

152  
205



# Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

"ESTUDIO DEL GENOMA DEL VIRUS DEL DENGUE:  
ANÁLISIS DE LA ESTRUCTURA PRIMARIA DE LAS  
PROTEÍNAS DE NUCLEOCÁPSIDE Y DE MEMBRANA  
DEL VIRUS"

*TESIS PROFESIONAL*

Que para obtener el Título de  
B I O L O G O  
p r e s e n t a  
GUSTAVO JESUS ORTEGA LULE



México, D. F.

1991

FALLA DE ORIGEN



## **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

Introducción.....	1
Planteamiento del problema.....	9
Objetivos.....	10
Material y Métodos.....	11
a) Identificación de clonas positivas.....	11
b) Purificación de plásmidos y análisis del tamaño de los insertos.....	13
c) Secuenciación nucleotídica.....	15
I Purificación de ADN de cadena sencilla.....	15
II Reacciones de secuencia.....	17
ch) Amplificación del ADNc por reacción en cadena de la polimerasa (PCR).....	19
d) Predicción de la estructura secundaria de las proteínas.....	23
Resultados.....	31
a) Análisis de las clonas positivas.....	31
b) Secuenciación nucleotídica de los insertos.....	31
c) Síntesis y amplificación de regiones no representadas en el banco genómico.....	32
ch) Análisis de estructura primaria.....	32
d) Predicción de la estructura secundaria.....	37
Discusión.....	41
Conclusión.....	45
Lista de figuras.....	46
Bibliografía.....	49

## INTRODUCCION

El dengue es una enfermedad causada por el virus del mismo nombre, el cual pertenece a la familia Flaviviridae (1), que está compuesta de aproximadamente 66 especies de virus divididos en tres grupos dependiendo de su forma de transmisión (mosquitos, garrapatas o sin vector conocido) y 8 subgrupos definidos antígenicamente (tabla 1) (2,49). El virus del dengue es transmitido principalmente por el mosquito Aedes aegypti, y con menos eficiencia por otros miembros del Suborden Nematocera de la familia Culucidae, e infecta preferencialmente a las células del sistema reticuloendotelial.

DIVISION DE LOS FLAVIVIRUS: POR VECTOR Y COMPLEJO ANTIGENICO		
VECTOR		
GARRAPATA	MOSQUITO	SIN VECTOR
COMPLEJO ANTIGENICO		
ENCEFALITIS TRANSMITIDA POR GARRAPATA.  TYULENIY.	ENCEFALITIS JAPONE- SA.  NTAYA  UGANDA S  DENGUE	RIO BRAVO  MODOC

[TABLA 1] Complejos antígenicos de los Flavivirus definidos por ensayos de neutralización cruzada con anti sueros policlonales

La enfermedad se consideró durante muchos años limitada a regiones de Africa, Asia y Australia. Sin embargo, durante la segunda guerra mundial se aislaron los serotipos 1 y 2 en el

Continente Americano (2,3). En 1957 debido a la existencia de fiebre amarilla en América, la Organización Panamericana de la Salud auspició una campaña de erradicación del Aedes y es así que el 20 de septiembre de 1963 se declaró erradicado el mosquito (3,4), por lo que en nuestro país no se tuvo notificación de dengue de 1963 a 1978, en este último año reaparece la enfermedad en el Sur de México. En 1980 alcanza su morbilidad máxima (51,506 casos) ocupando el sexto lugar dentro de las enfermedades notificables (figura 1). En los años siguientes invade por la costa del Golfo hasta el norte del país afectando 23 estados (figura 2) (4,5).

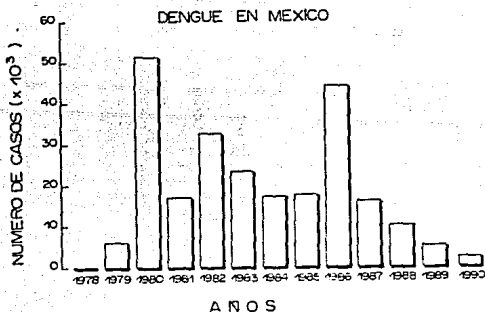


Fig 1 Número de casos de dengue en México a partir de 1978.

Actualmente circulan 3 de los 4 serotipos y las principales áreas endémicas son: Guerrero, Oaxaca, Chiapas, Yucatán, Puebla y Veracruz. En la población mundial se ha estimado que ocurren arriba de 100 millones de casos anuales (5,6,7).

## DENGUE EN LAS AMERICAS



Fig 2 Mapa de la distribución de los distintos serotipos virus del dengue en las Americas.

La enfermedad tiene dos formas clínicas: el dengue clásico, que generalmente se presenta de manera benigna y el dengue hemorrágico que tiene dos variantes; el síndrome de choque por dengue (SCD) y la fiebre hemorrágica por dengue (FHD), con una letalidad del 10 al 40 % de los casos.

El dengue nos preocupa en México porque ha causado grandes epidemias que han incapacitado temporalmente a miles de individuos y, aunque sólo tenemos notificación en forma aislada de síndromes graves, existe la amenaza de que se presenten epidemias y pueda volverse endémico, con altas tasas de mortalidad, ya que existen todas las condiciones epidemiológicas para el desarrollo de este tipo de manifestaciones como son: una población susceptible (primoinfectada), densidad adecuada del vector, tres de los cuatro serotipos circulando, clima propicio para el desarrollo del mosquito, etc. Los datos epidemiológicos muestran que el virus afecta principalmente a personas jóvenes (menores de 15 años de edad) y mujeres (16,14).

Actualmente se proponen dos hipótesis para tratar de explicar el desarrollo de las formas graves de la enfermedad. La primera propone la variación antigénica (de la proteína de envoltura viral) de los virus mientras circulan en la naturaleza (15). Y la segunda una facilitación de la infección mediada por anticuerpos heterotípicos no neutralizantes, esto es, una infección secuencial por distintos serotipos del virus del dengue (14). Esta última hipótesis es la más aceptada ya que se ha demostrado que durante las distintas epidemias ocurridas tanto en países Asiáticos como del Caribe (16) con FHD y SCD, los pacientes habían



Fig 4 Microfotografía del virus del dengue, la barra blanca mide aproximadamente 100 nm.

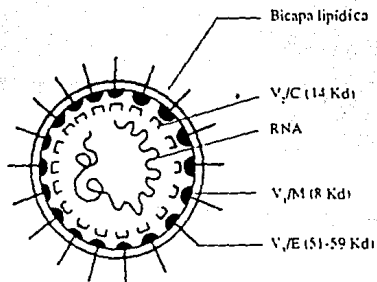


Fig 5 Modelo hipotético de la organización de los componentes del virus del dengue: lípidos, proteínas virales estructurales y ARN.



estado previamente infectados con un serotipo distinto al que ocasionó el brote epidémico (figura 3).

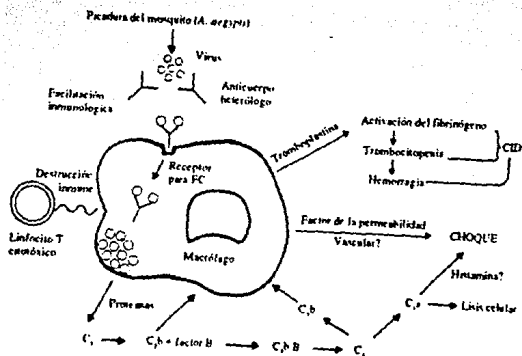


Fig 3 Representación esquemática del fenómeno de facilitación por anticuerpos y de los mecanismos efectorios y de amplificación que participan en la patogénesis de la FHD/SCD.

La variación antigénica y biológica en flavivirus ha sido demostrada en Encefalitis Japonesa (50), Fiebre amarilla (51) y Encefalitis de San Luis (52).

El virus del dengue es de aproximadamente 50 nm de diámetro (figura 4), y al igual que los demás flavivirus está compuesto de 6% de ARN monocatenario de sentido positivo, 66% de proteínas, 17% de lípidos y 9% de carbohidratos (5,6) (figura 5). El ARN contiene 10 712 bases (7,57,22), presenta una estructura de m<sup>7</sup>GppAmp en el extremo 5' (Cap tipo I)(9) y carece de una secuencia de poli A en el extremo 3' (8). Entre los distintos flavivirus en la región no traducible del ARN tanto del extremo 5' como 3', existen estructuras secundarias conservadas



dando a la interacción ácido nucleico-proteína (7). Este péptido no forma anticuerpos (Ac) con actividad neutralizante (18), y posee una región de aminoácidos (aa) muy conservada que va del residuo 38 al 55 (13), ó 42 a 59 según diferentes autores, la cual está probablemente asociada a la interacción proteína-proteína, proteína-ARN o ambas. En esta región predominan aa apolares. En algunos flavivirus este péptido carece de la metionina inicial, en dengue no se ha comprobado pero se piensa que la falta de este aa es común en el grupo (13,18). En el extremo carboxilo terminal de la proteína se encuentra una región hidrofóbica de aproximadamente 14 aa, la cual ha sido propuesta como secuencia señal para la translocación de la segunda proteína, la precursora de membrana o prM, al retículo endoplasmático (13,17,20,21,22). Sabemos que esta secuencia y la presencia de microsomas son necesarios tanto para la traducción de la región que codifica para las proteínas estructurales, como para la glucosilación. Durante el procesamiento postranscripcional no se requiere de la región que codifica a las proteínas no estructurales (23).

En el extremo 5', de acuerdo al orden en el genoma, después de la proteína de nucleocápside se encuentra la región que codifica para el precursor de membrana o prM, el cual tiene un peso molecular de aproximadamente 11 400 daltones, que es un péptido glucosilado (7), cuyo procesamiento enzimático da lugar a la proteína de membrana o M. Esta proteína tiene un peso molecular de 7 000 a 8 000 daltones (13), y está asociada a la bicapa lipídica. Su extremo carboxilo terminal, al igual que la proteína C, posee una secuencia hidrofóbica que sirve de señal para la

translocación de la segunda glucoproteína (proteína E) viral (13,25). Se ha propuesto que el corte de prM a M esté dado por una proteasa del Aparato de Golgi (7). Sin embargo se ha encontrado una región conservada de seis aa que incluye la secuencia cisteína-triptofano relacionada con el sitio activo de las tiol-proteasas, en base a lo cual, se ha postulado que pueda presentar actividad tanto autocatalítica como proteolítica en algunas proteínas no estructurales (13,19).

Por su capacidad de inducir la formación Ac neutralizantes se propone que una región de la proteína M interactúe con el exterior del virus (24). Asimismo, ha sido encontrada principalmente en células infectadas y en menor proporción en el virión maduro (13,26). Se ha observado que el corte de prM a M se da en condiciones ácidas, semejantes a las que existen en el Aparato de Golgi durante el procesamiento postraducciona (26).

La adsorción de los viriones a las células permisivas del sistema retículo endotelial se lleva a cabo por la interacción de la glucoproteína de capa externa o E con receptores para los componentes del sistema inmune, como son los receptores para Fc de inmunoglobulina (complejo virus-anticuerpo) o receptores para algunos componentes del sistema del complemento. Una vez unido el virión, o el complejo virus-anticuerpo, al receptor, se lleva a cabo la penetración y el desnudamiento viral, eventos que parecen ser simultáneos y dependientes de pH (18). Los virus que contienen la proteína prM presentan una disminución de la capacidad infectiva (13,26), hecho que ha sido relacionado con la desnudación del virión durante la penetración a la célula huésped (26).

La proteína E es un péptido glucosilado transmembranal que vista al microscopio electrónico, se observa como espículas de 2-5 nm (figuras 4), es de aproximadamente 60 000 daltones, contiene un 7 % de azúcares constituidos principalmente por: manosa, glucosamina, galactosa y fucosa. Antigénicamente es la proteína más importante, ya que en ella residen los epítopes específicos para grupo (flavivirus), subgrupo o complejo (dengue), tipo y/o subtipo (dengue 1,2,3,4), determinados serológicamente tanto por neutralización como por inhibición de la hemaglutinación. Aunque en esta proteína residen las principales actividades biológicas del virus como son: unión a receptores, neutralización, hemaglutinación, fijación del complemento, facilitación de la infección, neurotropismo, etc., aún no esta bien caracterizada a nivel biológico y bioquímico.

## PLANTEAMIENTO DEL PROBLEMA

En base a los antecedentes anteriormente expuestos y con objeto de caracterizar un aislado de serotipo 2 del virus del dengue que ocasionó problemas epidemiológicos en nuestro país, se llevó a cabo el análisis de la región del genoma que codifica para las proteínas estructurales.

El interés particular de caracterizar esta cepa, es porque parece ser una variante antigénica de este serotipo (comunicación personal Dr. Duane Gubler, Center for disease control). La observación anterior esta basada en el hecho de que utilizando anticuerpos monoclonales tipo-especificos la cepa fue claramente identificada como serotipo 2; sin embargo, utilizando anticuerpos provenientes de líquido ascítico hiperinmune, dirigidos contra la cepa prototipo de referencia para este serotipo (Nueva Guinea C) no neutralizaban, o lo hacían muy débilmente.

En este trabajo se presenta el análisis de la estructura primaria y secundaria para las proteínas de nucleocápside y membrana del virus del dengue serotipo 2 genotipo mexicano (dengue-2 mexicano).

## OBJETIVOS

1. Determinar la estructura primaria de las proteínas de nucleocápside (C) y de membrana (M) del virus del dengue, serotipo 2 genotipo mexicano (D2-M), a partir de la secuencia nucleotídica
2. Analizar la secuencia deducida de aminoácidos de las proteínas estructurales de este virus.
3. Analizar y predecir la estructura secundaria de las mismas por métodos estadísticos y fisicoquímicos

## MATERIAL Y METODOS

### -IDENTIFICACION DE CLONAS POSITIVAS

La genoteca de ADN complementario (ADNc) al ARN del virus del dengue, previamente establecida en el laboratorio (63), fue rastreada con objeto de obtener clonas que contuvieran secuencias homólogas a las proteínas de membrana y nucleocápside viral. La identificación de las mismas se hizo por hibridación en colonia utilizando una sonda específica para la región estructural del dengue 2 S1 (C 38) la cual esta clonada en el plásmido pGEM1 en el sitio para Eco RI (figura 7). Esta sonda que fue donada por el Dr Young S. Hahn del Tecnológico de California, Pasadena, CA, contiene la región del nucleótido 16 al 2103 que abarca la información para las proteínas C, prM, M, y E.

#### OBTENCION DE LA SONDA PARA Gp60

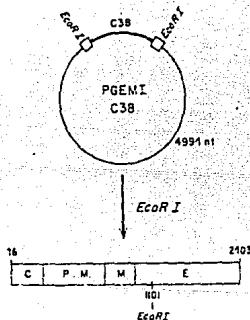


Fig. 7 La clona C38 fue generada por ligación en el sitio Hind III de la clona C3 y clona 8 de dengue 2S1. Contiene de los nucleótidos 16 al 2103 que representan casi en su totalidad la región estructural. Esta clona esta insertada en el sitio de EcoRI del plásmido pGEM1. Donada por el Dr Young S. Hahn de CALTECH, California.



El escrutinio se hizo por hibridación en colonia al 70 % de homología; se sembraron 500 clonas tomadas al azar sobre filtros de nitrocelulosa colocados en medio sólido Luria-Ampicilina (LA) contenido en cajas de Petri, y se dejaron crecer 2 mm aproximadamente (12 h) a 37 °C. Las bacterias fueron lisadas con NaOH 0.5 M y NaCl 1.5 M, la solución se utilizó dos veces durante 5 minutos (min), posteriormente se fijó el ADN a 80 °C con vacío durante 2 h. Se lavaron los detritus celulares con una solución SSPE 2X (NaCl 0.36 M, NaPO<sub>4</sub> 20 mM, Na<sub>2</sub> EDTA 2 mM pH 7.2) durante 5 min 2 veces.

Los filtros se prehibridizaron a 42 °C por 3 h en una bolsa sellada con 8 ml de solución prehibridizadora [amortiguador de citratos 9 mM pH 7.4, NaCl 0.9 M, SDS 0.1 %, ADN desnaturalizado proveniente de esperma de salmón 100 µg/ml, ficoll 0.1 %, albúmina serica bovina (ASB) 0.1 %, polivinilpirrolidona (PVP)]. Transcurrido este tiempo se cortó la esquina de la bolsa, se extrajo la solución prehibridizadora con una pipeta Pasteur, y se agregó la misma solución con 200 ng de la sonda previamente desnaturalizada por ebullición durante 10 min y marcada radiactivamente por la técnica de traducción de cortes (Nick translation), con las enzimas DNasa I y el fragmento mayor de la ADN polimerasa I (Klenow). La reacción se llevo a cabo en presencia de Tris 100 mM, NaCl 50 mM, MgCl<sub>2</sub> 10 mM, ditrioteitol (DTT) 1 mM, gelatina 0.1 mg/ml, dCTP, dGTP y dTTP 20 µM de cada uno y 15 µCi de <sup>32</sup>P-dATP (3000 Ci/mMol).

La reacción de marcado se hizo a 14 °C por 2 h, y se detuvo a 70 °C durante 10 min. La marca libre se eliminó por cromatografía en sephadex G-50. Los filtros con la marca se incubaron a 42 °C

durante un tiempo equivalente a  $3 C_{0.5}$  [que determina el tiempo medio de renaturalización del ADN =  $(1/X) (Y/5) (Z/10)$ , (donde X es la concentración en  $\mu\text{g}$  de la sonda, Y es la complejidad en kilobases (kb) y Z es el volumen de la reacción en ml)], posteriormente se lavaron en SSPE 2X una vez y cuatro veces con SSPE 0.2X durante 15 min cada una, finalmente los filtros fueron expuestos a una película Kodak (X-Omat) toda la noche a  $-70\text{ }^{\circ}\text{C}$ , al día siguiente se reveló la película.

#### -PURIFICACION DE PLASMIDOS Y ANALISIS DEL TAMAÑO DE LOS INSERTOS

La purificación de los plásmidos se hizo a partir de las clonas que dieron señal positiva para las proteínas estructurales. Con este propósito se sembró un inóculo de cada colonia seleccionada en medio LA, y se dejó crecer en agitación continua rápida a  $37\text{ }^{\circ}\text{C}$  hasta alcanzar la fase estacionaria; a continuación se inocularon 500 ml de medio LA con 5 ml de cultivo y se incubaron a  $37\text{ }^{\circ}\text{C}$ , permitiendo su crecimiento hasta obtener una densidad optica (DO) de 0.5 a una longitud de onda de 500 nm. Posteriormente se agregó el cloranfenicol ( $150\text{ }\mu\text{g/ml}$ ) para favorecer la multiplicación del plásmido y se dejó en agitación por 18 h a  $37\text{ }^{\circ}\text{C}$ .

Las células se separaron por centrifugación a 2 500 g por 20 min, retirándose el sobrenadante y resuspendiendo el paquete celular en 30 ml de solución TES (Tris 50 mM pH 7.5, EDTA 80 mM y sacarosa al 10 %) con 10 mg/ml de lisozima, incubándose a temperatura ambiente 10 min.

La lisis celular se llevó a cabo mediante ciclos de congelación (etanol y hielo seco) y descongelación sucesiva, por tres

veces, y posteriormente fueron centrifugadas a 120 000 g durante 45 min. A continuación se llevó a cabo una precipitación diferencial a partir de los sobrenadantes con medio volumen de cloruro de litio al 20 %, y un volumen de isopropanol al 100 %, congelándose a -70 °C y descongelándose suavemente.

Posteriormente la muestra se centrifugó a 15 000 g por 10 min y se recuperó el precipitado resuspendiéndose en 500 µl de solución TE (TrisHCl 10 mM pH 7.5 y EDTA 1 mM), con 30 µl de RNasa A 1 mg/ml (previamente hervida), a 37 °C por 1 h. El material insoluble se precipitó por centrifugación a 17 000 g 5 min. Los plásmidos fueron recuperados a partir del sobrenadante por extracción con solventes orgánicos (fenol/cloroformo) y precipitación con etanol absoluto más acetato de sodio 3 M pH 5.

Para analizar el tamaño de los insertos, se hizo una doble digestión (Eco RI/Hind III) de los plásmidos (10 µg) que tenían insertos en el sitio para Sma I y con Eco RI (10 µg) para los que estaban clonados en el adaptador de Eco RI.

Los insertos se separaron del vector por electroforesis en agarosa al 0.8 %, utilizando voltaje constante (100 volts). En el gel se incluyó como marcador de peso molecular al bacteriófago  $\phi$ X 174 digerido con Hae III. En el amortiguador de corrida TBE 1X se agregó bromuro de etidio a una concentración de 0.5 µg/100 ml. Al finalizar la corrida las bandas se observaron con luz ultravioleta de onda larga y se cortaron con un bisturí. Los fragmentos de agarosa se colocaron en bolsa de diálisis (previamente hervida durante 10 min con bicarbonato de sodio al 2 % y 1 mM de EDTA, lavada con agua destilada y hervida en agua megapura) con 200 µl

de TBE (Tris-Acido Bórico y EDTA) 0.1 X y se extrajo el ADN por electroelución a 150 volts durante 30 min, invirtiendo por 40 segundos los polos para despegar el ADN adherido al papel de la bolsa de diálisis. Finalmente la muestra se colocó en un tubo eppendorf y se extrajo el ADN con fenol/cloroformo. Para retirar los restos de bromuro de etidio, presentes en la preparación antes de precipitarse se hizo una extracción con isopropanol saturado con TNE (Tris 100 mM, NaCl 5 M, EDTA 1 mM).

#### - SECUENCIACION NUCLEOTIDICA

La secuenciación del ADN se llevó a cabo por el método de la cadena terminal, descrito por Sanger (27) y modificado por Tabor y Richardson (28). El método está basado en la síntesis in vitro de una cadena de ADN con ayuda de la ADN polimerasa del bacteriófago T7 la cual tiene baja actividad de exonucleasa 3'→5', utilizando como molde ADN de cadena sencilla (ADNcs). Cuando se usa una mezcla apropiada de 5' trifosfatos de desoxinucleótidos (dNTPs), incluyendo uno marcado radioactivamente y uno de los cuatro trifosfatos de dideoxinucleótidos (ddNTPs), la enzima cataliza la polimerización de la cadena hasta que un ddNTP análogo es incorporado. En reacciones individuales para cada 2'-3' ddNTP se completa la información de la secuencia y las cadenas marcadas de diferentes tamaños se resuelven por electroforesis en gel de poliacrilamida, revelándose por autoradiografía. La secuencia se obtiene por lectura directa de la película.

#### -A. PURIFICACION DE ADN DE CADENA SENCILLA.

Los insertos a secuenciar se subclonaron en el bacteriófago

M13 mp 18 en el sitio de Sma I. Con este propósito se alinearon los extremos protruyentes con ayuda del fragmento mayor de la ADN polimerasa I (Klenow, 12 U) en presencia de Tris HCl 10 mM, pH 7.5, MgCl<sub>2</sub> 40 mM, DTT 1.5 mM, dCTP, dGTP, dTTP, dATP 35 mM de cada uno, la reacción se incubó durante 60 min a temperatura ambiente. Posteriormente el ADNc se extrajo con solventes orgánicos [fenol (25)/cloroformo-isoamílico(24:1) v/v] y se precipitó con 2  $\frac{1}{2}$  volúmenes de etanol y 1/10 de volumen de NaCl 2M. Estos insertos se ligaron con ADN ligasa del bacteriófago T4, los fagos recombinantes se transfectaron en la cepa JM101 de E. coli [F'tra D36 Pro A<sup>+</sup> pro B lac I<sup>q</sup> lac 2 AM 151 Sup E Thi (lac-pro AB)] que se hizo competente para su transformación, de acuerdo a la siguiente técnica: Un volumen de medio Luria (Tryptona 1 %, extracto de levadura 0.5 %, NaCl 0.5 %) se inoculó con 1/10 del volumen de cultivo bacteriano crecido hasta su fase estacionaria. El cultivo se mantuvo a 37 °C con agitación media, hasta que alcanzó una DO de 0.5 a 550 nm. Se centrifugó y resuspendió en la mitad del volumen original con una solución de CaCl<sub>2</sub> 100 mM.

Posteriormente se incubaron 15 min a 4 °C, se recuperaron por centrifugación a 3 000 g durante 5 min, y se resuspendieron en 1/15 del volumen original. Por cada 50-100 ng de ADN recombinante (ADNr) se agregaron 300  $\mu$ l de células competentes y se incubaron 40 min en hielo/agua (0 °C) dando un choque térmico a 42 °C por 2 min y a 0 °C durante 5 min. A cada tubo se añadieron 20  $\mu$ l de isopropil  $\beta$ -D Tiogalactopiranosido (IPTG) 100 mM, 50  $\mu$ l de 5-Bromo-4 cloro 3 indolil  $\beta$ -D galactopiranosido (X-Gal) al 2 % en dimetilformamida (DMF), 200  $\mu$ l de células JM101 en crecimiento exponencial y 3 ml de agar en medio mínimo esencial. La mezcla se

homogeneizó y se vertió en cajas de Petri con agar en medio mínimo. Se permitió que el gel solidificara y a continuación se incubaron a 37 °C hasta observar el desarrollo de placas blancas de crecimiento lento (12-18 h). Estas placas se "picaron" con la punta de un palillo estéril, se sembraron en 3 ml de bacterias JM101 previamente diluidas 1:100 en medio fresco y se incubaron a 37 °C por 6 h. Posteriormente se centrifugaron en tubos eppendorf y se recuperó el sobrenadante guardándose a 4 °C toda la noche.

Al día siguiente se añadieron 220 µl de polietilenglicol (PEG) al 20 % en NaCl 2.5 M, dejándose 30 min a temperatura ambiente. A continuación las muestras se centrifugaron y se resuspendieron en 100 µl de TE (Tris 10 mM, EDTA 0.1 mM). Finalmente el ADN se extrajo con solventes orgánicos (fenol/cloroformo) y se precipitó con etanol y NaCl 2 M. El ADNcs se resuspendió en agua megapura estéril y se observó en minigel de agarosa al 1 %.

-B. REACCIONES DE SECUENCIA; ALINEACION DEL INICIADOR (PRIMER)  
CON EL MOLDE, REACCION DE MARCAJE, ELONGACION DE LA CADENA Y  
TERMINACION DE LA SINTESIS.

La alineación del iniciador de M13 con el molde (ADNcs) se llevó a cabo en un solo tubo de reacción utilizando 0.5 pmoles de iniciador universal (5' GTAAAACGACGGCCAGT 3'), 200 mM Tris HCl pH 7.5, 100 mM MgCl<sub>2</sub>, 250 mM NaCl y 2 µg de ADN en un volumen de 10 µl, la mezcla se incubó en un multibloque a 65 °C por 2 min y se dejó que la temperatura descendiera lentamente hasta 35 °C (en no

menos de 30 min). Después de que la temperatura fue menor a 35 °C se procedió con la reacción de marcado. A la reacción anterior (10 µl) se le agregó una mezcla de 7.5 µM de dGTP, dCTP y dTTP, 19 µCi/µl [ $\alpha^{35}\text{S}$ -dATP] 1 000-1 500 Ci/mmol, 0.1 M de DTT y 10 U de ADN polimerasa del fago T7 en un volumen de 15.5 µl, y se incubó a temperatura ambiente durante 5 min.

Para la terminación de la reacción, se transfirieron 3.5 µl de esta mezcla a 4 tubos eppendorf marcados para cada nucleótido A, T, G y C. A cada uno de los tubos se le agregó la mezcla de terminación con cada uno de los análogos de la manera siguiente:

A; dGTP, dATP, dCTP y dTTP 80 µM de cada uno, 80 µM de ddATP, NaCl 50 mM.

C; dGTP, dATP, dCTP y dTTP 80 µM de cada uno, 80 µM de ddCTP, NaCl 50 mM.

G; dGTP, dATP, dCTP y dTTP 80 µM de cada uno, 80 µM de ddGTP, NaCl 50 mM.

T; dGTP, dATP, dCTP y dTTP 80 µM de cada uno, 80 µM de ddTTP, NaCl 50 mM.

Se incubó la reacción a 38 °C por 5 min y se agregaron 4 µl de solución para detener la reacción, que contenía 95 % de formamida, 20 mM de EDTA, 0.05 % de azul de bromofenol, y 0.05 % de xilencianol FF. Los fragmentos del ADN obtenidos se separaron en un gel (de 80 cm de largo y 0.4 mm de espesor) al 5 % de acrilamida, bisacrilamida (19:1) y urea 7 M (el vidrio chico previamente se trató con 8-metacriloxipropiltrimetoxilano al 2 % en etanol, y el grande se siliconizó con sigmacote).

El gel se recorrió una hora en amortiguador de TBE pH 8.3 a 1 600 volts y 40-50 watts, hasta que el gel alcanzó una temperatura entre 50 y 55 °C. Las muestras se desnaturalizaron a 80 °C durante 2 min inmediatamente antes de cargarlas en el gel. En cada corrida se permitió que el xilencianol presente en la muestra migrara 80 cm, o llegara al frente de la corrida. Al término de la misma se fijaron en solución de ácido acético al 10 % con metanol al 12 % durante 1 h, posteriormente se horneó a 80 °C toda la noche y al día siguiente se sometió a autorradiografía con película Kodak X-Omat a temperatura ambiente durante 18 h

- AMPLIFICACION DEL ADNc POR REACCION EN CADENA DE LA POLIMERASA (PCR).

Las secuencias que no estaban representadas en el banco de ADNc y de los genes en estudio, ya fuera del extremo 5' o 3' se sintetizaron directamente a partir de ARNm con oligonucleótidos sintéticos específicos de diferentes tamaños (no menos de 12 nt), utilizando transcriptasa reversa (AMVRT). Para la síntesis de la primera cadena se utilizaron 3 µg de ARN monocatenario en un volumen de reacción de 25 µl en solución de Tris-HCl 100 mM pH 8, ditiotreitól (DTT) 4mM, 30 U de inhibidor de ribonucleasas de placenta humana (HPRI), una mezcla 20 mM de cada uno de los 4 desoxiribonucleótidos trifosfatos, 45 ng de hexanucleótidos sintetizados al azar (Amersham) como indicadores, 10 µCi de [32P]-dATP (3 000 Ci/mM) y 40 U de transcriptasa reversa del virus que ocasiona la mieloblastosis aviaria (AMVRT).

La reacción se incubó a 42 °C durante 60 min, se pasó a hielo/agua (0 °C) y se tomaron alícuotas de 1 µl para determinar



la cantidad de ADN de cadena sencilla (ADNccs) polimerizada. Con este propósito se fijaron estas muestras y las que se tomaron al tiempo 0 de reacción en filtros de papel Whatman (DEAE-81), los cuales se lavaron de marca no incorporada con  $\text{Na}_2\text{HPO}_4$  0.5 M, 6 veces durante 5 min, con agua 2 veces 1 min y con etanol 2 veces 1 min. Posteriormente se estimó la radiactividad incorporada en cada una de las muestras utilizando Tolueno-PPO como sistema de centelleo líquido.

En la síntesis de la segunda cadena se utilizó como templado la molécula híbrida (ARN+ADNccs) resultante de la reacción de síntesis de cadena sencilla y la enzima RNasa H (E. coli), para digerir parcialmente el ARN, el cual funcionó como iniciador y fué reemplazado eficientemente utilizando la enzima ADN polimerasa I (E. coli). Consecutivamente se incorporó a la reacción la enzima T4 ADN polimerasa, con actividad de exonucleasa 3'-5' con objeto de remover cualquier residuo remanente 3' de la reacción de la primera cadena y de esta forma corregir posibles errores en la síntesis de la doble cadena.

La reacción se llevó a cabo en un volumen de 125  $\mu\text{l}$  en una solución que contenía: Tris-HCl 100mM pH 8, DTT 2 mM, 2-mercapto-etanol 5 mM, KCl 60 mM, HEPES 35 mM,  $\text{MgCl}_2$  7 mM, una mezcla de dATP, dCTP, dGTP, dTTP 0.4 mM, ADNccs 60 ng/ml, RNasa H 2 U, ADN polimerasa I 57 U y [ $^{32}\text{P}$ ]-dATP 15  $\mu\text{Ci}$  (3 000 Ci/mmol); la reacción se incubó consecutivamente a 12 °C durante 60 min, 22 °C por 60 min, 70 °C por 10 min y agua/hielo (0 °C); finalmente se agregaron 5 U de T4 ADN polimerasa, se incubó 10 min a 37 °C y se paro añadiendo 4  $\mu\text{l}$  de EDTA 0.25 M pH 8.

Se estimó el rendimiento de síntesis de ADNccs de igual forma que en la reacción anterior, tomando alícuotas de 1  $\mu$ l al inicio y término de la misma. El ADNccs se extrajo con solventes orgánicos (fenol/cloroformo-isoamílico) de la manera ya descrita y se purificó mediante cromatografía de sephadex G50. Posteriormente se midió la radiactividad de una de las fracciones y el material del pico radiactivo se precipitó con 5 volúmenes de etanol y un volumen de acetato de sodio 3 M pH 5, conservándose así hasta su uso.

El ADNc sintetizado de esta manera se amplificó para su posterior utilización en la síntesis de molde en el análisis de secuenciación, mediante el método de PCR, descrito originalmente por Mullis (62), el cual permite la amplificación in vitro de segmentos de ADN a través de una serie de reacciones a diferentes temperaturas, en las cuales primero hay una temperatura que permite la desnaturalización del ADN, otra que permite la alineación de los iniciadores y una adicional que ayuda a la polimerización de la cadena sintetizada; estos tres pasos consecutivos se refieren a un ciclo y la reacción esta basada en la repetición de estos ciclos amplificando el segmento de  $10^5$  a  $10^9$  veces. Se utilizó una ADN polimerasa de 94 KD termoestable que posee una alta procesividad (taq ADN polimerasa). La reacción se llevo a cabo en un volumen final de 100  $\mu$ l que contenía 10 mM de Tris-HCl pH 8.3, 50 mM de KCl, 1.5 mM de  $MgCl_2$ , 0.01 % de gelatina (peso/volumen) una mezcla de 200  $\mu$ M de cada uno de los dNTP's, 1  $\mu$ M de iniciador en sentido 5'-3', 1  $\mu$ M de iniciador en sentido 3'-5', 1 ng de ADN molde y 2.5 U de taq ADN polimerasa. Si la muestra contiene proteasa, se calienta a 95 °C durante 5 min y

despues se agrega la enzima; para evitar la evaporación de la muestra a cada uno de los tubos se les agregó de 50 a 100  $\mu$ l de aceite mineral.

La programación del multibloque se hizo con el siguiente protocolo: el primer ciclo se llevo a cabo de una sola vez a 92 °C 2 min, 50 °C 3 min y 72 °C 2 min. El segundo ciclo se repitio 28 veces a 94 °C 2 min, 50 °C 2 min, 72 °C 5 min y 4 °C 8 horas. Al termino de la amplificación las muestras se limpiaron del aceite mineral con cloroformo y se analizaron tanto por electroforesis en gel de agarosa al 1.2 % como por Southern blot, de la siguiente manera: los geles se sumergieron en NaCl 1.5 M con NaOH 0.5M 60 min a temperatura ambiente para desnaturalizar el DNA, luego se neutralizaron en una solución de Tris-HCl 1M pH 8.0 y NaCl 1.5M durante 1 h a temperatura ambiente, con inversión lenta; a continuación el gel se saturó con una solución de SSPE 10X durante 30 min y se colocó sobre un papel filtro Whatman de 3 mm que hacia contacto con el amortiguador de transferencia (SSPE 15X); éste se cubrió con un filtro de nitrocelulosa previamente puesto en SSC 2X y otro papel Whatman; para favorecer el flujo se colocó encima una capa de 8 cm de papel absorbente (sanitas), un vidrio y un peso de aproximadamente 400 g, se cubrió con plástico Kleen-pack y se dejó transferir 18 h a temperatura ambiente.

Al día siguiente se recuperó la nitrocelulosa y se sumergió en SSC 6X 5 min a temperatura ambiente; posteriormente se retiró el exceso de SSC y se dejó secar en una hoja de papel filtro; ya seco el DNA se fijó por calor a 80°C con vacío durante 2 h y se

guardó en papel Whatman hasta su uso.

La reacción de marcaje de los oligonucleótidos sintéticos específicos se llevó a cabo en un volumen de 100  $\mu$ l con 50mM de Tris HCl pH 7.6, 1 mM de espermidina, 0.1M  $MgCl_2$ , 1 mM EDTA, 50 mM DTT, 300 ng de oligonucleótido, 20  $\mu$ Ci de 5' $\gamma$ <sup>32</sup>P- ATP (3000 Ci/mmol) y 10 U de la enzima polinucleotido cinasa. La reacción se incubó a 37°C por 30 min y posteriormente se paró con 0.5M EDTA. Finalmente se hizo una extracción con fenol, otra con cloroformo y se precipitó con etanol y NaCl 0.5M. La hibridación se llevó a cabo bajo el procedimiento de la sección V utilizando como solución hibridizadora 2 X SSC, 0.2% SDS (w/v), 1mM EDTA, 50% de formamida y 2 X solución Denhardt's. Las muestras positivas se subclonaron en bacteriófago M13mp18, para el análisis de secuencia nucleotídica de la manera descrita en la sección C.

#### -PREDICCIÓN DE LA ESTRUCTURA SECUNDARIA DE LAS PROTEINAS

A la utilización de varios programas (estereoquímicos y estadísticos-empíricos) para predecir las estructuras secundarias de una proteína se denomina predicción conjunta, este método nos da una exactitud del 70 al 75 % a lo largo de la secuencia de aa.

Los métodos estadísticos-empíricos son la agrupación de los estadísticos y heurísticos, los cuales están basados en parámetros obtenidos del análisis de secuencias y estructuras conocidas. Los parámetros son físicoquímicos y la comparación se hace con las estructuras observadas por cristalografía de rayos X, por ejemplo, hélices  $\alpha$ , se ven cuales aa se presentan entre estas estructuras, cuales son casi exclusivas de las mismas, que aa son los que rompen la estructura, y la combinación de varios aa que

estructura tienden a formar. No se estudia su disposición en el espacio ni su posible empaquetamiento.

Los métodos estereoquímicos toman en consideración la arquitectura del posible empaquetamiento de la cadena polipeptídica y la interacción de las proteínas con moléculas de agua. Estos métodos están basados principalmente en los ángulos  $\phi$ ,  $\psi$ ,  $\tau$ ,  $\omega$  los cuales corresponden a los cuatro substituyentes del  $C_{\alpha}$ .

$\psi$  corresponde al substituyente entre el  $C_{\alpha}$  y el grupo carboxílico,  $\phi$  entre el  $C_{\alpha}$  y el grupo amino,  $\tau$  es el ángulo de  $C_{\alpha}$  y  $C_{\beta}$  y  $\omega$  es el enlace entre el grupo carboxilo y el amino. Basado en la arquitectura de estos substituyentes, se observa en aa con cadena lateral larga que el  $C_{\beta}$  también juega un papel importante en la arquitectura del posible empaquetamiento de la cadena peptídica.

Para predecir la estructura secundaria de las proteínas de membrana y nucleocápside del virus del dengue se utilizaron distintos métodos estadísticos y fisicoquímicos (diagrama 1).

Los programas fueron: Chou-Fasman (30) complementado con la predicción para giro  $\beta$  de los mismos autores (31). Este programa calcula los aa con alta probabilidad de formar hélices  $\alpha$  (Pa), los aa formadores de lámina  $\beta$  (PB) y los aa con alta probabilidad de formar giros (Pt). Los aa de la secuencia que no estén en los tres grupos anteriores se denominan Estructuras No Repetidas (ENR) y se calcula como Pc. Este programa está clasificado en el grupo de los estadísticos, ya que, en la predicción de regiones estructurales toma como base, los estudios realizados en proteínas cuya estructura terciaria ya es conocida por cristalografía

de rayos X. Las magnitudes relativas (expresadas en porcentajes) de los valores de cada aa los clasifica en formadores, indiferentes y rompedores de cada estructura.

El algoritmo de Garnier y col. (32) se clasifica dentro de los métodos heurísticos, el cual está basado en las propiedades fisicoquímicas de cada aa, tomando en cuenta a los aa vecinos, esto es, lee por grupos de aa, a los cuales se les denomina "ventanas". Los autores recomiendan que si la proteína es soluble se debe utilizar una ventana de 7 a 9 aa y si la proteína es de membrana la ventana debería ser de 11 a 19 aa. El número de ventana (NV) es el mismo número de aa que se va a agrupar, esto es, que si tenemos un aa  $i$  se van a leer el mismo número de aa tanto hacia atrás de  $i$  ( $i-x$ ) como hacia adelante ( $i+x$ ) donde  $x$  es el número de la ventana menos 1 dividido entre 2  $[(NV-1)+2]$ . Este programa además requiere una información adicional, con el objeto de no subpredecir estructuras, la cual los autores denominaron constante de decisión (CD), ésta es tomada en cuenta según el porcentaje que la proteína posee de hélice  $\alpha$  y lámina  $\beta$ . La elección correcta de la CD aumenta la exactitud del programa en un 7 %. Garnier y col agrupan a las proteínas en 4 tipos diferentes de la manera siguiente: Las ricas en hélice  $\alpha$  que tienen más del 30% de esta estructura y menos o igual del 10 % de lámina  $\beta$ . Las ricas en lámina  $\beta$  que poseen más del 30 % de  $\beta$  y menos o igual del 10 % de hélices  $\alpha$ . Las de estructura secundaria regular  $\alpha$  y  $\beta$  con porcentajes menores al 25 % y las mixtas que son aquellas que no entran en ninguna de las anteriores.

Gascuel y Goldmar (33) utilizan un método estadístico que explora rigurosamente la composición de aa en cada tipo de

estructura secundaria. La distribución de los aa a lo largo de cada motivo estructural es analizado según el porcentaje con que aparece en la estructura, esto es, se observó que hay aa que tienen cierta preferencia por aparecer al inicio de las hélices  $\alpha$  o al final, estas observaciones son válidas para los demás motivos estructurales. Este método posee un parámetro equivalente a las constantes de decisión de Garnier y col, denominándolo parámetro  $N_g$ , que tiene la finalidad de aumentar la exactitud de la predicción. Este programa solo predice la probabilidad en tres de los cuatro motivos estructurales (Pa, PB y Pc).

El algoritmo de Deleage y Roux (34) está basado principalmente en los principios de Chou y Fasman (30) complementado por los parámetros propuestos por Nakashima y col (46). Estos autores proponen que la adición de estos parámetros eleva la exactitud de la predicción (7 a 8 %) hecha por el algoritmo de Chou y Fasman únicamente. Este modelo de predicción se ha llamado método de doble predicción (34).

Levitt (35) desarrolló un programa estadístico que analiza las coordenadas atómicas de un gran número de proteínas globulares con objeto de identificar regiones de hélice  $\alpha$ , lámina  $\beta$  y giro, y está basado en los patrones de enlace de hidrógeno, distancias entre  $C_\alpha$  y ángulos de torsión entre  $C_\alpha$ .

Los métodos modificados de Chou-Fasman y Garnier para proteínas de membrana fueron desarrollados por Croft, A. (54) en base a los algoritmos originales de estos autores complementados con los parámetros propuestos por Rao y Argos (36) para la predicción de hélices de membrana. Se utilizaron estos métodos en la predic-

ción de las estructuras transmembranales que podrían servir de anclaje a la proteína M.

Para la predicción de hélices de membrana de la proteína M se utilizaron los algoritmos de: Rao y Argos (36) este se basa en las características físicas de cada uno de los aa, y perfeccionado por la comparación de la estructura propuesta para la bacteriorodopsina. El esquema fue ideado para delinear las regiones que probablemente estén asociadas a membrana, a partir de secuencias primarias de proteínas que se sabe interaccionan con la membrana lipídica.

Klein y col (37) encontraron un análisis discriminativo que clasifica a las superfamilias de proteínas en seis grupos que pueden ser distinguidas sobre la base de cuatro variables, caracterizando la composición de aa y datos locales de la secuencia. Estas variables son promedio hidrofóbico, carga neta, longitud de la secuencia y variación periódica en residuos hidrofóbicos a lo largo de la cadena.

Eisenberg (38) analiza la estructura en términos de "momento hidrofóbico" de la proteína completa y de los segmentos de estructura secundaria. El momento hidrofóbico es el análisis conjunto de dos parámetros, el momento cero y el momento uno. El momento cero está definido como la suma de la hidrofobicidad de cada residuo y el momento uno o momento dipolar hidrofóbico que es análogo al momento dipolar eléctrico de un grupo de cargas (38).

Los métodos fisicoquímicos utilizados fueron los siguientes: Hidrofilicidad (Hl) de Hoop y Woods (39) que mide la capacidad de la molécula proteínica para interaccionar con el agua a lo largo



de la secuencia de aa dependiendo de las características individuales de cada aa. Hidrofobicidad (Hb) de Miyazawa (40) la cual mide la repulsión de la molécula proteínica al agua. Hidropatía (Hp) de Kyte y Doolittle (41), que evalúa la hidrofóbicidad y la hidrofiliicidad simultaneamente a lo largo de la cadena de aminoácidos. Flexibilidad (Flex) de Karplus y Schulz (42) este programa mide la correlación entre las láminas B, hidropatía máxima y flexibilidad mínima, y entre asas predichas, alta flexibilidad de cadena y segmentos proteínicos que aceptan inserción de longitud variable de secuencias individuales. Antigenicidad (Ag) de Hoop y Wood (39) estos autores enfatizan el caracter polar promedio de los segmentos de la secuencia, el cual esta relacionado con los grupos antigenicos o capaces de producir una respuesta inmune. Polaridad (Pol) de Graham (44). Este autor desarrollo una formula para ponderar las diferencias entre aa que combinan las propiedades que mejor correlacionan (como son composición, polaridad y volumen molecular) con la frecuencia de substitucion de residuos en las proteínas.

Los análisis de las predicciones se hicieron en una computadora PC/AT compatible y se utilizarón los siguientes programas: Predict7 de R.S. Carmenes (53); Seqanal de A. Crofts (54) versión 1.03, Universidad de Illinois, EUA; y PC/GENE Intelligenetics versión 6.26 (55).

En los métodos de Chou-Fasman, Deleage y Roux, y Levitt, se utilizo una ventana de 7 aa para calcular los valores  $P_{\alpha}$  y  $P_{\beta}$  tanto en la proteína de nucleocápside (C) como en la precursora de membrana (PrM). En la proteína de membrana (M) se utilizó una

ventana de 16 residuos de aa en los cálculos de Pa, P $\beta$ , y Pt. Para los métodos modificados de Chou-Fasman y Garnier, así como en los programas de Rao y Argos, Eisenberg y Klein, se utilizó una ventana de 19 aa.

Para todas las proteínas se usaron inicialmente las constantes de decisión (CD) siguientes CD $\alpha$ =0, CD $\beta$ =0, CDt=0 y n=1. Después del análisis primario para observar los porcentajes de  $\alpha$ + $\beta$  mayores de 50 % se tomaron las siguientes constantes de decisión; CD $\alpha$ =0, CD $\beta$ =20 CDt=0, n=1. En el método de Gascuel y Goldmar también se utilizaron inicialmente los parámetros estándares y posteriormente bajo el mismo criterio se modificaron a N[S] $\alpha$ =1.10; N[S] $\beta$ =1.545 y N[S] coil=1.00.

La predicción de la estructura para cada proteína se hizo tomando en cuenta los siguientes parámetros:

- 1.-Predicción de cada proteína por los distintos métodos
- 2.-Asignación de hélices alfa con un requerimiento mínimo de 6 aa consecutivos formadores de hélice  $\alpha$ ; para estructura  $\beta$  4 aa. consecutivos formadores de estructura  $\beta$ ; y para giro un mínimo de 3 aa consecutivos con alta probabilidad de formar giros.
- 3.-La discrepancia en longitud para las hélices alfa dada por los diferentes métodos fue corregida por los parámetros estadísticos propuestos por Richardson (47), tomando en cuenta la preferencia de los distintos aa. por los extremos de la hélice.
- 4.-Corrección de las estructuras asignadas utilizando índices fisicoquímicos. Los índices de flexibilidad de cadena e hidropatía se tomaron como indicadores recíprocos y comple-

mentarios de la posición de segmentos polipeptídicos con respecto al interior hidrofóbico de las proteínas (47). Los segmentos con alta hidropatía y baja flexibilidad fueron considerados como estructuras  $\beta$ ; segmentos con baja hidropatía y alta flexibilidad fueron tomados como estructuras no repetidas según Crawford (59).

- 5.-Las hélices alfa anfipáticas fueron predichas tomando en cuenta un promedio de hidrofobicidad de 0-0.2, con un momento hidrofóbico de 0.3-0.4, usando un ángulo de 100 grados. Los datos positivos fueron correlacionados con la probabilidad de superficie e hidrofiliidad.

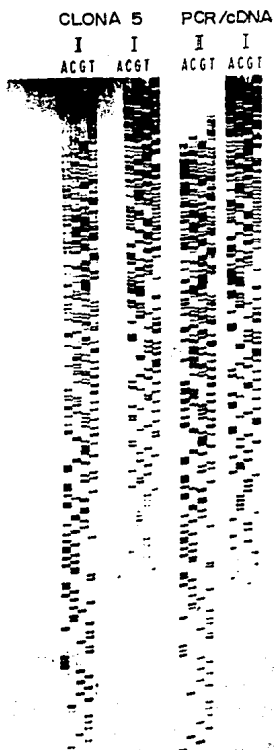
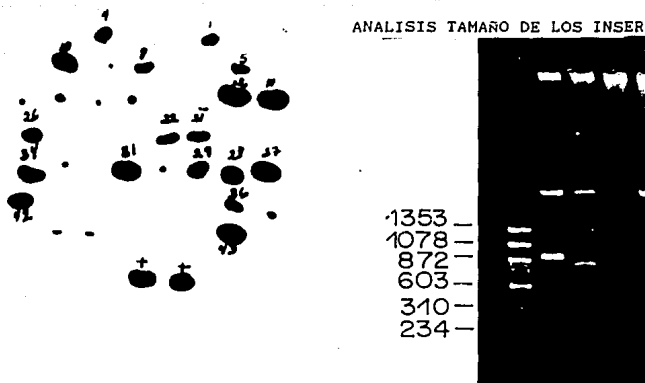


Fig. 10 Secuenciación nucleotídica de la clona 5 y un ADMc amplificado por PCR

## RESULTADOS

### ANALISIS DE LAS CLONAS POSITIVAS.

El tamaño de los insertos del ADNc de las 45 clonas con señal positiva (figura 8a) se determinó por electroforesis de los productos de digestión de los plásmidos con las enzimas de restricción EcoRI/Hind III, con objeto de proteger el sitio de SmaI; el tamaño de los insertos varió de 300 a 1000 pares de bases (pb) (figura 8b).



### a) CLONAS POSITIVAS PARA PROTEINAS ESTRUCTURALES DEL VIRUS DEL DENGUE b) $\lambda$ X 8 9 10 1

Fig. 8 a) Clonas positivas por hibridación en colonia para proteínas estructurales del virus del dengue 2.

b) Análisis del tamaño de los insertos por electroforesis en gel de agarosa. En el primer carril se muestra un marcador de peso molecular ( $\lambda$ X/Hae III), en los otros carriles las clonas según el número asignado durante el escrutinio a la genoteca.

### SECUENCIACION NUCLEOTIDICA DE LOS INSERTOS.

El análisis de la secuencia nucleotídica se llevó a cabo por el método de la cadena terminal (27). La ubicación de las regiones secuenciadas se hizo por analogía con la secuencia reportada previamente para la cepa prototipo del virus del dengue-2



(Nueva Guinea C) (figura 9). En ocasiones hubo que hacer una segunda e incluso una tercera hibridización en el banco genómico con objeto de identificar clonas de los extremos 5' y 3' que no estaban representados en la genoteca.

#### ESTRATEGIA DE SECUENCIACION DE LA REGION ESTRUCTURAL DE D2M

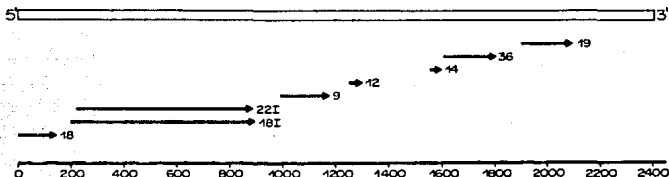


Fig. 9 Estrategia de secuenciación de la región estructural de dengue 2 mexicano.

#### -SINTESIS Y AMPLIFICACION DE REGIONES NO REPRESENTADAS

Las secuencias que no estaban representadas en el banco, se sintetizaron directamente a partir de ARNm con oligonucleótidos sintéticos específicos. El material sintetizado se amplificó por PCR para posteriormente utilizarlo en la obtención del molde en el análisis de secuenciación (figura 10). El ADNc amplificado se observó mediante electroforesis en gel de agarosa.

Para el análisis posterior de la secuencia nucleotídica, las muestras positivas se subclonaron en M13mp18.

#### ANALISIS DE ESTRUCTURA PRIMARIA.

La región estructural del virus del dengue 2 mexicano contiene 2409 nucleótidos (nt) incluyendo 81 nt en el extremo 5' que representan la región no traducible 5' (RNT5') y 2328 nt que corresponden a un producto poliproteínico de 776 aa (figura 11).

La RNTS' presenta 4 codones de término en las posiciones 42 (UAA), 48 (UGA), 57 (UGA) y 60 (UAA) que preceden al codón de inicio (AUG) a partir del cual se encuentra un marco abierto de lectura ininterrumpida y 35 codones de inicio sin marco adecuado de lectura.

En el análisis de composición de bases observamos que la región estructural de D2M contiene un 33 % de adenina, 25 % de guanina, 21.2 % de timina y 20.8 % de citosina. Tanto la composición de codones utilizados como la frecuencia de aparición de los aa se muestran en la tabla 2.

CODONES UTILIZADOS EN LA SECUENCIA Y EL PORCENTAJE DE CADA UNO.

TTT Phe 13 1.6%	TCT Ser 5 .6%	TAT Tyr 5 .6%	TGT Cys 10 1.2%
TTC Phe 18 2.3%	TCG Ser 9 1.1%	TAC Tyr 7 .9%	TOC Cys 8 1%
TTA Leu 7 .9%	TCA Ser 18 2.3%	TAA --- 0 0%	TGA --- 0 0%
TTG Leu 11 1.4%	TCG Ser 1 .1%	TAG --- 0 0%	TGG Trp 16 2%
CTT Leu 7 .9%	CCT Pro 7 .9%	CAT His 13 1.6%	CGT Arg 3 .3%
CTC Leu 7 .9%	CCC Pro 4 .5%	CAC His 7 .9%	CCG Arg 8 1%
CTA Leu 6 .7%	CCA Pro 16 2%	CAA Gln 14 1.8%	CGA Arg 2 .2%
CTG Leu 30 3.8%	CCG Pro 2 .2%	CAG Gln 12 1.5%	CCG Arg 1 .1%
ATT Ile 12 1.5%	ACT Thr 7 .9%	AAT Asn 13 1.6%	AGT Ser 5 .6%
ATC Ile 20 2.5%	ACC Thr 16 2%	AAC Asn 10 1.2%	AGC Ser 4 .5%
ATA Ile 22 2.8%	ACA Thr 40 5.1%	AAA Lys 37 4.7%	AGA Arg 21 2.7%
ATG MET 35 4.5%	ACG Thr 11 1.4%	AAG Lys 21 2.7%	AGG Arg 8 1%
GTT Val 12 1.5%	GCT Ala 7 .9%	GAT Asp 4 .5%	GGT Gly 4 .5%
GTC Val 14 1.8%	GCC Ala 8 1%	GAC Asp 22 2.8%	GCC Gly 12 1.5%
GTA Val 6 .7%	GCA Ala 13 1.6%	GAA Glu 31 3.9%	GGA Gly 42 5.4%
GTC Val 19 2.4%	GCC Ala 7 .9%	GAG Glu 14 1.8%	GGG Gly 12 1.5%

Ala => 35	Gln => 26	Leu => 68	Ser => 42	--- TAA 0
Arg => 43	Glu => 45	Lys => 58	Thr => 74	TGA 0
Asn => 23	Gly => 70	MET => 35	Trp => 16	TAG 0
Asp => 26	His => 20	Phe => 31	Tyr => 12	
Cys => 18	Ile => 54	Pro => 29	Val => 51.	

Tabla 2. PORCENTAJE DE CODONES UTILIZADOS, DE BASES NUCLEOTIDICAS Y NUMERO DE AA EN LA SECUENCIA DE LAS

PROTEINAS ESTRUCTURALES DEL VIRUS DEL DENGUE.



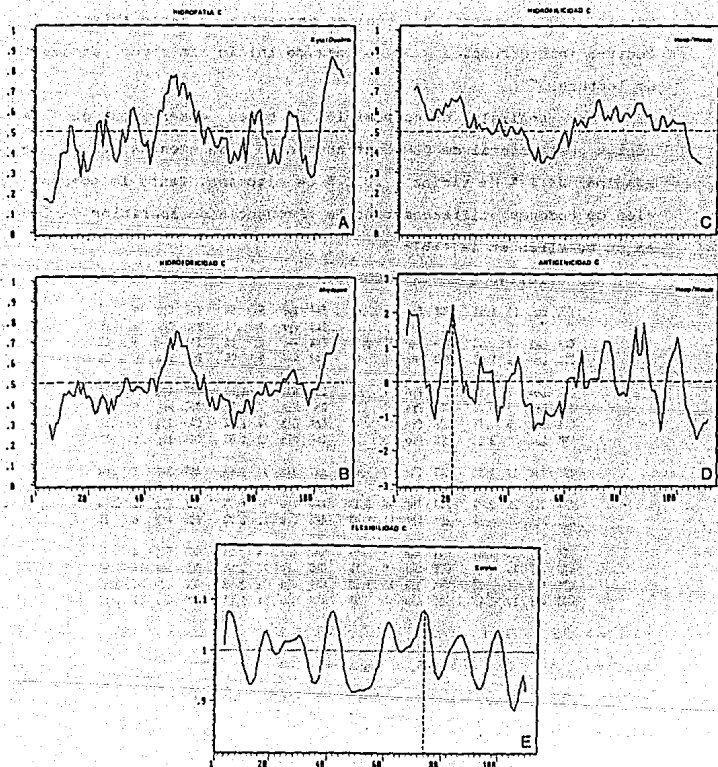


Fig. 13 Graficas de análisis fisicoquímicos de la proteína C. a) hidropatía. b) hidrofobicidad. c) hidrofilicidad. d) antigenicidad. e) flexibilidad.



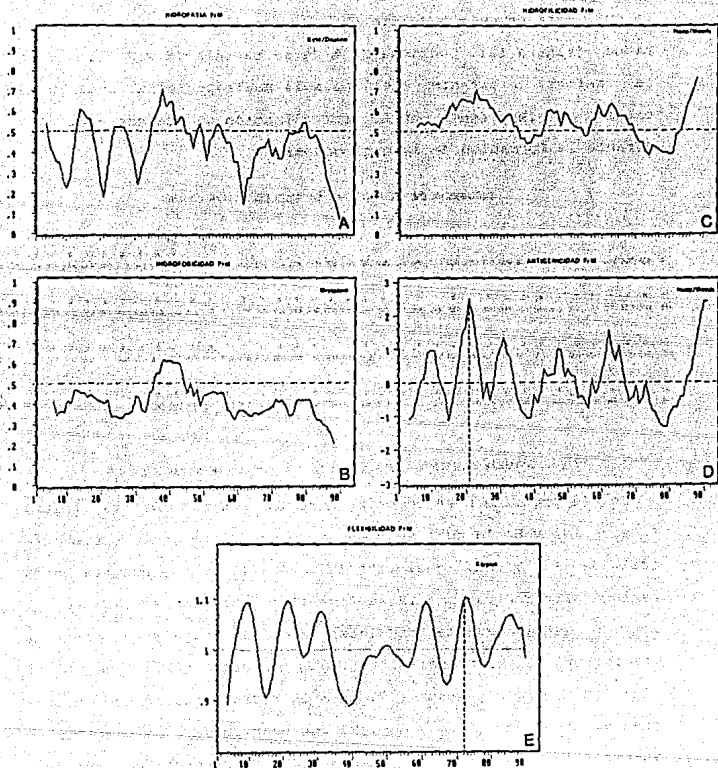


Fig. 15 Graficas de los perfiles fisicoquimicos de la proteina PRM. a) hidropatía. b) hidrofobicidad. c) hidrofilicidad. d) antigenicidad. e) flexibilidad.

La siguiente proteína secuenciada fue el precursor de membrana o prM y el péptido M (figura 14). El precursor de membrana contiene 91 aa que corresponden a 273 nucleótidos. Esta proteína inicia con dos aa hidrofóbicos y el resto de la misma es altamente hidrofílica (figura 15C) lo que le confiere la característica de ser muy soluble.

SECUENCIA DE LA PROTEINA DE MEMBRANA DENGUE-2M

```

1 5' A T T T T G C E P N I V I R E E S G C L L F T T K D G I N H E S L M 37
  T MAU GAU CUG ACC ACA GCG AAC GGA GAA GCA CAC AUG AUC GAU MAU ACA GAA GAA AAA GGA AAA AGC GAU CUG UAU AAG ACA AAG GAC GGC ACC AAC AUG UAU ACC GAU AUG 111

56 A H D L C E L C E D T I T T E C P F L K O N E P E N I O C M S H E T S F M 74
112 GCG AUG CAC GAU GAU CAG CAG UAG GAU GAA CAC ACA AUG ACC UAU AAA UGU GCG UAU GAU AAG CAG AAC GAA GAA GAA AAC MAU GAU UGU UGU GCG AAC AAC UGC ACC GCG ACA UGU 222

75 V I T G Y C T T T C E R R R E E R T Y A L V P S V E H L E T R I E I N R 223
225 GAU GAU UAU GCG ACA UGU ACC ACC ACA GGA GAA GAC CAC ACA ACA GAA AAA ACA UCA GAU GCA GAU GAU GCA CAC GAG GGA AUG GGA UAG GAG ACA GCA GAU GAA ACA UGC MAU 333

132 S S E G A V E H V Q R I E T W I L V S D C F T I N A A L A Y T T E I T M 343
334 UCA UCA GAA GCG GCG UAG AAA GAU GAU CAG GAA AAG CAA ACA UGU UAG AUG CAG ACA CAU CCU GCG UAU ACC MAU AUG GCA GCA CAC CUG GCA UAC ACC MAU GCA ACC ACC CAU 444

149 F R E V I I F I L L A S P E R T 160
445 UAG GAA ACA CUG CUG MAU UAC AUG GAA CAG ACA GCG AAC GAU GAU UCA ATG ACA 490

8 4 5 5  S I T I O P R O B A B L E D E G L U C O S I L A C I O N
  - - - -  S I T I O P R O B A B L E D E C O T E
  
```

Fig. 14 Secuencia de nucleótidos y deducida de aa de la proteína de membrana o M.

Además presenta 5 regiones antigénicas y flexibles (aa 6-12, 18-25, 29-34, 44-49, 58-65 y 82-91 (figura 15D y E) con una pequeña región hidrofóbica (aa 35-43) (41) (figura 15A y B) que se podría encontrar en el centro de la proteína. En el aa 66 se encuentra un sitio de glucosilación que presenta la siguiente secuencia Ser-Asn-Ser-Thr. En el extremo carboxilo de prM se encuentra la secuencia Arg-Arg-Glu-Lys-Arg que podría ser el sitio de corte reconocido por enzimas procesadoras para producir la proteína de membrana o M.

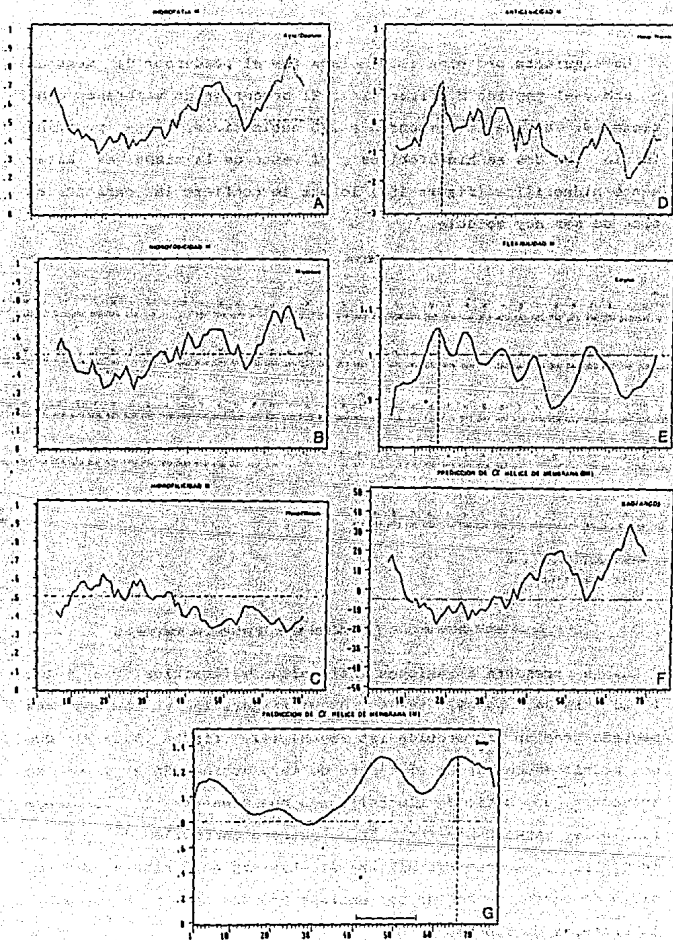


Fig. 16 Graficas de los perfiles fisicoquimicos de la proteina M. a) hidropatia. b) hidrofobicidad. c) hidrofilicidad. d) antigenicidad. e) flexibilidad. f) helice de membrana (Pao/Argos). g) helice de membrana (Lein y Col).

El péptido M es una proteína integral de membrana, que contiene 225 nt que corresponden a 75 aa (figura 14) la cual alterna dos regiones hidrofílicas (aa 9-36, 54-56) con dos hidrofóbicas (aa 37-42, 57-75). Estas regiones podrían funcionar como sitios de interacción con la nucleocápside viral. Se propone que la principal función de la proteína M sea el ensamblaje del virus. La proteína M en su extremo amino terminal es principalmente hidrofílica (figura 16C), presenta 3 regiones antigénicas y flexibles (figura 16D y E). Por otro lado el extremo carboxilo no contiene regiones antigénicas ni flexibles, en este extremo se encuentra la región transmembranal (36,37,38) de la proteína (figura 16F y G). Hacia el extremo carboxilo encontramos una región altamente hidrofóbica de 19 aa (figura 16A y B) que funcionaría como señal para la translocación de la siguiente proteína la proteína E la cual no fue utilizada en este trabajo.

Al alinear la secuencia deducida de aa para dengue 2 mexicano con las secuencias previamente reportadas para dengue 1, 3, 4 y diferentes topotipos del serotipo 2 (Nueva Guinea, Jamaica, Puerto Rico) (figura 17) observamos que en general la organización del genoma en la región estructural es similar a la de otros flavivirus reportados a la fecha, y en particular que existen regiones muy conservadas en todas las proteínas.



GUIPRO	VIRVOYEGDGSCKIPFEIMDLKRRHVLGRLLITVNPVITKXDSPVNIEAEPFGDSYIIIGVEPQGLKLMWFKKSSIGOMIETTMRGAKRMAILGDTAN	700
PROTED1	LVQVYKYGTDAPCKIPFSSODEKGVTONRGLITANPIVTDKXKPVNIEAEPFGESYIVVWAGERAKLMSFKKSSIGOMFEATARGAREMAILGDTAN	700
JAMPRO	VIRVOYEGDGSCKIPFEIMDLKRRHVLGRLLITVNPVITKXDSPVNIEAEPFGDSYIIIGVEPQGLKLMWFKKSSIGOMFETTRGAKRMAILGDTAN	700
PRICOPRO	VIRVOYEGDGSCKIPFEIMDLKRRHVLGRLLITVNPVITKXDSPVNIEAEPFGDSYIIIGVEPQGLKLMWFKKSSIGOMFETTRGAKRMAILGDTAN	700
MEXICANA	VIRVOYEGDGSCKIPFEIMDLKRRHVLGRLLITVNPVITKXDSPLNIEAEPFGDSYIIIGVEPQGLKLMWFKKSSIGOMFETTRGAKRMAILGDTAN	700
PROTED3	LIVVETKGEDAPCKIPFSTEDQCKAHNGRLITANPVVYTKKEEPVNI EAQPPFGESNIVIGSGDKALKINWYRKGSSIGOMFOATARGAREMAILGDTAN	698
PROTED4	VVKVYEGAGAPCKVPIEIRDVNREKVVGRIIISSTPLAENTNSVTNIELERPL-DSYIVIGVGNLSALTLHWFKKSSIGOMFESTYBGAKRMAILGDTAN	699
	xxx'x'xxx''x'xx 'x      '''xxx x'xxxxxxx      ''x'x x' 'xx' x 'x'x 'xx'x''''x'      '''      *****x'''	
GUIPRO	HFGSLGGVFTSICKALHQVFGAIYGAAFSGVSWIMKILIGVIITWIGMNSRSTSLSVSLVVGVVTLVLYGVWQA	775
PROTED1	DFGSGGVFTSVGKLIHQIFGTAYGVLFSGVSWIMKIGICILLTWGLNSRSTSLSMTCIAVGMVTLVLYGVWQA	775
JAMPRO	DFGSLGGVFTSICKALHQVFGAIYGAAFSGVSWIMKILIGVIITWIGMNSRSTSLSVSLVLYGVVTLVLYGVWQA	775
PRICOPRO	DFGSLGGVFTSICKALHQVFGAIYGAAFSGVSWIMKILIGVIITWIGMNSRSTSLSVSLVLYGVVTLVLYGVWQA	775
MEXICANA	DFGSLGGVFTSICKALHQVFGAIYGAAFSGVSWIMKILIGVIITWIGMNSRSTSLSVSLVLYGVVTLVLYGVWQA	775
PROTED3	DFGSGVGLKSLGKQVHVFSGSANTALFSGVSWIMKIGICVLLTWIGLNSRSTSLSMFSSTIATGIIITVLYGVWQA	773
PROTED4	DFGSGVGLFTSLGKAVHVFSGSYTTMFGGVSIMRILIGFLVWIGTNSRNTSHAHTCIAVGGITLFLGFTVQA	774
	x'''x''xxx'x' x''x''x' 'xx 'x'''' xx '' xx 'x' ''x''xx x x x' x''x''      '''	

FIGURA 17 ALINEAMIENTO MULTIPLE DE LOS AMINOACIDOS DE LA REGION ESTRUCTURAL DE LOS DIFERENTES SEROTIPOS DEL VIRUS DEL DENGUE: LAS x MUESTRAN LOS SITIOS CONSERVADOS DE LA SECUENCIA Y LOS ' MUESTRAN LOS SITIOS IDENTICOS DE LA SECUENCIA. GUIPRO (Dengue 2 topotipo Nueva Guinea) PROTED1 (Dengue 1) JAMPRO (Dengue 2 topotipo Jamaica) PRICOPRO (Dengue 2 topotipo Puerto Rico) MEXICANA (Dengue 2 topotipo México) PROTED3 (Dengue 3) PROTED4 (Dengue 4)



**TABLA 3 PORCENTAJE DE ESTRUCTURA SECUNDARIA DE LAS PROTEINAS ESTRUCTURALES DE DENGUE 2 MEXICANO POR DIFERENTES METODOS DE PREDICCION Y LA PREDICCION CONJUNTA**

PROTEINA DE NUCLEOCAPSIDE C					
MET	% $\alpha$	% $\beta$	%GIROS	%ENR	% $\alpha+\beta$
Ch&F	22	31	19	28	53
D&R	45	13	4	38	58
L	42	18	5	35	60
G	34	39	6	21	73
G&G	31	23	46		54
MCF	44	42	12	2	86
GM	43	32	8	17	75

PROTEINA PRECURSORA DE M					
MET	% $\alpha$	% $\beta$	%GIROS	%ENR	% $\alpha+\beta$
Ch&F	14	22	41	23	36
D&R	13	14	32	41	27
L	16	14	3	67	30
G	34	14	30	22	48
G&G	19	16	65		35
MCF	33	16	23	28	49
GM	39	8	45	8	47

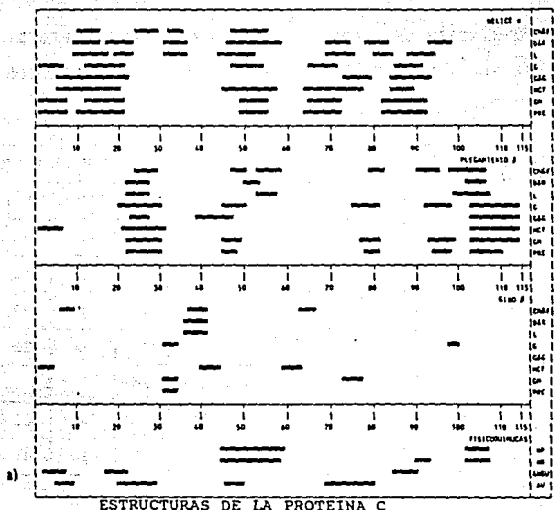
PROTEINA DE MEMBRANA M					
MET	% $\alpha$	% $\beta$	%GIROS	%ENR	% $\alpha+\beta$
Ch&F	12	53	24	11	65
D&R	31	25	0	44	56
L	28	37	0	35	65
G	19	24	9	48	43
G&G	19	19	62		38
MCF	52	39	7	2	91
GM	60	23	0	17	83

**PREDICCION CONJUNTA DE LAS PROTEINAS ESTRUCTURALES DE DENGUE 2 MEXICANO**

	% $\alpha$	% $\beta$	%GIROS	%ENR	% $\alpha+\beta$
C	38	27	4	31	65
PrM	44	5	32	19	49
M	33	21	0	45	54

MET = METODO, %  $\alpha$  = PORCENTAJE DE HELICE  $\alpha$ , %  $\beta$  PORCENTAJE DE LAMINA  $\beta$ , % GIROS = PORCENTAJE DE GIROS  $\beta$ , % ENR = PORCENTAJE DE ESTRUCTURA NO REPETIDA, C = PROTEINA DE NUCLEOCAPSIDE, PrM = PROTEINA PRECURSORA DE MEMBRANA, M = PROTEINA DE MEMBRANA, Ch&F = CHOU Y FASHAN (30,31), D&R = DELEAGE Y ROUX (36), L = LEVITT (37), G = GARNIER (32), G&G = GASCUEL Y COLMARD (33), MCF = CHOU Y FASHAN MODIFICADO, GM = GARNIER MODIFICADO.

PREDICCIÓN ESTRUCTURA SECUNDARIA  
DE LA PROTEINA C



	Numero de Residuo	Estructura Predicha	Longitud	Comentarios
a)	1-7	$\alpha$	7	*Hl, *Po, *Flex
b	8-11	ENR	4	*Hl
c	12-21	$\alpha$	10	*Hl, *Po, *Flex
d	22-30	$\beta$	9	*Hl, *Flex
e	31-34	G	4	*Hl, *Flex
f	35-44	ENR	10	*Hl, *Flex
g	45-48	$\beta$	4	*Hp
h	49-55	$\alpha$	7	*Hp
i	56-63	ENR	8	*Hp, *Flex
j	64-72	$\alpha$	9	*Hl
k	73-77	ENR	5	*Hl, *Flex(pico)
l	78-81	$\beta$	4	*Hl
m	82-92	$\alpha$	11	*Hl, *Flex
n	93-98	$\beta$	6	*Hl
ñ	99-102	ENR	4	*Hl, *Flex
b)	o 103-110	$\beta$	8	*Hp
p	111-114	ENR	4	*Hp

Fig. 18 a) Predicción por cada uno de los métodos para hélice  $\alpha$ , lámina  $\beta$  y para giro  $\beta$ . Análisis físico químico de la proteína C. CH&F= CHOU Y FASMAN (30,31), D&R= DELEAGE Y ROUX (36), L = LEVITT (37), G= GARNIER (32), G&G= GASCUEL Y GOLMARD (33), MCF= CHOU Y FASMAN MODIFICADO, GM= GARNIER MODIFICADO, PRE= PREDICCIÓN, KR&D= KLEIN, KANEHISA & DELISI [SOAP (37)], RA= RAO Y ARGOS (36), HP= HIDROPATIA (41), HB= HIDROFOBICIDAD (40), HL= HIDROFILICIDAD (39), AM= ANTIGENICIDAD DE WELLING (45), AH&W= ANTIGENICIDAD DE HOPP & WOODS (39).

b) Tabla de predicción de cada una de las estructuras predichas de la proteína C con los comentarios físicoquímicos de estas estructura (Hl= Hidrofilicidad, Hp= Hidropatía, Flex=flexibilidad, Po=polaridad).

## PREDICCIÓN DE LA ESTRUCTURA SECUNDARIA.

En la tabla 3 se muestra la composición y localización de las regiones estructurales predichas para las proteínas C, prM y M del virus del dengue 2 mexicano obtenidas por diferentes métodos, así como su predicción conjunta.

En la predicción conjunta de la proteína C observamos que está constituida por 38 % de hélice  $\alpha$ , 27 % de estructura  $\beta$ , 4 % de giro y 31 % de ENR. Encontramos a lo largo de la secuencia 5 regiones para hélice  $\alpha$ , 5 para estructura  $\beta$ , 1 giro y 6 ENR (figura 18A), las cuales están divididas por sus características fisicoquímicas de la manera siguiente (figura 18B): Hélice  $\alpha$ , 1 anfipática (aa 82-92), 1 hidrofóbica (aa 49-55) y 3 hidrofílicas (aa 1-7, 12-21, 64-72), las dos primeras hélices  $\alpha$  hidrofílicas son polares. Las estructuras  $\beta$  fueron 1 hidrofóbica (aa 45-48) y 4 hidrofílicas (aa 23-30, 78-81, 93-98, 103-110). El único giro es parcialmente hidrofílico y flexible (aa 31-34). De los 8 picos de flexibilidad 3 son hélices  $\alpha$ , uno es  $\beta$ , uno es giro  $\beta$  y 4 son estructuras al azar (figura 19).

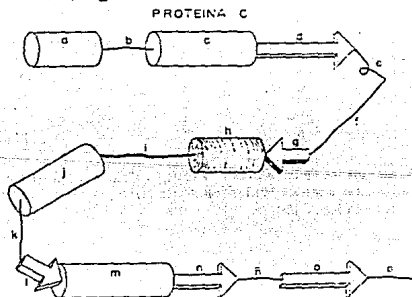
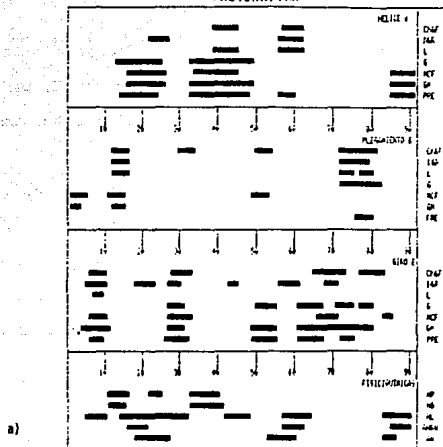


Fig. 19 Esquema de la estructura secundaria de la proteína C. Los cilindros son hélices  $\alpha$ , las flechas son láminas  $\beta$ , los rizados son giros  $\beta$  y las líneas son estructuras no repetidas.

PREDICCIÓN ESTRUCTURA SECUNDARIA  
PROTEINA PrM



ESTRUCTURAS DE LA PROTEINA PrM

	Numero de Residuo	Estructura Predicha	Longitud	Comentarios
a)	1-5	ENR	5	▲H1, Hp
b)	6-9	G	4	▲H1, ▲Flex
c)	10-13	ENR	4	▲H1, ▲Flex
ch)	14-25	α	12	▲H1, ▲Po, ▲Flex
d)	26-32	G	7	▲Flex
e)	33-48	α	16	†H1, †Hp
f)	49-55	G	7	▲H1, ▲Flex
g)	56-62	α	7	▲H1, ▲Po, †Flex
h)	63-67	G	5	▲H1
i)	68-71	ENR	4	▲Flex(pico)
j)	72-75	G	4	▲Flex
k)	76-80	β	5	▲Hp
l)	81-84	ENR	4	†Hp, H1
m)	85-91	α	7	▲H1, ▲Po, ▲Flex

Fig. 20 a) Predicción por cada uno de los métodos para hélice  $\alpha$ , lámina  $\beta$  y para giro  $\beta$  en la proteína PrM.  
b) Tabla de estructuras predichas con los comentarios fisicoquímicos de cada estructura. Abreviaturas ver pie de la figura 12-I

La proteína prM está constituida por un 44 % de hélice  $\alpha$ , 5 % de estructura  $\beta$ , 32 % de giro y 19 % de ENR (tabla 3). De acuerdo con el análisis estadístico encontramos en la secuencia 4 hélices  $\alpha$ , 1 estructura  $\beta$ , 5 giros y 4 ENR (figura 20A). De acuerdo a sus propiedades fisicoquímicas se encontraron (figura 20B) que de las 4 hélices  $\alpha$ , 3 son hidrofílicas (aa 14-25, 56-62, 85-91) y 1 anfipática (aa 33-48). La estructura  $\beta$  presente es hidrofóbica (aa 76-80). Los giros son hidrofílicos (aa 6-9, 26-32, 49-55, 63-67, 72-75) y las ENR son 3 hidrofílicas (aa 1-5, 10-13, 68-71) y una anfipática (aa 81-84). De los 8 picos de flexibilidad 2 son hélice  $\alpha$  y 4 giros, 2 ENR (figura 21).

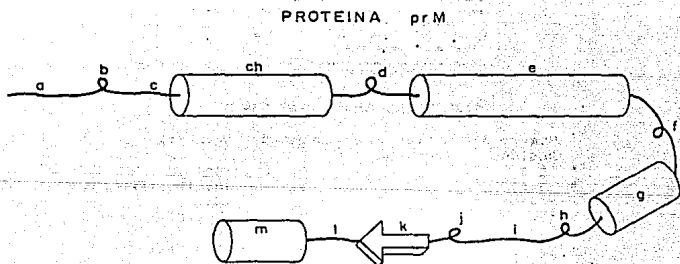
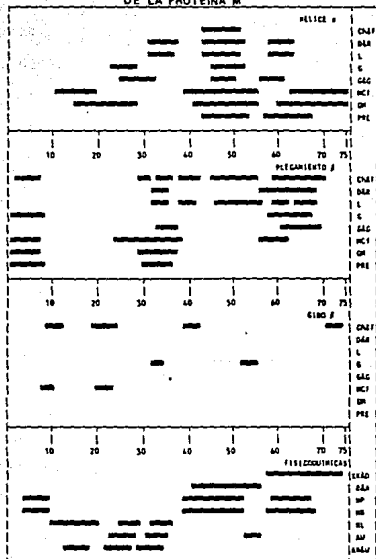


Fig. 21 Esquema de la estructura secundaria de la proteína prM. Los cilindros son hélices  $\alpha$ , las flechas son láminas  $\beta$ , los rizados son giros  $\beta$  y las líneas son estructuras no repetidas.

PREDICCIÓN ESTRUCTURA SECUNDARIA  
DE LA PROTEÍNA M



a)

ESTRUCTURAS DE LA PROTEÍNA M

	Numero de Residuo	Estructura Predicha	Longitud	Comentarios
a	1-8	$\beta$	8	*Hp
b	9-29	ENR	21	*H1, *Po, *Flex(pico)
c	30-36	$\beta$	7	*H1
ch	37-42	ENR	5	*Hp, *Flex
d	43-53	$\alpha$	11	*Hp
e	54-56	ENR	3	*Flex, *H1
f	57-67	$\alpha$	11	*Hp
g	68-75	ENR	8	*Hp

Fig 12-III a) Predicción por cada uno de los métodos para hélice  $\alpha$  lámina  $\beta$  y para giro  $\beta$  en la proteína M.  
b) Tabla de estructuras predichas con los comentarios fisicoquímicos de cada estructura. Abreviaturas ver pie de figura de la 12-1

La proteína M se predijo con un 33 % de hélice  $\alpha$ , 21 % de estructura  $\beta$  y 45 % de ENR (tabla 3). Repartidas a lo largo de la secuencia encontramos 2 hélices  $\alpha$ , 2 estructura  $\beta$  y 4 ENR (figura 22A). Las 2 hélices  $\alpha$  son hidrofóbicas (aa 43-53, 57-67) (figura 22B). Las estructuras  $\beta$  son una hidrofílica (aa 30-36) y una hidrofóbica (aa 1-8). En cuanto a las ENR dos son hidrofílicas (aa 9-29, 54-56) y dos son hidrofóbicas (aa 37-42, 68-75) (figura 23)

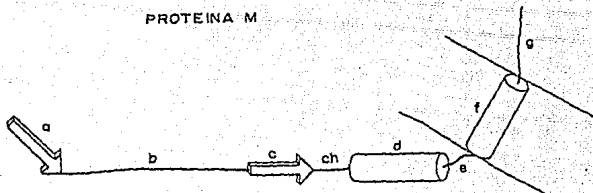


Fig 23 Esquema de la estructura secundaria de la proteína M. Los cilindros son hélices  $\alpha$ , las flechas son láminas  $\beta$ , los rizados son giros  $\beta$  y las líneas son estructuras no repetidas.

La clasificación dentro de las familias de proteínas se hizo por medio de dos algoritmos, el primero de ellos propuesto por Nakashima y col. (46), que utiliza la composición de aminoácidos para su clasificación en familias y el segundo propuesto por Taylor y Thorton (58), que propone el porcentaje de regiones estructurales  $\alpha$  y  $\beta$  (figura 24). Al analizar la predicción secundaria de nuestras proteínas observamos que el método de Taylor (58) fue más acertado, ya que esta clasificación ubica a nuestras proteínas en la familia que estructuralmente corresponde.

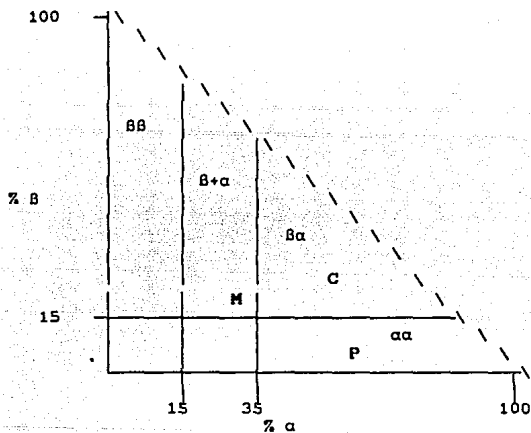


Fig 24 Grafica de la clasificación en distintas familias estructurales de las proteínas C (C), prM (P) y M (M) del virus del dengue-2 mexicano.



## DISCUSION

### **ESTRUCTURA PRIMARIA**

Por analisis de homologia, encontramos que la secuencia nucleotidica de la región estructural del virus del dengue-2 mexicano, es similar a las secuencias reportadas para otros aislados del serotipo 2 (Nueva Guinea, Jamaica, Puerto rico), existiendo una similitud del 90 % al 96%. En la secuencia de aa la homologia entre los diferentes serotipos fué del 63.37 % al 70.55 % y de un 92.86 % al 96.91 % dentro del mismo serotipo. Las variaciones en la secuencia de aa no están distribuidas uniformemente a lo largo de las proteínas, sino que existen regiones muy conservadas y otras con mayor variabilidad (figura 17).

La organización del genoma encontrada es igual a la de los flavivirus en general (7,19,22,25,57). La composición de los codones en la región traducible (tabla 2) es similar a la reportada en esta familia, no encontrandose codones sin sentido. En porcentaje de bases encontramos que la frecuencia más elevada es de adenina con un 33 % y el aa más abundante es la treonina con un 9.4 %.

La región estructural analizada comprende 776 aa y en el procesamiento postraduccional de esta poliproteína podrían estar involucradas por lo menos 3 proteasas; una que removería la metionina inicial en la proteína C (21); otra proteasa ubicada en el lumen del retículo endoplasmático (6,13,18), la cual podría cortar el extremo carboxilo de C, M y E y finalmente una proteasa del aparato de Golgi la cual ha sido implicada en la proteólisis de numerosas proteínas virales, en secuencias con aa básicos y

probablemente asociada a la maduración de prM a M, aunque hay evidencias de que prM es autocatalítica (23).

La proteína C tiene 4 regiones bien definidas que alternan en hidrofiliidad e hidrofobicidad, posee más del 20 % de aa básicos, lo cual se piensa este relacionado con su interacción con el ARN viral. Por la cantidad de picos de flexibilidad (figura 13E) que presenta, se puede proponer que esta proteína tenga cierto movimiento. La primer región hidrofóbica (figura 13A y B) presenta valores muy altos por lo que se cree, que esta es la región que forma la parte central de la proteína, alrededor de la cual se estructura el resto de la molécula.

La proteína prM es principalmente hidrofílica lo cual esta relacionado con el hecho de que es un peptido que se "pierde" antes del ensamblaje viral, esto indica que posiblemente sea liberado al citoplasma o sea secretado al exterior celular cuando el virus sale al medio extracelular.

La proteína M es claramente de membrana como se pudo constatar por los algoritmos específicos para este fin. Esta proteína presenta dos zonas bien definidas: una hidrofílica que se propone sea la región de interacción con la proteína C, y un segmento hidrofóbico que se postula sea el segmento que tiene la interacción con la proteína E.

#### ESTRUCTURA SECUNDARIA

La utilización conjunta de varios métodos de predicción de estructura secundaria nos dio la posibilidad de aumentar del 10 al 15 % la exactitud de la misma con respecto al porcentaje de los programas utilizados en forma individual, logrando acercar nuestra predicción a un 65 a 70 % de exactitud.

El análisis de nuestros datos dio los siguientes resultados:

C. Es una proteína soluble con un índice de hidropatía (GRAVY [GRand AVERAGE of HYdropathy])<sup>1</sup> de -1. La cual pertenece a la familia  $\alpha/\beta$ , esto es que posee más del 50 % de la suma de hélice  $\alpha$  y estructura  $\beta$ , la cual tiene una alternancia de estas estructuras ( $\alpha$ ,  $\beta$ ). Llama la atención el elevado porcentaje de ENR (31 %), del cual el 25 % corresponde a zonas muy hidrofóbicas, lo que sugiere la posibilidad de que se encuentren en el interior de la proteína.

prM. Es una proteína soluble con un índice GRAVY de - 7.17, pertenece a la familia de proteínas  $\alpha/\alpha$ . Esto es que posee más del 35 % de hélice  $\alpha$  y menos del 15 % de estructura  $\beta$ . Su principal característica son grandes zonas hidrofílicas las cuales están principalmente en giros y ENR. Posee casi el 50 % de giros y ENR lo que le da potencialmente una gran flexibilidad.

M. Proteína de membrana con un índice GRAVY de 4.5, pertenece a la familia de proteínas  $\alpha+\beta$ . Esto indica que tiene agrupaciones de hélice  $\alpha$  en un extremo de la molécula y agrupaciones de estructura  $\beta$  en el otro extremo. Posee hélices  $\alpha$  muy hidrofóbicas que pueden corresponder a los segmentos transmembranales y regiones hidrofílicas en ENR, con gran flexibilidad, lo que permitiría su interacción con otras moléculas.

Los perfiles fisicoquímicos son una herramienta importante en la correlación de la predicción de estructura secundaria de una proteína, con la predicción de su función, mediante los paráme-

<sup>1</sup> Los valores negativos del índice de hidropatía son valores hidrofílicos

tros propuestos. nos permitio por ejemplo verificar, si las regiones hidrofílicas de la proteína presentaban las condiciones necesarias para ser consideradas como regiones antigénicas. Este tipo de análisis ha sido propuesto principalmente por Hoop/Wood (39) y por Welling (45) en base a los estudios sobre proteínas ya caracterizadas.

En referencia al aislado viral analizado cabe mencionar que éste nos fue enviado como una cepa probablemente variante a nivel antigénico, en base al antecedente de que reaccionaba con anticuerpos monoclonales contra serotipo 2 y no era neutralizado por anticuerpos provenientes de líquido ascítico de ratón, dirigidos contra la cepa prototipo de referencia para dengue 2 (Nueva Guinea C). Posteriormente Monath y col (61) encontraron que esta cepa mexicana era antigénicamente única por radioinmunoanálisis, utilizando anticuerpos monoclonales que reconocen diferentes epitopes del serotipo 2, analizaron la curva de unión con un modelo matemático, y con predicciones estadísticas que determinaron que la cepa en estudio está relacionada con una "firma" (signature) común, localizaron además los epitopes diferentes. Los autores encontraron que tanto con la cepa Nueva Guinea C, como con cepas aisladas en Puerto Rico y el Caribe, difería por lo menos en 3 epitopes, y de todos los demás topotipos de dengue 2 por lo menos en 5 epitopes (comunicación personal). Esta técnica fue comparada por estos mismos autores con el análisis de "huella digital" de ARN (fingerprint), observando que este virus presentaba un patrón genómico parecido a los aislados de dengue 2 antes mencionados. El hecho de que difiera el análisis de firma

(antigénico) y de huella digital de ARN, no es sorprendente, ya que los genes que codifican las proteínas estructurales comprenden el 15 % del total del genoma y en la técnica de huella digital del ARN se examina sólo del 10 al 15 % del ARN total, sin que signifique que es la misma región la que se analiza, además de que no revela diferencias a nivel nucleotídico/antigénico o de localización en el genoma.

#### CONCLUSION

La secuencia nucleotídica obtenida para dengue 2 mexicano posee la misma organización general del genoma que los flavivirus reportados a la fecha.

La secuencia deducida de aa tiene gran similitud con diferentes aislados del serotipo 2. De igual manera el porcentaje de codones y bases así como sus perfiles fisicoquímicos son muy similares.

La estructura secundaria de cada una de estas proteínas nos permitió ubicarlas en las diferentes familias estructurales y conocer algunas de sus propiedades fisicoquímicas

## LISTA DE FIGURAS

- TABLA 1 Complejos antigénicos de los flavivirus definidos por ensayos de neutralización cruzada con antisueros policlonales.
- FIGURA 1 Grafica del número casos de dengue en Mexico en distintos años.
- FIGURA 2 Mapa de la distribución de los distintos serotipos virus del dengue en las Americas.
- FIGURA 3 Representación esquemática del fenómeno de facilitación por anticuerpos, de los mecanismos efectores y de amplificación que participan en la patogénesis de la FHD/SCD.
- FIGURA 4 Microfotografía del virus del dengue.
- FIGURA 5 Modelo hipotético de la organización de los componentes del virus del dengue: lípidos, proteínas virales estructurales y ARN.
- FIGURA 6 Esquema de la estructura general del genoma del virus del dengue.
- DIAGRAMA 1 Estrategia para la predicción de la estructura secundaria de las proteínas C, prM y M.
- FIGURA 7 Mapa de la sonda C38 clonada en pGEMI. Análisis del tamaño de los insertos.
- FIGURA 8 a) Clonas positivas para proteínas estructurales del virus del dengue. b) Análisis del tamaño de los insertos de las clonas positivas para proteínas estructurales.
- FIGURA 9 Estrategia de secuenciación de la región estructural de dengue 2 mexicano.
- FIGURA 10 Secuenciación nucleotídica de la clona 5 y un ADNc amplificado por PCR.

FIGURA 11 Secuencia de nucleótidos y secuencia deducida de aminoácidos.

TABLA 2 Composición de bases, codones y aminoácidos de la región estructural del aislado de dengue 2 mexicano.

FIGURA 12 Secuencia nucleotídica y deducida de aa de la proteína de nucleocápside o C.

FIGURA 13 Graficas de análisis fisicoquímicos de la proteína C.  
a) hidropatía. b) hidrofobicidad. c) hidrofiliicidad. d) antigenicidad. e) flexibilidad.

FIGURA 14 Secuencia de nucleotidos y aa de las proteínas, precursor de membrana o prM y de membrana o M.

FIGURA 15 Graficas de los perfiles fisicoquímicos de la proteína prM. a) hidropatía. b) hidrofobicidad. c) hidrofiliicidad. d) antigenicidad. e) flexibilidad.

FIGURA 16 Graficas de los perfiles fisicoquímicos de la proteína M. a) hidropatía. b) hidrofobicidad. c) hidrofiliicidad. d) antigenicidad. e) flexibilidad. f) hélice de membrana (Rao/Argos). g) hélice de membrana (Klein y Col).

FIGURA 17 Alineación múltiple de aa de la región estructural de diferentes serotipos.

TABLA 3 Porcentaje de predicción de estructura secundaria por los distintos métodos y la predicción conjunta.

FIGURA 18 a) Predicción por cada uno de los métodos para hélice  $\alpha$ , lámina  $\beta$  y para giro  $\beta$ . Análisis fisicoquímicos de la proteína C. b) Tabla de predicción de cada una de las estructuras predichas de la proteína C con los comentarios fisicoquímicos de las mismas.

FIGURA 19 Esquema de la estructura secundaria de la proteína C.

FIGURA 20 a) Predicción por cada uno de los métodos para hélice  $\alpha$ , lámina  $\beta$  y para giro  $\beta$  en la proteína prM. Tabla de estructuras predichas con los comentarios fisicoquímicos de cada estructura.

FIGURA 21 Esquema de la estructura secundaria de la proteína prM

FIGURA 22 a) Predicción por cada uno de los métodos para hélice  $\alpha$  lámina  $\beta$  y para giro  $\beta$  en la proteína M. b) Tabla de estructuras predichas con los comentarios fisicoquímicos de cada estructura.

FIGURA 23 Esquema de la estructura secundaria de la proteína M.

FIGURA 24 Grafica de la clasificación en distintas familias.



ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA

BIBLIOGRAFIA

- 1 Westaway, E.G., Brinton, M.A., Gaidamovich, S.Y., Horzinek, M.C., Igarashi, A., Kaariainen L., Ivov, D.K., Porterfield, J.S., Rusell, P.K. y Trent, D.W. 1986. Flaviviridae. Inter-virology. 24:183-187.
- 2 Monath, T.P. 1986. Pathobiology of the flaviviruses. In: Schlesinger S., Schlesinger M.J. eds. The Togaviridae and Flaviviridae. New York: Plenum Press. : 375-440.
- 3 Storer, T.I., Usinger, R.L., Stebbins, R.C. Nybakken, J.W. 1982. Zoologia General. Barcelona, España. Ed. Omega: 587-638.
- 4 Trent, D.W., Naeve, C.W. 1980. Biochemistry and replication. In: Monath, T.P. St Louis Encephalitis. Washington, D.C.: American Public Health Association. 159-199.
- 5 Deubel, V. Crouset, J., Benichas, D., Digoutte, J.P. Bouloy, M., and Girard, M. 1983. Preliminary characterization of the ribonucleic acid of Yellow Fever virus. Ann Virol 134E:581-588.
- 6 Westaway, E.G. 1980. Replication of Flaviviruses. In: Schlesinger, S.W. ed. The Togaviruses. New York. Academic Press. 531-581.
- 7 Rice, C.M., Lenches, E.M., Dalgardo, L., Eddy, S.R., Shin, S.J., Sheets, R.L., Trent, D.W., y Strauss, J.S 1985. Nucleotide Sequence of Yellow Fever virus: Implications for Flavivirus gene expression and evolution. Science. 229: 726-733.

- 8 Wengler, G., Wengler, G. 1981. Terminal Sequences of the Genome and Replicative-form RNA of the Flavivirus West Nile Virus. Virology. 113: 544-555.
- 9 Cleaves, G.R., Dubin, D.T. 1979. Methylation status of intracellular dengue type 2 40S RNA. Virology. 96: 159-165.
- 10 Brinton, M.A., Fernández, A.V. y Amato, J. 1986. The 3'-Nucleotides of Flavivirus genomic RNA form a conserved secondary structure. Virology. 153: 113-121.
- 11 Brinton, M.A., Dispoto, J.H. 1988. Sequence and secondary structure analysis of the 5' terminal region of Flavivirus genome RNA. Virology. 162: 290-299.
- 12 Hahn, C.S., Hahn, Y.S. y Rice, C.M. 1987. Conserved element in the 3' untranslated region of Flavivirus RNA and potential cyclization sequences. J. Mol. Biol. 198: 33-41.
- 13 Rice, C.M. 1986. Structure of the Flavivirus genome. In: Schlesinger, S., Schlesinger, M.J. eds. The Togaviridae and Flaviviridae. New York. Plenum Press. 279-326.
- 14 Halstead, S.B. 1988. Pathogenesis of dengue: Challenges to molecular Biology. Science. 239: 476-481.
- 15 Rosen, L. 1977. The emperor's new clothes revisited, or reflections on the pathogenesis of dengue hemorrhagic fever. Amer. J. Trop. Med. Hyg. 26 (3): 337-343.
- 16 Guzman, M.G., Kouri, M.G., Bravo, J. 1990. Dengue Hemorrhagic Fever in Cuba, 1981-A Retrospective Seroepidemiologic Study. Am. J. Trop. Med. Hyg. 42(2): 179-184.

- 17 Schwartz, R.M. & Dayhoff. 1978. Matrices for detecting distant relationships. In: Atlas of protein sequence and structure. Vol. 5. Dayhoff M.O. ed. Natl. Biomed. Res. Found. Washington, D.C. 353-358.
- 18 Brinton, M.A. 1986. Replication of flaviviruses. In: Schlesinger, S., Schlesinger, M.J. eds. The Togaviridae and Flaviviridae. New York. Plenum Press. 327-374.
- 19 Dalgardo, L., Trent, D.W. Strauss, J.H. Rice, C.M. 1986. Partial nucleotide sequence of the Murray Valley Encephalitis Virus genome. J. Mol. Biol. 184: 99-105.
- 20 Von Heijne, G. 1985. Signal sequences. The limits of variation. J. Mol. Biol. 184: 99-105.
- 21 Bell, J.R., Kinney.R.M., Trent, D.W., Lenches, E.M., Dalgardo, L., Strauss, J.H. 1985. Amino-terminal aminoacid sequences of structural proteins of three flaviviruses. Virology. 143: 224-229.
- 22 Zhao, B. Macokow, .A., Buckler-White, L., Markoff, R.M., Chanock, C., Laic, J y Makino, Y. 1986. Cloning full-length dengue type 4 DNA sequenses: Analysis of genes coding for structural proteins. Virology. 155: 77-88.
- 23 Markoff, L. 1989. In vitro processing of dengue virus structural proteins: Cleavage of the pre-membrane protein. J. Virol. 63 (8): 3345-3352.
- 24 Kaufman, B.M. Summer, P.L., Dubois, D.R., Houston-Cohen, W. Gentry, M.K., Timchak, R.L., Burke, D.S. y Eckels, K.H. 1989. Monoclonal antibodies for dengue virus prM glycoprotein protect mice against lethal dengue infection. Am. J. Trop. Med. Hyg. 41 (5): 576-580.

- 25 Castle, E., Nowak, T., Leidner, V. Wengler, G. y Wengler, G. 1985. Sequence analysis of the viral core protein and the membrane-associated proteins V1 and V2 of the Flavivirus West Nile Virus and of the genome sequence for these proteins. Virology. 145: 227-236.
- 26 Randolph, V.B., Winkler, G. y Stollar, V. 1990. Acidotropic amines inhibit proteolytic processing of Flavivirus prM protein. Virology. 174: 450-458.
- 27 Sanger, F., Miklen, S. y Coulson, A. R. 1977. DNA sequencing with chain termination inhibitors. Proc. Nat. Acad. Sci. USA. 74: 5463-5467.
- 28 Tabor, S. y Richardson, C.C. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Nat. Acad. Sci. USA. 84 (14): 4767-4771.
- 29 M.J.Rooman & S.J.Wodak (1988). Identification of predictive sequence motif limited by protein structure data base size. Nature. 335, 45-49.
- 30 Chou P.Y., Fasman G.D. 1978. Prediction of the secondary structure of proteins from their amino acid sequence Adv. Enzym. 47: 45-148.
- 31 Chou P.Y., Fasman G.D. 1979. prediction of  $\beta$  turns. Biophys. J. 26:367-384.
- 32 Garnier, J., Osguthorpe, D.J., Robson, B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97-120.

- 33 Gascuel, O., Goldmar, J. L. 1988. A simple method for predicting the secondary structure of globular proteins: implications and accuracy. CABIOS 4: 357-365.
- 34 Deleage, G. y Roux, B. 1987. An algorithm for protein secondary structure prediction based on class prediction Protein Engineering. 1 (4): 289-294.
- 35 Levitt, M. 1978. Conformational preferences of aminoacids in globular proteins. Biochemistry. 17: 4277-4285.
- 36 Rao, M.J.K. y Argos, P. 1986. A conformational preference parameter to predict helices in integral membrane proteins. Biochem. Biophys. Acta. 869: 197-214.
- 37 Klein, P., Kanehisa, M., Delisi, C. 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta. 815: 468-476.
- 38 Eisenberg, D., Schwarz, E., Komarony, M., Wall, R. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179: 125-142.
- 39 Hopp, T.P., Woods, K.R. 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA. 78: 3824-3828.
- 40 Miyazawa, S., Jernigen, R.L. 1985. Estimation of effective intraresidue contact energies from protein cristal structures: Quasi-Chemical approximation. Macromolecules. 18: 534-552.
- 41 Kyte, J. y Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein J. Mol. Biol. 157: 105-132.

- 42 Karplus, P.A. y Schulz, G.E. 1985. Prediction of chain flexibility in proteins. Naturwissenschaften 72: 212-213.
- 43 Van Regenmortel, M.H.V. 1986. Which structural features determine protein antigenity. TIBS 11:36-39
- 44 Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. Science. 185: 862-864.
- 45 Welling, G.W., Weijer, W.J., Van der Zee, R., Welling-Wester, S. 1985. prediction of sequential antigenic region in proteins FEBS. Letters. 188 (2): 215-218.
- 46 Nakashima, H., Nishikawa, K., Ooi, T. 1986. The folding type of a protein is relevant to the aminoacid composition J. Biol. Chem. 99: 153-162.
- 47 Richardson, J.S. y Richardson, D.C. 1988. Amino acid preferences of specific locations at the end of the  $\alpha$  helices. Science 240: 1648-1652.
- 48 Myer, E.W. y Miller, W. 1988. Optimal alignment in linear space. CABIOS. 4 (1): 11-17.
- 49 Monath, T.P. Flaviviruses.1990. In Virology. Fields, B.N. y Knipe, D.M. ed. USA. Raven Press: 763- 814.
- 50 Okuno, T., Okada, T., Kondo, A., Susuki, M. y Oya, A. 1968. Immunotyping of different strains of japanese encephalitis virus by antibody absorption, haemagglutination-inhibition and the complement fixation test. Bull. WHO. 38: 547-563
- 51 Clarke, D.H. 1960. Antigenic analysis of certian group B artrophod borne viruses by antibody absorption. J. Exp. Med. 111: 21-32.

- 52 Monath, T.P., Cropp, C.B., Bowen, G.S. y Gardner, B. 1980. Variation in virulence for mice and rhesus monkey among St Louis encephalitis strains of different origin. Am. J. Trop. Med. Hyg. 29: 948-962.
- 53 Carmenes, R.S. 1989. A program for protein structure prediction. Biophys. Res. Comm. 159: 687-693.
- 54 Croft, A.R. 1987. A set of programs to aid in the prediction of protein structure from sequence data by anatomy croft. Copyright, A.R. Crofts, University of Illinois.
- 55 Moore, J., Engelberg, A. y Bairoch, A. 1988. Using PC/GENE for protein and nucleic acid analysis. Biotechniques. 6: 566-572.
- 56 Pustell, J. y Kafatos, F.C. 1982. A high speed, high capacity homology matrix: Zooming through SV 40 and Polyoma Nucleic. Acid. Res. 10 (15): 4765-4782.
- 57 Hahn, Y.S., Galler, R., Hunkapiller, T., Dalrymple, J.M., Strauss, J.H. y Strauss, E.G. 1988. Nucleotide sequence of Dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. Virology. 162: 167-180.
- 58 Taylor, W. R. y Thornton, J. M. 1984. Recognition of super secondary structure in proteins. J. Mol. Biol. 173: 487-514.
- 59 Thornton, J.M y Chakauya, B.L. 1982. Conformation of the terminal region in protein Nature. 298: 296-297.
- 60 Altamirano, M.M., Plumbridge, A.P., Hernández-Arana, A. y Calcagno, M. 1991. Secondary Structure of Escherichia coli glucosamine-6-phosphate deaminase from amino acid sequence and circular dichroism spectroscopy. Biochem. Biophys. Acta. 1076: 266-272.

- 61 Monath, T.P., Wands, J.R. Hill, L.J., Brown, N.V., Marciniak, R.A., Mong, M.A., Gentry, M.K., Burke, D.S., Grant, J.A. y Trent, D.W. 1986. Geographic classification of dengue-2 virus strains by antigen signature analysis. *Virology*. 154: 313-324.
- 62 Mullis, K.B. y Faloona, F.A. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Meth. Enzymol. 155: 355-350.
- 63 Ruiz, B. 1991. Estructura primaria y predicción de la estructura secundaria de las proteínas estructurales del virus del dengue, genotipo-2 mexicano. Tesis Doctoral. Instituto de Investigaciones Biomedicas. U.N.A.M. 121 pp.
- 64 Fasman, G. 1989. The development of the prediction of protein structure. In: Fasman, G. ed. Prediction of protein structure and the principles of protein conformation. Plenum press. New York. 193-316.
- 65 Taylor, W.R. 1987. Protein structure prediction. In: Bishop, M. J. & Rawling, C. J. eds. Nucleic acid and protein sequence analysis a practical approach. IRL Press. Oxford. 285-322.
- 66 Argos, P. 1989. Predictions of protein structure from gene and amino acid sequences. In Creighton, T. E. ed. Protein structure a practical approach. IRL Press. Oxford. 169-190.
- 67 Thornton, J. M. y Taylor, W. R. 1989. Structure prediction. In: Findlay, J. B. C. & Geisow, M. J. eds. Protein sequencing a practical approach. IRL Press. Oxford. 147-190.