conlleva un aumento en la proporción de aminoácidos como la alanina y la arginina, que estabilizan la estructura de la proteínas. Por otra parte, el incremento de G-C, también estabiliza la molécula de DNA. Por ello, Bernardi et al. postulan que el fenotipo tiene un componente dado por los resultados de la producción de proteínas además del fenotipo del propio genoma.<sup>(6)</sup> Esta diferenciación es relevante porque distingue la existencia de restricciones selectivas no relacionadas con la codificación de secuencias peptídicas.

Ya que existen varios compartimientos con proporciones de GC características, las restricciones de composición son diferentes para cada uno de ellos. Esto plantea, por ende, que la distribución de los genes en los distintos compartimientos depende de las restricciones de composición. Se ha demostrado que, en los vertebrados de sangre caliente, los genes relacionados con las funciones generales del organismo ('housekeeping genes') se distribuyen en los componentes pesados del genoma, mientras que otras partes como las secuencias repetidas y algunos pocos genes, se encuentran en los componentes ligeros.<sup>(6,62)</sup> A partir de las restricciones de composición, por consiguiente, Bernardi et al.<sup>(5,6)</sup> afirman que la distribución no aleatoria de los codones

resulta de las restricciones de composición.

### 5 Factores físicos en la evolución molecular.

Nuestro proyecto de trabajo surge de un intento de definir la existencia de restricciones físicas en la evolución de la secuencia de nucleótidos en el DNA. No obstante que la aplicación de algunas formas de análisis indicaba que las secuencias tenían una estructura subyacente, no era posible discernir los eventos involucrados en la determinación de las secuencias, especialmente factores físicos. Como una primera aproximación, se puso atención a la distribución de los pares G-C y A-T, que pueden relacionarse a un parámetro físico de estabilidad, y a la distribución en la secuencia de las purinas y las pirimidinas, que puede relacionarse con la estructura.

La síntesis de oligonucleótidos ha permitido la obtención de información en cuanto a la estabilidad y la estructura local del DNA y, con ello, la posibilidad de estudiar la secuencia de nucleótidos en relación con algunos factores físicos de manera concreta.

Dickerson,<sup>(20)</sup> ha estudiado las desviaciones locales en los ángulos estructurales de la molécula de DNA de doble hélice con respecto al valor promedio y ha encontrado que estas desviaciones son considerables. Si se establece que el factor estructural más importante es el impedimento al acomodo de purinas vecinas situadas en cadenas opuestas, se obtiene una buena aproximación a los valores reales en los ángulos de la estructura.

Recientemente, Breslauer et al.<sup>(12)</sup> han hecho estimaciones bastante precisas de la estabilidad del DNA en términos de la entalpia, entropía y energía libre de distintos digramas. De acuerdo a estos resultados, la estabilidad del DNA, determinada por estas secuencias de amarre, depende de la secuencia más que de la proporción de guanina y citosina. El cuadro II nos muestra que los digramas en los que hay pares G-C contiguos son más fuertes, lo que explica el efecto de estabilización al aumentar la proporción de estas bases. Sin embargo, los digramas que están formados por un par G-C junto con uno de A-T son débiles. Un

resultado importante es que estos valores de energía no se modifican por efecto de distintas secuencias aledañas, es decir, que estos valores se pueden usar en el análisis de diferentes secuencias de DNA para calcular parámetros globales sin necesidad de hacer correcciones que modifiquen sustancialmente el resultado de un cálculo dado.

Así, se pueden representar las secuencias de DNA en dos formas binarias. Primero, asignar el valor de cero para adenina y timina, y el valor de uno para guanina y citosina. Esto refleja la estabilidad en el DNA definida por los valores de energía en los digramas.<sup>(12)</sup> La segunda representación consiste en asignar el valor de cero para las pirimidinas y el valor de uno para las purinas, y esto refleja valores en los ángulos estructurales determinados por Dickerson.<sup>(20)</sup>

Un análisis estadístico de estas representaciones, comparadas a lo que se esperaría por azar,<sup>(20)</sup> mostró que el DNA de los eucariontes es más flexible y con menos desviaciones, en sus ángulos estructurales con respecto al promedio, que el DNA de los procariontes. Este resultado es consistente con la presencia de restricciones funcionales asociadas a la transcripción y la replicación. El DNA de los virus, por su parte, es similar al DNA de las células en que se replican, esto es, que los virus de los eucariontes son más flexibles y con menos variaciones de los ángulos estructurales, que los virus de los procariontes. Esto posiblemente se relacione con el proceso de integración del genoma viral en el genoma huésped.

Los distintos valores en las energías de amarre implican diferencias en la flexibilidad local del DNA, lo cual a su vez, puede dar lugar a diferentes posibilidades de eventos de mutación. ¿Existe una relación de la flexibilidad local con la frecuencia de mutación en un sitio determinado del DNA? Esta pregunta no es fácil de responder en virtud de que existen varios factores involucrados en los eventos de mutación. Debemos, consiguientemente, plantearnos si existe alguna forma de estudiar el efecto de la flexibilidad local en la molécula de DNA en la mutagénesis.

Hemos visto que las características físicas de la doble



hélice como la flexibilidad local y los ángulos estructurales, posiblemente están relacionadas con restricciones funcionales, pero ¿es posible que, en ausencia de factores selectivos intensos, la secuencia de nucleótidos en el DNA refleje restricciones físicas en el proceso de mutación?, es decir, ¿es la selección natural el fenómeno que por sí solo, determina la evolución molecular?

## PROCEDIMIENTOS DE ANALISIS

### § Generalidades.

En esta investigación, se desarrollaron dos líneas de estudio. Primero, la mayor parte del trabajo se centra en el análisis de la distribución de la estabilidad local en secuencias homólogas. La intención de este enfoque es la de poder hacer inferencias sobre la evolución de genes cuya diferencia se debe esencialmente a sustituciones nucleotídicas. Se eligieron los genes de las globinas cuya filogenia y arreglo en los cromosomas se conocen bien.<sup>(41,94)</sup> También se hizo este análisis en algunos genes de inmunoglobulinas. Las secuencias se obtuvieron, en su mayoría, del banco de genes<sup>(10)</sup> y algunas directamente de la literatura. El cuadro I contiene la lista de estas secuencias. Al comparar las secuencias del banco de genes y de algunos reportes originales, se encontraron en algunos casos, diferencias que no afectan nuestros resultados. Los valores termodinámicos que se usaron son los que reportan Breslauer et al.<sup>(14)</sup> y se muestran en el cuadro II.

La representación energética de las secuencias de DNA estudiadas se hizo en términos de la entalpía ya que nuestro interés es la estabilidad local del DNA como doble hélice y no la termodinámica de las transformaciones conformacionales de la molécula. Los cálculos se realizaron en computadoras IBM-PC con programas en lenguaje Pascal desarrollados por el grupo que dirige el Dr. Germinal Cocho en el Instituto de Física de la UNAM.

La segunda línea de estudio consistió en un análisis inicial del espectro de mutaciones espontáneas en el gene *lacI* de las cepas *MutH*, *MutS*, y *MutL* de *E. coli* que carecen de sistemas de reparación reportado por Schaaper y Dunn,<sup>(88)</sup> y en un estudio de los sitios mutables que se detectan al comparar las secuencias de las globinas y en el hombre y el gorila.<sup>(86)</sup>

CUADRO I. SECUENCIAS DE DNA ANALIZADAS.

	Gen	Región	Longitud (pb)	Referencia
<b>Globinas</b>				
<i>Glycine max</i> (soya)				
1	Leghemoglobina I	Exón 1	99	15
2		Exón 2	108	
3		Exón 3	105	
4		Exón 4	126	
5		Intrón 1	119	
6		Intrón 2	99	
7	Leghemoglobina II	Intrón 3	229	15
8		Exón 3	107	
9		Exón 4	144	
10	Leghemoglobina LBA	Intrón 3	778	51
11		Exón 1	95	
12		Exón 2	109	
13		Exón 3	105	
14		Exón 4	120	
15		Intrón 1	119	
16		Intrón 2	233	
17	Intrón 3	680		
<i>Xenopus laevis</i> (sapo africano)				
18	Globina $\beta$	mRNA maduro	441	105
<i>Gallus gallus</i> (pollo común)				
19	Globina $\alpha^D$	Exón 1	93	27
20		Exón 2	204	
21		Exón 3	129	
22		Intrón 1	148	
23	Globina $\epsilon$	Intrón 2	260	28
24		Exón 1	92	
25		Exón 2	223	
26		Exón 3	129	
27		Intrón 1	108	
28		Intrón 2	970	
<i>Oryctolagus cuniculus</i> (conejo común)				
29	Globina $\beta_1$	Exón 1	91	98
30		Exón 2	222	
31		Exón 3	129	
32		Intrón 1	127	
33		Intrón 2	573	

34	Globina $\psi\beta 2$ *	Exón 1	92	45
35		Exón 2	222	
36		Exón 3	128	
37		Intrón 1	100	
38		Intrón 2	753	
<i>Mus musculus</i> (ratón común)				
39	Globina $\alpha$	Exón 1	95	78
40		Exón 2	203	
41		Exón 3	128	
42		Intrón 1	122	
43		Intrón 2	135	
<i>Capra hircus</i> (cabra)				
44	Globina $\psi\beta^*$	Exón 1	92	19
45		Exón 2	223	
46		Exón 3	129	
47		Intrón 1	125	
48		Intrón 2	624	
<i>Halichoerus grypus</i> (foca gris)				
49	Mioglobina	Exón 1	94	12
50		Exón 2	223	
51		Exón 3	216	
<i>Gorilla gorilla</i> (gorila)				
52	Globina $\gamma$ *	Exón 1	92	86
53		Exón 2	223	
54		Exón 3	129	
55		Intrón 1	122	
56		Intrón 2	907	
<i>Pongo pygmaeus</i> (orangután)				
57	Globina $\theta 1$ *	Exón 1	96	66
58		Exón 2	205	
59		Exón 3	129	
60		Intrón 1	84	
61		Intrón 2	109	
<i>Homo sapiens</i>				
62	Globina $\alpha 1$	Exón 1	95	70
63		Exón 2	204	
64		Exón 3	128	
65		Intrón 1	118	
66		Intrón 2	150	
67	Globina $\zeta$	Exón 1	95	82
68		Exón 2	205	
69		Exón 3	128	
70		Intrón 1	842	
71		Intrón 2	239	

72	Globina $\beta$	Exón 1	93	60		
73		Exón 2	221			
74		Exón 3	128			
75	Globina $\epsilon$	Intrón 1	131	3		
76		Intrón 2	851			
77		Exón 1	92			
78		Exón 2	222			
79		Exón 3	129			
80		Intrón 1	123			
81		Intrón 2	855			
Inmunoglobulinas						
<i>Caiman crocodylus</i> (caimán)						
82	VIII	Péptido maduro	294	65		
<i>H. sapiens</i>						
83	Ig $\gamma_2$	Bisagra	36	34		
84		CH1	294			
85		CH2	327			
86		CH3	321			
87		Intrón 1	392			
88		Intrón 2	118			
89		Intrón 3	97			
90	Ig $\epsilon$	CH1	309	10		
91		CH2	321			
92		CH3	324			
93		CH4	443			
94		Intrón 2	207			
95		Intrón 3	88			
96		Intrón 4	83			
97	Ig $\kappa$ (alelo inv 3)	Cx	321	49		
98		Jx	99			
99	Ig $\lambda$ (isotipo Mcg) (isotipo Ke-OZ-)	CA	315	48		
100		CA	318			
101	VIII	Péptido maduro	291	68		
102	Vx	Exón 1	54	5		
103		Exón 2	297			
104		Intrón	125			
105	D	D1	31	92		
106		D2	31			
107		D3	28			
108		D4	31			
109		Ig $\mu$	DHQ52		11	84
110			J1		52	
111			J2		52	
112			J3		46	
113			J4		46	
114			J5		49	
115			J6		61	

116		$\psi J_1$	54	
117		$\psi J_2$	61	
118		$\psi J_3$	52	
119		Intrón i	503	
120		Intrón 1	63	
121		Intrón 2	91	
122		Intrón 3	158	
123		Intrón 4	335	
124		Intrón 5	160	
125		Intrón 6	327	
126		Intrón 7	393	
127		Intrón 8	341	
128		Intrón 9	150	
129		Intrón f	173	

Notas.

- \*: Secuencias copiadas directamente de la literatura.
- Intrón i: intrón inicial.
- Intrón f: intrón final.
- Los pseudogenes llevan antepuesta la letra  $\psi$ .

CUADRO II. VALORES TERMODINAMICOS ENTRE PARES VECINOS DE NUCLEOTIDOS EN EL DNA.

Interacción	$\Delta H^\circ$	$\Delta G^\circ$	$\Delta S^\circ$
A-A / T-T	9.1	1.9	24.0
A-T / T-A	8.6	1.5	23.9
T-A / A-T	6.0	0.9	16.9
C-A / G-T	5.8	1.9	12.9
G-T / C-A	6.5	1.3	17.3
C-T / G-A	7.8	1.6	20.8
G-A / C-T	5.6	1.6	13.5
C-G / G-C	11.9	3.6	27.8
G-C / C-G	11.1	3.1	26.7
G-G / C-C	11.0	3.1	26.6

Los valores están expresados en Kcal/mol para  $\Delta H^\circ$  y  $\Delta H^\circ$ , y en cal/°K mol para  $\Delta S^\circ$ . Estos se refieren a la desaparición de la interacción en una doble hélice en 1 mol/L de NaCl a 25 °C y pH = 7. (14)

§ Análisis de las globinas y del hombre y el gorila y del gene *lacI* de *Escherichia coli*.

Una forma de relacionar los eventos mutacionales en una determinada secuencia, con la energía local, se podría realizar a partir del alineamiento de secuencias homólogas y, subsecuentemente, relacionar los sitios de mutación detectados con la energía local en tales sitios. Sin embargo, en este tipo de aproximación, no es posible conocer el evento mutacional que ha ocurrido en un sitio en el que las secuencias homólogas son diferentes. Desarrollamos este enfoque con el estudio de las globinas y del hombre y del gorila. La mayoría de las diferencias en las cinco secuencias que se estudiaron se deben a transiciones. Si suponemos, que estas diferencias se deben a la transición G-C  $\longrightarrow$  A-T por desaminación de la 5-metil, citosina, encontramos que las mutaciones se asocian a la secuencia ..NCG.. donde N es cualquier base (Cuadro III). Este resultado es consistente con la idea de que en la secuencia ..NCG.., la citosina se metila y forma un sitio con alta probabilidad de mutación. De acuerdo a este primer resultado, entonces, no hay en principio relación de la mutabilidad con la entalpía local en la secuencia.

Otra forma de estudiar la relación de la estabilidad local en la doble hélice con los eventos de mutación, consiste en observar los eventos espontáneos de mutación en una secuencia determinada. Esto se ha logrado en las cepas de *E. coli* *MutH*, *MutS* y *MutL* en las que los sistemas de reparación son deficientes. Si se hace una representación de la secuencia en términos de las entalpías locales, se observa que los puntos con alta incidencia de mutaciones se encuentran en regiones donde la estabilidad local es menor. Esto se muestra en la Figura 1.

Ya que el número de transversiones reportado en el trabajo de Schaaper y Dunn<sup>(88)</sup> es muy pequeño, solamente nos dedicamos a estudiar las transiciones. Como medida de la estabilidad local se usó el promedio de las energías de amarre con las dos bases vecinas a la base mutante. En la Figura 1 se observa que la incidencia de mutaciones con respecto a la energía local no guarda la misma relación para transiciones distintas. En el caso de las transiciones A-T  $\longrightarrow$  G-C, es claro que la mutabilidad es mayor en



CUADRO III. FRECUENCIA DE TRANSICIONES EN LAS GLOBINAS Y DEL HOMBRE Y EL GORILA EN DISTINTAS SECUENCIAS LOCALES.

SECUENCIA	FSM/FST
A C A	0.055
A C T	0.077
T C A	0.024
T C T	0.030
A C G	0.472
A C C	0.126
T C G	0.083
T C C	0.068
G C A	0.040
G C T	0.065
C C A	0.072
C C T	0.064
G C G	0.540
G C C	0.000
C C G	0.300
C C C	0.089

La secuencia se define al suponer que al observar una transición en las secuencias alineadas, el par original es G-C. Las 16 secuencias locales posibles se agrupan de acuerdo al patrón de energía local como: débil-débil, débil-fuerte, fuerte-débil y fuerte-fuerte.

FSM: Frecuencia de sitios mutables con una secuencia local determinada.

FST: Frecuencia total de determinadas secuencias locales en la secuencia del gen.

Las secuencias y el alineamiento son datos de Scott et al. (86). Las secuencias son: Globina  $\gamma$  y  $\gamma^A$  del gorila,  $\gamma$  de los cromosomas A y B del hombre y la globina  $\gamma$  del hombre.

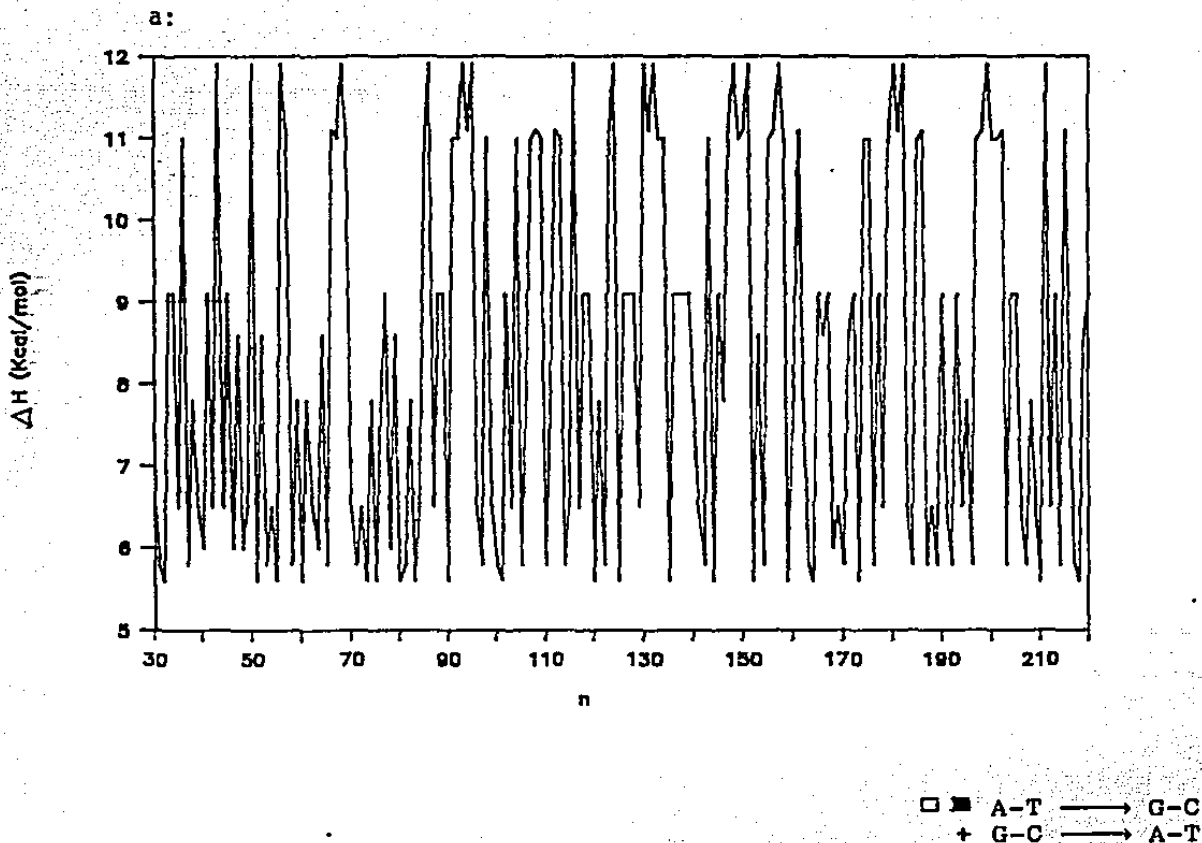


FIGURA 1.

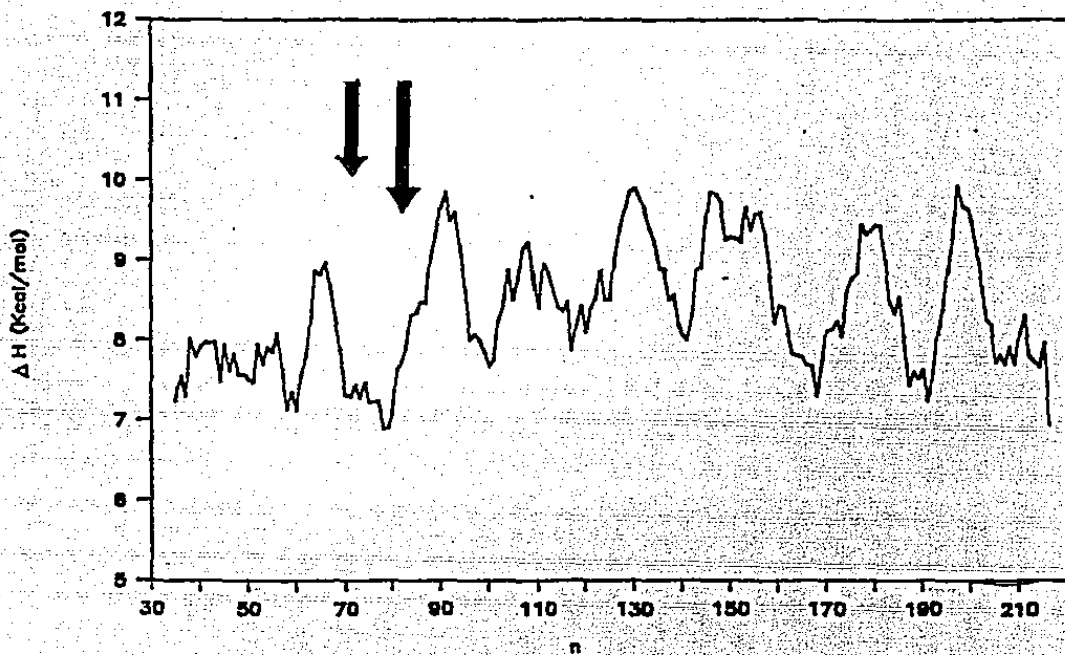
Análisis de la primera porción del gene *lacI* en cepas de *E. coli* carentes de sistemas de reparación. Los datos son de Schaaper y Dunn. (88)

- a: Representación en términos de la entalpia entre cada dos nucleótidos consecutivos.

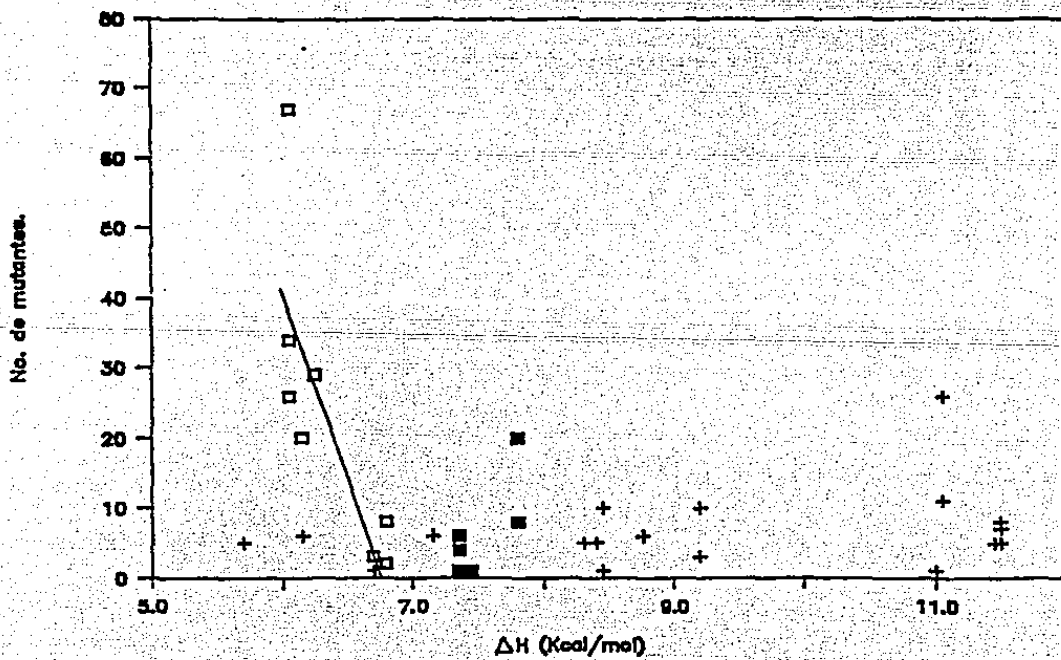
- b: Representación suavizada de la gráfica 1a. La entalpia en cada punto es el promedio de los 5 valores vecinos a cada lado de una posición determinada en la secuencia de energías. Las flechas indican los puntos con incidencia alta de transiciones A-T → G-C.

- c: Frecuencia de mutaciones en función de la entalpia promedio de los dos digramas vecinos a las bases mutables. Se muestra la línea de regresión para las sustituciones A-T → G-C cuando los enlaces vecinos son de la forma débil-débil ( $r = -0.81$ ). Los cuadros negros corresponden a la misma sustitución cuando los enlaces tienen la forma débil-fuerte o fuerte-débil.

b:



c:



sitios donde la estabilidad local es menor y existe una relación lineal cuando los enlaces alrededor de la base mutable son ambos enlaces débiles. En el caso de las transiciones G-C  $\longrightarrow$  A-T, esta relación parece inversa, sin embargo, la frecuencia de mutaciones se asocia en gran medida a la secuencia ..GCN.. similar al caso de las globinas y del hombre y el gorila. Otro resultado interesante de este estudio es que todas las transiciones ocurren en la primera y segunda posición del codón. Posiblemente, en los experimentos de Schaaper y Dunn,<sup>(88)</sup> ocurran mutaciones en la tercera posición del codón pero, como estas mutaciones no tendrían efectos fenotípicos evidentes, quizá estos mutantes no han podido detectarse. De todos modos, aunque estos mutantes en la tercera posición del codón puedan existir, al parecer no abundan demasiado, pues de lo contrario, se hubiera esperado que algunos de ellos si pudieran haber sido identificados.

Como una primera aproximación, lo anterior muestra que desde el punto de vista del cambio de entalpía, es posible considerar dos formas de mutación. El primer tipo abarca las mutaciones como la transición A-T  $\longrightarrow$  G-C, que ocurren en los sitios más débiles de la doble hélice y llevan a un aumento en la rigidez del DNA. La segunda clase de mutaciones son del tipo de la transición G-C  $\longrightarrow$  A-T, las cuales, al parecer ocurren por desaminación de la 5-metil,citosina e implican un aumento en la flexibilidad de la molécula de DNA.

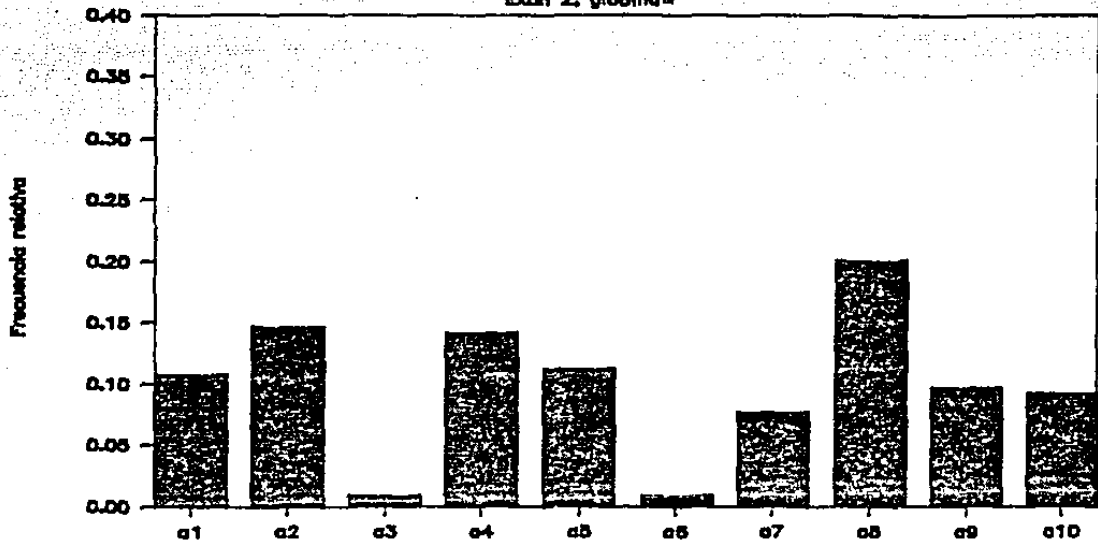
#### § Distribución estadística de las energías de amarre en distintas secuencias de DNA.

Ya que no es posible determinar por el método de alineamiento los eventos mutacionales subyacentes a las diferencias que puedan encontrarse en la comparación de dos secuencias, se planteó como alternativa el estudio de la distribución de la energía en la molécula de DNA. La intención de este enfoque es estudiar las características energéticas del DNA en sitios sujetos a diferentes restricciones selectivas.

El primer paso es la obtención de la distribución estadística de los distintos digramas en las secuencias del Cuadro I. En la Figura 2 se ilustran algunos ejemplos de estas

## Homo sapiens

Exón 2, globina $\alpha$



## Homo sapiens

Intrón 2, globina $\alpha$

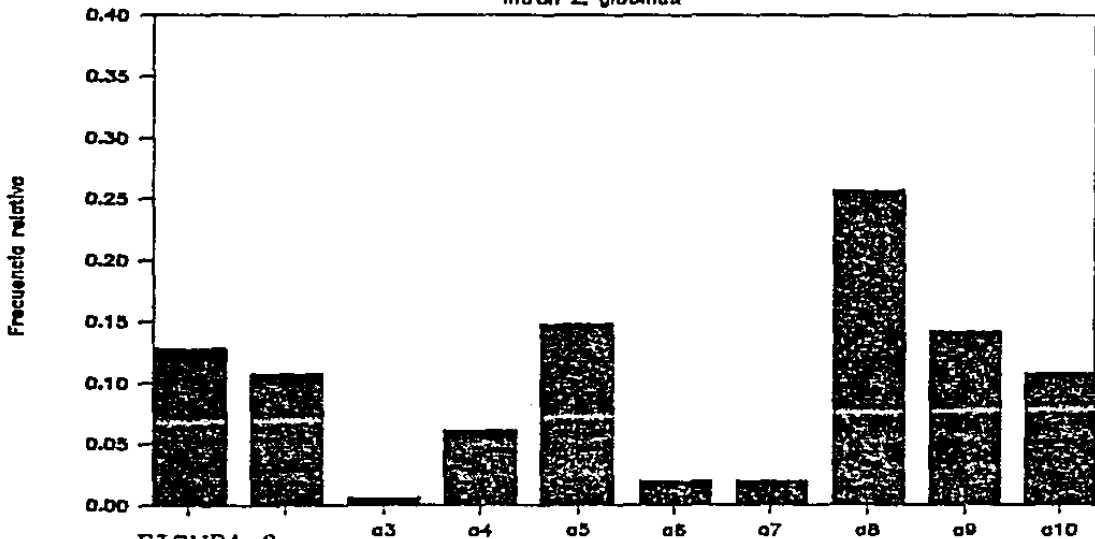
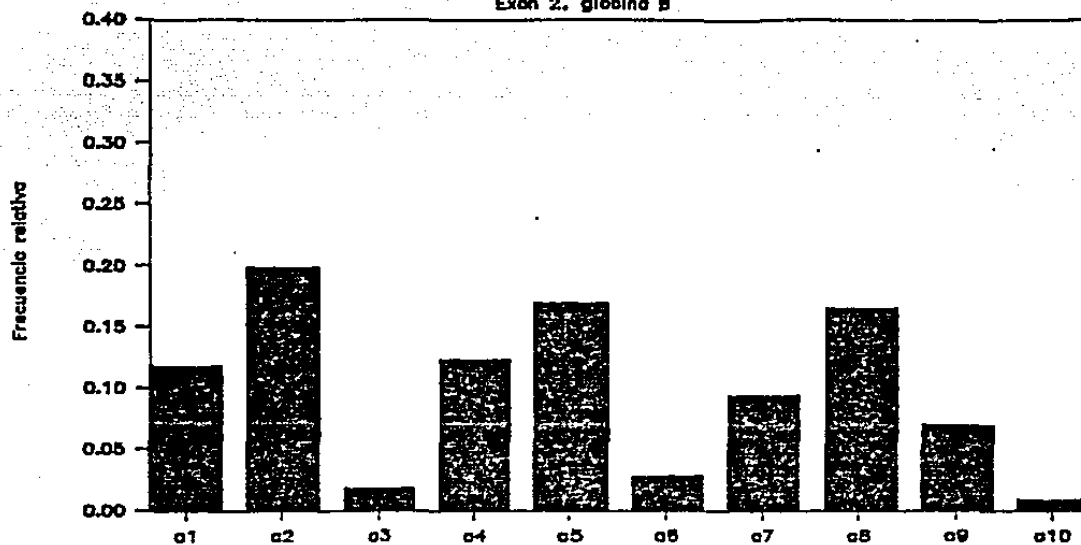


FIGURA 2.

Frecuencias relativas de los 10 digramas (ai) en algunas secuencias de DNA del Cuadro I. Los digramas se ordenan de acuerdo a la energía: a1 (GA/CT), a2 (CA/GT), a3 (TA/AT), a4 (GT/CA), a5 (CT/GA), a6 (AT/TA), a7 (AA/TT), a8 (GG/CC), a9 (GC/CG) y a10 (CG/GC).

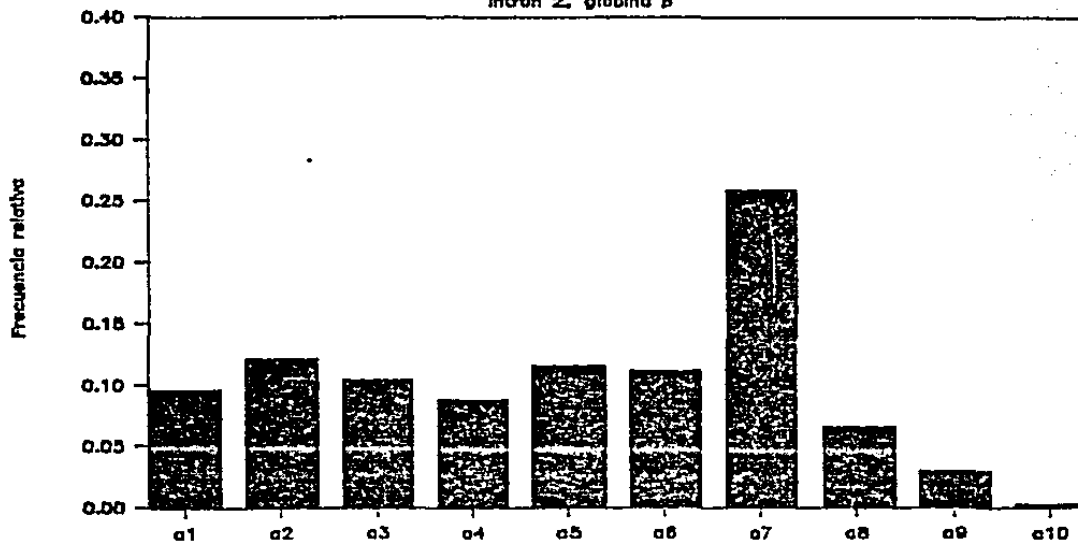
# Homo sapiens

Exón 2, globina  $\beta$



# Homo sapiens

Intrón 2, globina  $\beta$



distribuciones. Un examen general indica que, en principio, los digramas se pueden clasificar en tres grupos de energía. La división de estos tres grupos está dada por los digramas especulares TA/AT y AT/TA, los cuales tienen una abundancia relativa menor a los otros en parte debido a su simetría.

Puede observarse en la Figura 2 que las diferentes secuencias se distinguen generalmente por la abundancia relativa en el segundo y el tercer grupo. El primer grupo de energías es más o menos constante en su abundancia. Por ejemplo, en las secuencias de las globinas  $\alpha$ , la mioglobina y las inmunoglobulinas, abundan más los enlaces fuertes del tercer grupo mientras que en las globinas  $\beta$  y las leghemoglobinas, son más frecuentes los enlaces del segundo grupo.

Una forma de estudiar los eventos mutacionales que originan estas distribuciones estadísticas consiste en buscar correlaciones entre las abundancias de los diferentes digramas. Sin embargo, esto no proporciona información clara ni fácil de analizar si no se tiene una visión de conjunto que sirva como guía.

### § Clasificación y agrupamiento de digramas.

A partir de las distribuciones estadísticas de los digramas de las distintas secuencias, hemos planteado que éstos pueden agruparse en tres clases de energía. Con ello, se tiene la posibilidad de desarrollar un análisis con pocas variables y, por consiguiente, obtener resultados cuya interpretación es más sencilla.

El grupo de enlaces que podemos reconocer con claridad es el formado por los digramas GG/CC, GC/CG y CG/GC. La existencia de esta clase es evidente tanto en términos de composición de bases como por el hecho de que forman un cúmulo de alta energía de acuerdo al Cuadro II. La clasificación de los digramas restantes se hizo con base en el enlace TA/AT. La razón de ello es la discontinuidad que este digrama implica en la distribución estadística y en que, a partir del digrama TA/AT, no se obtiene el siguiente digrama en energía por una sustitución (véase la Figura 6). Se observa que esta forma de clasificar los digramas también

genera una distinción en lo que concierne a la composición de las bases. Así, el grupo de energías débiles está constituido por los digramas: GA/CT, CA/GT y TA/AT, y el grupo de energías medias por: GT/CA, CT/GA, AT/TA y AA/TT.

Una vez clasificados los enlaces en tres tipos de energía, se obtuvo, para cada clase, la abundancia promedio en las secuencias mostradas en el Cuadro I de la siguiente manera:

$$d_1 = (GA + TC + CA + TG + TA) / 5$$

$$d_2 = (GT + AC + CT + AG + AT + AA + TT) / 7$$

$$d_3 = (GG + CC + GC + CG) / 4$$

en donde  $d_1$ ,  $d_2$  y  $d_3$  son las abundancias promedio de las tres clases de enlaces.

A partir de lo anterior, se obtuvo una medida de la abundancia con respecto al azar, la cual, se obtiene de la siguiente expresión:

$$D_i = 3 \left( \frac{d_i}{\sum_{i=1}^3 d_i} \right)$$

La medida  $D_i$  toma valores desde  $D_i=0$  a  $D_i=3$  siendo  $D_i=1$ , la abundancia que se espera por azar. La Figura 3 muestra las relaciones de abundancia entre las tres clases de digramas.

Es posible en estas gráficas ajustar relaciones exponenciales pero se consideró mejor analizarlas como rectas en los intervalos  $D_3 < 1$  y  $D_3 > 1$ , en virtud de que el ajuste como relación exponencial se aleja de los puntos reales en los valores extremos de las variables  $D_1$ ,  $D_2$  y  $D_3$ . Se eligió la variable  $D_3$  como referencia porque esta refleja directamente la composición de la secuencia y la rigidez de la molécula como se ve en la Figura 4. Posteriormente, se verá que la variable  $D_3$  es muy útil en ese sentido. Los genes de las globinas  $\alpha$  y las inmunoglobulinas en general tienen valores  $D_3 > 1$  mientras que los correspondientes a las globinas  $\beta$  y las leghemoglobinas tienen valores  $D_3 < 1$ . En el caso de las secuencias con  $D_3 > 1$ , los exones se encuentran cerca del valor  $D_3=1$  mientras que los intrones y los pseudogenes tienen



valores en D3 mayores. En las secuencias con  $D3 < 1$ , también los exones se encuentran cerca del valor  $D3 = 1$ , mientras que, en sus intrones y pseudogenes, se tienen valores menores.

En la comparación D2 vs D3 se observa que estas variables guardan una relación inversa, lo cual indica la existencia de mutaciones que originan digramas de una clase a partir de los digramas de la otra. Lo mismo ocurre con la comparación D1 vs D3 en el intervalo  $D3 > 1$ , mientras que casi no hay cambio en D1 cuando  $D3 < 1$ . Cuando  $D3 = 1$ , se tiene que  $D2 = 0.93$  y  $D1 = 1.07$ . La relación D1 vs D2 muestra que estas variables mantienen una proporcionalidad directa en el intervalo  $D2 < 1$  lo cual indica que, en forma neta, no hay mutaciones que cambien los digramas entre D1 y D2. Así, en la disminución del valor D3 por abajo de 1, ocurren casi exclusivamente mutaciones  $D3 \longrightarrow D2$ .

La Figura 4 ilustra la relación de las variables D1, D2 y D3 con la entalpía promedio de las secuencias. Nuevamente, se ha preferido ajustar las curvas como rectas en los intervalos  $D3 < 1$  y  $D3 > 1$ , aunque también es posible ajustar relaciones exponenciales. La variable D3 es una medida directa de la entalpía promedio, sobre todo en el intervalo  $D3 > 1$ . Cuando  $D3 = 3$ , la entalpía promedio tiende a ser  $\Delta H \approx 11$  Kcal/mol, lo que nos indica que la formación de secuencias rígidas, involucra principalmente la formación del digrama GG/CC. De acuerdo con la Figura 3, el aumento en D3 es casi igual por la disminución en D2 y D1, de lo cual se deduce que pueden ocurrir las mutaciones;  $GA/CT \longrightarrow GG/CC$ ,  $GT/CA \longrightarrow GG/CC$  y  $CT/GA \longrightarrow GG/CC$ .

Aunque se pueden deducir algunas características globales de los eventos mutacionales a partir de las variables D1, D2, D3 y  $\Delta H$ , no se puede decir mucho sobre las mutaciones subyacentes especialmente en el intervalo  $D3 < 1$ . En la Figura 6 se esquematizan todas las vías posibles de transformación de digramas por sustitución, diagrama al que hemos denominado red mutacional y que nos sirve como guía para establecer correlaciones entre las abundancias de los digramas. Al plantear la red mutacional, se hacen 23 comparaciones en lugar de 90 como se haría normalmente al comparar todos los digramas entre sí. Las comparaciones entre los digramas se muestran en el Cuadro IV, y se obtiene una serie de

CUADRO IV. PARAMETROS DE LA CORRELACION ENTRE LAS ABUNDANCIAS DE LOS 10 DIGRAMAS EN DIFERENTES SECUENCIAS.

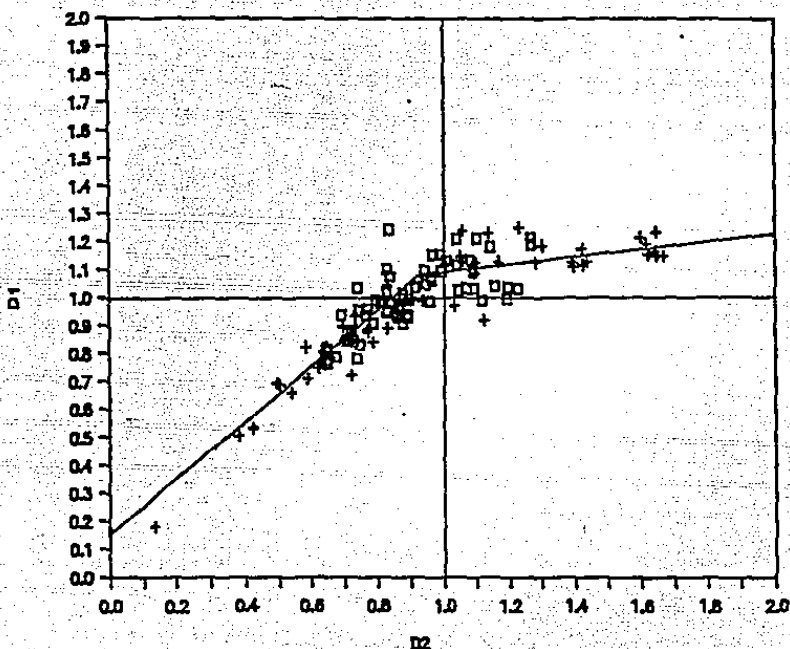
	INTRONES					
	D3 < 1			D3 > 1		
ay ↔ ax	r	m	b	r	m	b
1 ↔ 2	0.61	0.59	0.03	0.12	0.10	0.09
1 ↔ 3	-0.83	-0.58	0.16	-0.22	-0.53	0.12
2 ↔ 3	-0.86	-0.54	0.18	0.43	1.25	0.14
1 ↔ 4	0.10	0.16	0.10	0.09	0.11	0.10
1 ↔ 7	-0.76	-0.29	0.18	0.01	0.00	0.11
2 ↔ 5	0.54	0.32	0.09	0.57	0.55	0.07
2 ↔ 7	-0.72	-0.25	0.19	0.29	0.24	0.14
3 ↔ 7	0.79	0.43	-0.02	0.08	0.03	0.01
4 ↔ 5	0.22	0.10	0.09	0.12	0.08	0.09
4 ↔ 6	-0.48	-0.22	0.12	0.39	0.60	0.09
4 ↔ 7	-0.20	-0.05	0.11	-0.02	-0.02	0.10
5 ↔ 6	-0.90	-0.92	0.22	-0.45	-0.85	0.17
5 ↔ 7	-0.91	-0.52	0.25	0.14	0.12	0.15
6 ↔ 7	0.82	0.45	-0.02	0.32	0.12	0.01
4 ↔ 8	0.17	0.07	0.09	-0.46	-0.19	0.15
4 ↔ 9	0.06	0.04	0.09	-0.65	-0.55	0.15
5 ↔ 8	0.87	0.80	0.07	-0.70	-0.41	0.26
5 ↔ 10	0.52	6.95	0.08	-0.83	-0.74	0.18
8 ↔ 9	0.83	1.93	0.01	0.58	1.21	0.16
8 ↔ 10	0.73	14.70	-0.02	0.72	1.10	0.22
8 ↔ 1	0.71	1.00	-0.04	-0.51	-1.05	0.38
9 ↔ 1	0.45	0.27	-0.00	-0.67	-0.66	0.16
10 ↔ 2	-0.11	-0.02	0.01	-0.85	-0.99	0.20

		EXONES					
ay ↔ ax		D3 < 1			D3 > 1		
1 ↔ 2		-0.04	-0.03	0.12	-0.42	-0.31	0.18
1 ↔ 3		-0.54	-0.48	0.13	-0.00	-0.00	0.12
2 ↔ 3		-0.80	-1.12	0.23	-0.27	-0.36	0.17
1 ↔ 4		0.09	0.09	0.11	-0.54	-0.33	0.16
1 ↔ 7		-0.42	-0.20	0.14	-0.38	-0.25	0.14
2 ↔ 5		-0.29	-0.32	0.23	-0.33	-0.18	0.19
2 ↔ 7		-0.72	-0.40	0.24	0.59	0.50	0.13
3 ↔ 7		0.75	0.29	-0.00	-0.15	-0.06	0.02
4 ↔ 5		-0.08	-0.08	0.12	-0.72	-0.58	0.20
4 ↔ 6		-0.62	-0.83	0.15	-0.75	-1.59	0.16
4 ↔ 7		-0.72	-0.36	0.16	-0.45	-0.59	0.16
5 ↔ 6		-0.47	-0.49	0.18	0.11	0.23	0.14
5 ↔ 7		-0.22	-0.07	0.16	-0.02	-0.03	0.15
6 ↔ 7		0.55	0.19	0.03	0.65	0.65	-0.00
4 ↔ 8		0.13	0.06	0.10	-0.34	-0.35	0.18
4 ↔ 9		-0.08	-0.12	0.12	-0.20	-0.24	0.14
5 ↔ 8		-0.20	-0.06	0.16	-0.24	-0.27	0.20
5 ↔ 10		-0.37	-0.58	0.17	-0.80	-0.62	0.18
8 ↔ 9		0.41	1.16	0.04	-0.33	-0.40	0.23
8 ↔ 10		0.26	0.75	0.10	-0.07	-0.06	0.20
8 ↔ 1		0.26	0.21	0.10	0.06	0.07	0.19
9 ↔ 1		-0.89	-0.35	0.11	-0.02	-0.00	0.09
10 ↔ 2		0.01	0.00	0.02	-0.37	-0.40	0.12

Se muestran los valores de correlación directa entre la abundancia del enlace ax en las abscisas y ay en las ordenadas, r, es el índice de correlación, m, la pendiente y b, la ordenada al origen. Las secuencias que se tomaron en cuenta son las del CUADRO I excepto las globinas y del gorila,  $\theta_1$  del orangután y  $\psi\beta_2$  del conejo.

En las secuencias de exones no se consideraron los genes D y J ni tampoco los exones de los pseudogenes. En los intrones, se incluyeron todas las secuencias de este tipo a excepción de las globinas mencionadas.

Las secuencias se analizaron en dos grupos: con valor  $D3 < 1$  y  $D3 > 1$ .

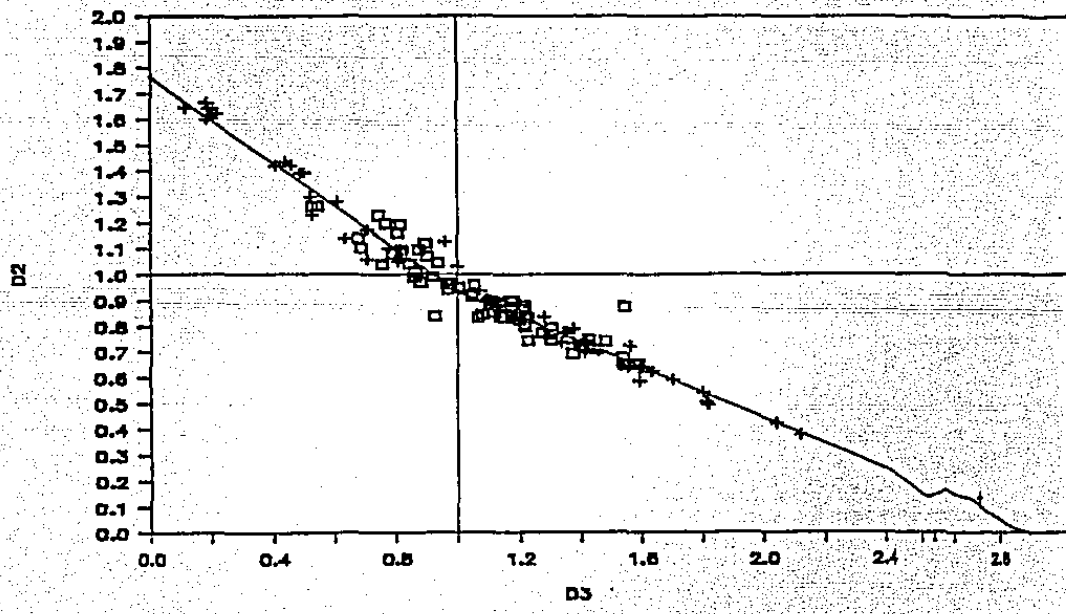
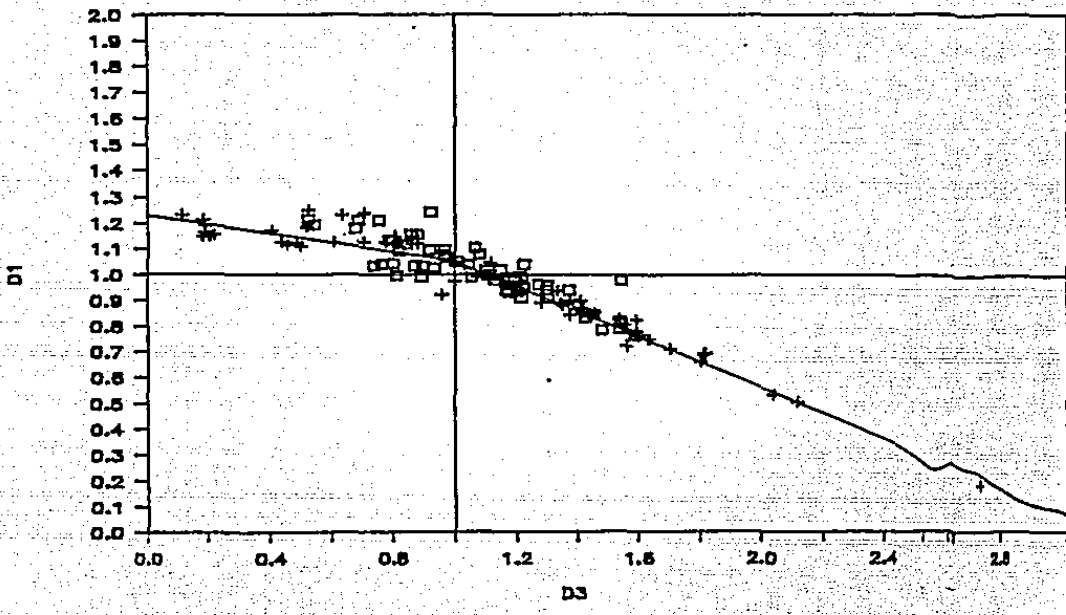


□ Exones  
+ Intrones

	r	
	D3 < 1	D3 > 1
D2 vs D1	0.41	0.90
D3 vs D1	-0.54	-0.97
D3 vs D2	-0.96	-0.96

FIGURA 3.

Relaciones entre los valores D1, D2 y D3 para las secuencias del Cuadro 1, excepto la globina y del gorila. Se muestran las rectas obtenidas por regresión lineal directa en las tres gráficas para los casos D3 < 1 y D3 > 1. Se indican los índices de correlación (r) para cada caso.



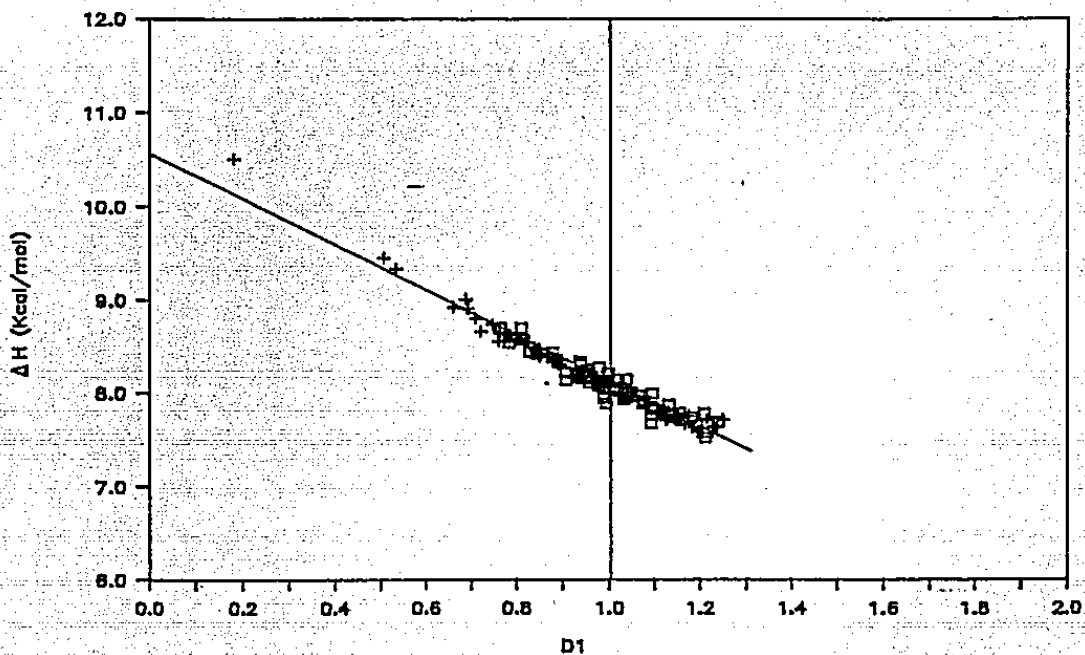
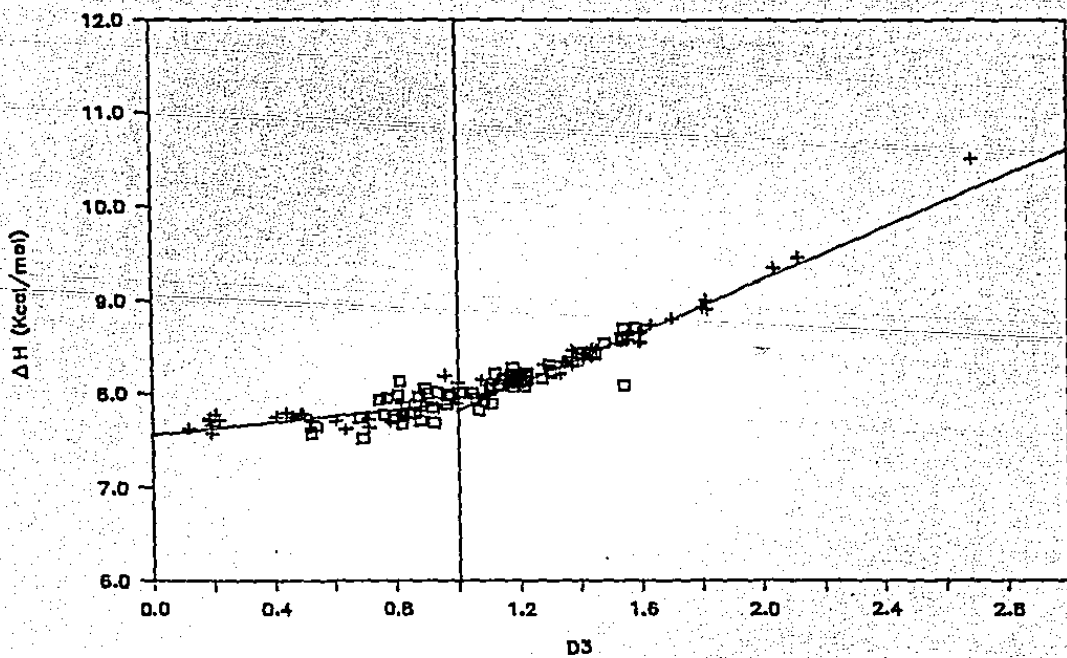
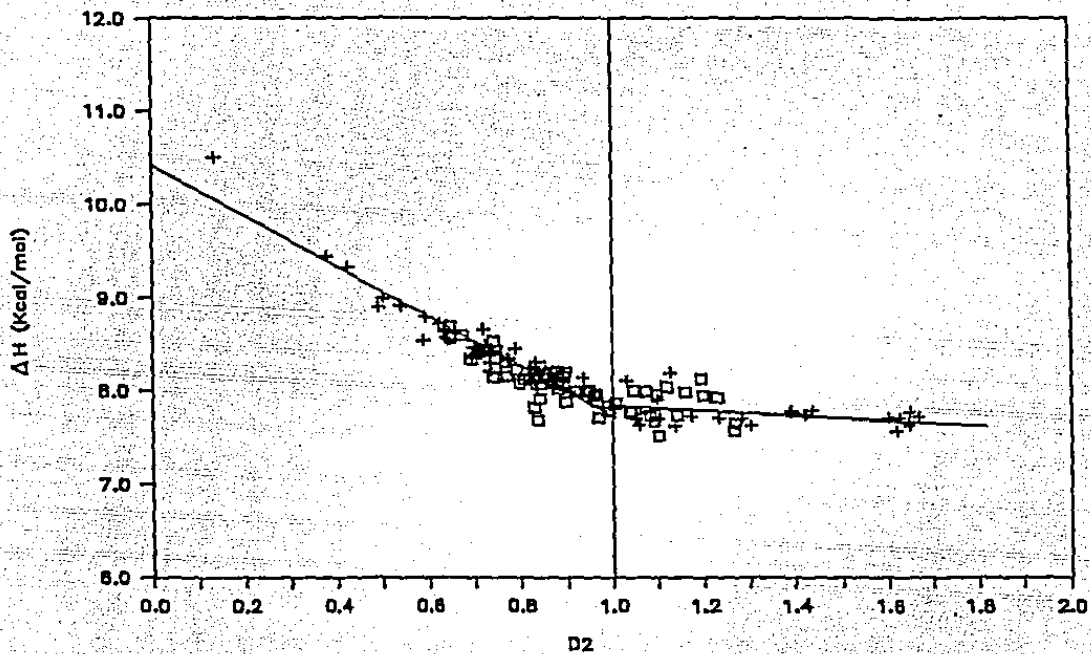
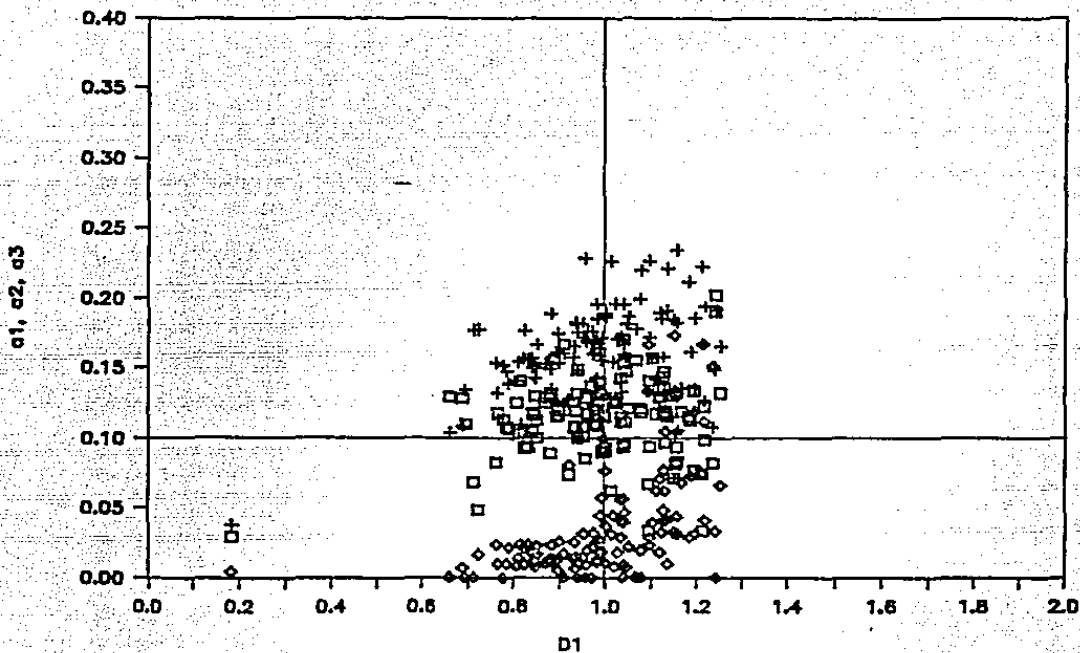


FIGURA 4.

Relación de los valores D1, D2 y D3 con la entalpía promedio ( $\Delta H$ ) de las secuencias del Cuadro I, excepto la globina y del gorila. Se muestran las rectas obtenidas por regresión lineal directa para los casos  $D3 < 1$  y  $D3 > 1$  y los índices de correlación ( $r$ ). En el caso  $D1$  vs  $\Delta H$ , se ajustó una recta en todo el intervalo de valores.



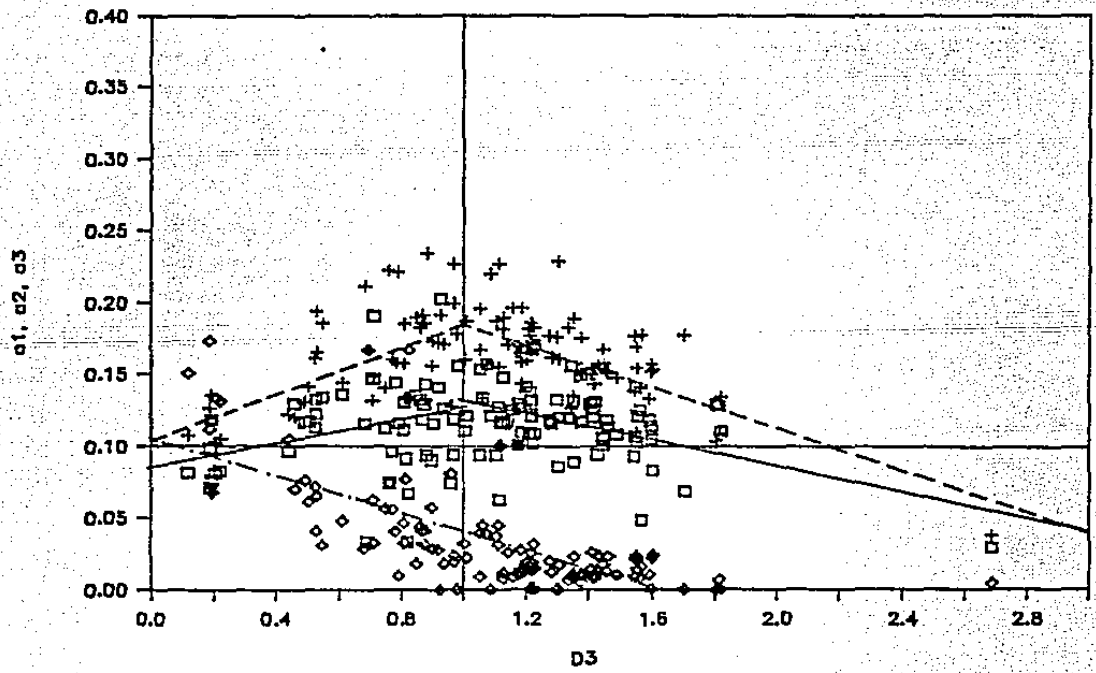
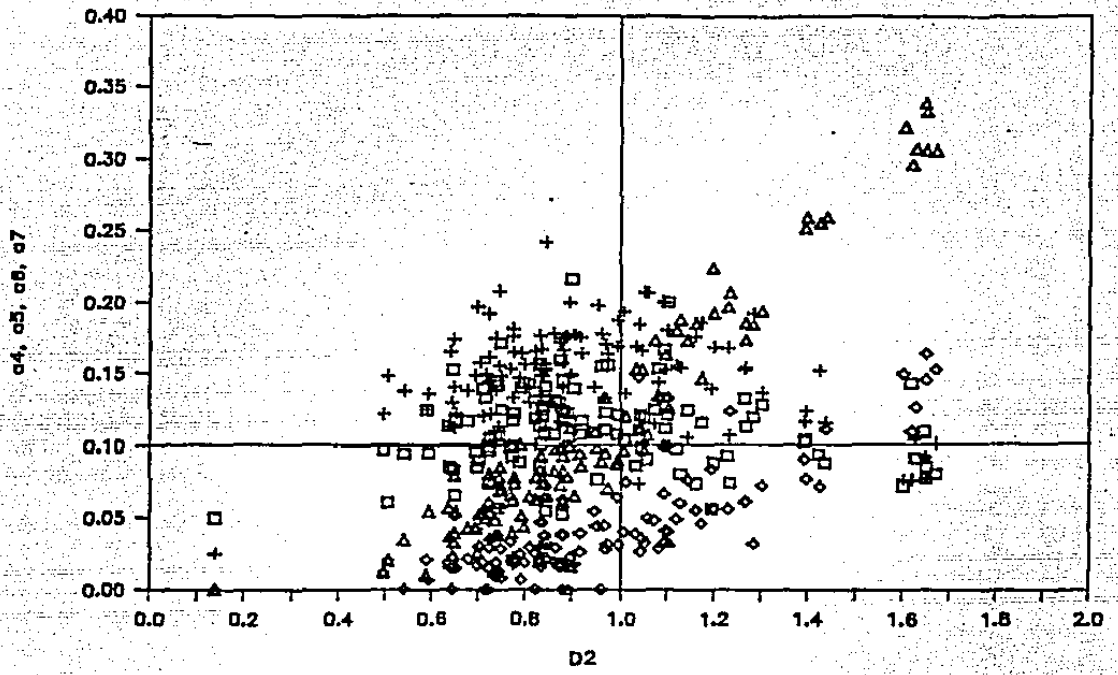


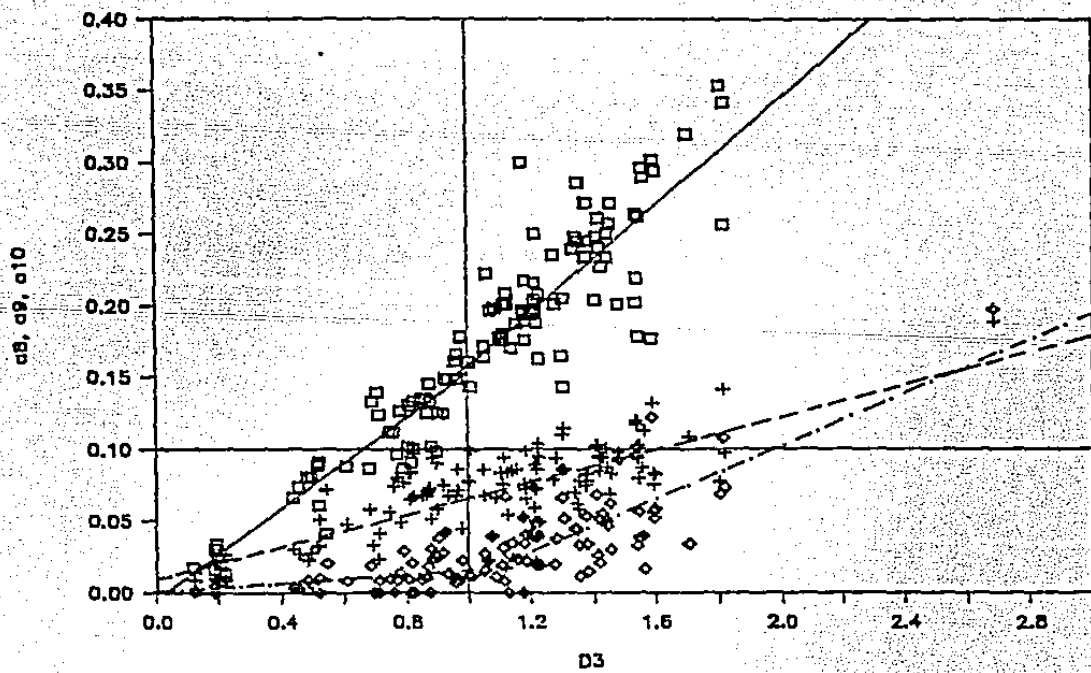
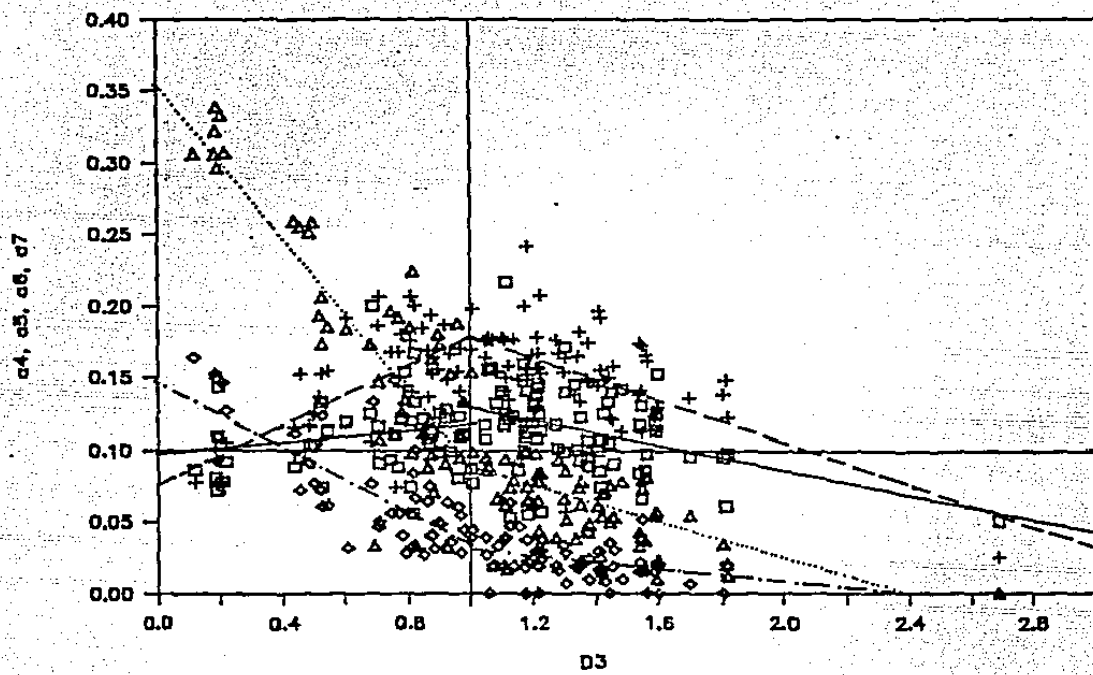
—	□	D3 vs a1	0.25	-0.46
- -	+	D3 vs a2	0.61	-0.60
- . -	◇	D3 vs a3		-0.69
—	□	D3 vs a4	0.22	-0.38
- -	+	D3 vs a5	0.66	-0.58
- . -	◇	D3 vs a6	-0.75	-0.40
.....	△	D3 vs a7	-0.82	-0.59
—	□	D3 vs a8		0.94
- -	+	D3 vs a9		0.76
- . -	◇	D3 vs a10	0.41	0.74

FIGURA 5.

Relación de las abundancias de los 10 digramas ( $a_i$ ) con los valores  $D_1$ ,  $D_2$  y  $D_3$ . Se indican las rectas obtenidas por regresión lineal directa y los índices de correlación ( $r$ ) para los casos  $D_3 < 1$  y  $D_3 > 1$ . Las secuencias analizadas son las del Cuadro I a excepción de lsd globinas y del gorila,  $\theta_1$  del orangután y  $\psi\beta_2$  del conejo.







CLASE 1

CLASE 2

CLASE 3

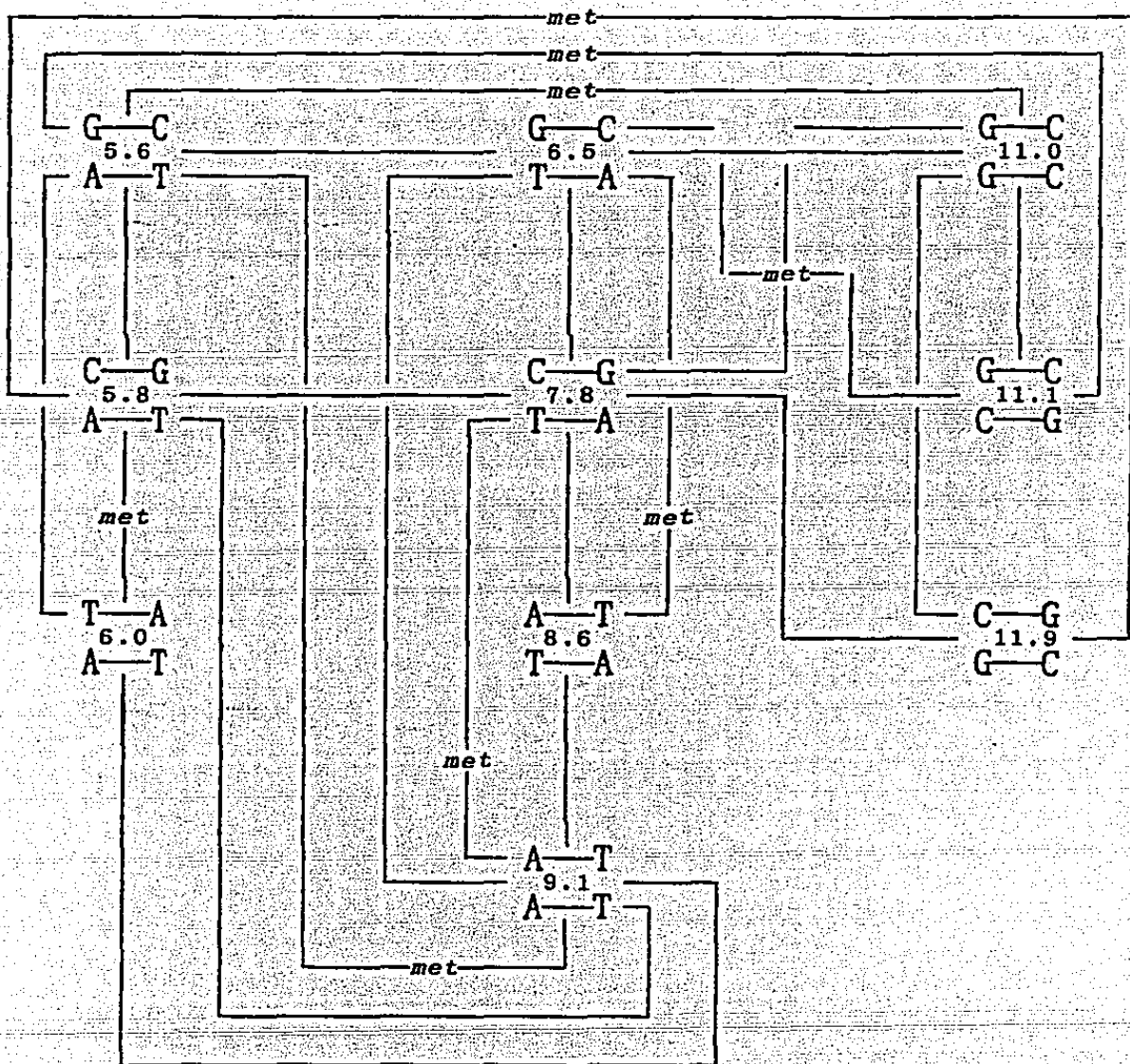


Figura 6.

Red mutacional. Se muestran las vías posibles de transformación por sustitución simple en los diez digramas del DNA agrupados en tres clases. Las bases superior izquierda e inferior derecha de cada interacción corresponden al extremo 5' de cada cadena. Los números indican el valor de  $\Delta H^\circ$  en Kcal/mol. La abreviatura *met* señala la posibilidad de la transición 5-mC  $\rightarrow$  T.

datos cuya interpretación no es fácil. En muchos casos la correlaciones entre digramas no coinciden con el comportamiento global que hemos descrito de las variables D1, D2 y D3.

La red mutacional muestra que un digrama tiene varias posibilidades de mutación, así que difícilmente pueden hallarse correlaciones claras entre cualesquiera dos de ellos. Una forma posible de solucionar este problema consiste en ver si existe alguna relación entre las abundancias de cada uno de los 10 digramas con las variables D1, D2 y D3, las cuales nos dan información global de las secuencias. En la Figura 5 se observa que para el caso de D1 y D2 no hay una relación definida con los enlaces que constituyen dichas variables. Al considerar D3 se obtiene para los 10 digramas una serie de curvas en general bien definidas. Esta información ahora puede complementarse con el Cuadro IV e interpretarse tomando como guía las gráficas de las Figuras 3 y 4.

La Figura 6 ilustra las vías mutacionales que ocurren cuando aumenta o disminuye la proporción de los enlaces fuertes, lo que se define por la variable D3. El incremento en los valores de D3 ocurre fundamentalmente por las sustituciones A-T  $\longrightarrow$  G-C o A-T  $\longrightarrow$  C-G especialmente como D1  $\longrightarrow$  D3. La disminución en el valor de D3 implica las mutaciones C-G  $\longrightarrow$  T-A y C-G  $\longrightarrow$  A-T en los cambios D3  $\longrightarrow$  D1, D2. Subsecuentemente, dentro de D1 y D2, estas mutaciones llevan hacia la formación de los enlaces TA/AT, AT/TA y AA/TT, especialmente este último.

### § Distribución de energías locales en secuencias específicas.

Con lo expuesto, se pueden describir de forma general los eventos mutacionales involucrados en la evolución de distintos tipos de secuencias. Cabe preguntarse si, en una secuencia dada, las mutaciones tienen las mismas implicaciones en términos de estabilidad a lo largo de esa secuencia. Hemos visto que, en relación con la entalpía, los exones se distinguen de las secuencias que no se expresan como los intrones y los pseudogenes. Las secuencias que no se expresan son o más rígidas o más flexibles que los exones los cuales, en general, tienen

composiciones equilibradas.

Los exones, tienen una estructura que podemos llamar semántica en el sentido de que la secuencia de bases tiene un significado funcional. Esta estructura consiste en la sucesión de los codones, formados a su vez, por tres bases que representamos como ..XYZ.. En los exones, por lo tanto, podemos distinguir los enlaces xy, yz y zx. ¿Está la energía distribuida igualmente en estos tres enlaces? En caso de existir una cierta estructura, ¿Es esta igual para diferentes tipos de secuencias?

Hemos visto que las tres bases que conforman un codón no están sujetas por igual a las restricciones selectivas. De acuerdo a ello, se esperaría que el enlace yz, en virtud de ser el menos restringido tuviera características energéticas parecidas a las de las secuencias, del tipo al que pertenece, que no se expresan. El enlace zx, en el otro extremo, tendría valores energéticos similares a los de los exones del tipo al cual pertenece la secuencia.

La Figura 7 ilustra los valores de  $D_1$ ,  $D_2$ ,  $D_3$  y  $\Delta H$  de los de los enlaces xy, yz y zx con respecto a los valores globales de la secuencia completa en los exones de las globinas. Se observa que la distribución de la energía en la secuencia no es igual en las tres posiciones del codón. Como habíamos supuesto, la entalpía del enlace yz es mayor que el promedio de la secuencia en las secuencias de entalpía alta, y menor en las secuencias de entalpía baja. La energía en el enlace xy, parece independiente del valor promedio de energía.

La Figura 8 muestra las relaciones mencionadas en los exones de las regiones constantes y variables de las inmunoglobulinas. En general, se puede apreciar la misma situación que en las globinas pero con algunas diferencias que discutiremos posteriormente. Se decidió estudiar los segmentos D y J de las inmunoglobulinas para estudiar las características energéticas de las secuencias hipermutables. La Figura 9 ilustra los perfiles que se obtuvieron en la distribución energética de estas secuencias y se observa igualmente una diferenciación entre los enlaces xy, yz y zx pero el patrón de las curvas no es igual al caso de las globinas y las regiones constantes y variables de las inmunoglobulinas.

Finalmente, las Figuras 10 y 11 muestran los resultados obtenidos con los pseudogenes y los intrones respectivamente. En el caso de los pseudogenes, la posición en la que hay una diferencia mayor con respecto al promedio, es en el enlace xy lo cual puede concordar con la observación de Miyata y Hayashida<sup>(71)</sup> de que la base X es la más mutable en estas secuencias. Además, el enlace zx, al contrario de los exones de las globinas, es el más independiente con respecto a los valores globales de  $\Delta H$ , D1, D2 y D3. En la Figura 10 no se graficó el pseudogene  $\theta 1$  del orangután porque en éste, si hay una diferenciación clara entre los valores D1, D2, D3 y  $\Delta H$  para las tres posiciones en el codón de la misma manera que en las globinas  $\alpha$  comunes. Aunque no se ha identificado ningún péptido producto de este gene, se ha encontrado que en muchas especies tiene características de un gene normal. Se sospecha que posiblemente tenga alguna función durante el desarrollo embrionario.<sup>(37)</sup> De ahí, que sea mejor no tenerlo en cuenta como secuencia no funcional.

En el caso de los intrones, es evidente que, en términos de los tres enlaces de los codones, no hay casi diferenciación en los valores energéticos y de composición. Esto induce a pensar que las diferencias en las características de los tres enlaces en los exones, se deben a restricciones selectivas. El que los intrones estén sujetos a restricciones selectivas menos fuertes, por lo menos en lo referente a la expresión genética, también explica porqué las variables  $\Delta H$ , D1, D2 y D3 en estas secuencias tienen valores más extremos en comparación a los exones.

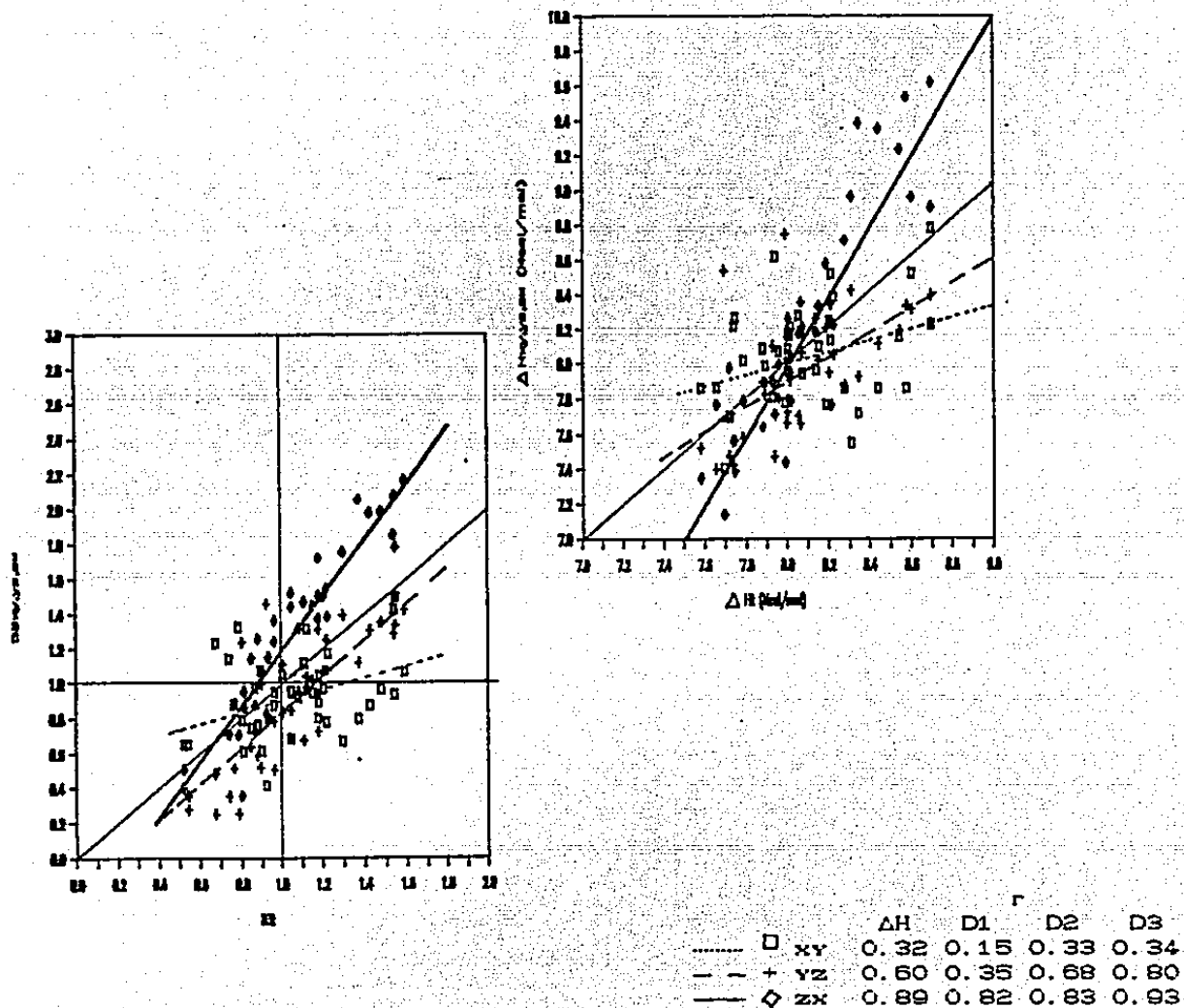
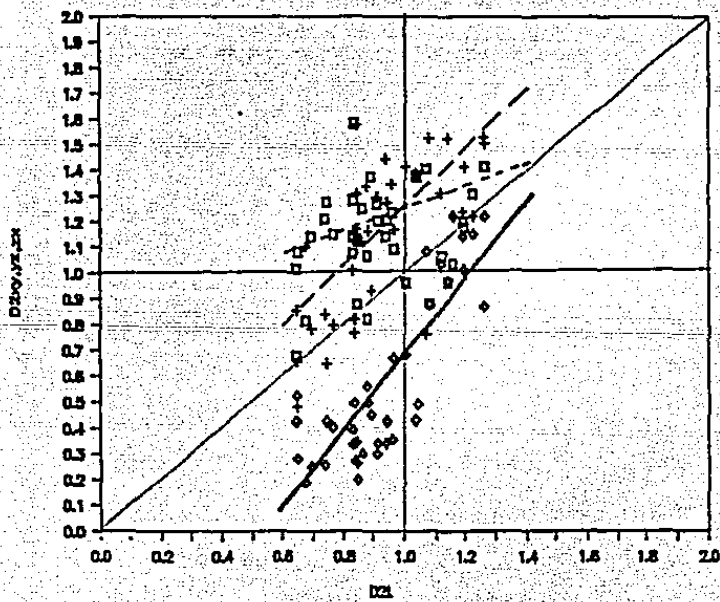
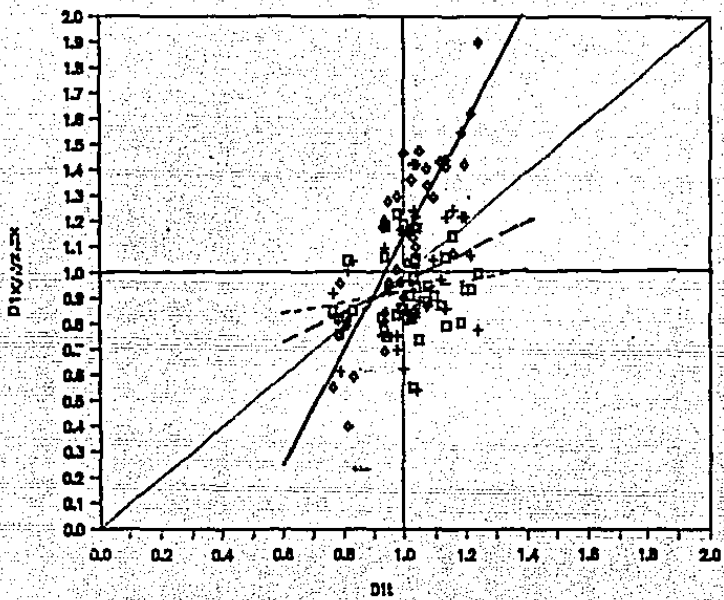
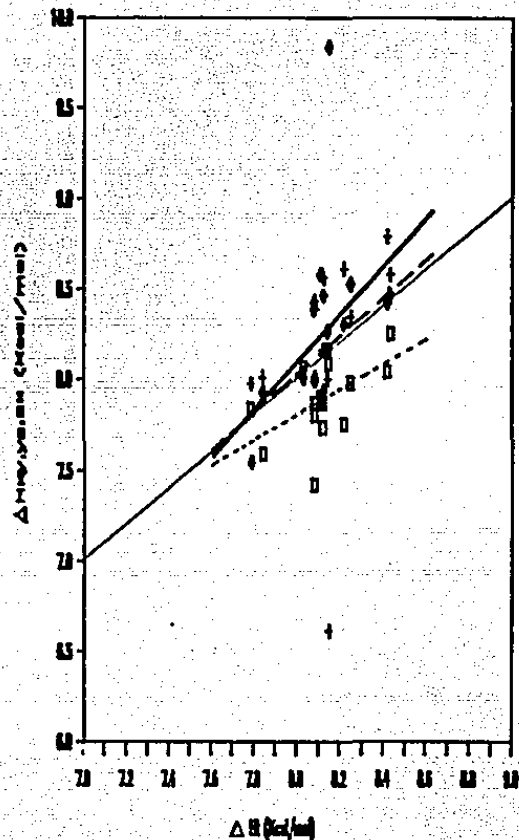
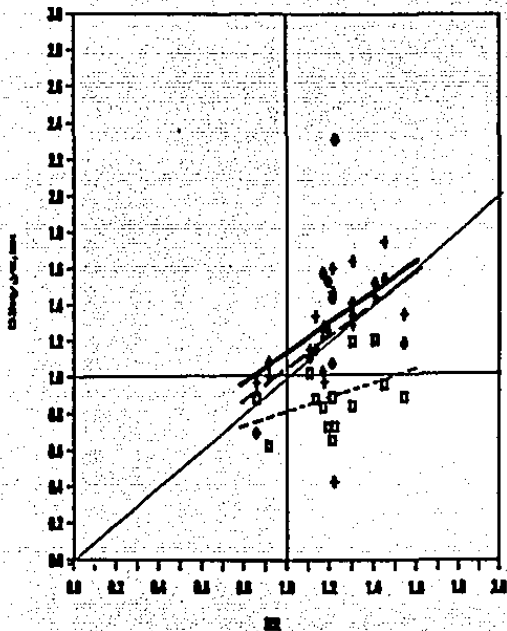


FIGURA 7.

Valores de las variables  $\Delta H$ , D1, D2 y D3 en los enlaces xy, vz y zx en los exones de las globinas (Cuadro I excepto la globina  $\gamma$  del gorila), en relación a los valores totales (t) de una determinada secuencia. Se indican las rectas obtenidas por regresión lineal directa y los índices de correlación (r). (n = 38).



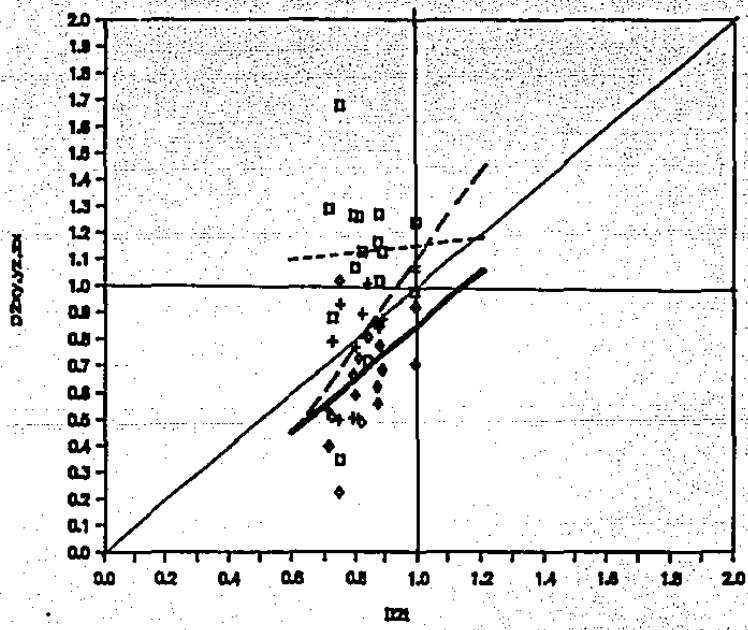
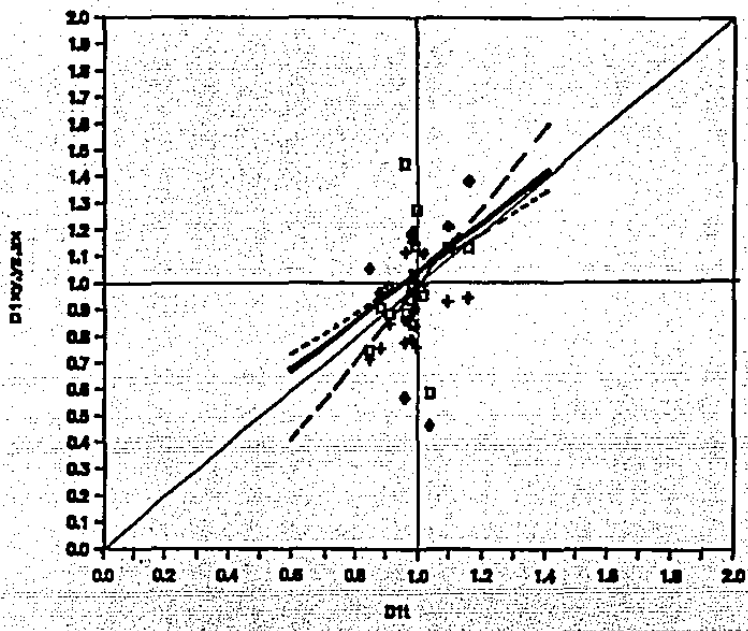


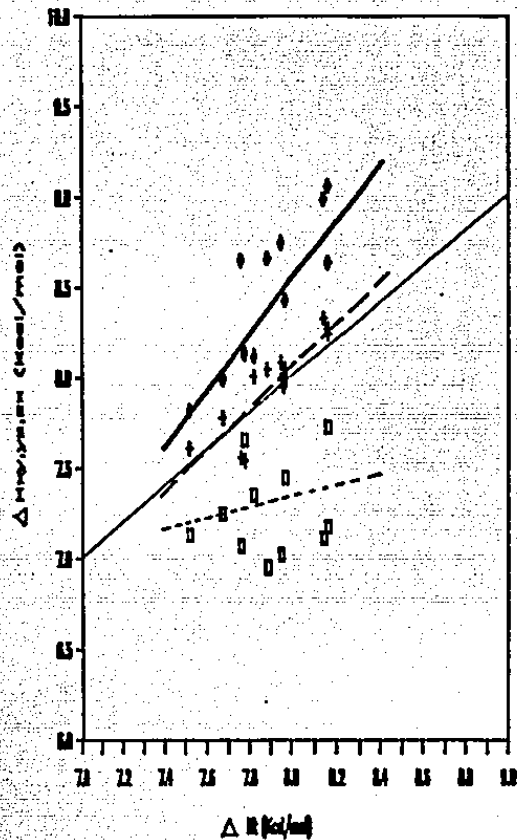
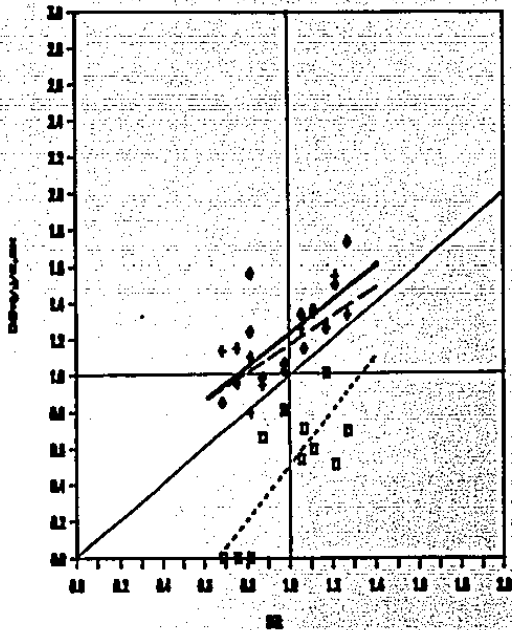


		$\Delta H$	D1	D2	D3
—	□ xy	0.52	0.27	0.03	0.33
- -	+ yz	0.35	0.34	0.68	0.48
—	◇ zx	0.44	0.28	0.45	0.43

FIGURA 8.

Valores de las variables  $\Delta H$ , D1, D2 y D3 en los enlaces xy, yz y zx en las regiones constantes y variables de las inmunoglobulinas (Cuadro I), en relación a los valores totales (t) de una determinada secuencia. Se indican las rectas obtenidas por regresión lineal directa y los índices de correlación (r). (n = 15).

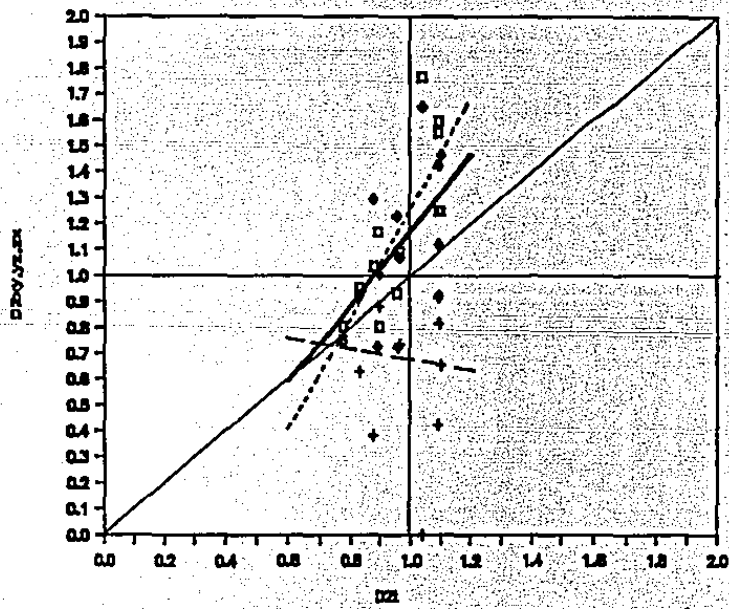
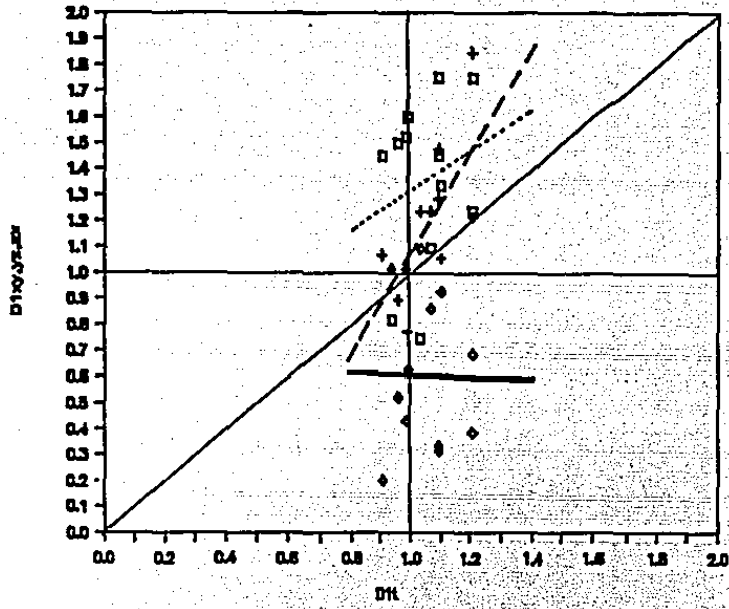


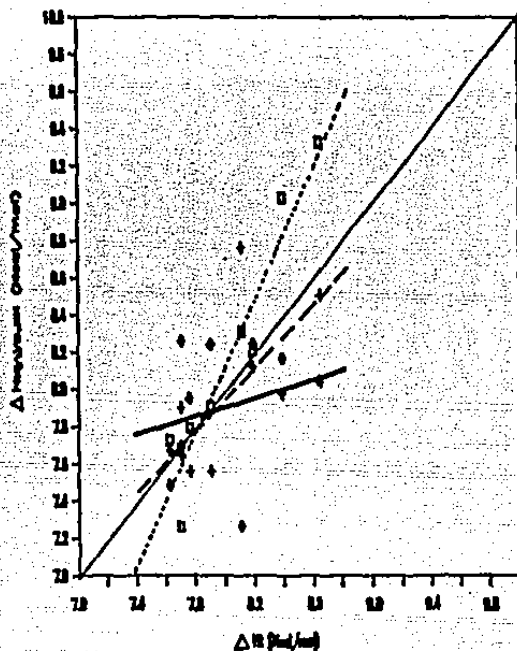
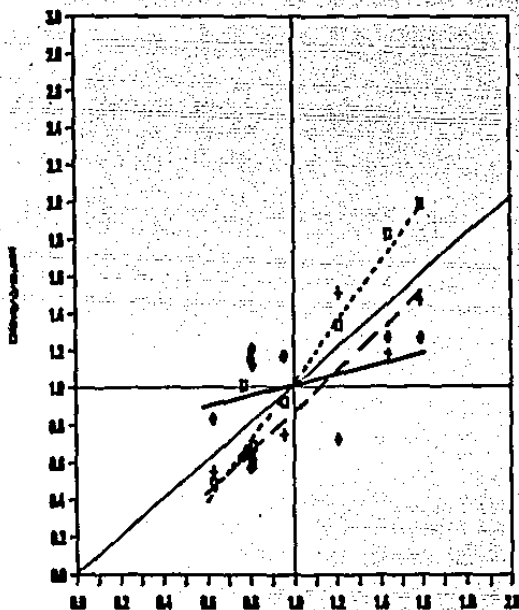


		$\Delta H$	$D1$	$r$	$D2$	$D3$
— □	xy	0.20	0.23	0.77	0.77	
- - +	yz	0.87	0.71-0.12	0.72	0.72	
— ◇	zx	0.75-0.02	0.59	0.67		

FIGURA 9.

Valores de las variables  $\Delta H$ ,  $D1$ ,  $D2$  y  $D3$  en los enlaces xy, yz y zx en los segmentos D y J de las inmunoglobulinas (Cuadro I), en relación a los valores totales (t) de una determinada secuencia. Se indican las rectas obtenidas por regresión lineal directa y los índices de correlación ( $r$ ). ( $n = 12$ ).

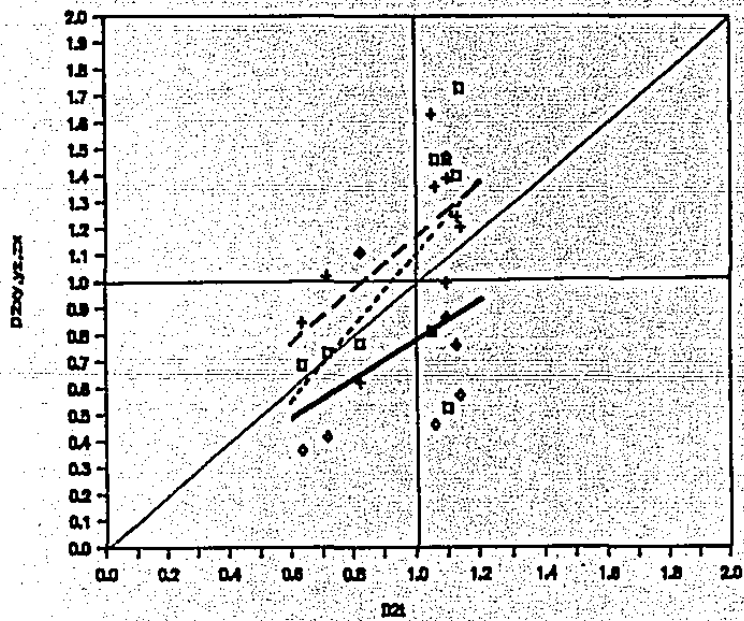
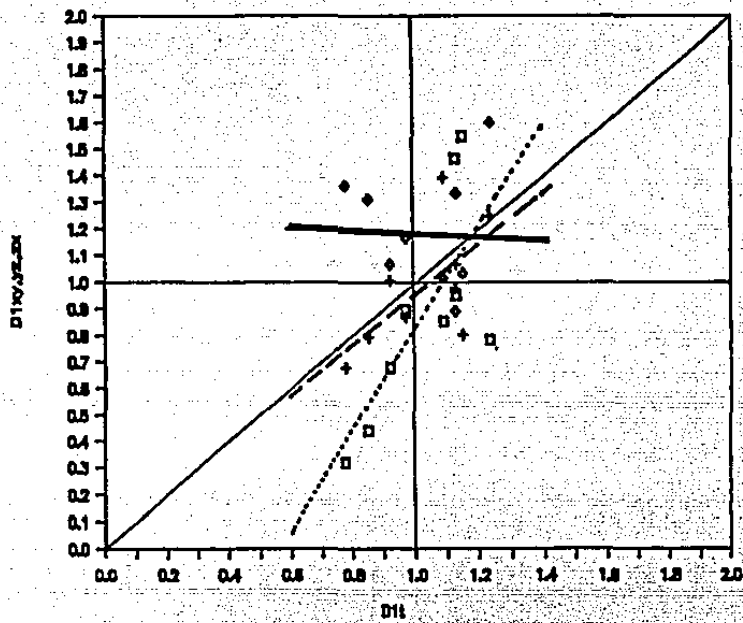


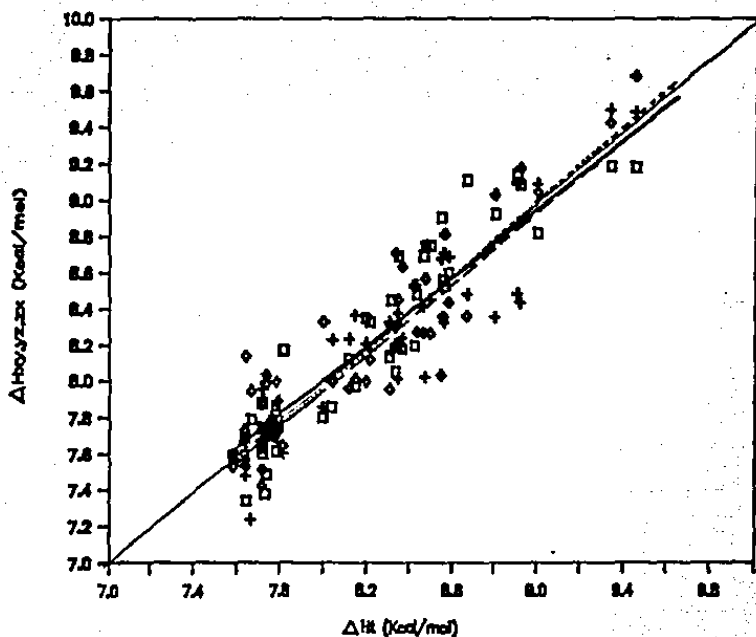


	$\Delta H$	$r$		
		D1	D2	D3
□ xy	0.95	0.72	0.61	0.97
+ yz	0.68	0.63	0.64	0.91
◇ zx	0.25	-0.04	0.39	0.41

FIGURA 10.

Valores de las variables  $\Delta H$ , D1, D2 y D3 en los enlaces xy, yz y zx en los pseudogenes (Cuadro I excepto el gene  $\theta 1$  del orangután que posiblemente es funcional), en relación a los valores totales (t) de una determinada secuencia. Se indican las rectas obtenidas por regresión lineal directa y los índices de correlación (r). (n = 9).

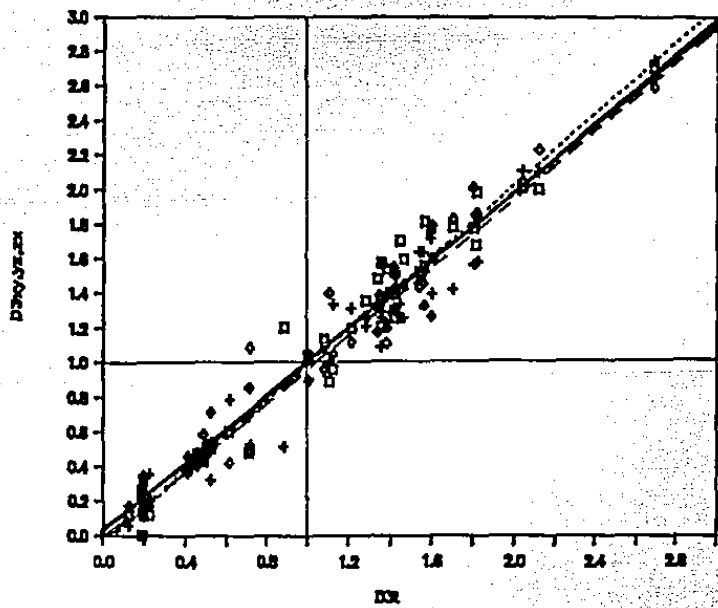
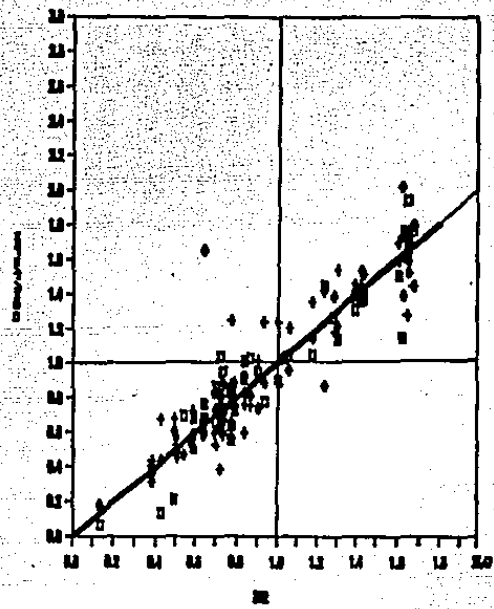
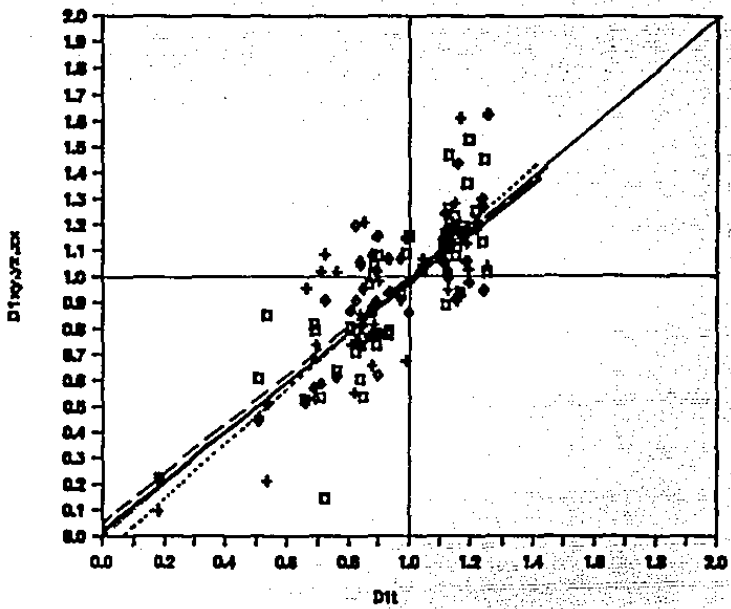




	$\Delta H$	$r$		
		D1	D2	D3
□ xy	0.95	0.81	0.94	0.98
+ yz	0.95	0.78	0.93	0.98
◇ zx	0.93	0.80	0.90	0.97

FIGURA 11.

Valores de las variables  $\Delta H$ , D1, D2 y D3 en los enlaces xy, yz y zx en todos los intrones del Cuadro I, en relación a los valores totales (t) de una determinada secuencia. Se indican las rectas obtenidas por regresión lineal directa y los índices de correlación (r). (n = 47). La fase de cada intrón se definió como la misma del exón que le precede.





## DISCUSION

### 5 Nociones sobre la relación de la energía local en el DNA y la mutagénesis.

Se ha sugerido la hipótesis de que la secuencia de nucleótidos en el DNA no está determinada únicamente por el proceso de la selección natural sino que existen restricciones físicas en la mutagénesis que definen condiciones específicas en el fenotipo del genoma sobre las que opera la selección natural.

Como una primera aproximación, se analizó el espectro de mutaciones en cepas de *E. coli* deficientes en los sistemas de reparación. Hemos visto, en la sección anterior (Figura 1), que las sustituciones del tipo A-T  $\longrightarrow$  G-C o A-T  $\longrightarrow$  C-G, ocurren en los sitios de entalpía menor. En la Figura 1 se observa, que bases en las que los enlaces vecinos tienen la conformación débil-fuerte o fuerte-débil, presentan una alta incidencia de mutaciones. Consideramos que esto puede deberse a que, en estas conformaciones, se tienen puntos de torsión que desestabilizan la estructura localmente. Como consecuencia, se tiene que las mutaciones a las que hacemos mención provocan, a largo plazo, un aumento en la entalpía promedio del DNA al formar enlaces más fuertes. Paralelamente, también se incrementa la fracción de los enlaces GG/CC, GC/CG y CG/CC.

En la Figura 1 también se ha hecho notar que las sustituciones G-C  $\longrightarrow$  A-T y G-C  $\longrightarrow$  T-A, al disminuir la fracción de los enlaces fuertes, provocan una disminución en la entalpía promedio. En el caso de las globinas y del hombre y el gorila, hemos visto que estas mutaciones se asocian a la secuencia ..NCG.. y particularmente ..GCG.. lo cual es acorde con la noción de que estas mutaciones ocurren debido a la desaminación de la 5-metil,citosina. En el caso de las cepas de *E. coli* sin sistemas de reparación, estas sustituciones se asocian a la secuencia ..GCN... Consideramos, de acuerdo a otros trabajos, <sup>(21,31)</sup> que las mutaciones en cuestión también ocurren por desaminación de 5-metil,citosina aunque no sabemos si en realidad las citosinas involucradas en estas sustituciones están metiladas o no.

A partir de la red mutacional que hemos planteado en la Figura 6, se puede concebir la mutagénesis como la transformación de un digrama determinado en otro. De acuerdo a los resultados obtenidos al analizar las sustituciones A ó T  $\rightarrow$  G ó C, podemos proponer intuitivamente que las transformaciones de los digramas se comportan como una transformación química simple, esto es:



donde  $e_i$  e  $e_j$  son los digramas  $i$  e  $j$  respectivamente. El que el índice ( $i$ ) sea menor al ( $j$ ) quiere decir que las mutaciones ocurren en los sitios más débiles de la molécula del DNA. Como se ve en la red mutacional, a largo plazo, este proceso llevaría a la acumulación de los enlaces fuertes si no hay restricciones selectivas para ello. Se puede entonces plantear una relación entre la abundancia de los distintos digramas con la energía a partir de la ecuación de Arrhenius. Del esquema (1), se tiene entonces que:

$$a_j/a_i = e^{(E_j - E_i)/RT} \quad (2)$$

donde  $a_j$ ,  $E_j$ ,  $a_i$  e  $E_i$  son las abundancias y las entalpias de los enlaces ( $j$ ) e ( $i$ ) respectivamente. En principio, esta relación se puede aplicar a todas las transformaciones en la red mutacional y obtener, a partir de la serie de ecuaciones resultantes, la distribución de energías en una secuencia en equilibrio. Si tenemos en cuenta el concepto de lectura de prueba de Hopfield,<sup>(50)</sup> la dependencia de la tasa de mutación con respecto a la temperatura,<sup>(85)</sup> e igualamos las expresiones que se desarrollan para la tasa de mutación, tenemos lo siguiente:

$$a_j/a_i = t_m = e^{(E_j - E_i)/RT} \quad (2.1)$$

donde  $t_m$ , es la tasa de mutación.

En las Figuras 3 y 4 hemos observado que en las secuencias que no se expresan y que, por lo tanto, están menos sujetas a las restricciones selectivas, se tienen valores altos en  $D_3$  y  $\Delta H$  como

se predice a partir del análisis de las mutaciones en sistemas carentes de reparación. De acuerdo a la relación entre la variable D3 y la abundancia de los diez tipos de digramas (Figura 5), el enlace GG/CC es el que más se acumula con el incremento de D3. Esto quizá sea resultado de que, como muestra la red mutacional, en el enlace GG/CC convergen tres transformaciones de enlaces débiles mientras que a los otros dos digramas de D3 únicamente convergen dos de estas transformaciones. En la Figura 5 se observa que, al incrementarse D3, disminuye la abundancia de los enlaces que originan los digramas que forman D3. También disminuyen los enlaces TA/AT, AT/TA y AA/TT los cuales no originan en un solo evento a los digramas de D3. Esto indica que, aunque no haya la correlación correspondiente, al comparar las abundancias de los digramas entre sí (Cuadro IV,  $D3 > 1$ ), si existen mutaciones en los enlaces mencionados que, llevan a la formación de los enlaces de D3.

Lo anterior muestra lo que denominaríamos como la tendencia natural de una secuencia determinada. Sin embargo, hemos visto que la mutagénesis abarca más eventos que los descritos. En el análisis de la globina y de los primates y en el gene *lacI* de *E. coli*, encontramos que las sustituciones G ó C  $\rightarrow$  A ó T, están relacionadas con la desaminación de la 5-metil,citosina. Como hemos visto, en los eucariontes la citosina se metila principalmente en la secuencia ..NCG. y particularmente en ..GCG.. Ya que el digrama CG/GC desaparece casi por completo cuando  $D3=1$  (Figura 5), se tiene que la metilación de la citosina y su consecuente desaminación ocurren en sitios donde la proporción de G-C es más o menos grande, lo que concuerda con el reporte de Li *et al.* <sup>(62)</sup> No disponemos de medidas de la entalpía en los enlaces con 5-metil,citosina pero, de acuerdo a lo expuesto, esta base es un sitio altamente mutable, especialmente por el aumento en la temperatura <sup>(29,85)</sup> lo que, en consonancia con la ecuación (2.1), incrementa la tasa de mutación hacia el digrama débil en una transformación cualquiera.

Cuando la citosina está metilada, los enlaces de D3 dejan de ser los elementos hacia cuya formación tienden las mutaciones en la red. Acorde con la hipótesis de que las secuencias de DNA

tienden a la formación de enlaces fuertes como ilustra el esquema (1), se tiene que las mutaciones llevarían a la formación de los enlaces TA/AT, AT/TA y AA/TT, en especial este último. De acuerdo al paradigma del esquema (1), en el intervalo  $D3 < 1$ , estos tres enlaces tienden a aumentar mientras que hay disminución en todos los demás (Figura 5). El enlace AA/TT muestra el incremento mayor mientras que el de menor proporción es el del digrama TA/AT. Los intrones y los pseudogenes, que están menos sujetos a restricciones selectivas, son los que presentan valores muy bajos en D3. A partir de la ordenadas al origen obtenidas en la Figura 5, se puede calcular la entalpia promedio en secuencias cuyo valor  $D3=0$ . Se tiene entonces que la entalpia promedio calculada ( $\Delta H_c$ ) para una secuencia en  $D3=0$  es:

$$\Delta H_c = \sum_{i=1}^{10} \Delta H_i^0 b_i$$

donde  $\Delta H_i^0$ , y  $b_i$  son respectivamente, la entalpia de enlace y la abundancia de un determinado enlace cuando  $D3=0$  (la ordenada al origen obtenida en la Figura 5 en la comparación de  $a_i$  vs  $D3$ ). Este cálculo da un valor de  $\Delta H_c \approx 7.6$  Kcal/mol que es igual al que se obtiene en la relación de las variables  $D1$ ,  $D2$  y  $D3$  vs  $\Delta H$  para  $D3=0$  (Figura 4). Así, la disminución en el valor de  $D3$  no lleva a una disminución drástica en la entalpia promedio de la secuencia sino que la entalpia se mantiene alrededor del valor mencionado. Esto se ve claramente en la Figura 4 cuando  $D3 < 1$ . Sin embargo, en nuestro concepto, esto no es producto de una restricción selectiva sino de las propiedades físicas del proceso de mutagénesis. Debemos notar que en el intervalo  $D3 \approx 1$ , la relación  $a_i$  vs  $D3$  muestra una dispersión considerable al contrario de los valores extremos en  $D3$ . Como ya se dijo, alrededor del valor  $D3=1$ , se localizan las diversas secuencias de exones. Es posible que la dispersión de los puntos se debe a los efectos de restricciones selectivas.

Cabe preguntarse entonces, si en efecto la distribución de los diez digramas en las secuencias con  $D3 < 1$ , son resultado de la desaminación de la 5-metilcitosina. Partimos de que, en los

vertebrados, la abundancia del digrama CG/GC es pequeña y está metilado en la citosina en una proporción de 60 a 90 %. No obstante, en regiones donde la cantidad de GC es alta, es decir, en las isocoras pesadas, no hay 5-metilcitosina en el enlace CG/GC. En los vertebrados, los genes involucrados en el mantenimiento de las funciones básicas del organismo ('Housekeeping genes'), en general, se localizan en estas zonas así como algunos otros, incluidas las globinas  $\alpha$ . Las globinas  $\beta$  y algunas otras secuencias, se encuentran en zonas donde la cantidad de GC es muy pequeña y, especialmente, la cantidad del digrama CG/GC. En estas zonas, entonces, la disminución del enlace en cuestión, se atribuye a la transición 5-mC  $\longrightarrow$  T. La metilación en estas zonas, se ha relacionado con el control de la transcripción que provoca una disminución en la actividad de expresión de los genes, y se cree que este fenómeno está involucrado en la diferenciación celular. La metilación en un sitio determinado, se puede transmitir por generaciones.<sup>(11)</sup> Entonces, las diferencias en los valores que hemos analizado entre las globinas  $\alpha$  y  $\beta$  se pueden explicar si consideramos los efectos de la metilación en las globinas tipo  $\beta$ . En lo que concierne a las leghemoglobinas, desconocemos realmente si en estas secuencias hay metilación de citosina. No obstante, se ha reportado que en las plantas, la metilación en general, es mayor que en los animales.<sup>(99)</sup> Además, de acuerdo a las distribuciones de las secuencias que analizamos, la cantidad del enlace CG/GC es muy pequeña como en las globinas  $\beta$ . En la Figura 2 se pueden ver las distribuciones de dos secuencias de lehemoglobinas como ejemplo.

La metilación de la citosina puede ocurrir en zonas ricas en GC,<sup>(11)</sup> y hemos dicho que en los pseudogenes, la transición 5-mC  $\longrightarrow$  T ocurre sobre todo en zonas ricas de GC. En nuestro trabajo, también hemos visto que el enlace CG/GC se incrementa a partir de D3=1, esto es, que la disminución de este digrama por desaminación de la citosina, ocurre en secuencias donde la cantidad de GC es más o menos alta. De esto se deduce que la desaminación de la 5-metilcitosina evita que una determinada secuencia llegue a ser muy rígida, sin embargo, no es posible afirmar con certeza si esto tiene, o ha tenido durante la

evolución, un valor selectivo.

## 5 Estructura energética de diferentes tipos de secuencias.

Hemos descrito de modo general la red mutacional como un fenómeno subyacente a la evolución de distintas secuencias. Es lógico preguntar si, en una cierta secuencia, todos los sitios son equivalentes en términos de la distribución de energía determinada por las mutaciones.

El primer elemento de estudio, en este contexto, se refiere a la distribución de la energía de amarre en las tres posiciones que conforman un codón. Ya que un codón está formado por tres bases XYZ, se tienen tres posiciones de los enlaces, las cuales son; xy, yz y zx. Hemos visto que las bases del codón no están sujetas por igual a la restricción selectiva definida por la secuencia de aminoácidos. Se tiene que el enlace más restringido es el xy mientras que el enlace zx es el más independiente de este factor de selección. Por lo tanto, se esperaría que el enlace zx tuviera características de secuencias no sujetas a las restricciones selectivas impuestas por la secuencia de aminoácidos en las proteínas, como los intrones.

En la Figura 7, se muestra la comparación de los valores totales de las variables  $\Delta H$ , D1, D2 y D3 con los valores de éstas en las tres posiciones de enlace para el caso de las globinas. El enlace xy, en el caso de las tres variables, casi no muestra correlación con el valor total de éstas, mientras que el enlace zx, para valores altos de entalpía y de D3, muestra valores mayores al promedio total. Esto significa que en el enlace zx, al estar menos sujeto a restricciones selectivas, se acumulan los enlaces fuertes de la clase D3. En el caso de secuencias con valores bajos de D3, el enlace zx presenta una entalpía promedio y un valor de D3 menor al valor total de las secuencias. Como muestra la Figura 7 (en la gráfica de D1), ello se debe al incremento en los valores de D1, lo cual, en consonancia con la red mutacional, puede deberse a la desaminación de la 5-metilcitosina. Las líneas de ajuste en el enlace yz, se mantienen cerca del valor promedio total de las secuencias con un

parecido ligero al comportamiento del enlace xy.

En la Figura 8 se ilustra el patrón de distribución de las variables AH, D1, D2 y D3 para las regiones constantes y variables de las inmunoglobulinas. Como en el caso de las globinas, la correlación más pequeña es la del enlace xy mientras que en el zx hay una correlación evidente. También, los valores de AH y D3 aumentan poco en el enlace xy al aumentar estos valores en toda la secuencia y se incrementan más en el enlace zx.

A diferencia de las globinas, en las inmunoglobulinas las diferencias de las pendientes en las rectas correspondientes a xy, yz y zx, no son tan grandes. Esto puede ser reflejo de que las inmunoglobulinas y las globinas no están sujetas a las mismas restricciones selectivas. Las proteínas menos sujetas a presiones selectivas evolucionan a una tasa de sustitución mayor de acuerdo al neutralismo y la tasa de sustitución en las globinas es menor a las de las inmunoglobulinas.<sup>(108)</sup> Así, las tres posiciones en los enlaces en las inmunoglobulinas se sustituyen a tasas no tan diferentes entre sí como en las globinas. En secuencias no restringidas en términos de la secuencia de aminoácidos como intrones y pseudogenes las cuales evolucionan más rápidamente (Figuras 10 y 11), se observa uniformidad o falta de regularidad en las tres posiciones. Esto implica que la distribución diferencial de la energía en las secuencias exónicas se debe a una diferenciación en la intensidad con la cual las restricciones en la secuencia de aminoácidos afectan las tres posiciones del codón.

El neutralismo afirma que la tasa de sustitución de nucleótidos tiende a ser la tasa de mutación en secuencias no restringidas selectivamente.<sup>(56,57)</sup> Hemos encontrado que en este tipo de secuencias las mutaciones tienden a lograr valores extremos en la composición de las bases (reflejado en D3) y en la energía de estabilidad del DNA. La ecuación (2.1) muestra que el valor de la tasa de mutación ( $t_m$ ) es proporcional al factor  $a_j/a_i$  que define el exceso de un digrama ( $a_j$ ) más estable que otro ( $a_i$ ). Estos resultados indican que en secuencias poco restringidas selectivamente, no sólo la tasa de sustitución tiende a ser la tasa de mutación sino que los eventos de sustitución tienden a reflejar la naturaleza no aleatoria de la mutación que hemos

empezado a definir con el concepto de la red mutacional.

En la Figura 8, los valores de  $\Delta H$  y  $D_3$  en el enlace  $zx$ , tampoco llegan a ser menores que el promedio total en ninguna parte de la gráfica. En el caso de las globinas, el que la línea de  $zx$  cruzase el valor promedio, se atribuye a la metilación lo cual no ocurre en las inmunoglobulinas. También se observa que las líneas  $xy$ ,  $yz$  y  $zx$ , en  $D_1$  coinciden casi con la línea del valor promedio para las secuencias completas y se observan algunas diferencias en las gráficas de  $D_2$  y  $D_3$ . El enlace  $zx$  es en el que hay el valor más alto en  $D_3$  y el valor más bajo en  $D_2$ . Esto significa que en las inmunoglobulinas, las mutaciones en  $zx$  ocurren más como  $D_2 \rightarrow D_3$ , esto es, en una fase más avanzada en la red mutacional con respecto a las globinas tipo  $\alpha$ . De ello se deduce que la incorporación de las globinas a las isocoras pesadas es reciente en la evolución de los mamíferos. El que las globinas tipo  $\beta$  se encuentren en los componentes ligeros del genoma, también apoya esta idea.

También se analizaron algunos segmentos  $D$  y  $J$  de los genes de inmunoglobulinas, los cuales codifican las regiones hipervariables de las inmunoglobulinas, con la intención de estudiar los parámetros físicos de nuestro interés en secuencias hipermutables. Se esperaría que las entalpías promedio de estas secuencias fueran muy bajas lo cual no es así. En la Figura 9 se ilustra la distribución de los valores promedio de las variables  $\Delta H$ ,  $D_1$ ,  $D_2$  y  $D_3$  en las posiciones  $xy$ ,  $yz$  y  $zx$ . En lo que se refiere a la gráfica de  $\Delta H$ , se observa un patrón muy parecido al de las regiones constantes y variables pero, en las secuencias  $D$  y  $J$ , las pendientes son muy diferentes entre sí. La entalpía promedio del enlace  $zx$  es considerablemente mayor al promedio al igual que la del enlace  $yz$  mientras que la del enlace  $xy$  es muy baja, incluso, menor a 7.6 Kcal/mol que es la energía promedio cuando  $D_3=0$  en el análisis global (Figura 4). Esta entalpía pequeña, se debe a una proporción alta de enlaces del tipo  $D_1$  y  $D_2$ , especialmente el primero. La correlación para el enlace  $xy$  en la gráfica de entalpía es muy pequeña lo que induce a pensar en la existencia de restricciones selectivas para el primer enlace del codón. El resultado es que el enlace  $xy$  es muy débil y está



rodeado por dos enlaces fuertes. De manera similar al caso del gene *lacI* de *E. coli*, las bases que forman este enlace (X y Y), son puntos de torsión en el sentido de que, alrededor de ellas, los enlaces tienen la conformación fuerte-débil o viceversa. Como se ve en la Figura 1, estos son puntos con incidencia alta de mutaciones. Con base en lo anterior, suponemos que en las secuencias D y J, existen mutaciones como  $D3 \longrightarrow D1$ ,  $D3 \longrightarrow D2$  y quizás  $D2 \longrightarrow D1$  en la posición xy. Esto posiblemente resulta de que la selección natural favorece los componentes menores de mutación ilustrados en el esquema (1).

Estos resultados, apoyan la idea de que la hipermutabilidad de las secuencias D y J depende de las propiedades físicas de la molécula de DNA que, en este caso, están determinadas por la selección natural. Se ha reportado que la mutación somática en las secuencias D y J no ocurre por inducción del antígeno y se presenta antes de la maduración del linfocito, (101) es decir, que ocurre mediante un mecanismo independiente de la maduración celular.

En la Figura 10 se muestra la distribución de los valores de las variables  $\Delta H$ , D1, D2 y D3 en las posiciones xy, yz y zx para las secuencias de pseudogenes del Cuadro I a excepción de la globina  $\theta 1$  del orangután. Se aprecia que las gráficas tienen el mismo patrón que en las globinas pero, a diferencia de éstas, en los pseudogenes, el enlace zx es el que se comporta de manera independiente a los valores promedio mientras que el enlace xy muestra las correlaciones y las pendientes mayores. En secuencias donde la entalpía promedio es alta, en el enlace xy la entalpía promedio es mayor así como el valor D3. En secuencias donde la entalpía promedio es baja, en el enlace xy hay valores menores en la entalpía y en D3. Como se ve en la figura en el intervalo  $D3 > 1$ , el aumento en los valores de D3 en el enlace xy se debe a una disminución del valor D1 y, en el intervalo  $D3 < 1$ , al disminuir D3 hay un incremento en los valores de D1 pero sobre todo en D2. Al igual que en el enlace zx de las globinas, en el enlace xy de los pseudogenes las mutaciones tienden a formar enlaces fuertes en dependencia del estado de metilación de la citosina. El que en los pseudogenes el enlace xy tenga un comportamiento similar al enlace

zx de secuencias que se expresan, apoya la idea de que los pseudogenes se originan de secuencias funcionales que han dejado de tener actividad. Ya que el enlace xy ha sido el más restringido en la historia de estas secuencias, se esperaría que, al ya no estar sujeto a restricciones selectivas, el enlace xy tuviera el componente de cambio más acentuado ya que los otros dos enlaces, especialmente el zx, han estado sujetos por más tiempo a las mutaciones.

En relación a los intrones, se espera que en estas secuencias, en virtud de que no se expresan, no exista ninguna forma de estructuración de la energía en periodos de tres enlaces. Se debe recordar que los extremos 5' y 3' de los intrones muestran secuencias de consenso relacionados con el sistema de edición del mRNA en el núcleo por lo que nuestra afirmación debe considerarse en forma muy general. La Figura 11 ilustra la distribución de los valores en  $\Delta H$ , D1, D2 y D3 en los enlaces xy, yz y zx definidos de acuerdo a la fase del exón precedente. Se ve que en este caso, no hay ninguna diferencia en la distribución de los valores de las variables mencionadas aunque debemos señalar que en el enlace xy las pendientes que se obtuvieron son ligeramente mayores y en el enlace zx son menores de forma consistente. Si consideramos lo expuesto en relación a los pseudogenes, podríamos pensar que los intrones tienen alguna relación de origen con secuencias que se expresan. Esto en cierta medida se opone a la idea de que los intrones se originaron como secuencias de intervención desde el origen mismo de las primeras células. Se ha postulado que la presencia de intrones en los primeros genomas era una forma de evitar la presencia de codones de terminación en una secuencia muy larga. (77,89)

Germinal Cocho (comunicación personal) ha desarrollado un método para analizar la periodocidad en el valor tres a lo largo de secuencias de DNA. Se ha observado que en zonas pequeñas de algunos intrones de eucariontes, existe una estructura energética con periodo tres. Esto sugiere que, por lo menos algunas partes de los intrones se originaron a partir de secuencias de tipo exónico. Esto no es incompatible con la idea de que los intrones se aparecieron desde el surgimiento de las primeras células.

Posiblemente, fenómenos como la transposición, la transfección viral y otras formas de recombinación pueden intercambiar segmentos de secuencias exónicas en los intrones.

## CONCLUSIONES

Se puede considerar que el proceso de la evolución molecular consiste de dos fases: La mutación y la fijación de este cambio en una población biológica determinada. La teoría de la evolución, ya sea en el marco del neutralismo o del seleccionismo, establece que el origen de las mutaciones tiene un comportamiento aleatorio con relación a la adecuación del organismo. Esto no debe interpretarse, como suele suceder, en el sentido de que las mutaciones ocurren por azar. La mutagénesis es un proceso físico y como tal está determinado por varios fenómenos que tienen un efecto específico en dicho proceso.

Esto significa que, en nuestro concepto, la evolución molecular depende por un lado de la situación biológica concreta en que ocurre una mutación y, por otra parte, de la naturaleza misma de esa mutación. Se plantea así el siguiente esquema:

$$P_i = f(B)_i \cdot f(F)_i$$

donde  $P_i$  es la probabilidad de que ocurra una sustitución dada en el sitio  $i$ ,  $f(B)_i$  define los factores biológicos que determinan la probabilidad de una cierta sustitución en  $i$ , y  $f(F)_i$  es el proceso físico de mutagénesis. Se ha visto que las mutaciones ocurren preferentemente en los sitios débiles de la molécula de DNA y que no todos los cambios son igualmente probables sino que, por ejemplo, ocurren más frecuentemente las transiciones que las transversiones. El  $f(B)_i$  no es igual para todos los sitios del genoma sino que hay lugares para los cuales existen fuertes restricciones biológicas, por ejemplo, las mutaciones tienen menos probabilidades de fijarse en los exones que en los intrones o en los pseudogenes. Dentro de los exones, también existe una restricción diferencial en las tres bases que conforman un codón.

Se plantea que en aquellos lugares del genoma en los que las restricciones biológicas son relativamente menos fuertes la secuencia de DNA revela más las características físicas de la mutagénesis. Este postulado generaliza la teoría neutralista según la cual la mayor parte de las variantes en una población son

neutrales de manera que la probabilidad de fijación de una mutación depende de la tasa de mutación. En los lugares del genoma menos restringidos biológicamente, la tasa de sustitución tiende a ser la tasa de mutación.

Se ha mencionado que esto puede explicar la distribución no aleatoria de los codones. Es claro que no se puede hacer una afirmación absoluta pues se ha demostrado en diversos sistemas la existencia de restricciones selectivas en diferentes sitios del genoma que no tienen que ver con la secuencia de aminoácidos de las proteínas, tales como la disponibilidad de tRNA, la estabilidad de los ácidos nucleicos y el proceso de síntesis de proteínas. Esto no contradice la idea de un proceso físico subyacente en la evolución molecular sino que distingue en  $f(B)_i$  dos factores según la concepción de Cocho (comunicación personal):

$$f(B)_i = f(Be)_i \cdot f(Bi)_i$$

donde  $f(Be)_i$  denota lo que Cocho denomina selección natural externa, esto es, la selección sobre las propiedades funcionales de las proteínas, y  $f(Bi)_i$  define la selección natural interna que se refiere a las restricciones biológicas sobre el genoma y los aspectos que le son propios tales como la interacción de los ácidos nucleicos en la síntesis de proteínas.

Dentro de  $f(B)_i$  podemos encontrar diversos factores físicos como, por ejemplo, las restricciones hidrodinámicas en la adaptación al medio acuático o la estabilidad del DNA. Estos factores no deben confundirse con la noción de un proceso físico subyacente en la mutagénesis.

ABREVIATURAS

DNA	Acido desoxirribonucleico.
RNA	Acido ribonucleico.
mRNA	Acido ribonucleico mensajero.
tRNA	Acido ribonucleico de transferencia.
A	Adenina.
G	Guanina
C	Citosina.
T	Timina.
5-mC	5-metil,citosina.
2-AP	2-amino,purina.
DDP	Diamino,dicloro,platino
EMS	Etil,metano,sulfonato.
MMS	Metil,metano,sulfonato.
CsCl	Cloruro de cesio.
X	Primera base del codón.
Y	Segunda base del codón.
Z	Tercera base del codón.
xy	Enlace entre las bases X e Y.
yz	Enlace entre las bases Y e Z.
zx	Enlace entre las bases Z y X.
$\Delta G^{\circ}$	Cambio de energia libre por el enlace entre dos pares de bases.
$\Delta H^{\circ}$	Entalpia de amarre entre dos pares de bases.
$\Delta S^{\circ}$	Cambio de entropia por la interacción entre dos pares de bases.
$\Delta H, \Delta H_t$	Entalpia promedio de una secuencia.
$\Delta H_{xy}$	Entalpia promedio en el enlace xy.
$\Delta H_{yz}$	Entalpia promedio en el enlace yz.
$\Delta H_{zx}$	Entalpia promedio en el enlace zx.
pb	Pares de bases.
Kb	Kilobase: mil pares de bases.
K	Grado Kelvin.
Kcal	Kilocaloria.

## REFERENCIAS

- 1.- Alvarez-Tostado J. M., Ruiz M. E. y Novaro O. (1986). Molecular aspects of antitumoral activity of Pt(NH<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub> complexes. *Acta sud Americana de Qu#mica*, 6: 33-46.
- 2.- Avery O. T., MacLeod C. M. y McCarty M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Expl. Med.*, 79: 137-158.
- 3.- Baralle F. E., Shoulders C. C. y Proudfoot N. J. (1980). The primary structure of the human ε-globin gene. *Cell*, 21: 621-626.
- 4.- Beadle G. W. y Tatum E. L. (1941). Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA*, 27: 499-506.
- 5.- Bentley D. L. y Rabbits T. H. (1980). Human immunoglobulin variable region genes-DNA sequences of two Vx genes and a pseudogene. *Nature*, 288: 730-733.
- 6.- Benzer S. (1961). On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. USA*, 47: 403-415.
- 7.- Bernardi G. y Bernardi G. (1985). Codon usage and genome composition. *J. Mol. Evol.*, 22: 363-365.
- 8.- Bernardi G. y Bernardi G. (1986). Compositional constraints and genome evolution. *J. Mol. Evol.*, 24: 1-11.
- 9.- Bernardi G., Olofsson B., Filipski J., Zerial M., Salinas J., Cuny G., Meunier-Rotival M. y Rodier F. (1985). The mosaic genome of warm-blooded vertebrates. *Science*, 228: 953-958.
- 10.- Bilofsky H.S., Burks C., Fickett J. W., Goad W. B., Lewitter F. I., Rindone W. P., Swindell C. D. y Tung Ch.-S. (1986). The GenBank genetic sequence databank. *Nucl. Acid. Res.*, 14: 1-4.
- 11.- Bird A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, 321: 209-213.
- 12.- Blanchetot A., Wilson V., Wood D. y Jeffreys A. L. (1983). The seal myoglobin gene: an unusually long globin gene. *Nature*, 301: 732-734.
- 13.- Bloomfield V. A., Crothers D. M. y Tinoco I. (1974). Physical chemistry of nucleic acids. Harper and Row, publishers, New York.
- 14.- Breslauer K. J., Frank R., Blöcker H. y Marky L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA*, 83: 3746-3750.

- 15.- Brisson N. y Verma D. P. S. (1982). Soybean leghemoglobin gene family: Normal, pseudo, and truncated genes. *Proc. Natl. Acad. Sci. USA*, 79: 4055-4059.
- 16.- Britten R. J. (1986). Rates of DNA sequence evolution differ between taxonomic groups. *Science*, 231: 1393-1398.
- 17.- Britten R. J. y Davidson E. H. (1969). Gene regulation for higher cells: a theory. *Science*, 165: 349-357.
- 18.- Cairns-Smith A. G. y Hartman H. (eds). (1986). Clay minerals and the origin of life. Cambridge University Press.
- 19.- Cleary M. L., Schon E. A. y Lingrel J. B. (1981). Two related pseudogenes are the result of a gene duplication in the goat  $\beta$ -globin locus. *Cell*, 26: 181-190.
- 20.- Cocho G., Rius J. L., Miramontes P. y Medrano L. (1988). Local energy density and DNA dynamics. En preparación.
- 21.- Coulondre C., Miller J. H., Farabaugh P. J. y Gilbert W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, 274: 775-780.
- 22.- Crick F. H. C. (1968). The origin of the genetic code. *J. Mol. Biol.*, 38: 367-379.
- 23.- Crick F. (1979). Split genes and RNA splicing. *Science*, 204: 264-271.
- 24.- Cheung S., Arndt K. y Lu P. (1984). Correlation of *lac* operator DNA imino proton exchange kinetics with its function. *Proc. Natl. Acad. Sci. USA*, 81: 3665-3669.
- 25.- Dickerson R. E. (1983). Base sequence and helix structure variation in B and A DNA. *J. Mol. Biol.*, 166: 419-441.
- 26.- Dobzhansky T., Ayala F. J., Stebbins G. L. y Valentine J. W. (1980). Evolución. Ediciones Omega, Barcelona.
- 27.- Dodgson J. B. y Engel J. D. (1983). The nucleotide sequence of the adult chicken  $\alpha$ -globin genes. *J. Biol. Chem.*, 258: 4623-4629.
- 28.- Dodgson J. B., Stadt S. J., Choi O-R., Dolan M., Fischer H. D. y Engel J. D. (1983). The nucleotide sequence of the embryonic chicken  $\beta$ -type globin genes. *J. Biol. Chem.*, 258: 12685-12692.
- 29.- Drake J. W. (1966). Spontaneous mutations accumulating in bacteriophage T4 in the complete absence of DNA replication. *Proc. Natl. Acad. Sci. USA*, 55: 738-742.
- 30.- Drake J. W. y Baltz R. H. (1976). The biochemistry of mutagenesis. *Ann. Rev. Biochem.*, 45: 11-37.



- 31.- Duncan B. K. y Miller J. H. (1980). Mutagenic deamination of cytosine residues in DNA. *Nature*, 287: 560-561.
- 32.- Eigen M. y Schuster P. (1979). The hypercycle. A principle of natural self-organization. Springer-Verlag. Berlin, Heidelberg, New York.
- 33.- Eldredge N. (1971). The allopatric model and phylogeny in paleozoic invertebrates. *Evolution*, 25: 156-167.
- 34.- Ellison J. y Hood L. (1982). Linkage and sequence homology of two human immunoglobulin  $\gamma$  heavy chain constant region genes. *Proc. Natl. Acad. Sci. USA*, 79: 1984-1988.
- 35.- Enver T. y Patient R. (1986). Importance of helical periodicity. *Nature*, 319: 99-100.
- 36.- Fiers W. y Grosjean H. (1979). On codon usage. *Nature*, 277: 328.
- 37.- Fischel-Ghodsian N. y Higgs D. R. (1987). Function of a new globin gene. *Nature*, 329: 397.
- 38.- Friedberg E. C. (1985). DNA repair. W. H. Freeman, San Francisco.
- 39.- Glickman B. W. y Ripley L. S. (1984). Structural intermediates of deletion mutagenesis: A role for palindromic DNA. *Proc. Natl. Acad. Sci. USA*, 81: 512-516.
- 40.- Goldschmidt R. B. (1948). Ecotype, ecospecies and macroevolution. *Experientia*, IV: 465-472.
- 41.- Goodman M., Moore W. y Matsuda G. (1975). Darwinian evolution in the genealogy of haemoglobin. *Nature*, 253: 603-608.
- 42.- Grantham R., Gautier C., Gouy M., Mercier P. y Pavé A. (1980). Codon catalog usage and the genome hypothesis. *Nucl. Acid. Res.*, 8: r49-r62.
- 43.- Grantham R., Gautier C., Gouy M., Jacobzone M. y Mercier P. (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucl. Acid. Res.*, 9: r43-r74.
- 44.- Grosjean H. y Fiers W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene*, 18: 199-209.
- 45.- Hardison R. C. y Margot J. B. (1984). Rabbit globin pseudogene  $\Psi\beta 2$  is a hybrid of  $\delta$ - and  $\beta$ -globin gene sequences. *Mol. Biol. Evol.*, 1(4): 302-316.

- 46.- Hasegawa M., Yasunaga T. y Miyata T. (1979). Secondary structure of MS2 phage RNA and bias in code word usage. *Nucl. Acid. Res.*, 7: 2073-2079.
- 47.- Hershey A. D. y Chase M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.*, 36: 39-56.
- 48.- Hieter P. A., Hollis G. F., Korsmeyer S. J., Waldmann T. A. y Leder P. (1981). Clustered arrangement of immunoglobulin  $\lambda$  constant region genes in man. *Nature*, 294: 536-540.
- 49.- Hieter P. A., Max E. E., Seidman J. G., Maizel J. V. Jr. y Leder P. (1980). Cloned human and mouse kappa immunoglobulin constant and J region genes conserve homology in functional segments. *Cell*, 22: 197-207.
- 50.- Hopfield J. J. (1974). Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. USA*, 71: 4135-4139.
- 51.- Hyldig-Nielsen J. J., Jensen E. Ø., Paludan K., Wiborg O., Garret R., Jørgensen P. y Marcker K. A. (1982). The primary structure of two leghemoglobins genes from soybean. *Nucl. Acid. Res.*, 10: 689-701.
- 52.- Ikemura T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.*, 151: 389-409.
- 53.- Jukes T. H. y Kimura M. (1984). Evolutionary constraints and the neutral theory. *J. Mol. Evol.*, 21: 90-92.
- 54.- Kimura M. (1968). Evolutionary rate at the molecular level. *Nature*, 217: 624-626.
- 55.- Kimura M. (1981). Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons. *Proc. Natl. Acad. Sci. USA*, 78: 5773-5777.
- 56.- Kimura M. (1986). DNA and the neutral theory. *Phil. Trans. R. Soc. Lond. B* 312: 343-354.
- 57.- Kimura M. y Ohta T. (1974). On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. USA*, 71: 2848-2852.
- 58.- King L. J. y Jukes T. H. (1969). Non-darwinian evolution. *Science*, 164: 788-798.

- 59.- Koespel R. R. y Khan S. A. (1986). Static and initiator protein-enhanced bending of DNA at a replication origin. *Science*, 233: 1316-1318.
- 60.- Lawn R. M., Efstratiadis A. y O' Conell C. (1980). The nucleotide sequence of the human  $\beta$ -globin. *Cell*, 21: 647-651.
- 61.- Lazcano A., Guerrero R., Margulis L. y Oró J. (1988). The evolutionary transition from RNA to DNA in early cells. *J. Mol. Evol.*, 27: en prensa.
- 62.- Li W.-H., Wu Ch.-I. y Luo Ch.-Ch. (1984). Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.*, 21: 58-71.
- 63.- Lindahl T. (1974). An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl. Acad. Sci. USA*, 71: 3649-3653.
- 64.- Lipman D. J. y Wilbur W. J. (1983). Contextual constraints on synonymous codon choice. *J. Mol. Biol.*, 163: 363-376.
- 65.- Litman G. W., Berger L., Murphy K., Litman R., Hinds K., Jahn C. L. y Erickson B. W. (1983). Complete nucleotide sequence of an immunoglobulin VH gene homologue from *Caiman*, a phylogenetically ancient reptile. *Nature*, 303: 349-352.
- 66.- Marks J., Shaw J-P. y James Shen C-K. (1986). Sequence organization and genomic complexity of primate  $\theta 1$  globin gene, a novel  $\alpha$ -globin-like gene. *Nature*, 321: 785-788.
- 67.- Martin F. H. y Tinoco I. Jr. (1980). DNA-RNA hybrid duplexes containing oligo (dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucl. Acid. Res.*, 8: 2295-2299.
- 68.- Matthyssens G. y Rabbitts T. H. (1980). Structure and multiplicity of genes for the human immunoglobulin heavy chain variable region. *Proc. Natl. Acad. Sci. USA*, 77: 6561-6565.
- 69.- Meunier-Rotival M., Soriano P., Cuny G., Strauss F. y Bernardi G. (1982). Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. *Proc. Natl. Acad. Sci. USA*, 79: 355-359.
- 70.- Michelson A. M. y Orkin S. H. (1980). The 3' untranslated regions of the duplicated human  $\alpha$ -globin genes are unexpectedly divergent. *Cell*, 22: 371-377.

- 71.- Miyata T. y Hayashida H. (1981). Extraordinarily high evolutionary rate of pseudogenes: Evidence for the presence of selective pressure against changes between synonymous codons. *Proc. Natl. Acad. Sci. USA*, 78: 5739-5743.
- 72.- Modiano G., Battistuzzi G. y Motulsky A. G. (1981). Nonrandom patterns of codon usage and of nucleotide substitutions in human  $\alpha$ - and  $\beta$ -globin genes: An evolutionary strategy reducing the rate of mutations with drastic effects? *Proc. Natl. Acad. Sci. USA*, 78: 1110-1114.
- 73.- Modrich P. (1987) DNA mismatch correction. *Ann. Rev. Biochem.*, 56: 435-466.
- 74.- Morgan T. H. (1935). The scientific basis of evolution. W. W. Norton and Company, Inc., publishers, New York.
- 75.- Muller H. J. (1927). Artificial transmutation of the gene. *Science*, 66: 84-87.
- 76.- Muzyczka N., Poland R. L. y Bessman M. J. (1972). Studies on the biochemical basis of spontaneous mutation. *J. Biol. Chem.*, 247: 7116-7122.
- 77.- Naora H., Miyahara K. y Curnow R. N. (1987). Origin of noncoding DNA sequences: molecular fossils of genome evolution. *Proc. Natl. Acad. Sci. USA*, 84: 6195-6199.
- 78.- Nishioka Y. y Leder P. (1979). The complete sequence of a chromosomal mouse  $\alpha$ -globin gene reveals elements conserved throughout vertebrate evolution. *Cell*, 18: 875-882.
- 79.- Nussinov R. (1980). Some rules in the ordering of nucleotides in the DNA. *Nucl. Acid. Res.*, 8: 4545-4562.
- 80.- Oparin A. I. (1970). Origen de la vida sobre la tierra. Editorial Tecnes, Madrid.
- 81.- Ponce T. M. (1984). Explicaciones teleológicas y explicaciones causales en Biología. En: Otero M. H. (comp). *Materialismo y ciencias naturales*. UNAM: 119-137.
- 82.- Proudfoot N. J., Gil A. y Maniatis T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell*, 31: 553-563.
- 83.- Radman M. y Wagner R. (1986). Mismatch repair in *Escherichia coli*. *Ann. Rev. Genet.*, 20: 523-538.
- 84.- Ravetch J. V., Siebenlist U., Korsmeyer S., Waldmann T. y Leder P. (1981). Structure of the human immunoglobulin  $\mu$  locus: Characterization of embryonic and rearranged J and D genes. *Cell*, 27: 583-591.

85. - Reaney D. C. y Pressing J. (1984). Temperature as a determinative factor in the evolution of genetic systems. *J. Mol. Evol.*, 21: 72-75.
86. - Scott A. F., Heath P., Trusko S., Boyer S. H., Prass W., Goodman M., Czelusniak J., Edward Chang L.-Y. y Slightom J. L. (1984). The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution. *Mol. Biol. Evol.*, 1(5): 371-389.
87. - Schaaper R. M., Danforth B. N. y Glickman B. W. (1986). Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. *J. Mol. Biol.*, 189: 273-284.
88. - Schaaper R. M. y Dunn R. L. (1987). Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: The nature of *in vivo* DNA replication errors. *Proc. Natl. Acad. Sci. USA*, 84: 6220-6224.
89. - Senapathy P. (1986). Origin of eukaryotic introns: a hypothesis based on codon distribution statistics in genes, and its implications. *Proc. Natl. Acad. Sci. USA*, 83: 2133-2137.
90. - Sheperd J. C. (1981). Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. *Proc. Natl. Acad. Sci. USA*, 78: 1596-1600.
91. - Sheperd J. C. W. (1984). Fossil remnants of a primeval genetic code in all forms of life? *TIBS*, 9: 8-10.
92. - Siebenlist V., Ravetch J. V., Korsmeyer S., Waldmann T. y Leder P. (1981). Human immunoglobulin D segments encoded in tandem multigenic families. *Nature*, 294: 631-635.
93. - Slightom J. L., Blechl A. E. y Smithies O. (1980). Human fetal  $\gamma$ - and  $\alpha$ -globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell*, 21: 627-638.
94. - Smithies O., Blechl A. E., Sixiang S., Slightom J. L. y Vanin E. O. (1981). Co-evolution and control of globin genes. En: Organization and expression of globin genes. Stamatoyannopoulos G. y Nienhuis A. W. (eds). Alan R. Liss Inc., New York.
95. - Stanley S. M. (1975). A theory of evolution above the species level. *Proc. Natl. Acad. Sci. USA*, 72: 646-650.

- 97.- Trifonov E. N. y Sussman J. L. (1980). The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. USA*, 77: 3816-3820.
- 98.- Van Ooyen A., Van Den Berg J., Mantei N. y Weissmann C. (1979). Comparison of total sequence of a cloned rabbit  $\beta$ -globin gene and its flanking regions with a homologous mouse sequence. *Science*, 206: 337-344.
- 99.- Vanyushin B. F., Tkacheva S. G. y Belozersky A. N. (1970). Rare bases in animal DNA. *Nature*, 225: 948-949.
- 100.- Volkenshtein M. V. (1981). *Biofísica*. Editorial Mir, Moscú.
- 101.- Wabl M., Burrows P. D., Von Gabain A. y Steinberg Ch. (1985). Hypermutation at the immunoglobulin heavy chain locus in a pre-B-cell line. *Proc. Natl. Acad. Sci. USA*, 82: 479-482.
- 102.- Wagner R. Jr. y Meselson M. (1976). Repair tracts in mismatched DNA heteroduplexes. *Proc. Natl. Acad. Sci. USA*, 73: 4135-4139.
- 103.- Watson J. D. y Crick F. H. C. (1953). Molecular structure of nucleic acids. *Nature*, 171: 737-738.
- 104.- Watson J. D. y Crick F. H. C. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171: 964-967.
- 105.- Williams J. G., Kay R. M. y Patient R. K. (1980). The nucleotide sequence of the major  $\beta$ -globin mRNA from *Xenopus laevis*. *Nucl. Acid. Res.*, 8: 4247-4256.
- 106.- Wilson A. C., Bush G. L., Case S. M. y King M.-C. (1975). Social structuring of mammalian populations and rate of chromosomal evolution. *Proc. Natl. Acad. Sci. USA*, 72: 5061-5065.
- 107.- Wilson A. C., Sarich V. M. y Maxson L. P. (1974). The importance of gene rearrangement in evolution: Evidence from studies on rates of chromosomal, protein, and anatomical evolution. *Proc. Natl. Acad. Sci. USA*, 71: 3026-3030.
- 108.- Wilson A. C., Steven S. C. y White T. J. (1977). Biochemical evolution. *Ann. Rev. Biochem.* 46: 573-639.