



INSTITUTO POLITÉCNICO NACIONAL

ESCUELA NACIONAL DE CIENCIAS BIOLÓGICAS
SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN
MAESTRIA EN BIOMEDICINA Y BIOTECNOLOGÍA
MOLECULAR

ESTUDIO EVOLUTIVO, ESTRUCTURAL Y FUNCIONAL DE ALGUNAS PROTEÍNAS RELACIONADAS CON EL CÁNCER HUMANO

T E S I S

QUE COMO UNO DE LOS REQUISITOS PARA
OBTENER EL GRADO DE MAESTRIA EN CIENCIAS EN
BIOMEDICINA Y BIOTECNOLOGÍA MOLECULAR.

PRESENTA:

Q.C. GEMA LIZBETH RAMÍREZ SALINAS

DIRECTOR DE TESIS:

DR ALFONSO MÉNDEZ TENORIO.



DICIEMBRE 2010



INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México siendo las 13:00 horas del día 26 del mes de noviembre del 2010 se reunieron los miembros de la Comisión Revisora de Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de la ENCB-IPN para examinar la tesis titulada:

Estudio evolutivo, estructural y funcional de algunas proteínas relacionadas con el cáncer humano

Presentada por la alumna:

Ramírez

Apellido paterno

Salinas

Apellido materno

Gema Lizbeth

Nombre(s)

Con registro:

B	0	8	1	5	5	5
---	---	---	---	---	---	---

aspirante de:

Maestría en Ciencias en Biomedicina y Biotecnología Molecular

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Director de tesis

Dr. Alfonso Méndez Tenorio

Dr. José Luis Muñoz Sánchez

Dr. Sergio Enrique Meza Toledo

Dr. Juan Carlos Santiago Hernández

Dr. Rogelio Maldonado Rodríguez

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Manuel Jesús Piñón López



Escuela Nacional de Ciencias Biológicas
Sección de Estudios de Posgrado e Investigación



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO


CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 7 del mes diciembre del año 2010, la que suscribe Gema Lizbeth Ramírez Salinas alumna del Programa de Biomedicina y Biotecnología Molecular con número de registro B081555, adscrito a la Escuela Nacional de Ciencias Biológicas, manifiesta que es autora intelectual del presente trabajo de Tesis bajo la dirección de Dr. Alfonso Méndez Tenorio y cede los derechos del trabajo intitulado: "Estudio evolutivo, estructural y funcional de algunas proteínas relacionadas con el cáncer humano", al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección amendezt@ipn.mx. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Atentamente,


Gema Lizbeth Ramírez Salinas
Nombre y firma de la alumna


Dr. Alfonso Méndez Tenorio
Nombre y firma del Director de tesis

Este trabajo se realizó en el Laboratorio de Bioinformática y Biotecnología genómica del Departamento de Bioquímica de la Escuela Nacional de Ciencias Biológicas, bajo la dirección del Dr. Alfonso Méndez Tenorio, con el apoyo económico del Consejo Nacional de Ciencia y Tecnología.

TABLA DE CONTENIDO

TABLA DE CONTENIDO	<i>i</i>
LISTA DE FIGURAS	<i>iv</i>
LISTA DE TABLAS	<i>v</i>
ÍNDICE DE ABREVIATURAS	<i>vi</i>
RESUMEN	<i>viii</i>
ABSTRACT	<i>ix</i>
1. INTRODUCCIÓN	<i>1</i>
1.1 DESARROLLO DEL CÁNCER	<i>1</i>
1.2 PROTO-ONCOGENES Y GENES SUPRESORES DE TUMORES	<i>1</i>
1.2.1. Oncogenes	<i>2</i>
1.2.3. Genes supresores de tumores	<i>3</i>
1.3 GEN TP53 EN LA ONCOGÉNESIS	<i>3</i>
1.3.1. p53 y mutaciones	<i>5</i>
1.3.1 Activación de la proteína p53	<i>6</i>
1.3.2 p53 y control del ciclo celular	<i>7</i>
1.3.3. Proteína p53 y reparación	<i>8</i>
1.4. FAMILIA RAS	<i>9</i>
1.5. EL GEN CDKN2A	<i>13</i>
1.6. EL GEN ABL1	<i>15</i>
1.7. BIOINFORMÁTICA	<i>17</i>
1.8 ESTUDIO DEL IMPACTO DE LAS MUTACIONES EN LA ESTRUCTURA / FUNCIÓN DE P53	<i>20</i>
1.9 JUSTIFICACIÓN	<i>21</i>
1.10 OBJETIVO	<i>22</i>

1.10.1 Objetivo general	22
1.10.2 Objetivos específicos	22
10.11 HIPÓTESIS	23
2. MATERIALES Y MÉTODOS	24
2.1. Recursos bioinformáticos	24
2.2 Obtención de las secuencias	26
2.3 Búsqueda de proteínas homólogas	26
2.4 Alineamiento de secuencias	26
2.5 Identificación de regiones conservadas	27
2.6 Elaboración de árboles filogenéticos	27
2.7 Búsqueda de estructuras de las proteínas	27
2.8 Modelado de la estructura tridimensional	27
2.9 Extracción de la información de estructura secundaria y otros parámetros estructurales	28
2.10 Búsqueda de dominios estructurales/funcionales	28
2.11 Búsqueda de mutaciones asociadas con cáncer	28
2.12 Modelado de las proteínas mutantes	29
2.13 Análisis del impacto mutacional	29
2.14. Diagrama de flujo	32
3. RESULTADOS	33
3.1 Análisis evolutivo de la proteína silvestre e identificación de las secuencias conservadas	34
3.1.1. Alineamiento de Hras, Kras y Nras	34
3.1.2. Alineamiento de Abl1	36
3.1.3. Alineamiento de p53	36
3.1.4. Alineamiento de p16	38
3.1.5. Árbol filogenético de Hras, Nras y Kras	39
3.1.6. Árbol filogenético de Abl1	42
3.1.7. Árbol filogenético de p53	43

.8.1.8. Árbol filogenético de p16 _____	44
3.2. Localización de mutaciones, patrones mutagénicos e identificación del tipo de cambio	44
3.2.1. Localización de las regiones conservadas y no conservadas de Hras _____	45
3.2.2 Patrones mutagénicos _____	52
3.2.3. Identificación del tipo del cambio _____	66
3.4. Análisis estructural de la proteína silvestre con sus mutantes _____	66
4. DISCUSION _____	79
5. CONCLUSIONES _____	86
6. PERSPECTIVAS _____	87
7. REFERENCIA _____	88
8. ANEXOS _____	94
8.1. Anexo I _____	94
8.2. Anexo II _____	95
8.3. Anexo III _____	96
8.4. Anexo IV _____	97
8.5. Anexo V _____	98
8.6. Anexo VI _____	100
8.7. Anexo VII _____	102
8.8. Anexo VIII _____	104
8.9. Anexo IX _____	106
8.10 Anexo X _____	109

LISTA DE FIGURAS

Figura 1. Dominios estructurales de la proteína p53 en humano.	4
Figura 2. Mutaciones puntuales en la proteína p53:	6
Figura 3. Proteína p53 y el ciclo celular.	8
Figura 4: Región hipervariable de las proteínas RAS.	10
Figura 5. Activación del oncogén RAS.	11
Figura 6. Dominios conservados de las proteínas Ras	12
Figura 7. Los productos del gen CDKN2A	13
Figura 8. Mutación y dominios de p16	14
Figura 9. Dominios estructurales de Abl y Arg	16
Figura 10. Alineamiento de las proteínas de la superfamilia Ras (Hras, Kras y Nras)	35
Figura 11. Alineamiento de la proteína Abl1	36
Figura 12. Alineamiento de la proteína p53	37
Figura 13. Alineamiento de la proteína p16	38
Figura 14. Fragmento 1 del árbol filogenético de las proteínas Hras, Kras y Nras	40
Figura 15. Fragmento 2 del árbol filogenético de las proteína Hras, Kras y Nras	41
Figura 16. Árbol filogenético de la proteína Abl1	42
Figura 17. Árbol filogenético de la proteína p53	43
Figura 18. Árbol filogenético de la proteína p16	44
Figura 19. Regiones conservadas y variables de Hras	46
Figura 20. Regiones conservadas y variables de Kras	47
Figura 21. Regiones conservadas y variables de Nras	48
Figura 22. Regiones conservadas y variables de Abl1	49
Figura 23. Regiones conservadas y variables de p53	50
Figura 24. Regiones conservadas y variables de p16	51
Figura 25. Patrones mutagénicos de Hras	54
Figura 26. Patrones mutagénicos de Kras	55
Figura 27. Patrones mutagénicos de Kras	56
Figura 28. Patrones mutagénicos de Abl1 isoforma A (Fragmento 1)	58
Figura 29. Patrones mutagénicos de de Abl1 isoforma A (Fragmento 2)	59
Figura 30. Patrones mutagénicos de de Abl1 isoforma A	60
Figura 31. Patrones mutagénicos de de Abl1 isoforma A	61

Figura 32. Patrones mutagénicos de p16	63
Figura 33. Patrones mutagénicos de p53	65
Figura 34. RMS Posicional de la mutante HrasV81M	69
Figura 35. RMS posicional de Nras y NrasA59T	69
Figura 36. RMS posicional Abl1 y Abl1IsoAQ252E	70
Figura 37. RMS posicional de la mutante KrasIsoBT35I	71
Figura 38. RMS posicional de p53 y p53T155P	71
Figura 39. RMS posicional de p16 y p16P48R	72
Figura 40. Comparación del potencial dope entre Hras y HrasG12C	73
Figura 41. Comparación del potencial dope entre Nras y NrasQ61E	73
Figura 42. Comparación del potencial dope entre Kras y KrasIsoBE63K	74
Figura 43. Comparación del potencial dope entre Abl1 y Abl1IsoAG250E:	75
Figura 44. Comparación del potencial dope entre p53 y p53C242	75
Figura 45. Comparación del potencial dope entre p16 y p16D108H	76
Figura 46. Comparación de energía de torsión entre Hras y HrasG12R	77
Figura 47. Comparación de energía de torsión entre Nras y NrasG13R	78

LISTA DE TABLAS

Tabla1. Oncogenes más importantes:	2
Tabla2: Genes supresores tumorales más importantes	3
Tabla 3. Bases de datos primarias y especializadas	24
Tabla 4. Programas bioinformáticos	25
Tabla 5. Proteínas de estudio	33
Tabla 6. Proteínas modeladas	67

ÍNDICE DE ABREVIATURAS

ABL1, Oncogen Abl1

Abl1, Proteína Abl1

AKT, Oncogen v-akt

AMPc, Monofosfato de adenosina cíclico

ANK, Anquirina

APC, Gen APC (de sus siglas en inglés “adenomatous polyposis coli”)

BAX, Gen BAX (de sus siglas en inglés “BCL2-associated X protein”)

BCL2, Gen BCL2 (de sus siglas en inglés “B-cell CLL/lymphoma 2”)

BRCA1, Gen BRCA1 (de sus siglas en inglés “breast cancer 1”)

CDK4, Gen cinasa dependiente de ciclina 4

cdk4, Proteína cinasa dependiente de ciclina 4

CDKN1A, Gen Inhibidor 1A de cinasa dependiente de ciclina

CDKN2A, Gen Inhibidor 2A de cinasa dependiente de ciclina

CDKN2B, Gen Inhibidor 2B de cinasa dependiente de ciclina

c-erbB2, Gen c-erbB2 (de sus siglas en inglés “v-erb-b2 erythroblastic leukemia viral oncogene homolog 2i”)

DB, Base de datos

DNA, Ácido desoxirribonucleico

E2F, Factor de transcripción E2F

EGF, Factor de crecimiento epidérmico

G6PD, Glucosa 6 fosfato deshidrogenasa

GAP, 3 fosfatos de gliceraldehido

GBD, Dominio de unión al glucogeno

GDI, Inhibidor de disociación GDP

GDP, Difosfato de guanosina

GEF, Factor intercambiador de nucleótido de guanina

GTP, Trifosfato de guanosina

HRAS, Oncogen Hras (de las siglas “v-Ha-ras Harvey rat sarcoma viral oncogene homolog”)

Hras, Proteína GTPase Hras

hsp70, Choque térmico de 70kDa

HVR, Región hipervariable

KRAS, Oncogen Kras (de las siglas “v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog”)

Kras, Proteína GTPase Kras

LTA, Receptor de linfotoxina alfa

MAPK, Proteínas cinasas activadas por nitrógenos

MDM2, Gen MDM2 (de sus siglas en inglés "murine doble minute 2")

mdm2, Proteína Mdm2 (de sus siglas en inglés "murine doble minute 2")

MET, Gen MET (“de sus siglas en inglés “hepatocyte growth factor receptor”)

MYC, Gen MYC (“de sus siglas en inglés “v-myc myelocytomatosis viral oncogene homolog”)

NER, Reparación por escisión de nucleótidos

NRAS, Oncogen Nras (de sus siglas en inglés “Neuroblastoma RAS viral (v-ras) oncogene homolog”)

Nras, Proteína GTPase NRas

p14^{ARF}, Proteína p14 isoforma 1

p16, Proteína p16 isoforma 1

p53, Proteína p53 isoforma a

PDB, Base de datos de proteínas

pRb, Proteína retinoblastoma

PTEN, Gen PTEN (de sus siglas en inglés “phosphatase and tensin homolog”)

PTKc_Abl, Dominio catalítico de la proteína tirosina cinasa

RAF, Gen RAF (de sus siglas en inglés “v-raf-1 murine leukemia viral oncogene homolog 1”)

RAS, Familia de genes Ras

Ras, Familia de proteínas Ras codificada por los genes RAS

RB1, Gen retinoblastoma 1

RMSD, Desviación de la raíz cuadrada de las medias

RPA, Proteína de replicación A1

TCL, Lenguaje de herramientas de comando

TERT, Gen TERT (de sus siglas en inglés “telomerase reverse transcriptase”)

TP53, Gen que codifica a la proteína de 53 kDa p53

VMD, Dinámica molecular visual

RESUMEN

En el presente proyecto realizamos un estudio evolutivo, funcional y estructural de las proteínas Hras, Kras, Nras, Abl1 isoforma A, p53 y p16 las cuales están relacionadas con el cáncer, este proyecto se divide en tres etapas: 1) el análisis evolutivo y funcional de la proteína silvestre e identificación de las secuencias conservadas, 2) la localización de mutaciones e identificación del tipo de cambio y 3) el análisis comparativo de la estructura de las proteínas silvestres con respecto a las mutantes. La primera etapa consiste en el análisis de la historia evolutiva de cada una de las proteínas silvestres estudiadas observando cómo dichas secuencias de proteínas evolucionaron, y encontrando residuos con un alto grado de conservación que son esenciales para la función biológica, así como regiones variables. En la segunda etapa se basa primordialmente en la relación entre la estructura primaria de la secuencia de las proteínas silvestres (destacando dominios funcionales o estructurales, así como el grado de conservación de cada uno de los residuos de la secuencia silvestre) con las posiciones de las mutaciones relacionadas con el cáncer humano, encontrando patrones mutagénicos interesantes en los cuales se destaca que muchas posiciones de residuos que interactúan directamente en sitios funcionales de la proteína silvestre son las posiciones de los aminoácidos en donde ocurren mutaciones relacionadas con el cáncer, también se encontró otro patrón mutagénico en el cual las posiciones de los residuos donde ocurren mutaciones relacionadas se encuentran distribuidas en toda la estructura primaria, sin embargo la mayoría de estas posiciones de los residuos se encuentran localizados en dominios estructurales. La última etapa consiste en el estudio en el impacto de las mutaciones puntuales por sustitución de un solo aminoácido relacionadas con el cáncer en la estructura terciaria de las proteínas, a partir de la estructura tridimensional de la proteína silvestre, se obtienen *in silico* la estructura tridimensional de todas las mutantes estudiadas mediante un script especializado, se compararon la estructura tridimensional de la proteína silvestre con cada una de las mutantes. Obteniendo comparaciones de RMSD, RMS posicional, potencial DOPE, energía de los ángulos de torsión, en dichas comparaciones se obtuvo información interesante en el impacto que tienen las mutaciones puntuales por sustitución un solo residuo que están relacionadas con el cáncer.

ABSTRACT

In this project we conducted evolutive, functional and structural study of proteins Hras, Kras, Nras, Abl1 isoform A, p53 and p16, which are related to human cancer, this project is divided into three stages: 1) the evolutionary and functional analysis of the wild-protein and identification of conserved sequences, 2) the location of mutations and identification of the exchange rate and 3) the comparative analysis of wild-protein structure with respect to mutants. The first step is the analysis of the evolutionary history of each of the wild-proteins studied watching these protein sequences have evolved, and finding waste with a high degree of conservation which are essential for biological function and as well as variable regions. In the second stage is based primarily on the relationship between the primary structure of the wild-protein sequence (emphasizing structural or functional domains, as well as the degree of conservation of each residue of the wild-protein sequence) with the positions of mutations associated with human cancer, finding mutagenic patterns are interesting which notes that many residue positions that interact directly in functional sites of the wild-protein are amino acid positions where mutations occur related to human cancer, also found another mutagenic pattern in which the residue positions where mutations occur related are distributed throughout the primary structure, however most of these residue positions are located in structural domains. The last stage is to study the impact of point mutations by substitution of a single amino acid related to human cancer in the tertiary structure of proteins from the three-dimensional structure, of wild-protein obtained *in silico* structure 3D of all mutants studied by a specialized script, we compared the three-dimensional structure of wild-protein for each of the mutants. Obtaining the comparisons of RMSD, RMS positional, DOPE potential and, energy of the torsion angles. In these comparisons was obtained interesting information on the impact of point mutations by a single residue substitution that is related to human cancer.

1. INTRODUCCIÓN

1.1 DESARROLLO DEL CÁNCER

A nivel celular, el desarrollo del cáncer es visto como un proceso en el que participan varias mutaciones que aumentan progresivamente la capacidad de proliferación, la supervivencia, la invasión y la metástasis celular. Se piensa que el primer paso en el proceso de iniciación del tumor, es el resultado de una alteración genética que conduce a la proliferación anormal de una única célula (Giuriato *et al*, 2003, Heeg *et al*, 2006). Algunas de estas mutaciones confieren una ventaja selectiva a la célula, tal como la de tener un crecimiento más rápido por lo que los descendientes de una célula que llevan este tipo de mutación en consecuencia, se convierten en dominantes dentro de la población tumoral. El proceso se llama selección clonal, ya que un nuevo clon de células tumorales se ha desarrollado debido a su mayor tasa de crecimiento u otro tipo de ventajas tales como la supervivencia, la invasión, o metástasis que les confieren una ventaja selectiva. La selección clonal continúa a lo largo del desarrollo del tumor, por lo que su crecimiento es cada vez más rápido y más maligno (Cooper *et al*, 2000).

1.2 PROTO-ONCOGENES Y GENES SUPRESORES DETUMORES

La transformación maligna de una célula acontece por acumulación de mutaciones en genes específicos, las cuales juegan un papel clave en la inducción de cáncer (Osborne *et al*, 2004). Estos genes están agrupados en dos grandes clases: Proto-oncogenes y Genes supresores de tumores (Kufe *et al*, 2003). Estas dos clases de genes juegan un papel predominante en el desarrollo del cáncer. En su configuración normal estos genes tienen la función de controlar el ciclo celular, es decir la serie de eventos que permite a la célula crecer y dividirse. Los proto-oncogenes y genes supresores de tumores ejercen diferentes funciones en las células normales. Estos genes controlan la proliferación celular actuando como factores de crecimiento, receptores de factores de crecimiento, transductores de la señal mitogénica y factores de transcripción, así como reguladores de la apoptosis (Kufe *et al*, 2003, Alberts *et al*, 2008). Los proto-oncogenes promueven tal crecimiento, mientras que los genes supresores lo inhiben (Lodish *et al*, 2008).

1.2.1. Oncogenes

Un oncogén codifica una proteína capaz de transformar a las células normales en cancerígenas (Weinberg *et al*, 1996, McManus *et al*, 2004, Baba *et al*, 2007). Son muchos los oncogenes conocidos, pero todos ellos son derivados de los genes presentes en células normales (es decir, proto-oncogenes), cuyos productos participan en el control de las vías de crecimiento celular. (Ver la tabla 1). La conversión o la activación de un proto-oncogén a oncogén, en general implica la ganancia de función causada por la mutación (Croce *et al*, 2008).

TABLA 1: ONCOGENES MÁS IMPORTANTES.	
ONCOGENES	FUNCION BIOLOGICA
β -catenina	Adhesión celular, controla el crecimiento y el comportamiento normal de las células.
MYC	Factores de transcripción (Proteínas nucleares).
CDK4	Controla el ciclo celular en la fase G1.
LT	Apoptosis.
MDM2	Inhibe la transactivación mediada por la proteína p53.
RTK	Tirosin cinasa de receptores de membrana.
AKT	Factor de crecimiento.
RAS	GTP-asas (Cascadas de señal mitogénica).
RAF	Serin-Treonin-cinasa (Cascadas de señal mitogénica).
ABL	Tirosin-cinasa, control de la dinámica del citoesqueleto.
<i>c-erbB2</i>	Receptor de membrana para EGF (Factor de crecimiento epidérmico).
BCL2	Senescencia y muerte celular.
TERT	Senescencia celular.
BAX	Apoptosis.
MET	Apoptosis.

Tabla1. Oncogenes más importantes: La tabla muestra una relación con los oncogenes más importantes y su función biológica (Croce et al., 2008).

1.2.3. Genes supresores de tumores

Los genes supresores de tumores ó simplemente genes supresores, controlan la proliferación celular en el organismo sano (Weinberg *et al*, 1996, Wee *et al*, 2006). Son reguladores negativos del crecimiento y cuando no están presentes en la célula o se encuentran inactivos a causa de mutaciones, las células dejan de crecer normalmente y adquieren las propiedades proliferativas anormales características de las células tumorales (Kufe *et al*, 2003). En la tabla 2 se resumen algunos de los genes supresores más importantes en el ser humano.

TABLA 2: GENES SUPRESORES TUMORALES MÁS IMPORTANTES.	
GENES SUPRESORES TUMORALES	FUNCION BIOLÓGICA
APC	Está involucrado en cánceres de colon y estómago.
CDKN2B	Regulador del ciclo celular en G1.
CDKN2A	Regulador del ciclo celular en G1.
RB1	Factor de transcripción.
PT53	Factor de transcripción.
CDKN1A	Regulador del ciclo celular en G1.
PTEN	Regulador negativo en la vía de señalización AKT.
PP24	Controla el crecimiento y la división celular.

Tabla2: Genes supresores tumorales más importantes. La tabla muestra una relación con los genes supresores de tumores más importantes y su función biológica (Croce et al., 2008).

Los oncogenes y genes supresores de tumores que se analizaron en este proyecto de tesis fueron TP53, HRAS, NRAS, KRAS, ABL1 y CDKN2A (Proteína p16) . Los cuales se describen a continuación:

1. 3 GEN TP53 EN LA ONCOGÉNESIS

El gen Tp53 o el gen p53, también llamado el "guardián del genoma", se encuentra en el brazo corto del cromosoma 17 (17p13.105-p12), en condiciones de normalidad, codifica a una fosfoproteína nuclear que actúa como un regulador negativo (inhibe) de la proliferación celular mediante una acción compleja, puesto que al mismo tiempo actúa

como factor de transcripción, interruptor del ciclo celular e inductor de apoptosis. Desde el descubrimiento de la proteína p53 (codificada por el gen TP53) en 1979, hasta la elucidación de sus funciones celulares, el interés por esta proteína ha aumentado continuamente (Martin *et al*, 2002) porque se le encontró alterado en más del 50% de los tumores. p53 está constituida por 393 aminoácidos y consta de tres grandes dominios estructurales, los cuales se describen a continuación (Ver la figura 1):

Dominio N-terminal: Tiene la función de activación de la transcripción, la cual está ubicada en la región amino-terminal que participa en las interacciones proteína-proteína que regulan su estabilidad y función (Ver la figura 1) (Martin *et al*, 2002). Por análisis de mutagénesis dirigida se lograron definir los residuos L22 y W23 como esenciales para su actividad transactivadora (Soussi *et al*, 1987). Esta región (específicamente los residuos 18 al 23) es el sitio de unión de la proteína mdm2, de RPA, hsp70 (Picksley *et al*, 1994), la proteína viral E1b del adenovirus y de los coactivadores transcripcionales TAFII40 y TAFII60, (Thut *et al*, 1995) además, se ha informado un polimorfismo R/P en la posición 72, donde, la R72 presenta una estructura más sensible a la degradación inducida por la proteína E6 del papilomavirus humano (Harris *et al*, 1986).

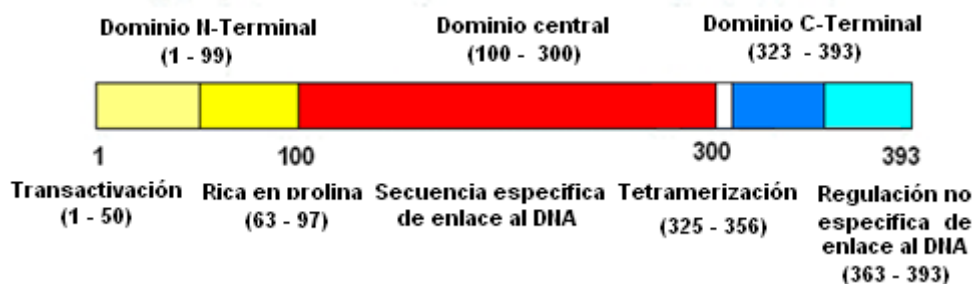


Figura 1. Dominios estructurales de la proteína p53 en humano. Se muestra los tres dominios estructurales con cada una de sus funciones biológicas (Alberts *et al.*, 2002).

Dominio central: Participa en la interacción proteína-DNA (ver la figura 1) ya que tiene contacto directo con secuencias específicas del DNA. La interacción se da en regiones del DNA con cuatro repeticiones de la secuencia PuPuPuCA/TA/TGPyPyPy a través del contacto con los residuos K120, S241, R248, R273, A276, C277, R283. Para que se lleve a cabo la interacción, es indispensable la presencia de una molécula de zinc unida a los residuos C176, C238, H179 y C242 de la proteína (Martin *et al*, 2002).

Dominio C-terminal o dominio de tetramerización: Interviene en la interacción entre unidades monoméricas de la proteína (Ver la figura 1). La proteína p53 se encuentra principalmente en forma de tetrámero y los residuos 323 al 353 son esenciales para su ensamblaje. Se le han atribuido otras dos funciones biológicas: señalización de la localización nuclear y reconocimiento del daño en el DNA. Las modificaciones post-traduccionales como fosforilaciones y acetilaciones sobre este dominio permiten potenciar la especificidad de la unión al DNA, razón por la cual también se le ha denominado región reguladora de la proteína. A pesar de ser una región vital para la estabilización del tetrámero, y por tanto, para la función de la proteína, se han detectado pocas mutaciones en la misma, posiblemente por ser una región que requiere mayor estudio (Martin *et al*, 2002).

1.3.1. p53 y mutaciones

Dada la importancia de p53 en el cáncer humano, esta es sin duda la proteína más estudiada en este tipo de padecimientos y se han realizado muchos estudios para evaluar el impacto de las mutaciones en su función. Las mutaciones en p53 han sido clasificadas de diversas maneras. Michalovitz propuso en 1991 una clasificación genética de mutaciones basada en la dominancia de su actividad (Martin *et al*, 2002). En otras clasificaciones las sustituciones encontradas en proteínas involucradas con el cáncer han sido clasificadas en dos grandes categorías en función del efecto sobre la estructura y función de la proteína:

1. **Mutaciones de contacto:** Afectan aminoácidos que participan en la interacción entre la proteína y el DNA, estas rompen los contactos que son cruciales entre las proteínas y el DNA, o entre p53 y otras proteínas.
2. **Mutaciones conformacionales:** Mutaciones que afectan el esqueleto de la proteína y que afectan su conformación y por lo tanto su función, en el caso de p53 estas mutaciones pueden afectar su afinidad por el DNA o la interacción con otras moléculas (Martin *et al*, 2002).

Una de las características más notables de la actividad oncogénica del gen TP53 es el interferir con la apoptosis por medio de un mecanismo de regulación negativa dominante. Por este motivo ofrece una ventaja selectiva a las células cancerosas para el crecimiento, la metástasis y la resistencia a la terapia antineoplásica. Dicho mecanismo es explicable por el hecho de que al mismo tiempo se traducen monómeros normales y anormales de la proteína p53 con capacidad de formar complejos no funcionales entre sí (heterotetramerización), alterando el reconocimiento de secuencias específicas del DNA (Martin *et al*, 2002, Soussi *et al*, 2001, Jiao *et al*, 2002). En contraste con muchos otros supresores de tumores, que a menudo son inactivados por la eliminación de aminoácidos o por mutaciones que causan cambio del marco de lectura, la mayoría de las mutaciones en TP53 son mutaciones puntuales por sustitución de nucleótidos que producen sustitución de aminoácidos (Ver la figura 2) (Martin *et al*, 2002, Verhagen *et al*, 2002, Soussi *et al*, 2006).

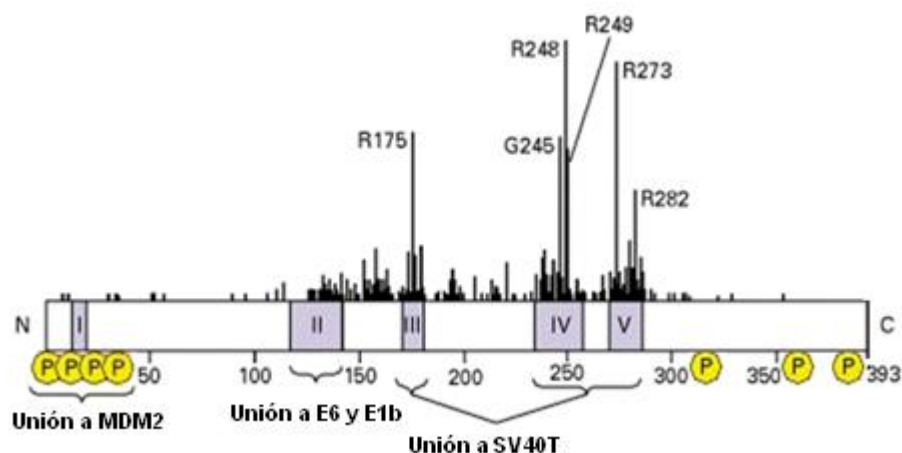


Figura 2. Mutaciones puntuales en la proteína p53: En esta imagen se muestran los principales sitios donde ocurren mutaciones puntuales relacionadas con el cáncer. Se puede observar 6 puntos de alta frecuencia, los cuales caen en los dominios conservados III, IV y V (Alberts *et al.*, 2002).

1.3.1 Activación de la proteína p53

El supresor tumoral p53 es una fosfoproteína nuclear con propiedades inhibitorias de cáncer. Este factor se mantiene en un estado reprimido en las células normales, pero es activado por modificaciones post-transcripcionales en respuesta a múltiples formas de estrés como los factores genotóxicos y los no genotóxicos (Martin *et al*, 2002):

- ⊙ **Genotóxico:** La irradiación, químicos cancerígenos, o agentes citotóxicos utilizados en el tratamiento del cáncer
- ⊙ **No genotóxicos:** Hipoxia, el agotamiento de los ribonucleótidos, y activación de cascadas de señalización del crecimiento.

Cuando se activa, la proteína p53 se acumula en niveles elevados en el núcleo y actúa como un factor multifuncional de transcripción, para aumentar o reprimir la expresión de varios conjuntos de genes implicados en la progresión del ciclo celular, apoptosis, respuesta adaptativa al estrés, diferenciación, y la reparación del DNA (Verhagen *et al*, 2002). De este modo, p53 controla y coordina la antiproliferación; evitando la replicación del DNA de células que sean expuestas a condiciones adversas (Martin *et al*, 2002)

1.3.2 p53 y control del ciclo celular

La proteína p53 regula negativamente la progresión del ciclo celular debido al efecto inhibitorio que realiza p21 (Martin *et al*, 2002) sobre diferentes complejos de cdk/ciclina en presencia de un daño en el DNA (Ver la figura 3), produciendo una detención en la transición G1/S y G2/ M del ciclo celular (Soussi *et al*, 2001). Esta inhibición genera una acumulación de la proteína pRb unida al factor de transcripción E2F. En condiciones normales, la pRb presenta diferentes estados de fosforilación (por complejos cdk/ciclina) hasta permitir la liberación del factor E2F, el cual es necesario para regular positivamente la transcripción de genes requeridos en el inicio y transcurso de la replicación (Herrera *et al*, 2004).

Por otra parte, tanto p21 como gadd45 pueden regular indirectamente la progresión por otra vía, interactuando con el antígeno nuclear de proliferación celular, de tal manera que no permite que se realice con éxito la fase de extensión de la cadena en la replicación del DNA; además de las inhibiciones indirectas del ciclo antes mencionadas (Wee *et al*, 2006), se ha demostrado que la proteína p21 puede llegar a formar complejos con subunidades de E2F induciendo, de igual manera, la detención del ciclo. Del mismo modo, la proteína de unión IGF-BP (cuyo gen es regulado positivamente por la proteína p53), al unirse con IGF impide la interacción con su receptor (IGFR) bloqueando, de esta manera, la vía de señalización que induciría el ingreso a mitosis. Recientemente se ha sugerido un papel para las proteínas Ras y MAPK en la regulación de p53 (Herrera *et al*, 2004). La sobre-expresión de la proteína Ras o activación de la vía

MAPK cinasa induce a la detención del ciclo dependiente de p53. De forma similar, se informa otro papel de la proteína p53 en el punto de control en la transición G2/M, (Martin *et al*, 2002) cuando se detecta una duplicación anormal de los centrosomas y alteración en el ensamble del huso mitótico.

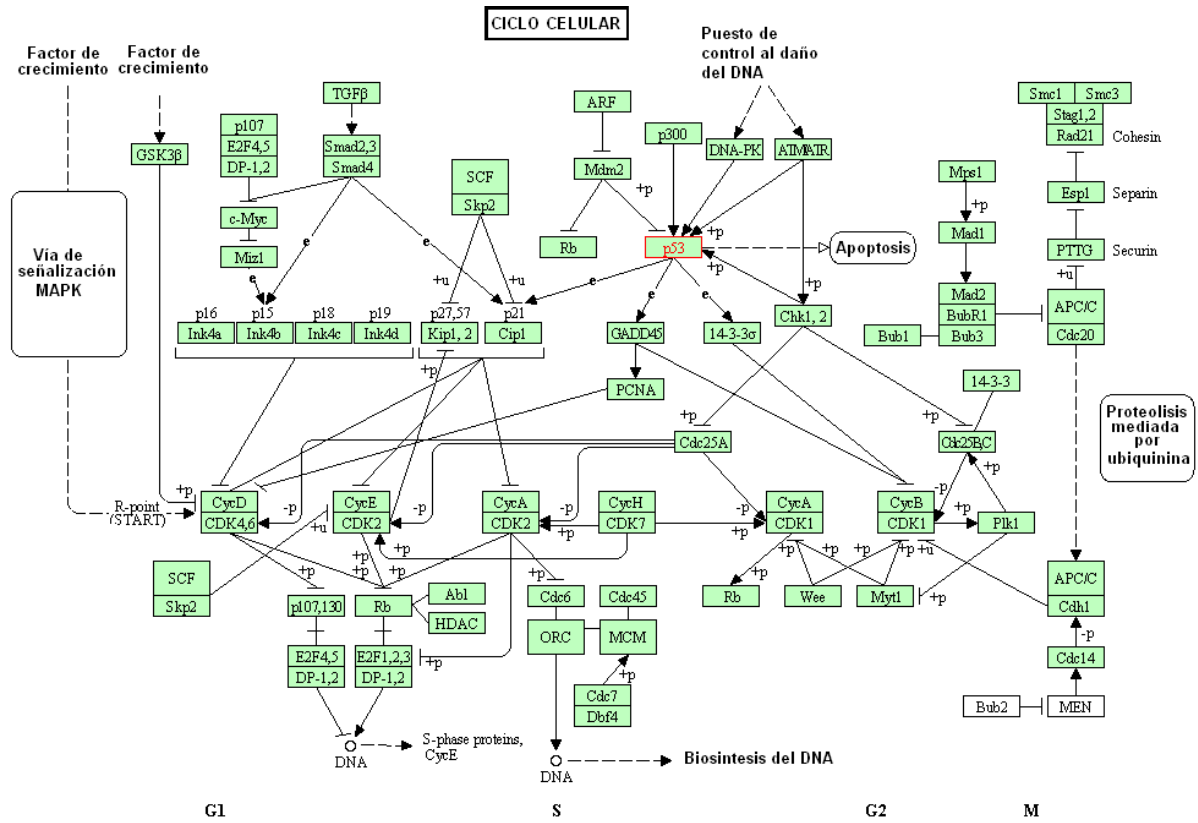


Figura 3. Proteína p53 y el ciclo celular. En esta imagen se muestra la interacción que tiene la proteína p53 con otros componentes y con algunas funciones celulares.

1.3.3. Proteína p53 y reparación

En la reparación resulta importante la actividad activadora en trans de la proteína p53 sobre los genes GADD45 y p21/Waf1. *in vitro* la proteína Gadd45 se une a PCNA para inducir la reparación de la lesión; por otra parte, la proteína p21 al unirse con PCN impide su interacción con la DNA polimerasa inhibiendo así la replicación. En consecuencia, un equilibrio entre estas dos proteínas controlará de manera eficiente dos mecanismos biológicos esenciales: reparación y replicación. Además, se ha propuesto que las proteínas ATM y NBS1 son componentes en la vía de señalización, corriente

arriba, de p53. La primera actúa en respuesta al daño del DNA por radiación gamma pero no así por radiación ultravioleta y la segunda en respuesta al daño por radiación ionizante; de manera similar, la polimerasa poli-ADP (PARP), importante en el reconocimiento del daño al DNA y su reparación, induce, *in vitro*, aumento en la cantidad de la proteína p53 en fibroblastos de ratón (Herrera *et al*, 2004).

En adición al mecanismo de detención del ciclo para reparar el daño y de los expuestos anteriormente, la proteína p53 se ha catalogado en la actualidad como “factor de accesibilidad a la cromatina” debido a su capacidad de reclutar a la proteína p300 para promover la acetilación de histonas y permitir el acceso de los componentes proteicos del sistema de reparación global del genoma, específicamente los que participan en NER (Reparación por escisión de nucleótidos) (Herrera *et al*, 2004).

1.4. FAMILIA RAS

Los miembros de la familia RAS son los proto-oncogenes que más frecuentemente se encuentran alterados en tumores sólidos humanos; la frecuencia varía alrededor del 30% y en algunos casos, como los adenocarcinomas pancreáticos y de pulmón supera el 50%.

En los humanos existen tres genes ras funcionales: Hras isoforma 1, Kras 2 y Nras que codifican proteínas homólogas, con un peso molecular de 21 kDa. Existen además dos pseudogenes: Hras-2 y Kras iso 1. Debido a que las proteínas Ras (Hras, Nras y Kras4B) tienen secuencia muy similar (> 90%), se puede deducir que son funcionalmente redundantes. Esta opinión se ve reforzada por las observaciones de que todas las isoformas de Ras comparten conjuntos comunes de efectores río abajo y río arriba de GEF. Tienen secuencias idénticas que cubren el efector, el factor de intercambio y el dominio GBD. La única región de las isoformas de Ras que exhibe divergencia significativa en las secuencias, se encuentra en la región hipervariable (HVR), ubicada en los últimos 24 residuos de las proteínas y que presenta aproximadamente el 10-15% de conservación en comparación con el 90% de identidad en los 165 residuos de la región N-terminal (ver la figura 4). Se demostró que la HVR contiene dos secuencias señal que cooperan en el destino membranal de Ras. La primera secuencia señal es CAAX (C= cisteína, A = aminoácidos alifáticos, X = serina o metionina) presentes en el extremo C-terminal (Hansen *et al*, 2002, Abankwa *et al*,

2010). Esta región es una secuencia que se modifica post-traduccionamente para hacerla más hidrofóbica: La secuencia AAX se elimina por proteólisis y luego la cisteína de la región C-Terminal es carboximetilada. La segunda secuencia señal consiste en un tramo de 6 residuos de lisina (175-180) en Kras, o la unión de ácido palmítico a cisteína (181) en Nras, o en las cisteínas 181 y 184 de Hras (Prior *et al*, 2001, Henis *et al*, 2009).

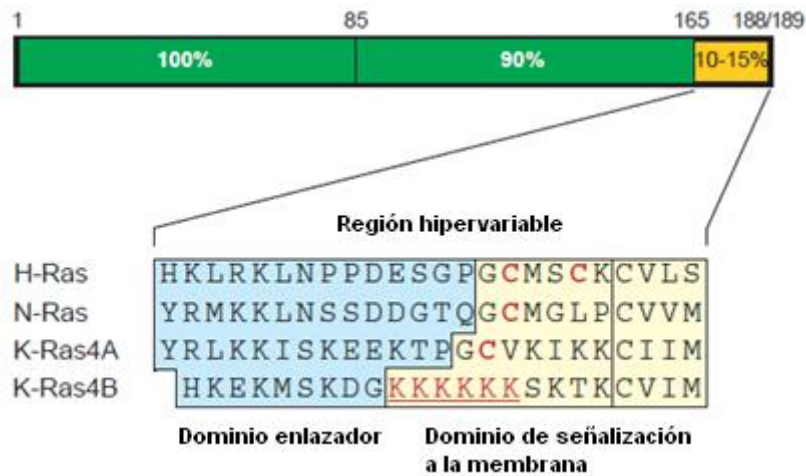


Figura 4: Región hipervariable de las proteínas RAS. En el diagrama se muestran las diferencias de conservación de las secuencias entre las isoformas de la proteína Ras. La región hipervariable es la única región con divergencia significativa entre las proteínas Ras (Prior *et al*, 2001).

Las interacciones entre las membranas celulares y la región hipervariable C-terminal de Ras determinan el tráfico y las proporciones relativas de cada isoforma en la señalización en la superficie celular, en el retículo endoplásmico, en el aparato de Golgi, en los endosomas y en los compartimentos de las mitocondrias (Quinlan *et al*, 2008, Omerovic *et al*, 2009).

Las proteínas Ras activan numerosas rutas de transducción de señales, pero es especialmente importante la de las proteínas cinasas activadas por mitógenos (MAPK del inglés *Mitogen-Activated Protein Kinases*), las cuales también a través de la transducción de señales activan a otras cinasas y genes reguladores de proteínas (Abankwa *et al*, 2010). Pertenecen a una familia extensa de GTPasas, la familia de proteínas G, que participa en el control de una gran variedad de procesos celulares e incluye, además de la superfamilia Ras, algunas proteínas involucradas en síntesis de proteínas (factores de iniciación y alargamiento) y las proteínas G heterotriméricas. Estas proteínas, como todas las GTPasas, son moléculas que unen e hidrolizan GTP y

funcionan como interruptores con dos conformaciones: un estado inactivo, unido a GDP, y a su estado activo, unido a GTP.

La activación de Ras por mutaciones somáticas ocurre en 15 a 20 % de todas las células malignas de humano. Se ha detectado en los genes HRAS, NRAS, y KRAS. Sin embargo, la gran mayoría de las mutaciones detectadas en los cánceres humanos se plantean en el gen KRAS. La activación de Kras por mutación se detectan frecuentemente en los tumores derivados de tejidos endodérmicos, tales como los de pulmón (35%), páncreas (95%), y colon (30%). Esto plantea la posibilidad que el Kras mutado activado ejerza su actividad oncogénica a nivel de las células precursoras endodérmicas o de las células madres (Quinlan *et al*, 2008).

Las proteínas Ras adquieren la capacidad de inducir la transformación celular por el cambio de uno solo de sus aminoácidos. Las mutaciones transformantes de estos genes ocurren en los codones 12, 13, 59, o 61, siendo los codones 12 y 61 los que más se afectan en los tumores humanos (ver la figura 5). Estas mutaciones provocan cambios de aminoácidos que se encuentran en el sitio activo de la proteína y por ello afecta profundamente su función, que resultan en la permanencia de las formas que impiden la hidrólisis del GTP (Cooper *et al*, 2000, Quinlan *et al*, 2008, Kalas *et al*, 2005).

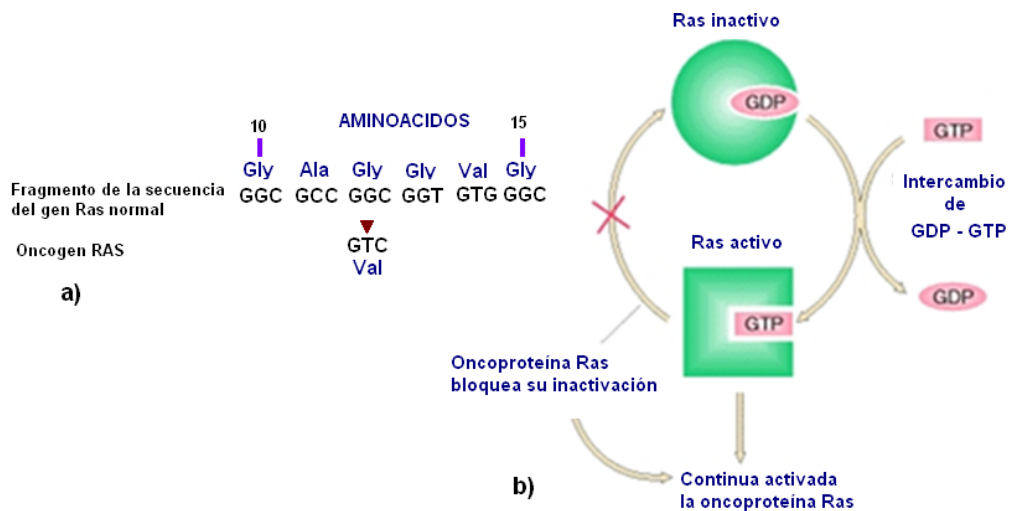


Figura 5. Activación del oncogén RAS. Las proteínas Ras constitutivas son proteínas mutadas que impiden la hidrólisis del GTP, lo que permite que Ras permanezca en estado activado. El codón 12 del gen está codificado por GGC. La mutación que sustituye el aminoácido glicina por valina, produce la inactivación por GAP, por lo que Ras permanece en estado activado. Ras requiere de GAP para ser inactivada y no puede inactivarse debido a que ha escondido el dominio G que contiene la subunidad alfa de las proteínas G heterotriméricas donde se une GAP (Cooper *et al*, 2000).

La estructura de las proteínas Ras está dividida en cuatro regiones o dominios (ver la figura 6):

Dominio de unión a los nucleótidos de guanina: Los primeros 165 aminoácidos unen el grupo purina del GTP y son causantes de la especificidad de la unión a los nucleótidos de guanina. Las mutaciones en este dominio disminuyen la afinidad de las proteínas ras, altera la especificidad de la unión de los nucleótidos y aumenta el AMPc.

Región efectora o dominio catalítico: El segmento cercano al extremo N-terminal que comprende los aminoácidos 32 al 40. Regula las actividades bioquímicas de las proteínas Ras y es necesario para la transmisión de señales y para la actividad transformante. Esta región interactúa con proteínas moduladoras y efectoras, es idéntica en diferentes especies y su conformación varía con el estado activo o inactivo. Las mutaciones en este dominio no activan a la adenilato ciclasa.

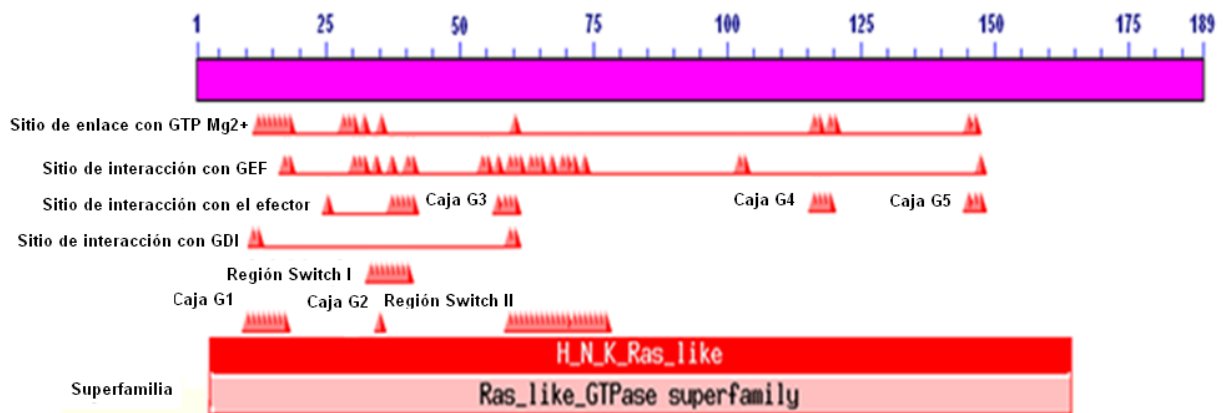


Figura 6. Dominios conservados de las proteínas Ras. En esta imagen se muestran los diferentes dominios y regiones conservadas de Hras, Nras y Kras. **FUENTE**

Dominio de unión a membrana: Es el extremo C-terminal, esencial para la localización de membrana y para la actividad oncogénica de Ras.

Región variable: Corresponde a los aminoácidos 165 al 185 y es variable entre distintas especies. No se modifica con el cambio conformacional de la proteína activa o inactiva.

1.5. EL GEN CDKN2A

El gen CDKN2A codifica a la proteína p16, es un gen supresor de tumores, que genera diversas variantes transcripcionales que difieren en sus primeros exones (Stone *et al*, 1995). Se han descrito al menos tres transcritos diferentes que codifican distintas isoformas de la proteína p16, dos de los cuales se sabe que actúan como inhibidores de la cinasa dependiente de ciclina 4 (Cdk4). El otro transcrito incluye un primer exón alternativo situado 20 kilobases corriente arriba respecto del resto del gen. Este transcrito contiene un marco abierto de lectura alternativo (p14^{ARF}) que codifica una proteína no relacionada estructuralmente con las otras isoformas. Este gen está compuesto por 4 exones: 1 α , 1 β , 2 y 3. La proteína p16 (ver la figura 7) es producto de los exones 1 α , 2 y 3 mientras que la proteína p14^{ARF} es producto de los exones 1 β , 2 y 3 (Robertson *et al*, 1999, Lin *et al*, 2007).

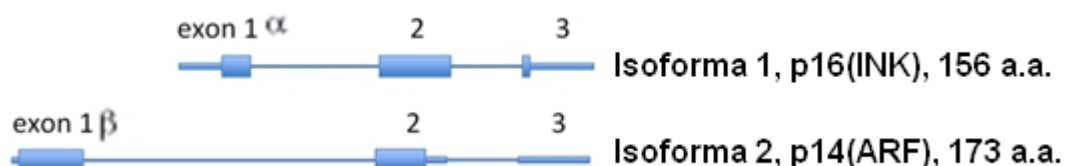


Figura 7. Los productos del gen CDKN2A. Este gen codifica para dos proteínas, la primera es p16, que es transcrita por los exones 1 α , 2 y 3; y la p14^{ARF}, que es transcrita por los exones 1 β , 2 y 3, pero con diferente lectura los fragmentos de los exones 2 y 3 (Lin *et al*, 2007).

La proteína p16 interviene en la etapa tardía G1 del ciclo celular, actuando en la vía regulatoria de las cinasas dependientes de ciclina (CDK), la proteína retinoblastoma (Forbes *et al*, 2008) y el factor de transcripción E2F (Tamm *et al*, 2002, Ghazizadeh *et al*, 2005, Han *et al*, 2007). El paso de la célula de la etapa G1 a la siguiente etapa de síntesis de DNA (etapa S), es determinado en parte por la fosforilación del Rb por la ciclina D y las cinasas dependientes de ciclina Cdk4 y Cdk6 (Okamoto *et al*, 1994, Campbell *et al*, 2000, Sharpless *et al*, 2002). Posteriormente la Rb actúa sobre otras dianas como el factor transcripcional E2F para promover la progresión del ciclo celular. Debido a que p16 inhibe a Cdk4, bloquea la progresión del ciclo celular. Debido a que p16 inhibe a Cdk4, bloquea la progresión de dicho ciclo, interfiriendo con la capacidad de Cdk4 de interactuar con la ciclina D y causando por lo tanto la detención del ciclo celular. La proteína p14^{ARF} es un inhibidor de MDM2, que de otra manera acelera la ubiquitinación de p53. La pérdida de p53 altera el control del ciclo celular y la capacidad de las células para sufrir la apoptosis.

La pérdida de CDKN2A por tanto contribuye a la génesis tumoral alterando la vía del pRB y del p53 (Bartkova *et al*, 1996, Lukas *et al*, 1999). Las anomalías en la metilación del DNA frecuentemente se han asociado con el desarrollo del cáncer. El silenciamiento de los genes supresores de tumor debido a hipermetilación del promotor es un papel primordial en la carcinogénesis. El gen CDKN2A puede ser silenciado no sólo por la hipermetilación del promotor, sino también por delección homocigota y mutación puntual (Marchini *et al*, 1997, Liggett *et al*, 1998).

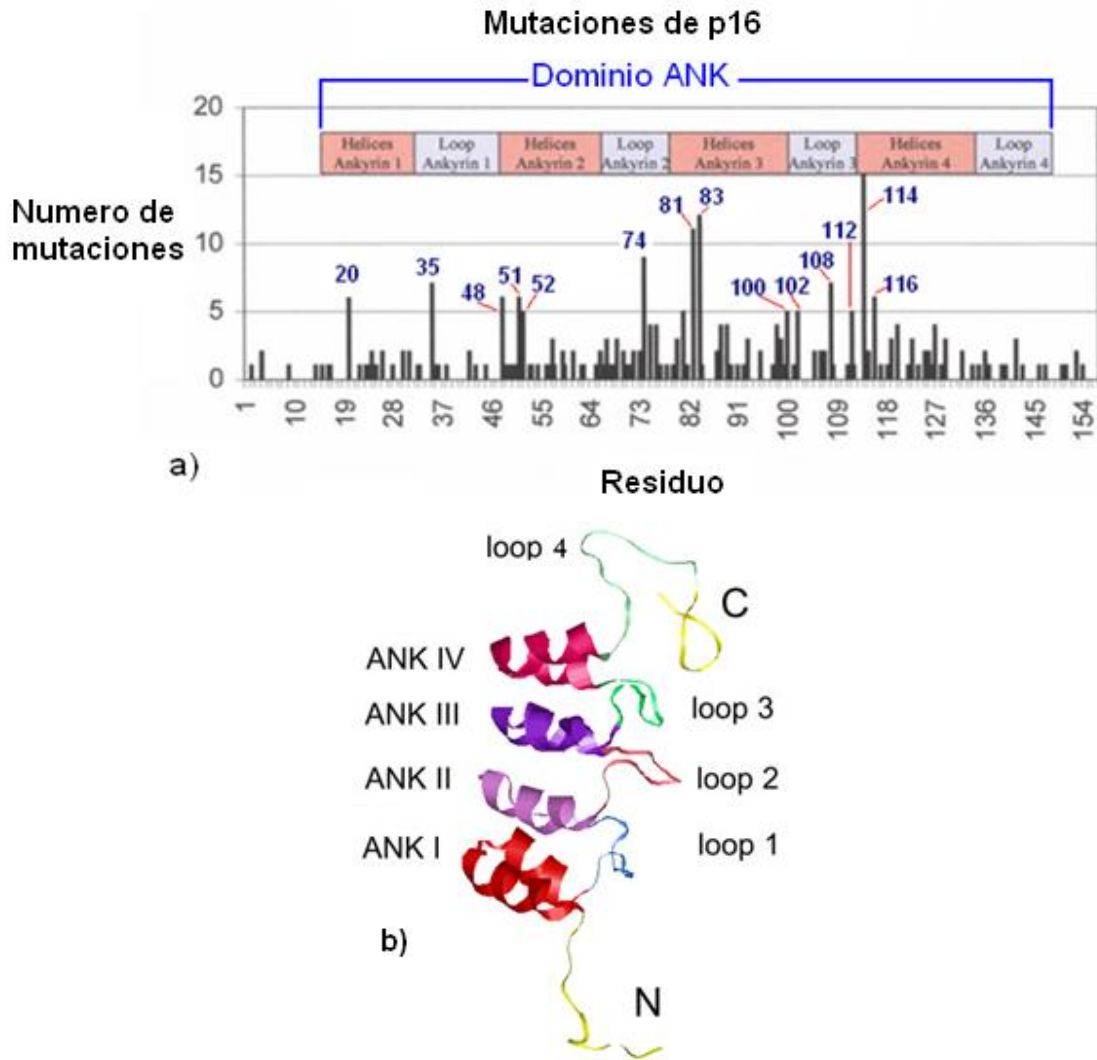


Figura 8. Mutación y dominios de p16. En el inciso a) se observa la frecuencia de las mutaciones por sustitución, en la cual se identifican los 14 puntos de alta frecuencia (“hotspot”), los cuales caen en el dominio conservado denominado ANK. En el inciso b) se muestra la estructura tridimensional de p16, visualizando las diferentes regiones del dominio ANK. La cual está comprendida por ANK I, ANK II, ANK III, ANK IV, loop 1, loop 2, loop3 y loop 4 (Greenblatt *et al*, 2003, Tang *et al*, 2003).

La proteína p16 contiene un solo dominio conservado denominado ANK (ankyrin), el cual está comprendido del aminoácido 15 al 146 (ver la figura 8). Este dominio contiene 4 repeticiones consecutivas de dos hélices paralelas y una horquilla beta. Esta región es muy importante para que pueda interactuar con cdk4 y cdk6 (Han *et al*, 2007, Greenblatt *et al*, 2003, Tang *et al*, 2003).

1.6. EL GEN ABL1

La familia Abl en los vertebrados consiste en dos proteínas Abl (Abl1) y Arg (Abl2). Las dos isoformas predominantes de Abl y Arg son generadas por empalmamiento alternativo. Abl posee dos isoformas ABL1-a y ABL1-b, obtenidos por procesamiento alternativo del primer exón (exones 1a y 1b). La isoforma ABL1-a esta presente tanto en el núcleo como en el citoplasma y la segunda sólo está en el citoplasma (Zheng *et al*, 2006). La isoforma ABL1-a se denomina con el nombre p145^{ABL} el cual indica su peso molecular en kDa (Smith *et al*, 2003). Esta proteína, en condiciones normales está involucrada en diversos procesos celulares tales como; regulación, reorganización del citoesqueleto, proliferación celular, inhibición del ciclo celular, diferenciación celular, transducción de señales implicadas en la adhesión celular y respuesta a estrés (Mughal *et al*, 2006, Gu *et al*, 2009).

Abl y Arg son proteínas (ver la figura 9), con tres dominios altamente conservados en la región N-terminal, llamados PTKc-Abl (también denominado cinasa del inglés “kinase” o SH1), SH2 y SH3 (Gu *et al*, 2009).

El dominio SH2 (del inglés *Src Homology 2*) reconoce específicamente el estado fosforilado de los residuos de tirosina. Este proceso constituye un evento fundamental en la transducción de señales a través de membrana. La fosforilación de la tirosina lleva a la activación de una cascada de interacciones proteína-proteína donde aquellas que contienen el dominio SH2 son reclutadas a sitios con tirosina fosforilada (Filippakopoulos *et al*, 2009, Nachman *et al*, 2010). El dominio SH2 es un regulador clave de la regulación alostérica cinasa y pueden desempeñar un papel tanto en estados inactivos, así como los estados activos de tirosina cinasa. La función del dominio SH2 es estabilizar el estado de inactividad de los miembros de la familia Src y se requiere con frecuencia para la actividad catalítica, lo que sugiere que una función fundamental es estabilizar el estado activo de la tirosina cinasa. Se han identificado mutaciones que comprometen la

función SH2 y como consecuencia la actividad cinasa (Filippakopoulos *et al*, 2009, Nachman *et al*, 2010). Los dominios SH2 y SH3 están implicados en los procesos de regulación, ayudan a mediar las interacciones proteína-proteína y modulan la activación de transducción de señales.

El dominio SH3 (del inglés *Src Homology 3*) reconoce secuencias ricas en prolina y es crucial para la autorregulación permitiendo una correcta actividad en su función como cinasa. En el dominio PTKc-Abl se encuentra la región catalítica que constituye el dominio cinasa, el cual abarca del aminoácido 235 al 509 (Smith *et al*, 2003) La actividad catalítica se encuentra en la Y393 y F401, los cuales se encuentran conservados en los diferentes grupos de cinasas tipo c-Src .

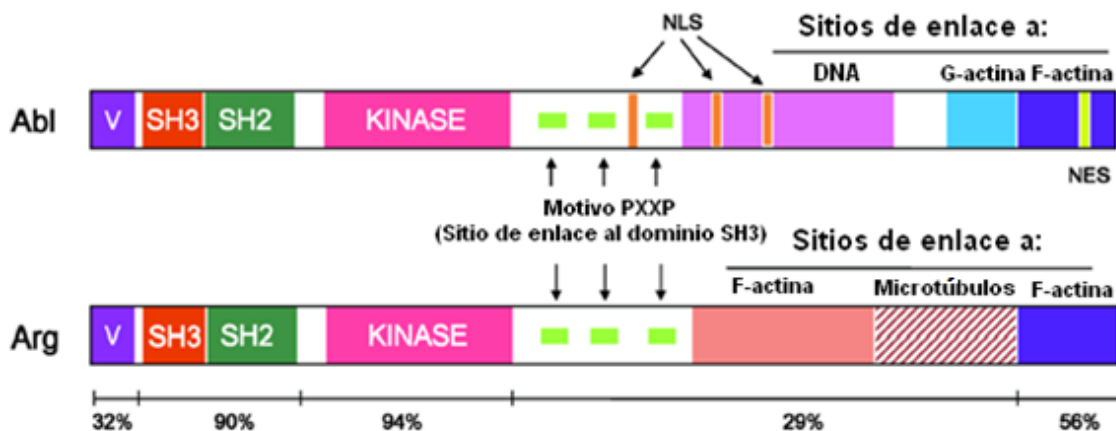


Figura 9. Dominios estructurales de Abl y Arg. Las Abl y Arg tienen una alta similitud en N-terminal en los dominios SH3, SH2 y SH1 (kinase). Pero en la región C-terminal contienen “motifs” ricos en Prolina, esta región se encuentra menos conservada (Gu *et al*, 2009).

Las secuencias C-terminal no están bien conservadas entre Abl y Arg, pero ambas cinasas exhiben buena conservación en secuencias ricas de prolina, que proporcionan un enlace para el dominio SH3 en proteínas que las contengan. Una característica única de la familia Abl es la presencia de dominios de unión a la actina, la proteína Abl1 también contiene señales NLS (del inglés *nuclear localization signals*), NES (del inglés *one nuclear export signal*) y un dominio de enlace al DNA (Gu *et al*, 2009).

1.7. BIOINFORMÁTICA

La bioinformática es un área de investigación interdisciplinaria que combina la informática y la bioestadística con las ciencias biológicas y biomédicas tales como la bioquímica, biología celular, la genética, la genómica, la fisiología, la inmunología y la biotecnología, entre otras (Moore *et al*, 2007). Estudia el flujo de información en todos los estadios del dogma central, como la organización y la regulación de los genes en la secuencia del DNA, la identificación de las zonas de transcripción del DNA, la predicción de la estructura de las proteínas a partir de su secuencia y el análisis de la función molecular de las biomoléculas, implicando la tecnología que utiliza las computadoras para el almacenamiento, recuperación, manipulación y distribución de información relacionada con macromoléculas. El énfasis aquí es sobre el uso de ordenadores, porque la mayor parte de las tareas de análisis de datos genómicos son muy repetitivos o complejos matemáticamente (Xiong *et al*, 2006, Forman *et al*, 2010, Kim *et al*, 2002).

La bioinformática surgió como una disciplina importante poco después del desarrollo de tecnologías de alto rendimiento de secuenciación de DNA en la década de 1970 (Moore *et al*, 2007), y no solamente se ha convertido en algo esencial para la investigación básica de biología genómica y molecular, sino que está teniendo un gran impacto en muchas áreas de la biotecnología y las ciencias biomédicas. Sus aplicaciones son, por ejemplo, en el diseño de fármacos basada en el conocimiento, el análisis forense de DNA y en la biotecnología agrícolas. Los estudios computacionales de interacciones de proteínas con sus ligandos proporcionan una base racional para la rápida identificación de nuevas moléculas potenciales para la síntesis y diseño de drogas. El conocimiento de las estructuras tridimensionales de las proteínas permite el diseño de moléculas que sean capaces de unirse al sitio receptor de una proteína diana con gran afinidad y especificidad. Este enfoque basado en la informática reduce significativamente el tiempo y el costo necesario para desarrollar fármacos con mayor potencia, menos efectos secundarios y menos toxicidad que con el enfoque tradicional de prueba y error. En la medicina forense, los resultados de los análisis filogenéticos moleculares han sido aceptados como prueba en los tribunales penales. Algunos métodos sofisticados como estadísticas bayesianas y aquellos basados en la probabilidad para el análisis de DNA se han aplicado en el análisis de identidad forense. Las herramientas bioinformáticas también se están utilizando en la agricultura, las bases de datos de genomas de los

vegetales y el análisis del perfil de expresión genética han desempeñado un papel importante en el desarrollo de nuevas variedades de cultivos que tienen una mayor productividad y más resistencia a las enfermedades (Xiong *et al*, 2006).

La bioinformática puede clasificarse en dos ramas: la bioinformática secuencial y bioinformática estructural. El primero se centra en las secuencias, con las principales aplicaciones de datos y análisis de los datos recogidos en los proyectos genoma, alineamiento de secuencias, redes metabólicas, morfometría y la evolución virtual, mientras que el segundo se centró en estructuras, con las principales aplicaciones para la predicción de estructuras 3D de proteínas y revelación de la relación entre estructura y función. Puesto que la estructura 3D de una proteína es determinada por su secuencia, el análisis de la secuencia y la alineación son también muy involucrados en la bioinformática estructural. La secuencia de aminoácidos de una proteína, estructura primaria, se puede determinar fácilmente a partir de la secuencia en el gen que lo codifica. Así, el número de secuencias de entrar en los bancos de datos ha aumentado rápidamente (Chou *et al*, 2004).

Herramientas bioinformáticas:

Las herramientas bioinformáticas han nacido de la necesidad de trabajar específicamente con una gran cantidad de secuencias de DNA y de proteínas que se encuentran almacenadas en bases de datos (Coulson *et al*, 1994), esta información se puede utilizar para buscar las propiedades de una secuencia, comparándolas con secuencias semejantes con propiedades conocidas o a través de reglas para predecir las propiedades de secuencias nuevas (Altschul *et al*, 1994, Thompson *et al*, 1994). Entre las herramientas bioinformáticas más comunes para la obtención de información tenemos (Dopazo *et al*, 2001):

- La obtención de secuencias de DNA o proteínas similares a la secuencia problema.
- La obtención de secuencias con palabras clave o información similar a la secuencia problema.
- La búsqueda de “motifs” funcionales o estructurales en la secuencia problema y obtención de secuencias que contienen los mismos “motifs”.

- El alineamiento múltiple de la secuencia problema con otras similares, y definición de regiones conservadas y variables.
- El ensamblaje de fragmentos de DNA y creación de mapas genómicos.
- La predicción de estructura y dinámica de macromoléculas.
- La reconstrucción de la filogenia a partir del alineamiento.
- La construcción de un motivo característico del alineamiento y el uso de éste para encontrar nuevas secuencias con “motifs” comunes con el alineamiento previamente realizado.
- Estudio de todos los genes y proteínas de un organismo: “Genómica y proteómica funcional”.

Análisis de patrones de expresión génica:

El desarrollo de la tecnología de los microarreglos, mediante la cual se pueden unir a pequeñas piezas de soportes sólidos decenas de miles de sondas capaces de detectar específicamente la expresión de un determinado gen en cada una de ellas, ha permitido comparar los niveles de expresión génica entre células normales y células tumorales (Scherf *et al*, 2000). Una de sus aplicaciones potenciales es la posibilidad de detectar rápidamente patologías por medio del análisis de las características genotípicas de distintos individuos, además de detectar, por ejemplo, qué tipo de medicina puede ser más útil para el perfil de expresión de un individuo concreto (Berns 2000).

1.8 ESTUDIO DEL IMPACTO DE LAS MUTACIONES EN LA ESTRUCTURA / FUNCIÓN DE P53

La caracterización de la proteína p53 de *Xenopus laevis* mostró que existen regiones con gran similitud a las de p53 de otras especies. Esto permite una clasificación estructural, según los dominios conservados evolutivamente y los dominios funcionales. Además, la identificación de los dominios conservados sugirió regiones proteicas con gran importancia funcional. Posteriormente, cuando se realizaron análisis de mutaciones en el gen TP53 se encontró que en efecto, existen sitios “hotspots” para las mutaciones asociadas con cáncer que comprometen principalmente 3 de los 5 dominios conservados (Soussi *et al*, 1987).

Andrew C.R y cols, (2000) realizaron un análisis bioinformático sobre el posible impacto que tienen las mutaciones en la estructura y función de la proteína p53. Realizaron un análisis estructural de la estructura cristalizada de p53, el cálculo de la estructura secundaria, el análisis de los ángulos de torsión, la accesibilidad al disolvente, y parámetros para la formación de puentes de hidrógeno. Los datos se almacenaron en una base de datos relacional, comparándola con otra base de datos de mutaciones asociadas con cáncer, para correlacionar los efectos estructurales con las mutaciones. Se encontró que 304 de las 882 (34,4%) mutaciones en el dominio central, pueden tener un impacto estructural en función de sus efectos, sobre el plegado de la proteína o el contacto DNA-proteína. La proporción de mutaciones “explicadas” aumentó a 55,6% cuando fueron incluidas todas las sustituciones de aminoácidos conservados (Martin *et al*, 2002).

Algo muy importante que se puede destacar de este trabajo, es que el análisis computacional que propusieron dichos autores se podría aplicar para estudiar otros genes frecuentemente mutados en este u otros padecimientos, tales, como la distrofina, BRCA1, y G6PD (Martin *et al*, 2002).

1.9 JUSTIFICACIÓN

El cáncer constituye la segunda causa de mortalidad en los países desarrollados y, debido al aumento paralelo de la incidencia de esta enfermedad con respecto al aumento de la esperanza de vida en la población, existe un gran interés médico y social por el cáncer. La mayoría de las neoplasias se desarrollan por mutaciones que ocurren en el DNA y que afectan específicamente a dos grandes grupos de genes: proto-oncogenes y genes supresores de tumores que codifican proteínas que están involucradas con el ciclo celular y que por lo tanto al estar mutadas pueden provocar que las células normales se conviertan en células cancerígenas. No obstante, existen relativamente pocos estudios encaminados a conocer cuál es el posible impacto que tienen estas mutaciones sobre la estructura o función de estas proteínas, excepto en algunos casos muy estudiados como son p53 y BRCA1. Otras proteínas son muy importantes en el cáncer y, por lo tanto, sería muy revelador conocer más sobre el mecanismo a nivel molecular en el que se ven afectadas por la presencia de mutaciones y establecer una asociación con el desarrollo del cáncer: Por ello ha surgido la necesidad de estudiar a estas proteínas a nivel molecular y comprender como las diversas mutaciones pueden impactar en la estructura secundaria, terciaria y en su función. Es importante, por otro lado, desarrollar una metodología general que pueda aplicarse a estas proteínas para estudiar las mutaciones a gran escala. Estos estudios, propios de la bioinformática estructural, pueden ser de gran relevancia en el desarrollo de fármacos o terapias para el tratamiento de esta enfermedad.

1.10 OBJETIVO

1.10.1 Objetivo general

Estudiar las variaciones en la secuencia de algunas proteínas proto-oncogénicas y supresoras de tumores potencialmente relacionadas con el cáncer humano e investigar “in silico” su posible participación en el cáncer mediante estudios evolutivos, estructurales y funcionales.

1.10.2 Objetivos específicos

- 1) Buscar proteínas humanas proto-oncogénicas y supresoras de tumores, para las cuales exista información de mutaciones que las asocien con diversos tipos de cáncer.
- 2) Buscar las mutaciones reportadas para las proteínas seleccionadas que estén asociadas con el desarrollo de cáncer en humanos.
- 3) Llevar a cabo estudios evolutivos en las proteínas seleccionadas, para identificar las regiones conservadas y variables que puedan estar involucradas con la estructura o función de las mismas.
- 4) Estudiar la distribución de mutaciones asociadas con cáncer en las regiones conservadas y variables de las proteínas en estudio.
- 5) Realizar estudios bioinformáticos estructurales para estimar el posible impacto de las mutaciones asociadas con cáncer, sobre la estructura o función de las proteínas en estudio.

10.11 HIPÓTESIS

Se postula que la mayoría de las mutaciones que afectan la secuencia de proto-oncogenes y genes supresores de tumores están asociadas con el desarrollo de cáncer, deben ocurrir en:

- zonas altamente conservadas durante el proceso evolutivo.
- dominios funcionales.
- regiones que desestabilizan la conformación estructural de las proteínas.
- regiones que afectan las interacciones con sus sustratos, con otras proteínas o con ácidos nucleicos.

2. MATERIALES Y MÉTODOS

2.1. Recursos bioinformáticos

Para poder realizar este trabajo se emplearon diferentes recursos bioinformáticos los cuales pueden clasificarse en programas informáticos y en bases de datos primarias o especializadas. A continuación se describen los recursos utilizados (ver tabla 3 y tabla 4):

TABLA 3. BASES DE DATOS PRIMARIAS Y ESPECIALIZADAS	
BASE DE DATOS	CARACTERÍSTICA
Protein Data Bank	Base de datos que contienen estructuras tridimensionales de proteínas y ácidos nucleicos. Estos datos, generalmente son obtenidos por cristalografía de rayos X o RMN. http://www.pdb.org
GenProt NCBI	Base de datos de las secuencias de proteínas en formato fasta o GenPet, Proporciona información general de las secuencias de proteínas. http://www.ncbi.nlm.nih.gov/protein/
CCD NCBI	Base de datos que contiene diferentes dominios conservados de las proteínas. http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
COSMIC Sanger Institute.	Catálogo de mutaciones somáticas en Cancer. http://www.sanger.ac.uk/genetics/CGP/cosmic/
PFAM Sanger Institute	Base de datos de familias de proteínas, cada una representada con alineamientos múltiples de secuencias, importante para identificar las regiones conservadas y sus funciones. http://pfam.sanger.ac.uk
IARC Tp53	Base de datos de las mutaciones relacionadas con cáncer del gen p53 humano. http://www-p53.iarc.fr

Tabla 3. Bases de datos primarias y especializadas: Se muestra la relación de bases de datos que fueron utilizados en este proyecto.

TABLA 4. PROGRAMAS BIOINFORMÁTICOS	
PROGRAMA	CARACTERISTICA
BLAST 2.0	Programa para la búsqueda de homólogos de DNA o proteínas basado en el alineamiento de secuencias de tipo local. http://blast.ncbi.nlm.nih.gov/Blast.cgi
BioEdit 7.0.9	Programa que contiene múltiples herramientas para el análisis de secuencias biológicas. http://www.mbio.ncsu.edu/BioEdit/bioedit.html
Clustal X 2.0	Programa para calcular alineamientos múltiples basados en métodos progresivos. ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX
Muscle EBI	Software para el alineamiento múltiple de secuencias tanto proteicas como nucleotidicas basado en métodos reiterativos y distancias entre composición de subsecuencias (palabras). http://www.ebi.ac.uk/Tools/muscle/index.html
T-Coffee	Programa para calcular alineamientos múltiples basado en métodos progresivos de alineamientos locales y globales. http://www.tcoffee.org/ .
PSI-Pred	Método para la predicción de estructuras secundarias en proteínas basado en información evolutiva de proteínas homólogas. http://bioinf.cs.ucl.ac.uk/psipred/
GENEDOC	Programa para la edición de alineamientos. http://www.nrbsc.org/downloads/
JALVIEW	Programa para visualizar alineamientos. http://www.jalview.org/Web_Installers/install.htm
MEGA 4	Programa para análisis filogenético. http://www.megasoftware.net/
Modeller 9v7	Programa para modelación por homología de estructuras tridimensionales de proteínas. http://www.salilab.org/modeller/modeller.html
Swiss-Model	Programa para modelación por homología de estructuras tridimensionales de proteínas. http://swissmodel.expasy.org/
Procheck v3.5.4	Programa para verificar la sanidad geométrica de la estructura de proteínas. http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
Perl 5.8	Lenguaje de programación. http://www.activestate.com/store/languages/register.plex?id=ActivePerl
Komodo 3.5	Editor de código de programación y soportes multilenguaje (P http://www.activestate.com/
VMD 1.8.7	Programa de visualización molecular para mostrar, animar y analizar sistemas biomoleculares. http://www.ks.uiuc.edu/Research/vmd/current/
ProSa 2003	Programa para el análisis de estructuras 3D de proteínas basado en el cálculo de z-score de potenciales estadísticos. https://prosa.services.came.sbg.ac.at/download/download.php
RasMol 2.7.5	Programa para la visualización molecular y modelación de estructuras de biomoléculas. http://www.umass.edu/microbio/rasmol/index2.htm
SSM	Programa para el análisis comparativo de estructuras 3D de proteínas. http://www.ebi.ac.uk/msd-srv/ssm/ssmstart.html .
QMEAN	Programa de análisis de estructuras tridimensionales de proteínas. http://swissmodel.expasy.org/qmean/cgi/index.cgi

Tabla 4. Programas bioinformáticos: Se muestra la relación de los programas bioinformáticos que fueron utilizados en este proyecto.

2.2 Obtención de las secuencias

En este trabajo se estudiaron las proteínas codificadas por los oncogenes (Kras, Nras, Hras, Abl1) y los genes supresores de tumores (p53, p16). Las secuencias de estas proteínas (en formato fasta) fueron obtenidas de la base de datos de la página del National Center for Biotechnology Information (NCBI).

2.3 Búsqueda de proteínas homólogas

Se buscaron los homólogos de estas proteínas en diferentes especies, utilizando la aplicación PSI-Blast del programa BLAST 2.0 del NCBI. Esta aplicación realiza la búsqueda de proteínas homólogas empleando una estrategia reiterativa. En la primera etapa del programa, se realiza una búsqueda de homólogos en una base de datos determinada empleando exclusivamente la secuencia problema (Bhagwat *et al*, 2007, Lee *et al*, 2008). A partir de los homólogos encontrados en esta primera iteración se calcula un alineamiento múltiple y una matriz de sitios con posiciones específicas denominada PSSM (por sus en inglés, Position-Specific Site Matrix), la cual es una representación estadística del alineamiento múltiple. En la siguiente iteración se realiza una segunda búsqueda en la base de datos pero ahora utilizando la PSSM como problema. De este modo se alcanza a localizar un mayor número de homólogos, los cuales se pueden integrar en la PSSM. Esta estrategia se repite hasta que no es posible localizar más homólogos. Este método tiene una alta sensibilidad para localizar homólogos remotos de las proteínas (Przybylski *et al*, 2007, Agrawal *et al*, 2009).

2.4 Alineamiento de secuencias

A partir de las secuencias homólogas encontradas se construyen alineamientos múltiples. Los alineamientos múltiples permiten localizar zonas variables y conservadas de las proteínas homólogas y son además el punto de partida para calcular árboles filogenéticos. No obstante, todos los algoritmos de alineamiento múltiple de secuencias emplean técnicas heurísticas por lo cual en general pueden rendir alineamientos subóptimos, con diversos tipos de errores según el algoritmo empleado (Gondro *et al*, 2007, Rausch *et al*, 2008). Por este motivo es conveniente emplear diversos algoritmos de alineamiento. En los últimos años los algoritmos que se han destacado por su

exactitud son Muscle EBI (Edgar *et al*, 2004), T-Coffee (Notredame *et al*, 2000) y Clustal W (Larkin *et al*, 2007), que fueron los empleados en este trabajo.

2.5 Identificación de regiones conservadas

La identificación de regiones conservadas que pueden ser identificadas aún en homólogos remotos de proteínas, juega un papel importante en la identificación de regiones de las proteínas que son importantes a nivel estructural. Las regiones conservadas de las proteínas son necesarias para mantener sus propiedades biológicas. Se puede postular que las regiones que se han conservado en homólogos que incluso llevan a cabo funciones distintas, son importantes para el mantenimiento de la estructura. Por otra parte, estos “motifs” existen debido a que estas regiones sufren fuertes restricciones estructurales a lo largo de la evolución de forma que pueden ser reconocibles mediante análisis de “motifs” y permiten tener una idea de cuál puede ser el papel funcional que una proteína desconocida está jugando en la célula (Xiong *et al*, 2006, Keith *et al*, 2008). La identificación de zonas conservadas se realizó por los programas GENEDOC y JALVIEW (Waterhouse *et al*, 2009).

2.6 Elaboración de árboles filogenéticos

La elaboración de los árboles filogenéticos se realizaron mediante los programas MEGA4, empleando métodos basados en distancias tales como UPGMA y neighbor joining o de unión de vecinos (Xiong *et al*, 2006, Saitou *et al*, 1987, Tamura *et al*, 2007, Kumar *et al*, 2008).

2.7 Búsqueda de estructuras de las proteínas

La búsqueda de estructuras 3D se llevó cabo con el programa Blast 2.0 del NCBI empleando las bases de datos PDB (Keith *et al*, 2008, Berman *et al*, 2008, Arnold *et al*, 2009). Se descargaron las estructuras normales de las proteínas en estudio.

2.8 Modelado de la estructura tridimensional

Para las proteínas de las cuales no se encontró su estructura 3D, se modeló su estructura. Utilizando el método por homología para la predicción de la estructura

tridimensional de las proteínas. Aunque este método puede considerarse como el más sencillo, es a su vez el más exitoso. Este análisis consiste en localizar un homólogo de la proteína de estructura conocida (empleando una herramienta como BLAST o PSI-BLAST) y emplear esta molécula como un molde para modelar la estructura mediante un alineamiento estructural 3D. Para esta tarea se emplearon los programas Modeller 9v7 y Swiss-Model (Schwede *et al*, 2003, Bordoli *et al*, 2009).

2.9 Extracción de la información de estructura secundaria y otros parámetros estructurales

Se utilizaron los programas DSSP y STRIDE para realizar el cálculo de la estructura secundaria a partir de los datos de coordenadas de la cadena principal de las proteínas. También se realizó el cálculo de accesibilidad del solvente y de otras propiedades geométricas tales como gráficas de Ramachandran, longitudes y ángulos de enlace, entre otras con Procheck 9v7. Para las estructuras obtenidas por modelación se llevo a cabo la predicción de estructuras secundarias mediante el programa: PSI-Pred. Se realizaron la evaluación de contactos incorrectos entre residuos con los programas Probe y Reduce (Xiong *et al*, 2006, Keith *et al*, 2008, Bourne *et al*, 2003).

2.10 Búsqueda de dominios estructurales/funcionales

La búsqueda de dominios estructurales se llevó a cabo con los programas Blast 2.0 del NCBI empleando modelos de Markov de los dominios funcionales reportados en las bases de datos Conserved Domain Database (CDD) del NCBI y Pfam del Instituto Sanger (Jefferson *et al*, 2006, Marchler-Bauer *et al*, 2009, Finn *et al*, 2010).

2.11 Búsqueda de mutaciones asociadas con cáncer

Se llevó a cabo la búsqueda de mutaciones reportadas en proto-oncogenes y genes-supresores asociadas con cáncer en humanos en diversas bases de datos. Entre las bases de datos que se consultaron fueron COSMIC del Sanger Institute (ver la tabla 3) y la base de datos IARC para mutaciones en p53 (Forbes *et al*, 2008, Olivier *et al*, 2002).

2.12 Modelado de las proteínas mutantes

El modelado de las proteínas mutantes se está realizó con el programas Modeller 9v7 mediante un script especializado de Modeller para modelar las proteínas mutantes mediante mutagénesis *in silico* de las estructuras de las proteínas silvestres (Eswar *et al*, 2008).

2.13 Análisis del impacto mutacional

Se realizaron análisis comparativos globales y locales entre estructuras de las proteínas silvestres y las mutantes. Estos estudios se llevaron a cabo para evaluar posibles perturbaciones estructurales o energéticas en la proteína originadas por la mutación. En la investigación de la estructura de las proteínas, a menudo se trata con la comparación de dos o más estructuras que es similar a la proteína: modelos estructurales alternativos, mutaciones puntuales, predicción de las proteínas y movimientos moleculares (Lema *et al*, 2005).

Algunas diferencias entre las estructuras son pequeñas, restringidas a una región local de vecindad (del inglés *local neighborhood*) y pueden estar enterradas por ruidos estructurales, derivados de artefactos experimentales. En tales casos las comparaciones estructurales, donde se superpone toda la estructura puede ser insatisfactoria y se tiene que investigar por fragmentos. Las perturbaciones locales en la estructura de la proteína pueden originarse por mutaciones puntuales, modificaciones químicas, interacciones entre moléculas, etc. Para evaluar tales perturbaciones estructurales de la proteína, usualmente deben ser comparadas de un modo detallado, para observar la posición y orientación de átomos individuales, residuos o de estructura secundaria. Es importante evaluar las perturbaciones en los estudios estructurales porque esto conduce a explicar cómo las modificaciones producen a cambios en la estructura y presumible en su función (Lema *et al*, 2005).

Comparaciones globales de la estructura tridimensional:

RMSD (Desviación de la raíz cuadrada de las medias). Se basa en el cálculo de la distancia en Angstroms (A) entre átomos equivalentes entre las estructuras que se desea comprar y que se han alineado en el espacio. Se utiliza la distancia cuadrática para compensar las fluctuaciones positivas y negativas de estas distancias y a su vez se utiliza un valor medio para toda la proteína (Maiorov *et al*, 1994, Bruschweiler *et al*, 2003). La fórmula es la siguiente:

$$RMSD = \sqrt{\frac{\sum (dist^2)}{n \text{ pares de atomos}}}$$

Donde *dist* corresponde a la distancia entre átomos equivalentes (según el alineamiento estructural de las dos proteínas, por ejemplo los C α de la cadena principal) y *n*, el número de parejas de átomos comparados entre las dos estructuras.

Comparaciones locales de la estructura tridimensional:

RMS posicional (Raíz media cuadrática por posición)

Son las distancias entre cada par de los aminoácidos que se encuentran alineados en la superposición de las dos estructuras que se desean comparar y se expresan en función de la posición de las parejas de residuos que se comparan, con respecto al alineamiento estructural de las dos proteínas (Yu *et al*, 2006).

DOPE (Discrete Optimized Protein Energy)

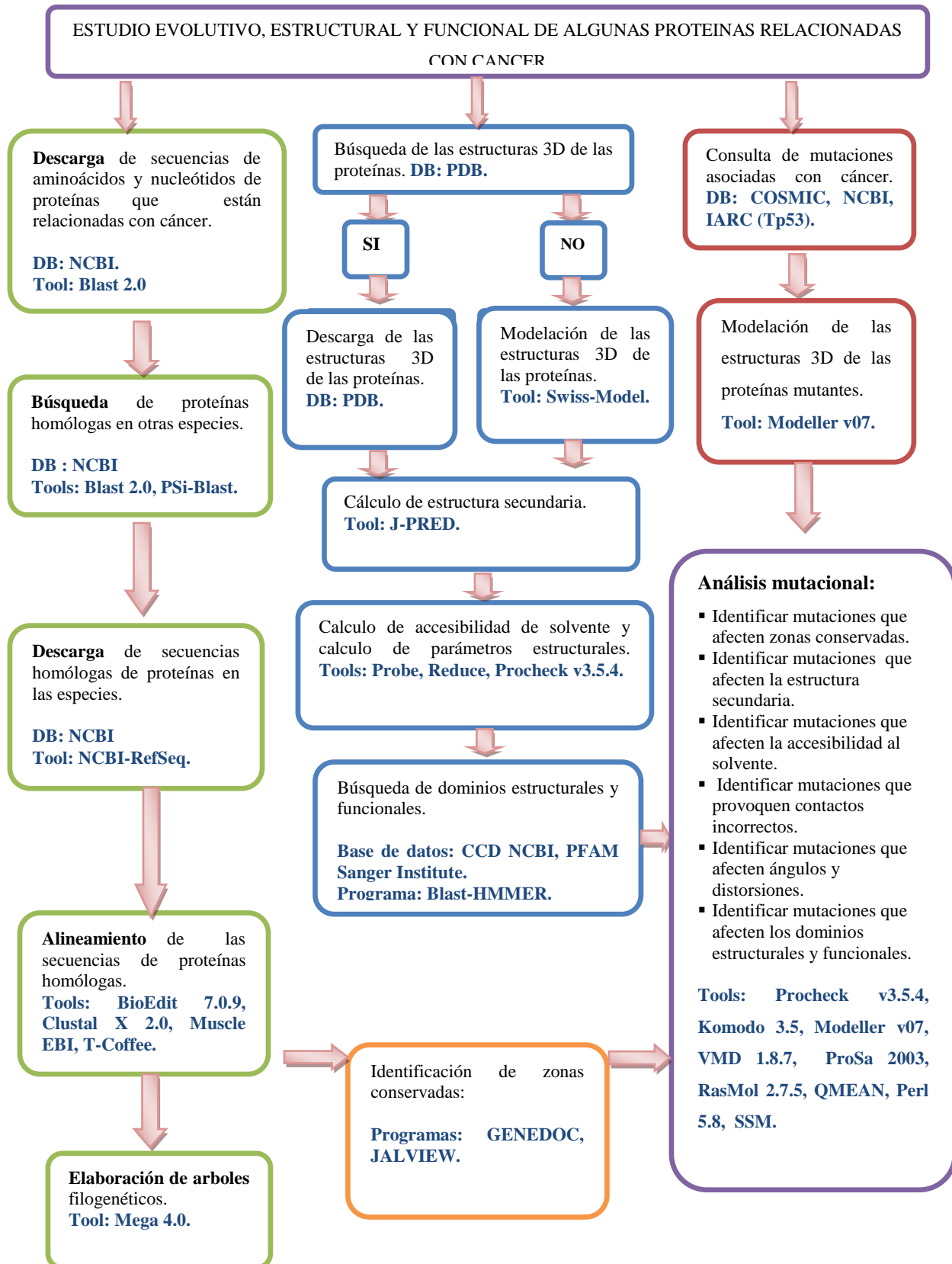
Es un potencial estadístico dependiente de las distancias atómicas pero independiente del ajuste de otros parámetros. (Discrete Optimized Protein Energy, or DOPE). Se basa en un estado de referencia optimizado que corresponde a átomos que no interactúan contenidos en una esfera homogénea cuyo radio depende del tamaño de las proteínas silvestres a partir de las cuales se derivó el potencial. Este potencial es útil para comparar proteínas silvestres y mutadas y se puede emplear como una medida local de las distorsiones entre ambas moléculas (Shen *et al*, 2006).

Potencial de ángulos de torsión

El cálculo de la energía potencial de ángulos torsión describe los cambios de la energía molecular producidos por la rotación alrededor de los enlaces y se expresa generalmente como una función del ángulo diedro o de torsión. El potencial de torsión está basado sobre las propensiones de los ángulos Φ y Ψ de un solo aminoácido (Benkert *et al*, 2008, Benkert *et al*, 2009).

El cálculo del RMSD de cada una de las proteínas silvestres, con sus correspondientes mutantes se realizó mediante el programa VMD 1.8.7. La obtención de la energía DOPE se realizó mediante el programa Modeller 9v7. El programa ProSa 2003 se utilizó para la obtención del valor Z-score y el programa QMEAN para obtener el potencial de ángulos de torsión.

2.14. Diagrama de flujo



3. RESULTADOS

Las proteínas que se escogieron para este estudio fueron seleccionadas por la cantidad de información disponible para ellas en cuanto a: 1) la función biológica, 2) la estructura tridimensional y 3) la disponibilidad de mutaciones asociadas con cáncer. En la tabla 5 se muestra la lista de las proteínas seleccionadas, con sus respectivos genes, claves de la secuencia de la proteína y de la secuencia de nucleótidos asignadas por el NCBI, de donde fueron descargadas, para su estudio.

TABLA 5. PROTEÍNAS DE ESTUDIO				
Símbolo del gen	Gene ID	Proteína	Clave de la proteína	Clave de nucleótido
ONCOGENES				
KRAS	3845	c-K-ras2 isoforma b	NP_004976	NM_004985.3
HRAS	3265	v-Ha-ras isoforma 1	NP_005334	NM_005343.2
ABL1	25	v-abl1 isoforma a	NP_005148	NM_005157.3
NRAS	4893	Neuroblastoma	NP_002515	NM_002524
SUPRESORES DE TUMORES				
TP53	7157	p53	NP_000537	NM_000546
CDKN2A	1029	p16	NP_000068.1	NM_000077.3

Tabla 5. Proteínas de estudio: Se muestran las proteínas que se estudiaron y sus claves que fueron asignadas en la base de datos de la NCBI.

El análisis se puede clasificar en tres etapas: 1) Análisis evolutivo de la proteína silvestre e identificación de las secuencias conservadas, 2) Localización de mutaciones e identificación del tipo de cambio y 3) Análisis comparativo de la estructura de las proteínas silvestres con respecto a las mutantes

Por motivos de espacio no se muestran todos los resultados generados en dichos estudios, por ello se creó una base de datos en línea donde se pueden mostrar dicha información, la dirección de la página principal de dicha base de datos es la siguiente: http://bioinformatica.homelinux.org/cancer_proteinas/. No obstante si se incluye un resumen de los principales tipos de cambios potencialmente asociados al cáncer.

Las cuales se describen a continuación:

3.1 Análisis evolutivo de la proteína silvestre e identificación de las secuencias conservadas

Este estudio se centró en la búsqueda de proteínas homólogas entre todas las especies posibles, las cuales dan un valor de expectación significativo en términos biológicos derivado de la puntuación del alineamiento local. La herramienta empleada para este análisis fue el programa PSI-Blast del NCBI y se utilizó un E-value de corte menor de 0.001 para identificar a las proteínas homólogas. Posteriormente se llevaron a cabo alineamientos múltiples por medio de los programas Clustal X 2.0, Muscle EBI y T-Coffee, los cuales se evaluaron mediante el programa GENDOC para encontrar el mejor alineamiento múltiple, y tomar lo de base, para identificar las regiones conservadas o las regiones variables y encontrar los dominios funcionales o estructurales. Y también utilizarlo para realizar el árbol filogenético y poder entender como evolucionaron dichas secuencias de proteínas.

3.1.1. Alineamiento de Hras, Kras y Nras

Las proteínas Hras, Kras y Nras en el humano son proteínas homologas, con un elevado grado de similitud, ya que al alinear dichas secuencias se encuentra que del aminoácido 1 al 165 tiene una similitud mayor del 90% y en la región hipervariable su grado de similitud desciende drásticamente siendo un 10 a un 15%. Por ello mismo al realizar la búsqueda de sus secuencias homologas, se encontró que compartían dichas secuencias. Por ello se realizó un solo alineamiento y un árbol filogenético para dichas proteínas.

Al observar dicho alineamiento múltiple (ver la figura 10 o en la página siguiente http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_h_k_nras.html) en el cual se alinearon 79 proteínas homologas podemos encontrar residuos que tiene un grado de conservación del 100%, dichos residuos son residuos críticos para la función biológica de los cuales se destacan los sitios de activación, sitios de interacción con otras moléculas y sitios efectores.

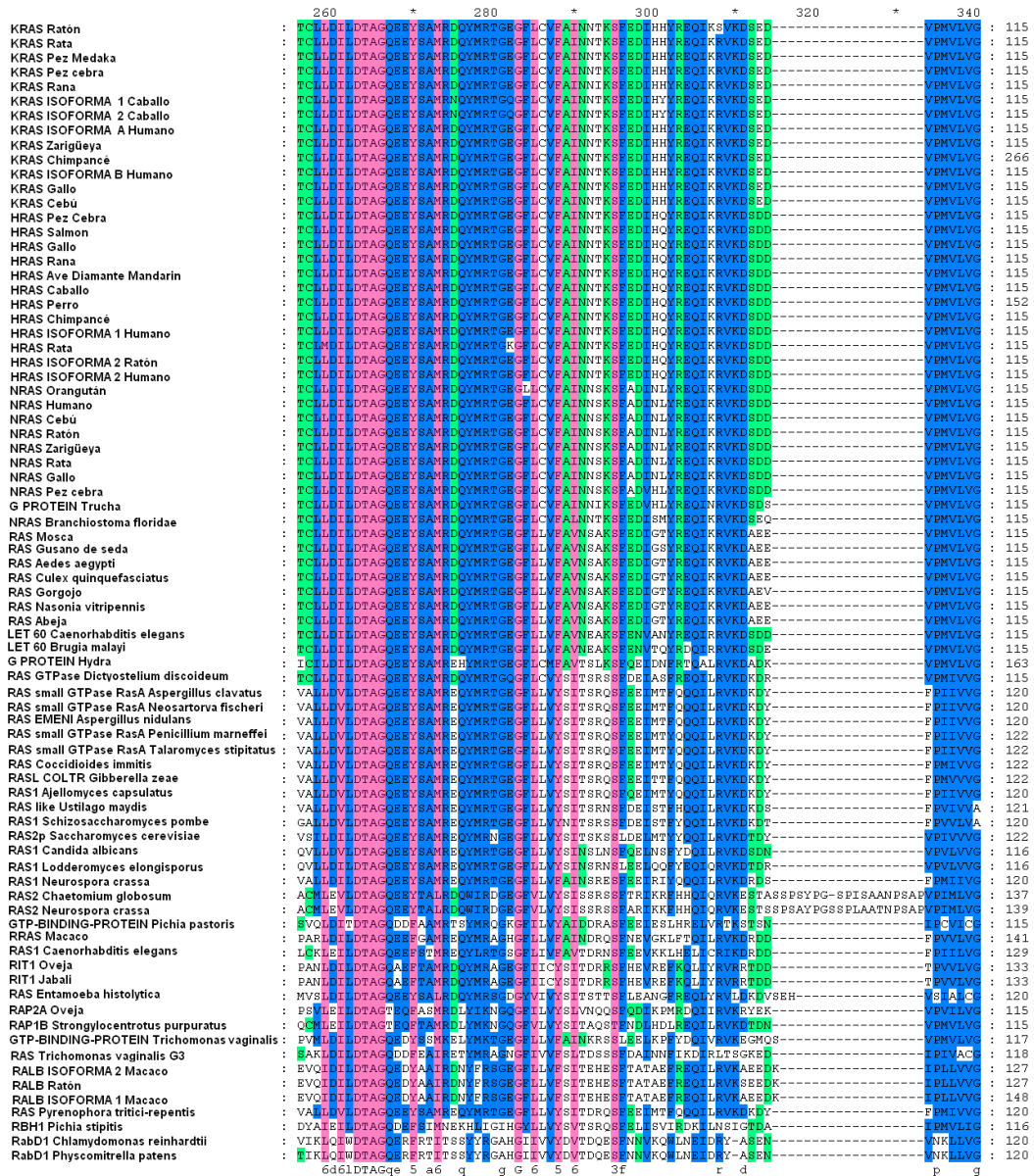


Figura 10. Alineamiento de las proteínas de la superfamilia Ras (Hras, Kras y Nras): Se muestra un fragmento representativo del alineamiento múltiple de Hras, Kras y Nras. Se alinearon 79 proteínas que se distribuyen en los reinos Animalia, Fungi y Protista. Todas estas secuencias pertenecen a la super-familia Ras. El fragmento del alineamiento que se muestra en la figura con respecto a la Kras humano va del residuo 50 al 115, se observan columnas de color rosa, las cuales indican que tienen un grado de conservación del 100%, una de ellas es la región DTAG en la secuencia de Kras humano pertenece a los residuos 57 al 60 (en el alineamiento esta región está en los residuos 263 al 266), esta columna es un sitio de interacción GDI (Se puede observar el alineamiento completo en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_h_k_nras.html).

3.1.2. Alineamiento de Abl1

En el caso de la proteína Abl1 isoforma A se encontraron homólogos en 12 especies todas ellas pertenecientes al reino animalia. Todas estas proteínas son tirosin-cinasas y sus funciones son muy parecidas. El alineamiento muestra (ver la figura 11 o la página siguiente http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_abl.html) que las regiones más conservadas son aquellas que coinciden con los dominios SH2, SH3 y PTKc_Abl, siendo el último dominio el que se encuentra con un grado mayor de conservación.

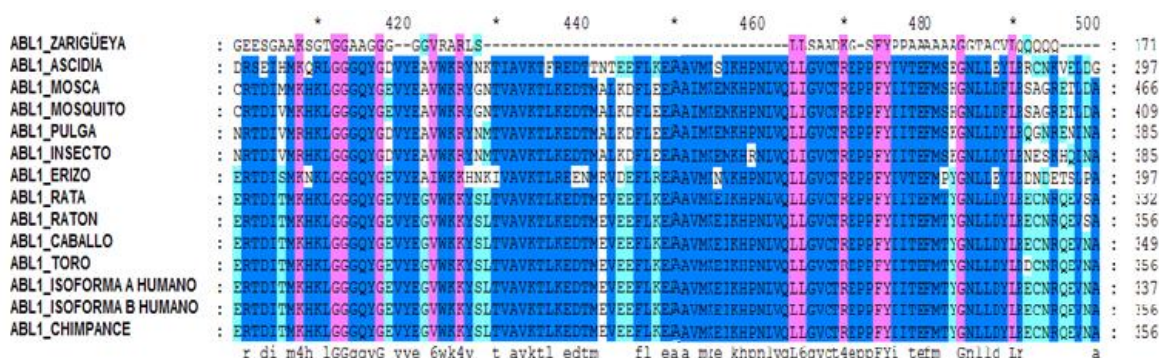


Figura 11. Alineamiento de la proteína Abl1: Se alinearon 12 especies (mediante el programa Muscle EBI) que pertenecen al reino animalia. El fragmento del alineamiento que se muestra en la figura con respecto a la Abl1 isoforma a humano va del residuo 237 al 337, se observan columnas de color rosa, las cuales indican que tienen un grado de conservación del 100%, las columnas de color azul tienen un grado mayor del 80% y el verde agua tienen un grado mayor del 60%. La región KLGGGQYG (En Abl1 isoforma a humano este localizado en los residuos 247 al 253), esta región pertenece al sitio de enlace del ATP y la región VWKKY (En Abl1 isoforma a humano este localizado en los residuos 259 al 263), la cual pertenece al enlace del dominio SH3/SH2. La región KHPNLVQLLG (En Abl1 isoforma a humano este localizado en los residuos 259 al 263), la cual pertenece al enlace del dominio SH3/SH2. Otra región importante es la secuencia ITEFMTYGNLL (En Abl1 isoforma a humano este localizado en los residuos 259 al 263) esta región pertenece al sitio de enlace del ATP. Estas zonas son regiones funcionales, por lo tanto tienen un alto grado de conservación, y aquellos residuos que son esenciales, tienen más del 80% de conservación (Se puede observar el alineamiento completo en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_abl.html).

3.1.3. Alineamiento de p53

Para la proteína p53 se realizó el alineamiento múltiple (ver la figura 12 o en la siguiente dirección http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_p16.html) con 27 especies, distribuidas principalmente en el reino animal en las que encontramos

primates, equinos, roedores, felinos, caninos, aves, anfibios, peces, parásitos, entre otros. El alineamiento resalta claramente que la región mejor conservada es la del dominio central.

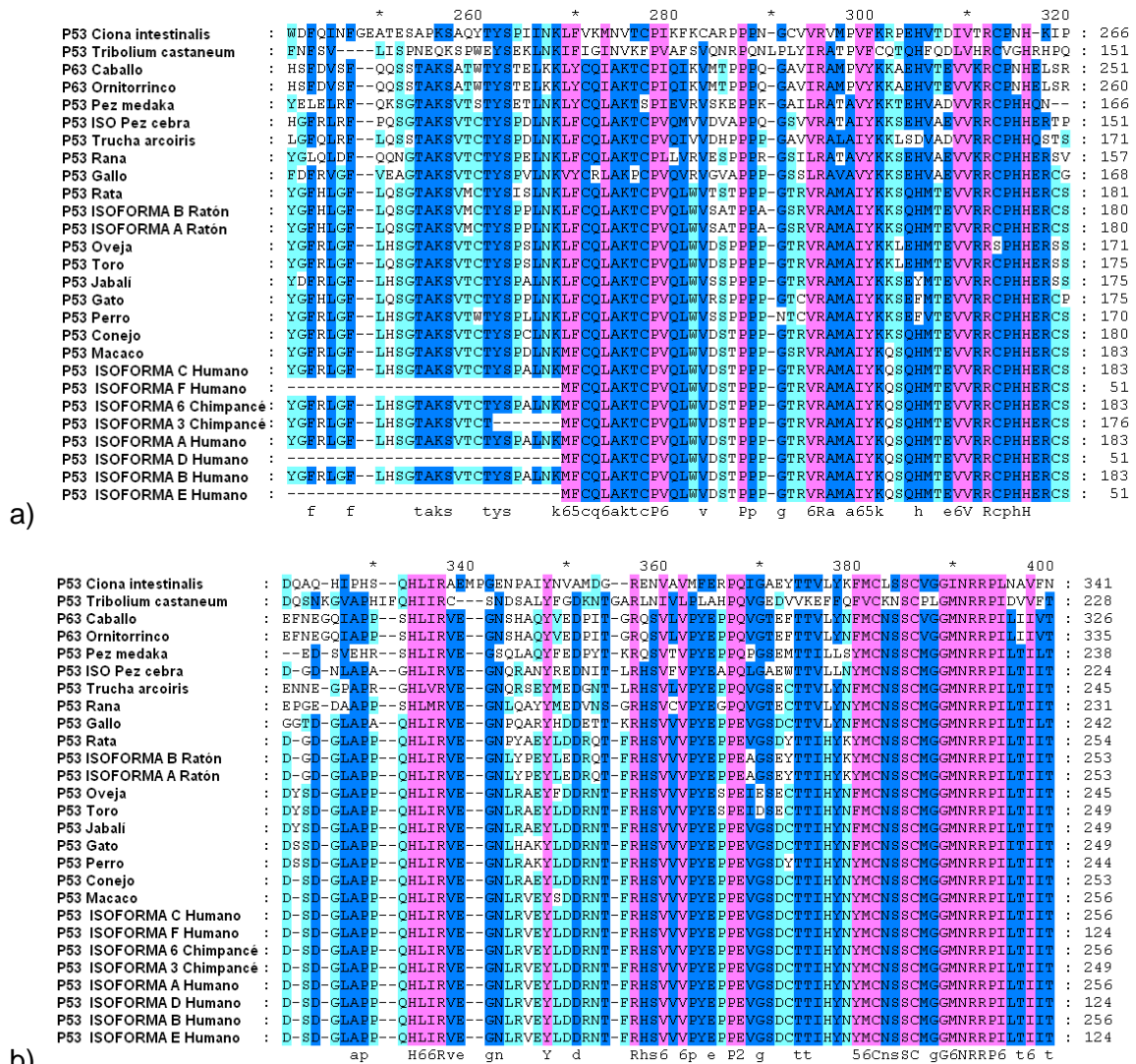


Figura 12. Alineamiento de la proteína p53: Se alinearon 27 especies que pertenecen al reino animalia. a)

En el fragmento del alineamiento se observan las secuencias: MECQAKTCFVQL (En el alineamiento estos residuos van del 269 al 278), PPP (En el alineamiento estos residuos van del 286 al 288), VRAMAIY (En el alineamiento estos residuos van del 294 al 298) y TEVVRCP (En el alineamiento estos residuos van del 307 al 314), en estas cuatro regiones se observan columnas con un grado de conservación elevado, estas regiones pertenece al dominio central, participando en el enlace la interacción de la p53 con el DNA. b) El fragmento del alineamiento se observan las secuencias: HLIRVE (En el alineamiento estos residuos van del 334 al 339), MCNSSCMGGMNRRP (En el alineamiento estos residuos van del 381 al 394), y TLED (En el alineamiento estos residuos van del 400 al 402), estas tres secuencias están muy conservadas ya que participan en la unión de p53 con el DNA. (Se puede observar el alineamiento completo en http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_p53.html).

3.1.4. Alineamiento de p16

El alineamiento de la proteína p16 mostró un bajo grado de conservación en la comparación de 20 secuencias homólogas (ver la figura 13 o en la siguiente dirección http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_p16.html), los cuales se encuentran distribuidas solamente en el reino animal. Estas especies son primates, roedores, equinos, peces, anfibios, aves y bovinos. La zona mejor conservada es la región ANK.

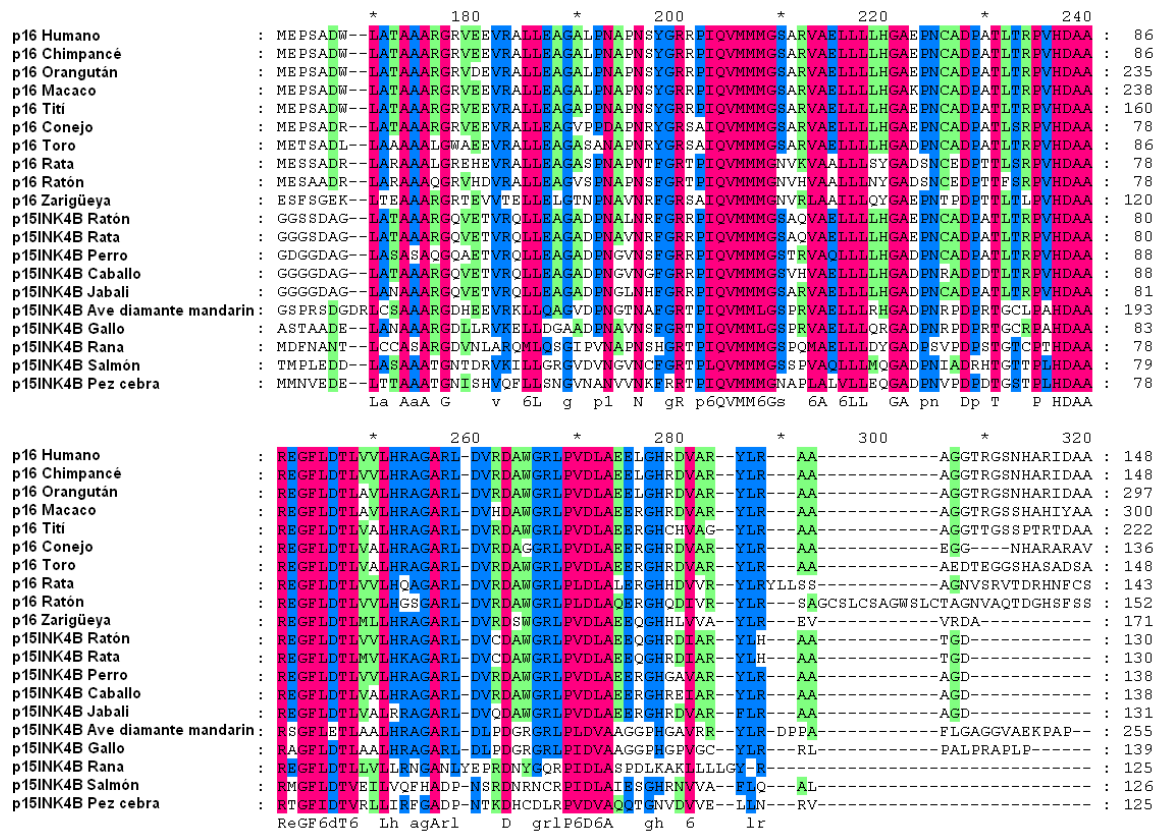


Figura 13. Alineamiento de la proteína p16: Se alinearon 20 secuencias que pertenecen al reino animalia. Este fragmento del alineamiento muestra el dominio ANK de p16, en el cual se observan zonas que están muy conservadas (columnas de color rosa 100%, color azul 80% al 99% y verde 60% al 79%). Estas secuencias son IQVMMMG (en el alineamiento estos residuos van del 203 al 209) que pertenecen a los subdominios loop1 y ANKII, la región VAELL (En el alineamiento estos residuos van del 213 al 218) está en el loop2, la secuencia PVHDAAREGFLDTL (En el alineamiento estos residuos van del 235 al 258) se localiza en el subdominio ANK III, y WGRLPVDLAEE (En el alineamiento estos residuos van del 265 al 275), está en los subdominios loop3 y ANKIV (Se puede ver el alineamiento completo en la siguiente dirección http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_p16.html).

3.1.5. Árbol filogenético de Hras, Nras y Kras

Hras, Kras y Nras tienen un grado de similitud elevado por ello mismo se realizó un árbol filogenético para las 3 proteínas (ver la figura 14 o la página siguiente http://bioinformatica.homelinux.org/cancer_proteinas/arbol_h_k_nras.html), en dicho árbol se encuentran 79 proteínas que pertenecen a la superfamilia RAS, las cuales están distribuidas en los reinos Protista, Fungí y Animalia, se puede dividir el dendograma en dos grandes regiones la primera está encerrada con un cuadro color amarillo y la segunda región está encerrada con un cuadro rojo, en grandes rasgos las especies de las proteínas del primer grupo pertenecen a mamíferos, invertebrados, insectos y nemátodos y el segundo grupo está compuesto primordialmente por especies que pertenecen al reino Fungí, y algunos mamíferos y protistas. Las proteínas Kras del ratón, la rata, el pez medaka, el pez cebra, la rana, el caballo, el humano, la zarigüeya, el chimpancé, el gallo y el cebú se encuentran agrupadas en un subnodo (ver la figura 14, las proteínas Kras se encuentran marcadas con un cuadro de color morado), el cual confirma su cercanía evolutiva, así como su función biológica, ya que todas estas proteínas provenientes de mamíferos, aves, anfibios y peces intervienen en la integridad del citoesqueleto, la proliferación, la diferenciación, la adhesión, la migración celular y la apoptosis, incluso están relacionadas con diferentes tipos de cáncer. El siguiente subnodo (ver la figura 14, las proteínas Hras se encuentran marcadas con un cuadro de color rosa), pertenece a las proteínas Hras del pez cebra, del salmón, del gallo, de la rana, del ave diamante mandarín, del caballo, del perro, del chimpancé, del humano, de la rata y del ratón, este subnodo se encuentra distribuido de acuerdo a su similitud, encontrándose juntos los peces, los anfibios y la del ave, así como la proteína Hras del chimpancé se encuentra a lado de Hras isoforma 1 del humano, todas estas proteínas tienen funciones biológicas similares como la adhesión, regulación de la actina en el citoesqueleto, endocitosis, participa en la vías de señalización como la de MAPK, la de ErbB, y la GnRH, así como la en la melanogénesis. El siguiente subnodo (ver la figura 14, las proteínas Nras se encuentran marcadas con un cuadro de color amarillo), pertenece a las proteínas Nras del orangután, del humano, del cebú, del ratón, de la zarigüeya, de la rata, del gallo y del pez cebra, estas proteínas tienen funciones biológicas como las vías de señalización de MAPK, de ErbB, de GnRH, entre otras, así como están relacionados con melanogénesis, cáncer endometrial, leucemia mieloide aguda, gliomas y melanomas.

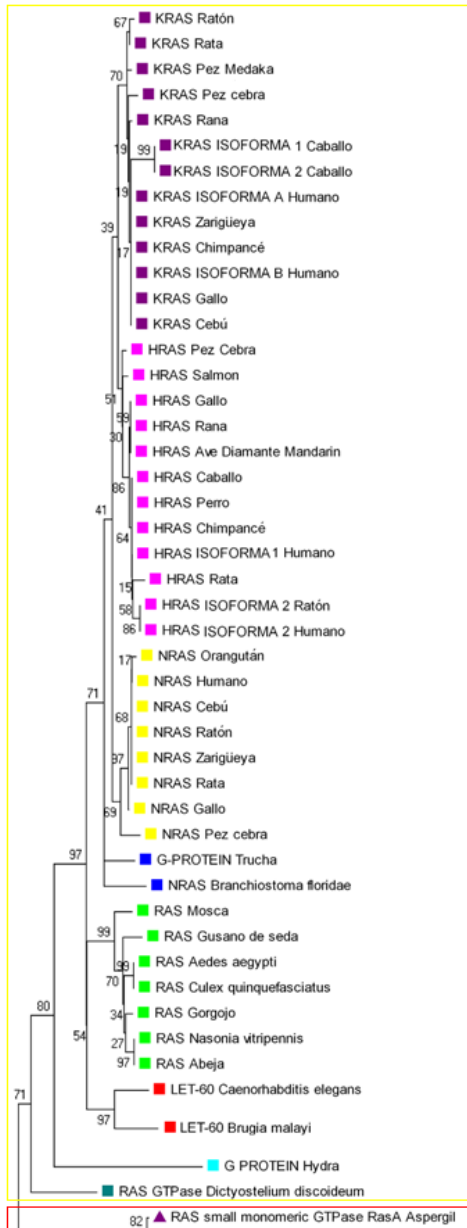


Figura 14. Fragmento 1 del árbol filogenético de las proteínas Hras, Kras y Nras: La historia evolutiva se infirió usando el método de Neighbor Joining, mediante el programa MEGA 4.0 (Se puede observar el árbol filogenético completo de las proteínas Hras, Kras y Nras en la siguiente dirección electrónica http://bioinformatica.homelinux.org/cancer_proteinas/arbol_h_k_nras.html).

Como podemos observar las proteínas Hras, Nras y Kras del humano son proteínas homologas, es interesante que separen en diferentes subnodos agrupándose con las mismas proteínas pero que provienen de diferentes especies. Esto se debe a que Hras, Nras y Kras son proteínas que participan en las mismas funciones en el ciclo celular pero son codificadas por diferentes genes, y estos genes se originaron por duplicación

de un gen ancestro en común, es por ello que comparten la misma función pero en diferentes tejidos. Es por ello que solamente los homólogos ortólogos se encuentran en el mismo subnodo del árbol filogenético. La proteína Nras de *Branchiostoma floridae* no se localiza en el mismo nodo de las demás Nras, debido a que esta especie es un invertebrado marino que no ha cambiado en 500 millones de años y se parece mucho al ancestro común de todos los vertebrados, el hecho de encontrar similitud significativa entre la secuencias la proteína Nras del invertebrado con las secuencias de Kras, Nras y Hras del humano confirman que la función biológica que desempeñan a nivel celular es muy importante, por lo tanto sus secuencias se conservan a lo largo de la evolución, y por lo tanto los residuos con alto grado de conservación deben de ser críticos en la función biológica o en la estructura de dichas proteínas.

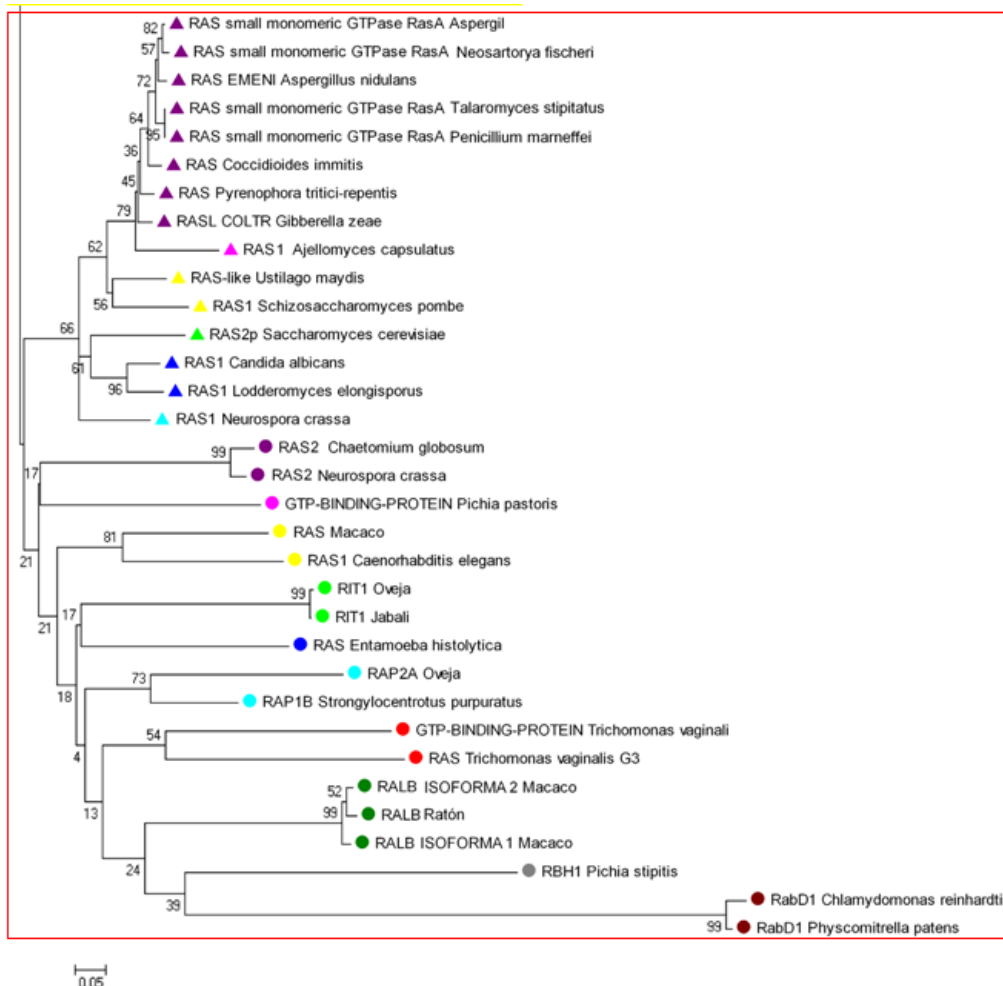


Figura 15. Fragmento 2 del árbol filogenético de las proteína Hras, Kras y Nras: El Árbol filogenético de la proteína Nras se realizo mediante el programa MEGA 4.0.La historia evolutiva fue inferida usando el método de Neighbor-Joining. (Se puede observar el árbol filogenético completo de las proteínas Hras, Kras y Nras en la siguiente dirección http://bioinformatica.homelinux.org/cancer_proteinas/arbol_h_k_nras.html).

En la segunda gran región del árbol se encuentran proteínas que pertenecen a la superfamilia ras, pero no pertenecen a la subfamilia HRas/NRas/KRas. Pero tienen función básica de hidrolizar el GTP, interviniendo en vías de señalización de MAPK, como en el caso de RAS2 de *Neurospora crassa* que pertenece a la subfamilia RAS2 (ver la figura 15), la cual se caracteriza de ser exclusiva para el reino Fungi.

3.1.6. Árbol filogenético de Abl1

En el árbol filogenético de Abl1 isoforma A (ver la figura 16 o en la siguiente dirección http://bioinformatica.homelinux.org/cancer_proteinas/arbol_abl.html), se forman los siguiente grupos: las especies más cercanas con respecto a la proteína humana, se encuentra el chimpancé, el caballo, el toro, después se encuentran la rata y el ratón, se encuentran juntas en un nodo, mostrando la cercanía evolutiva de estas dos proteínas, las cuales están marcadas con un cuadrado verde. Las proteínas en el insecto himenóptero y la pulga, se encuentran en otro nodo identificadas con un cuadrado azul y las proteínas del mosquito y la mosca también están localizadas en otro nodo que están señaladas con un triángulo rosa. La distribución de este árbol nos muestra que estas proteínas evolutivamente están más alejadas a la proteína Abl1 isoforma A humana, se observa cuales secuencias están más relacionadas entre ellas, las cuales se agrupan.

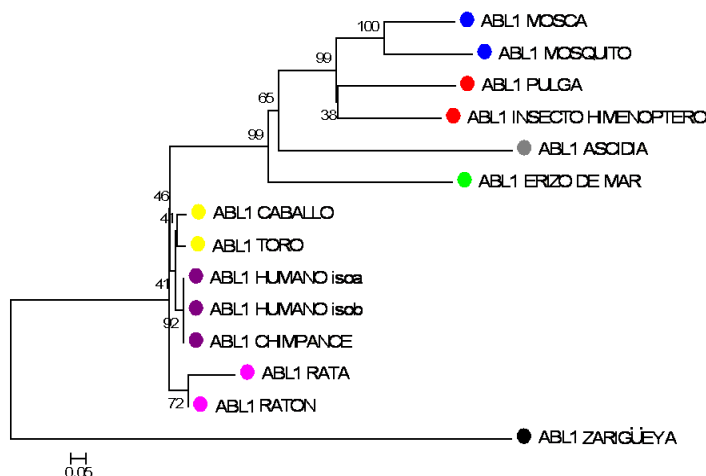


Figura 16. Árbol filogenético de la proteína Abl1: El Árbol filogenético de la proteína Abl1 se realizó mediante el programa MEGA 4.0 a partir del alineamiento previamente obtenido por el programa Muscle EBI. La historia evolutiva fue inferida usando el método de Neighbor-Joining. El porcentaje se obtuvo mediante el test bootstrap (100000 replicas), creando un árbol filogenético consenso. Este árbol está dibujado a escala, con longitudes de los brazos en unidades similares a las distancias evolutivas. Las distancias evolutivas fueron calculadas usando el método basado en la matriz de Dayhoff y en unidades del número de sustituciones de aminoácido por sitio (Se puede observar el árbol filogenético de las proteínas Abl1 isoforma A en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/arbol_abl.html).

.8.1.8. Árbol filogenético de p16

En el árbol filogenético (ver la figura 18 o en la siguiente dirección electrónica http://bioinformatica.homelinux.org/cancer_proteinas/arbol_p16.html) se puede visualizar la relación evolutiva de las proteínas p16 entre ellas. La proteína humana se encuentra en un nodo que es compartido con p16 del chimpancé, p16 del macaco, p16 del orangután y p16 del tití, son las proteínas con mayor cercanía evolutiva, teniendo funciones biológicas similares, El endograma nos muestra que hay una relación estrecha evolutiva entre la proteína p15INK4B de los roedores, el perro, caballo y jabalí.

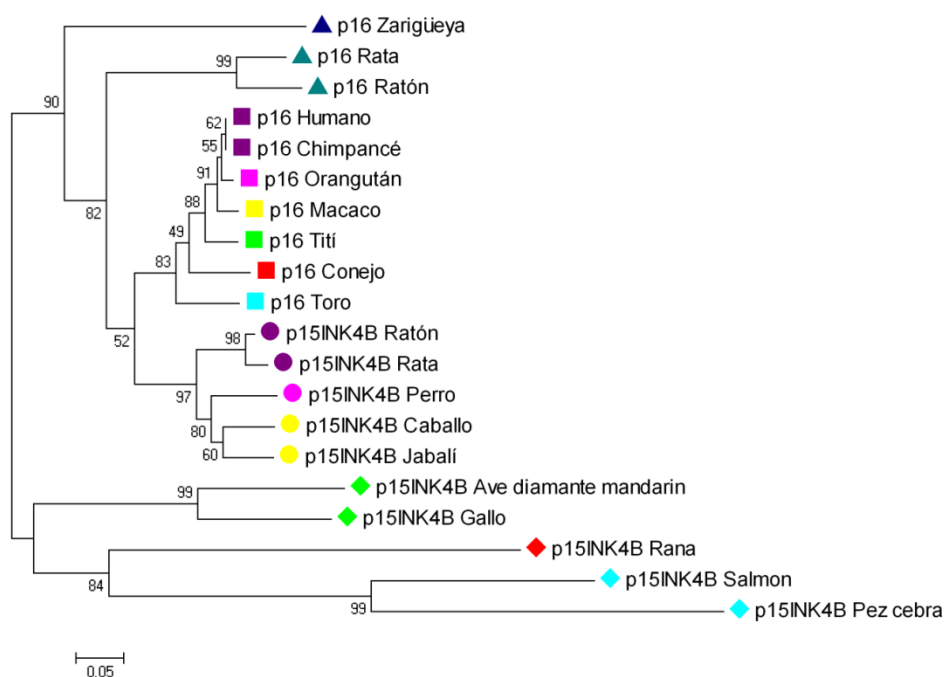


Figura 18. Árbol filogenético de la proteína p16: El Árbol filogenético de la proteína p16 se realizó mediante el programa MEGA 4.0. El porcentaje se obtuvo mediante el test bootstrap (100000 replicas), creando un árbol filogenético consenso. Este árbol está dibujado a escala, con longitudes de los brazos en unidades similares a las distancias evolutivas. Las distancias evolutivas fueron calculadas usando el método basado en la matriz de Dayhoff (Se puede observar el árbol filogenético de p16 en la siguiente dirección electrónica http://bioinformatica.homelinux.org/cancer_proteinas/arbol_p16.html).

3.2. Localización de mutaciones, patrones mutagénicos e identificación del tipo de cambio

La localización de los “motifs” conservados en los alineamientos locales de los homólogos tuvo otro fin importante en este estudio. Se investigó la posible presencia de

mutaciones puntuales implicadas en el desarrollo de cáncer en estas regiones altamente conservadas a lo largo de la evolución.

Para esto primero se hizo una búsqueda en varias bases de datos especializadas, lo cual permitió encontrar las mutaciones puntuales para cada una de las proteínas que se estudiaron asociadas con cáncer. Entre las bases de datos consultadas, una de las más relevantes fue COSMIC (Catalogue Of Somatic Mutations In Cancer) del instituto de Sanger, (<http://www.sanger.ac.uk/genetics/CGP/cosmic/>). A partir de estos datos se seleccionaron aquellas mutaciones que originan sustituciones en un solo aminoácido en dichas proteínas. Con los datos obtenidos tanto de la selección de las mutaciones puntuales, y su ubicación con respecto a las regiones conservadas y variables proporcionadas por los alineamientos se realizó la siguiente observación:

- Localización de las regiones conservadas y no conservadas de las proteínas.
- Patrones de mutagenicidad de las proteínas
- Identificación del tipo de cambio.

3.2.1. Localización de las regiones conservadas y no conservadas de Hras

Se realizaron alineamientos múltiples para cada una de las proteínas estudiadas con sus secuencias ortólogas.

3.2.1.1 Localización de las regiones conservadas y no conservadas de Hras

En el alineamiento de todas las secuencias de proteínas Hras que se encuentran en diferentes especies (ver la figura 19 o ver la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_hras.html) se puede observar que la proteína Hras isoforma 1, es una molécula con un grado de conservación alta. A partir de este alineamiento se obtuvo el score de conservación con el programa Clustal X 2.0, en el cual se identificó que en la proteína Hras isoforma 1 tiene 142 residuos con un grado de conservación del 100%, 16 columnas tienen una conservación del 50 al 99% y 31 aminoácidos tienen una conservación menor del 50%.

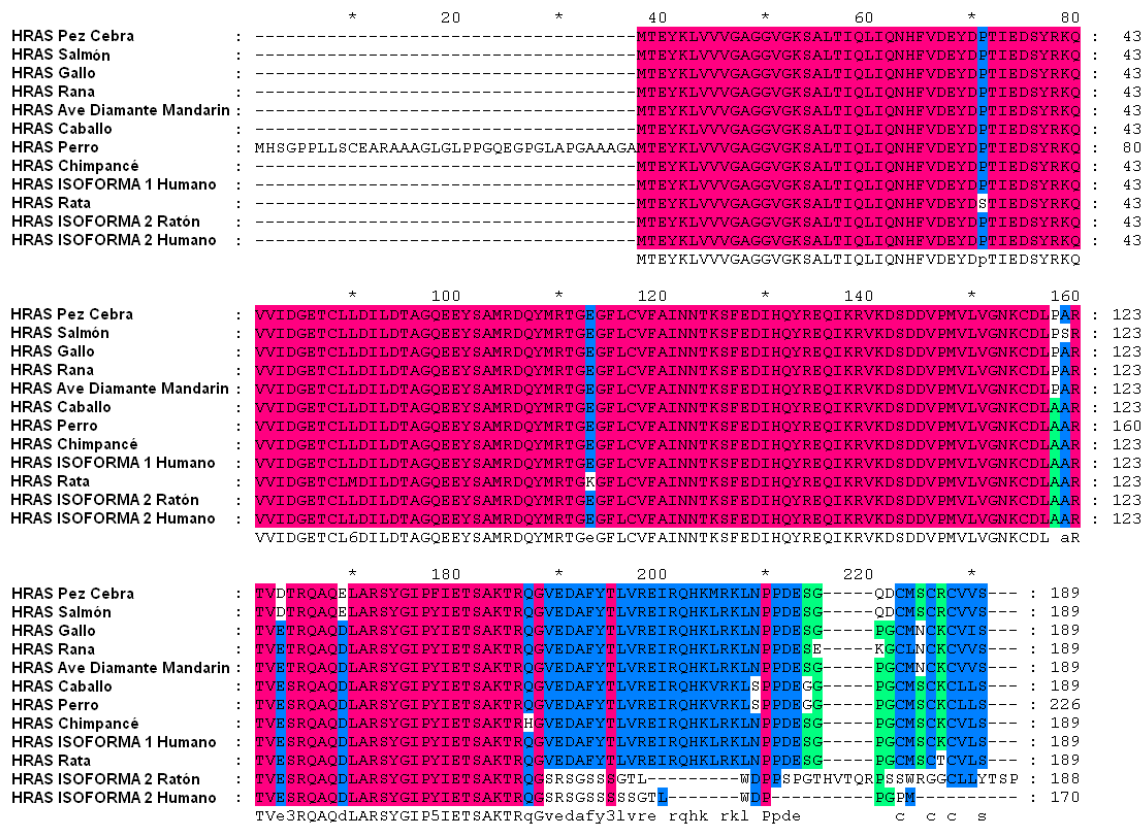


Figura 19. Regiones conservadas y variables de Hras: En esta figura se muestra el alineamiento múltiple de Hras. Se alinearon 12 proteínas que se distribuyen en los reinos Animalia. Todas estas secuencias son proteínas Hras que pertenecen a otras especies las cuales fueron localizadas mediante la herramienta PSI-Blast del NCBI, alineadas con el programa Muscle EBI. En el cual podemos observar las regiones variables y las regiones conservadas. Las columnas que se encuentran de color rosa indican que los aminoácidos tienen aproximadamente una conservación del 100%, las que están de color azul mayor al 80% y las de color verde mayor al 60% (Se puede observar el alineamiento completo de Hras en la dirección http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_hras.html).

3.2.1.2. Localización de las regiones conservadas y no conservadas de Kras

A partir del alineamiento de todas las secuencias de proteínas Kras que se encuentran en diferentes especies (ver la figura 20 o ver la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_kras.html) se obtuvo el score de conservación con el programa Clustal X 2.0, en el cual se identificó que en la proteína Kras isoforma B tiene 150 residuos con un grado de conservación del 100%, 26 columnas tienen una conservación del 50 al 99% y 12 aminoácidos tienen una conservación menor del 50%.

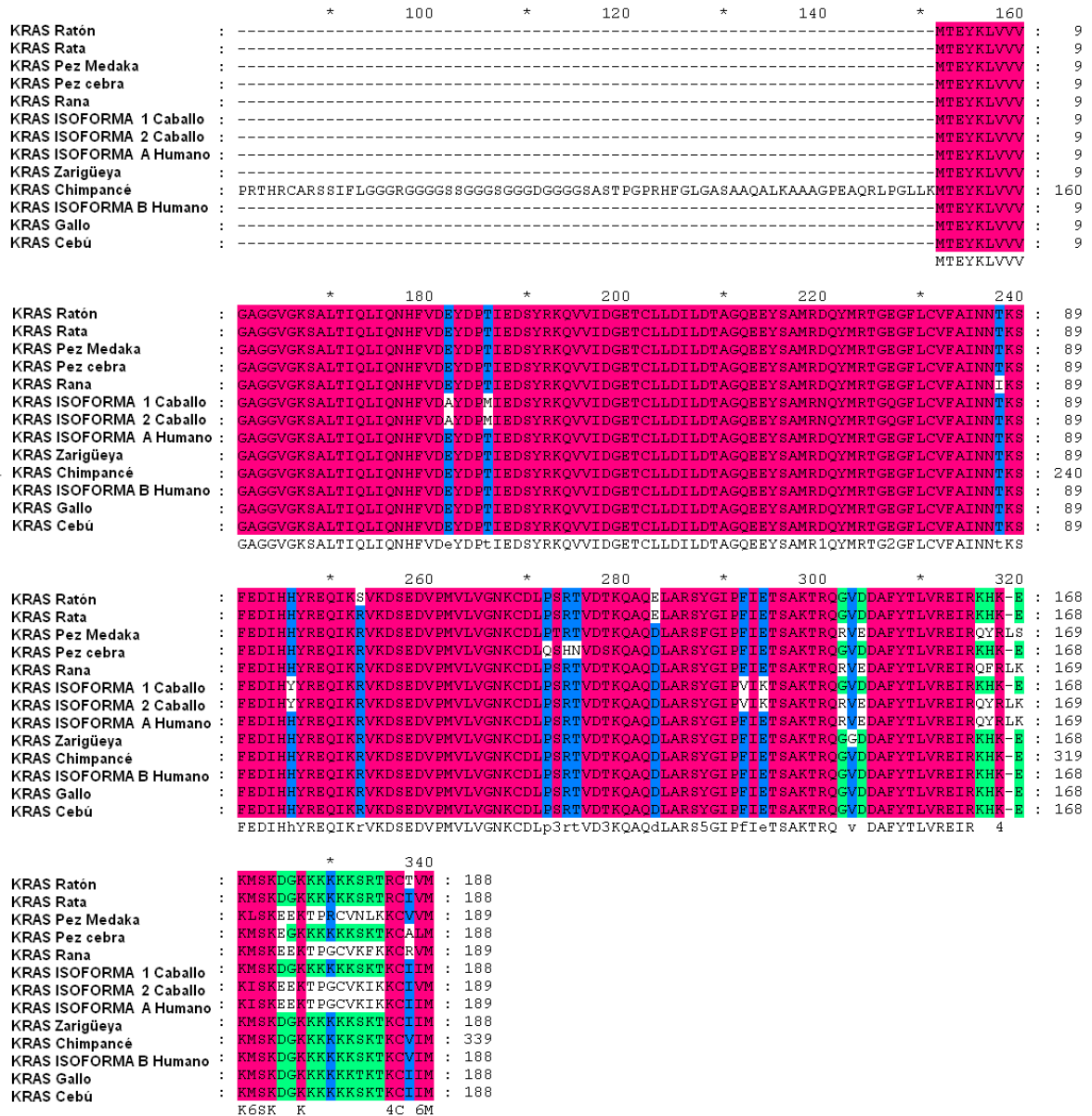


Figura 20. Regiones conservadas y variables de Kras: En esta figura se muestra un fragmento del alineamiento múltiple de Kras. Se alinearon 13 proteínas que se distribuyen en los reinos Animalia. Todas estas secuencias son proteínas Kras que pertenecen a otras especies. En el cual podemos observar las regiones variables y las regiones conservadas. Las columnas que se encuentran de color rosa indican que los aminoácidos tienen aproximadamente una conservación del 100%, las que están de color azul mayor al 80% y las de color verde mayor al 60%. (Se puede observar el alineamiento completo en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_kras.html).

3.2.1.3. Localización de las regiones conservadas y no conservadas de Nras

A partir del alineamiento de todas las secuencias de proteínas Nras que se encuentran en diferentes especies (ver la figura 21 o ver la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_nras.html) se obtuvo el score de conservación con el programa Clustal X 2.0, en el cual se identificó que Nras del humano tiene 106 residuos con un grado de conservación del 100%, 61 columnas tienen una conservación del 50 al 99% y 22 aminoácidos tienen una conservación menor del 50%.

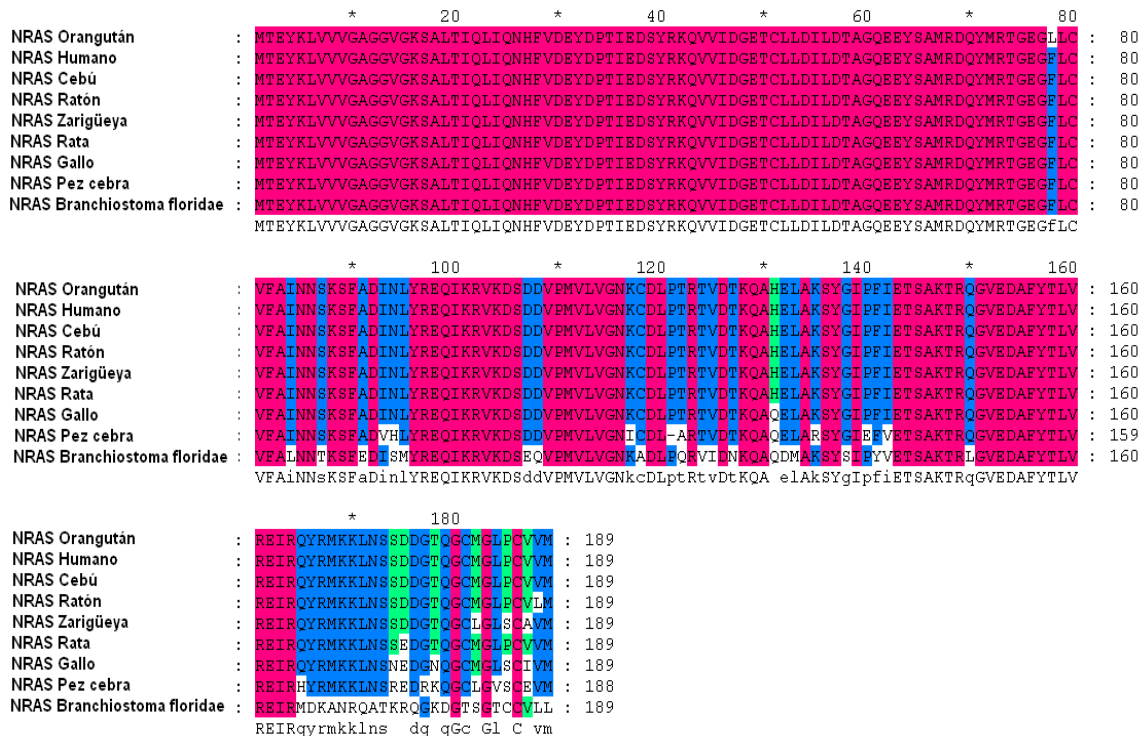


Figura 21. Regiones conservadas y variables de Nras: En esta figura se muestra un fragmento del alineamiento múltiple de Nras. Se alinearon 9 proteínas que se distribuyen en los reinos Animalia. Todas estas secuencias son proteínas Nras que pertenecen a otras especies. En el cual podemos observar las regiones variables y las regiones conservadas. Las columnas que se encuentran de color rosa indican que los aminoácidos tienen aproximadamente una conservación del 100%, las que están de color azul mayor al 80% y las de color verde mayor al 60% (Se puede observar el alineamiento completo de las proteínas Nras en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_nras.html).

3.2.1.4. Localización de las regiones conservadas y no conservadas de Abl1

A partir del alineamiento de todas las secuencias de proteínas Abl1 que se encuentran en diferentes especies (ver la figura 22 o ver la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_abl1.html) se obtuvo el score de conservación con el programa Clustal X 2.0, en el cual se identificó que Abl1 del humano tiene 62% de los residuos con un grado de conservación del 100%, 16% de sus columnas tienen una conservación del 50 al 99% y 22% de los aminoácidos tienen una conservación menor del 50%.

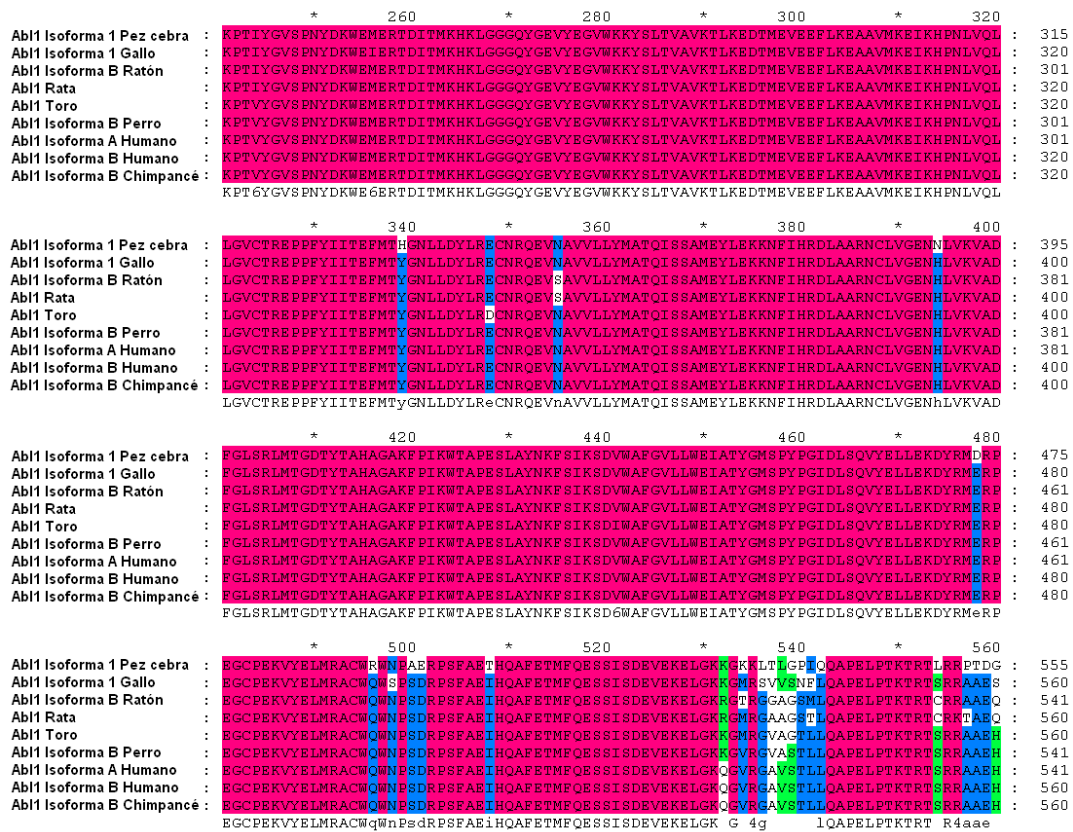


Figura 22. Regiones conservadas y variables de Abl1: En esta figura se muestra un fragmento del alineamiento múltiple de Abl1. Se alinearon 9 proteínas que se distribuyen en los reinos Animalia. Todas estas secuencias son proteínas Abl1 que pertenecen a otras especies. En el cual podemos observar las regiones variables y las regiones conservadas. Las columnas que se encuentran de color rosa indican que los aminoácidos tienen aproximadamente una conservación del 100%, las que están de color azul mayor al 80% y las de color verde mayor al 60% (Se puede observar el alineamiento de la proteína Abl1 en la página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_abl1.html).

3.2.1.5. Localización de las regiones conservadas y no conservadas de p53

A partir del alineamiento de todas las secuencias de proteínas p53 que se encuentran en diferentes especies (ver la figura 23 o ver la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_p53.html) se obtuvo el score de conservación con el programa Clustal X 2.0, en el cual se identificó que p53 del humano tiene 39% de los residuos con un grado de conservación del 100%, 23% de sus columnas tienen una conservación del 50 al 99% y 28% de los aminoácidos tienen una conservación menor del 50%.

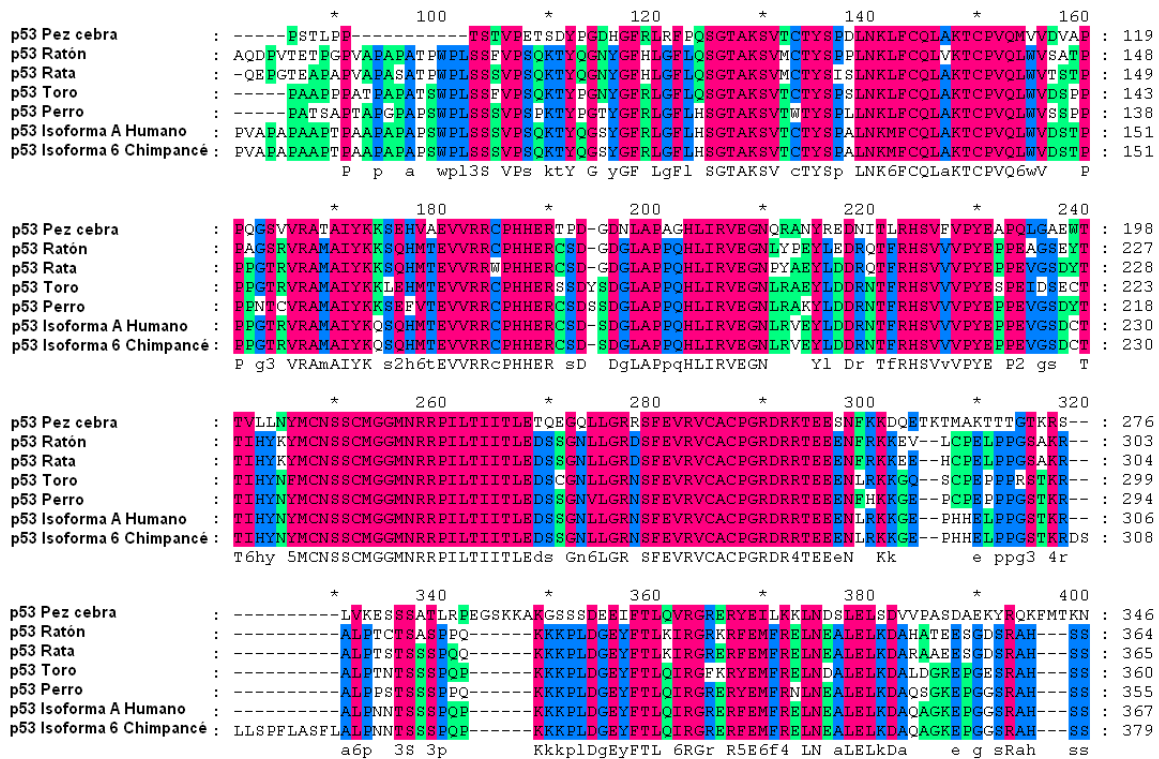


Figura 23. Regiones conservadas y variables de p53: En esta figura se muestra un fragmento del alineamiento múltiple de p53. Se alinearon 7 proteínas que se distribuyen en los reinos Animalia. Todas estas secuencias son proteínas p53 que pertenecen a otras especies. En el cual podemos observar las regiones variables y las regiones conservadas. Las columnas que se encuentran de color rosa indican que los aminoácidos tienen aproximadamente una conservación del 100%, las que están de color azul mayor al 80% y las de color verde mayor al 60% (Se puede observar el alineamiento de p53 en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_p53.html).

3.2.1.6. Localización de las regiones conservadas y no conservadas de p16

A partir del alineamiento de todas las secuencias de proteínas p16 que se encuentran en diferentes especies (ver la figura 24 o ver la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_p16.html) se obtuvo el score de conservación con el programa Clustal X 2.0, en el cual se identificó que p16 del humano tiene 102 de los residuos con un grado de conservación del 100%, 19 de sus columnas tienen una conservación del 50 al 99% y 35 de los aminoácidos tienen una conservación menor del 50%.

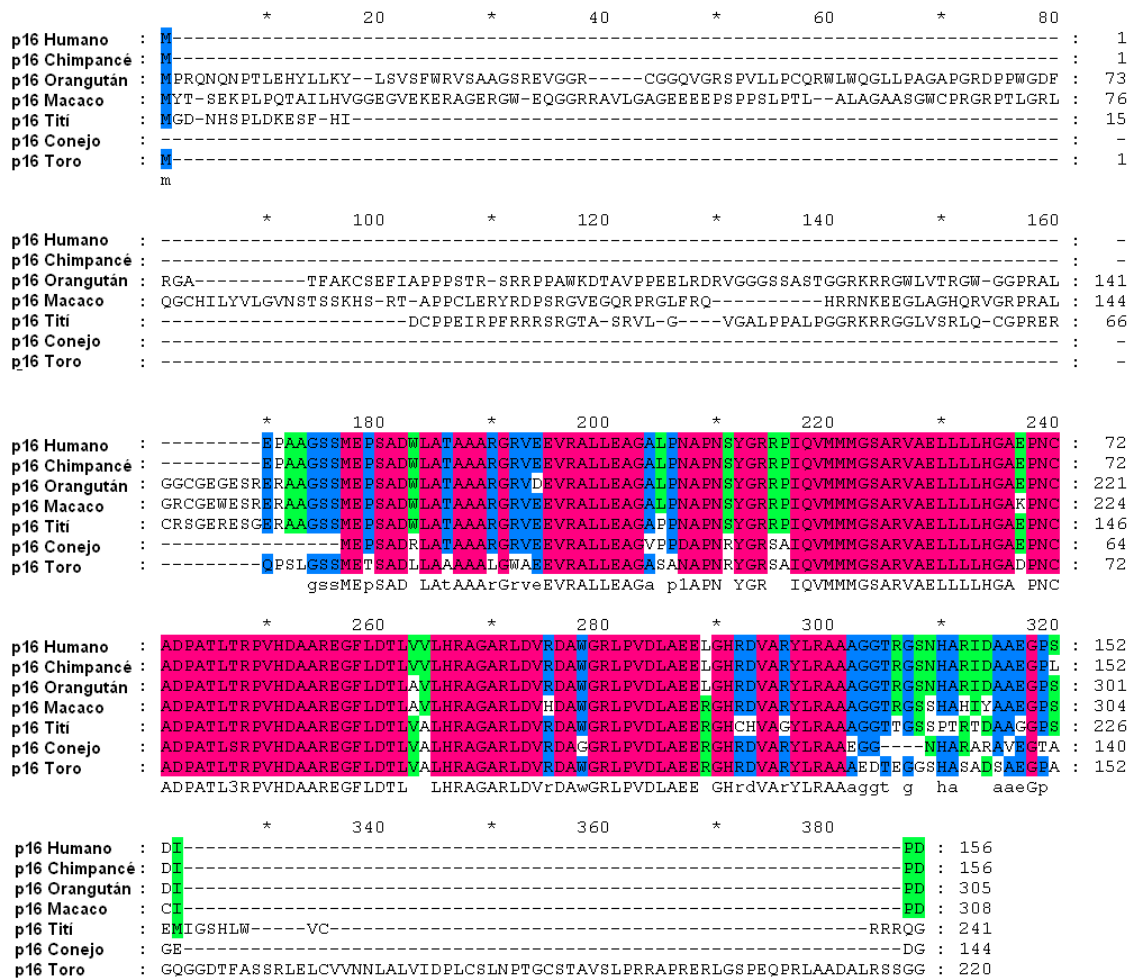


Figura 24. Regiones conservadas y variables de p16: En esta figura se muestra el alineamiento múltiple de p16. Se alinearon 7 proteínas que se distribuyen en los reinos Animalia. Todas estas secuencias son proteínas p16 que pertenecen a otras especies. En el cual podemos observar las regiones variables y las regiones conservadas. Las columnas que se encuentran de color rosa indican que los aminoácidos tienen aproximadamente una conservación del 100%, las que están de color azul mayor al 80% y las de color verde mayor al 60% (Se puede observar el alineamiento de la proteína p16 en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/alineamiento_estruc_p16.html).

3.2.2 Patrones mutagénicos

Después de obtener dichos alineamientos se cuantificaron las columnas que tenían un alto grado de conservación (100%), un intermedio grado de conservación (50 al 99 %), y un bajo grado de conservación (0 al 49%). Para después relacionar cuantos residuos de las mutaciones relacionadas con cáncer caen en las columnas con alto, mediano y bajo grado de conservación (ver la tabla 7).

PROTEÍNA	CONSERVACIÓN	PORCENTAJE DE REGIONES CONSERVADAS	PORCENTAJE DE MUTACIONES QUE CAEN EN LA REGION	PORCENTAJE DE LAS COLUMNAS DONDE OCURREN MUTACIONES
KRAS	ALTA	80	97	13
	INTERMEDIO	14	1.5	4
	BAJA	6	1.5	8
HRAS	ALTA	75	100	11
	MEDIA	8	0	0
	BAJA	17	0	0
NRAS	ALTA	56	100	12
	INTERMEDIO	32	0	0
	BAJA	12	0	0
P53	ALTA	39	80	40
	INTERMEDIO	23	13	11
	BAJA	38	7	3
P16	ALTA	65	74	84
	MEDIA	12	13	78
	BAJA	23	13	43
ABL1	ALTA	62	92	84
	INTERMEDIO	16	8	78
	BAJA	22	0	0

Tabla 7. Relación de las regiones conservadas con las mutaciones relacionadas con cáncer: En esta tabla se muestra cual es el porcentaje del grado de conservación de cada residuo de las proteínas estudiadas, el porcentaje de las mutaciones que caen en las regiones alta, intermedio y bajo grado de conservación, así como el grado de porcentaje de las mutaciones donde ocurren las mutaciones.

La relación que se obtuvo fue que la mayoría de las mutaciones caen en columnas con un alto grado de conservación, como en el caso de la proteína Kras, el 97 por ciento de los residuos que están implicados en mutaciones caen en regiones altamente conservadas, esta proteína tiene un 80% de columnas conservadas y en el 13% de dichas columnas altamente conservadas se producen mutaciones cancerígenas. Aun en la proteína p16 que es una secuencia que se encuentra menos conservada a lo largo de

la evolución, podemos observar la misma tendencia, la mayoría de las mutaciones relacionadas con cáncer caen en columnas altamente conservadas o con una conservación intermedia. Por lo tanto podemos sugerir que las mutaciones que caen en regiones altamente conservadas van a tener un impacto mutacional mayor a las que se encuentren en columnas con menor grado de conservación.

3.2.2.1. Patrones mutagénicos de Hras, Kras y Nras

Las mutaciones relacionadas con el cáncer en Hras (ver la figura 25) y Nras (ver la figura 27), tienen un 100%, para Kras el 97% (ver la figura 26) de las mutaciones se encuentran en columnas altamente conservadas, las mutaciones que caen el residuo E31 tiene un porcentaje del 58% de conservación y el residuo T35 tiene un porcentaje de conservación del 46%. Con respecto a las proteínas Hras, Kras y Nras el 100% están en regiones funcionales, también se identificó que la mayoría de las mutaciones puntuales se localizan en tres regiones principalmente, las cuales se encuentran en los aminoácidos 12, 13, 60, 61, 63, 32 y 35, estas regiones son importantes en la función de las proteínas, las cuales tienen que ver con sitios de enlace GTP/Mg²⁺ (Pacold ME *et al* 2000), sitio de interacción GEF (Margarit *et al.*, 2003), sitio de interacción GDI y la región switch I lo cual concuerda con la literatura, que las mutaciones en la familia ras produce una sobre activación en ellas (Marchler-Bauer A *et al.* 2009 y Marchler-Bauer A, Bryant SH *et al* 2004).

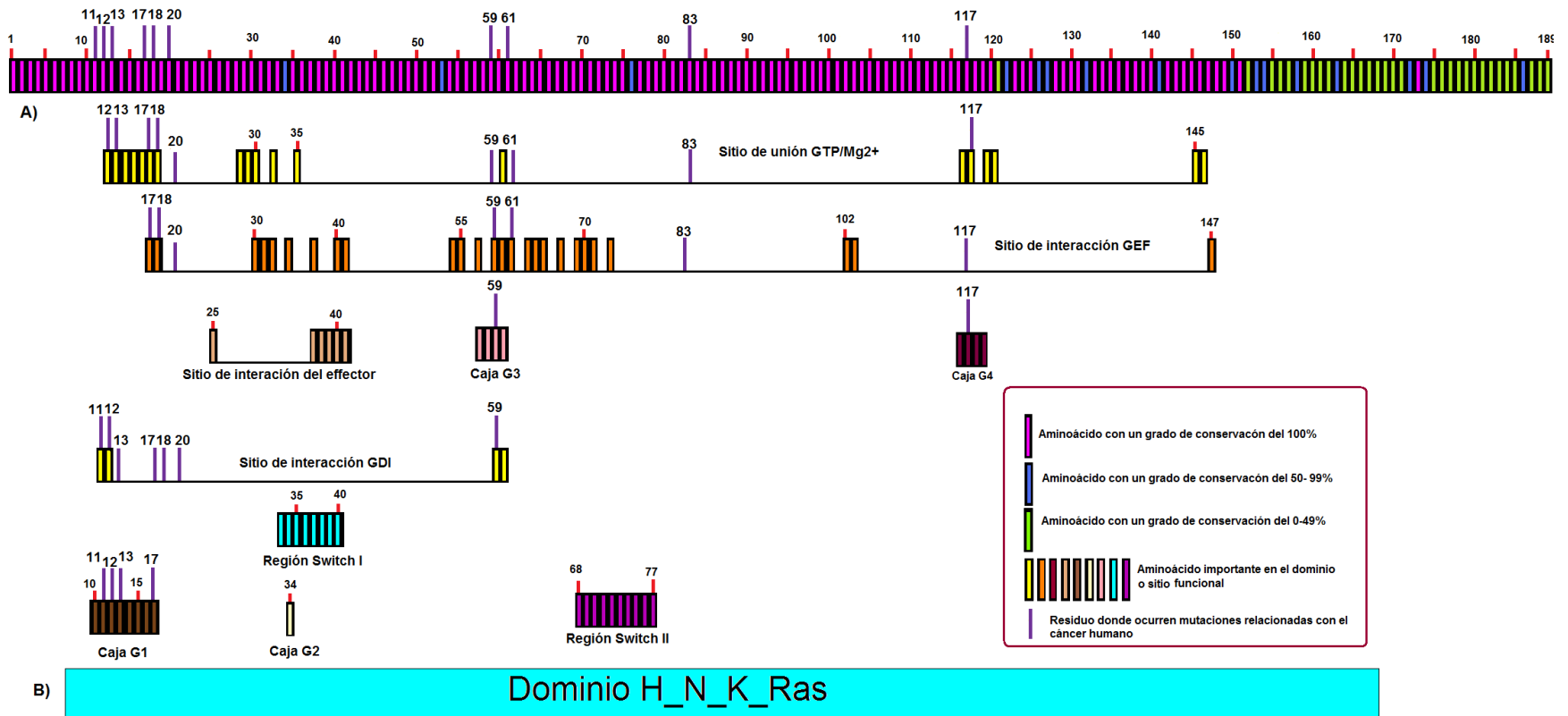


Figura 25. Patrones mutagénicos de Hras: En esta figura se muestra una relación entre las regiones conservadas de la proteína Hras y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína, destacando aquellos residuos que son importantes para dichos subdominios y cuales residuos que están relacionados con el cáncer humano se encuentran en residuos importantes para estos subdominios o aquellos residuos que solamente se encuentran en los subdominios, sin ser residuos esenciales.

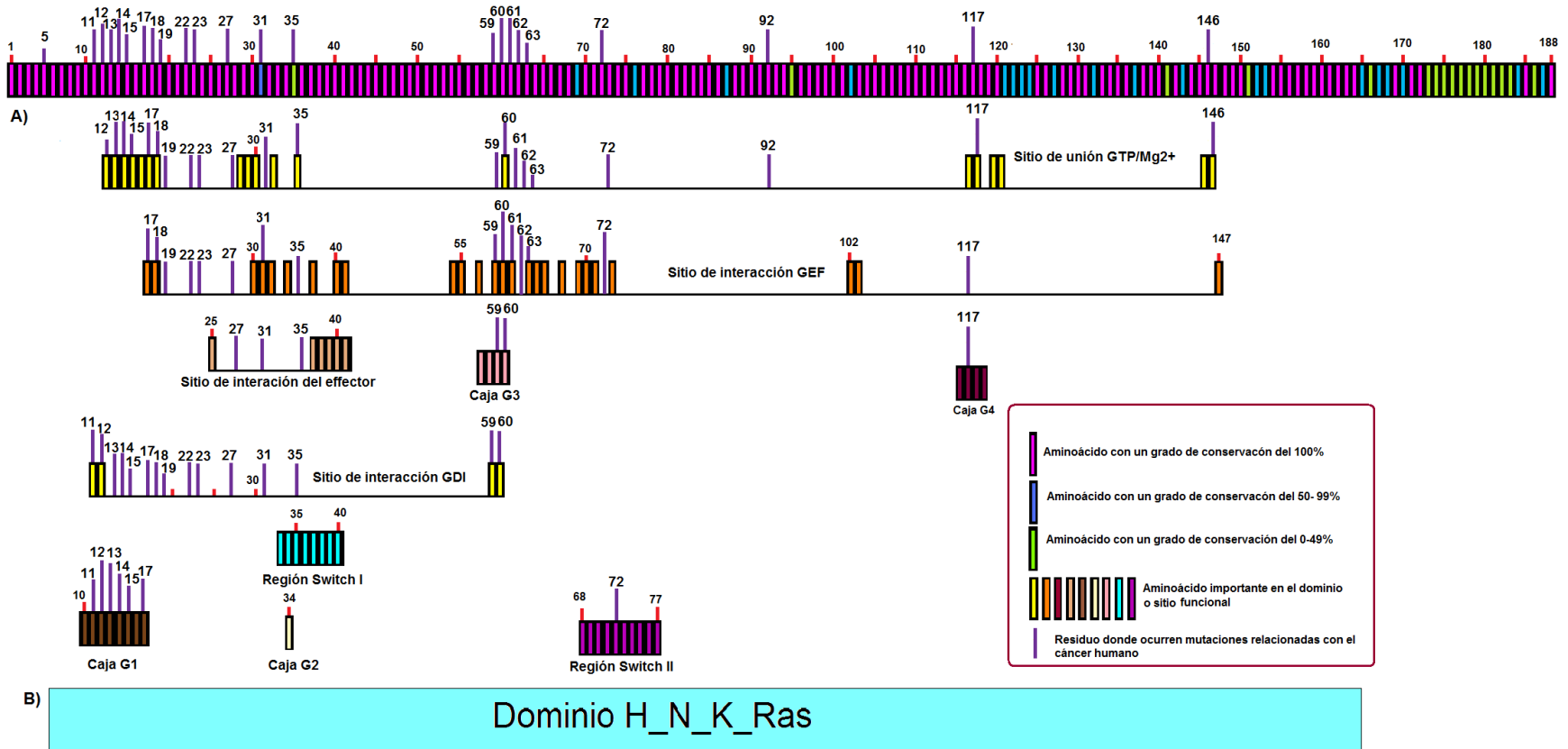


Figura 26. Patrones mutagénicos de Kras: En esta figura se muestra una relación entre las regiones conservadas de la proteína Kras y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína, destacando aquellos residuos que son importantes para dichos subdominios y cuales residuos que están relacionados con el cáncer humano se encuentran en residuos importantes para estos subdominios o aquellos residuos que solamente se encuentran en los subdominios, sin ser residuos esenciales.

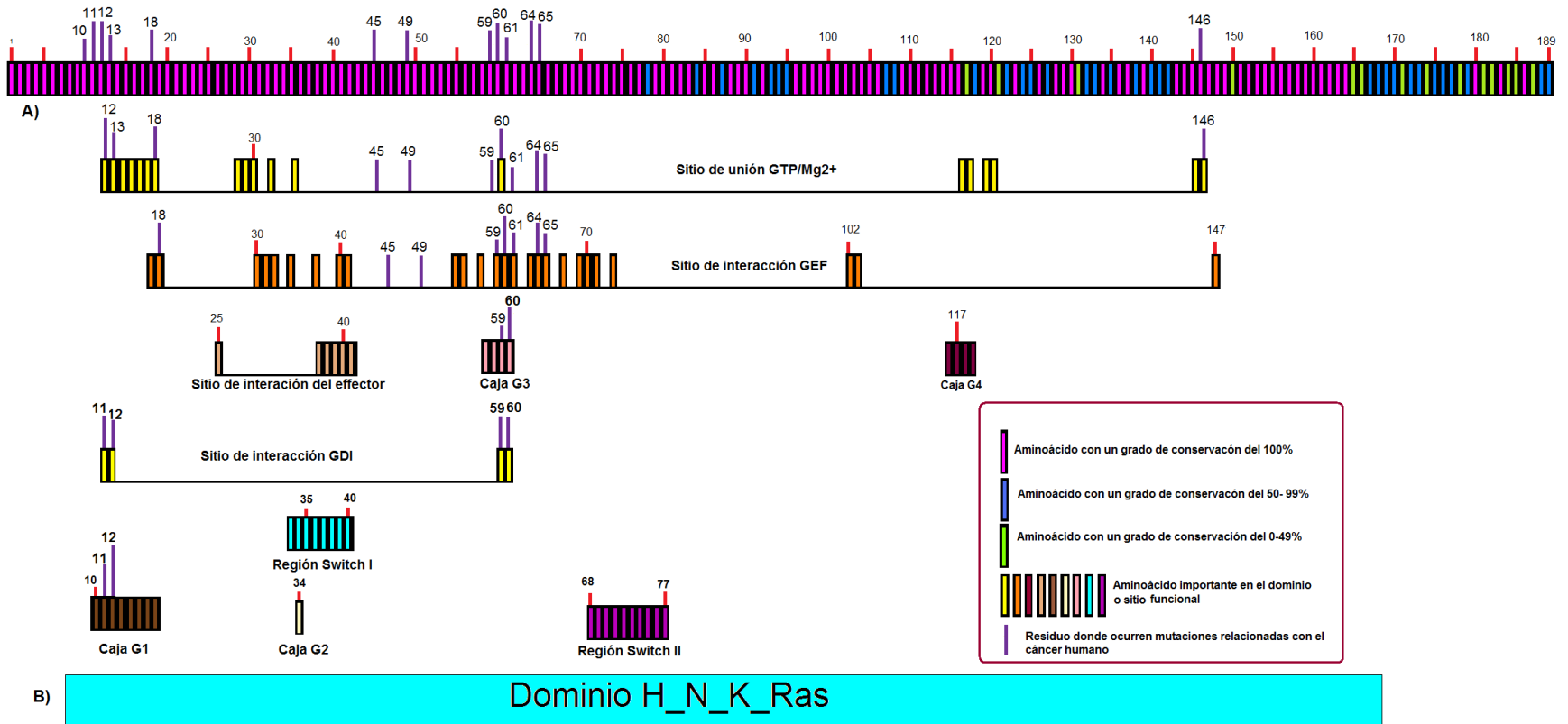


Figura 27. Patrones mutagénicos de Nras: En esta figura se muestra una relación entre las regiones conservadas de la proteína Nras y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína, destacando aquellos residuos que son importantes para dichos subdominios y cuales residuos que están relacionados con el cáncer humano se encuentran en residuos importantes para estos subdominios o aquellos residuos que solamente se encuentran en los subdominios, sin ser residuos esenciales.

3.2.2.2. Patrones mutagénicos de Abl1 isoforma A

En Abl1 isoforma A las mutaciones relacionadas con cáncer humano tienen un alto grado de conservación estructural, (ver las figuras 28, 29, 30 y 31) los residuos con un porcentaje de conservación del 100% son los que están localizados en las siguientes posiciones R47, R166, M244, L248, G250, Q252, E255, D276, V304, F311, T315, G321, M343, M351, E352, F359, F382, L387, T389, H396, S417 y F486. Solamente dos mutaciones tienen un 82 % de conservación los cuales son los residuos M237 y E459.

El 94% de las mutaciones puntuales analizadas de la proteína Abl1 isoforma A se encuentran localizadas en el dominio catalítico de las proteínas tirosinas cinasas Abl (PTKc_Abl), (Marchler-Bauer A *et al.* 2009, Marchler-Bauer A y Bryant SH *et al* 2004), de estas substituciones se destacan aquellas que son importantes a nivel de la función proteica, como son las mutaciones que se encuentran en los aminoácidos 248, 250 y 252 ya que son residuos que interactúan con el sitio de activación (Levinson *et al* 2006). Las mutaciones también se localizaron en sitio de enlace con el ATP (Cowan Jacob *et al* 2007), siendo los aminoácidos 248, 250, 253, 315, 371, 255 y 382. Un tercer sitio es el de la activación de los bucles en el cual se encontraron las mutaciones 382, 387 y 396 y el último sitio es el dominio de interfaz entre SH3/SH2 en el cual la mutación en el residuo 343, tiene relevancia funcional. El 3% de las mutaciones se encontró en el dominio SH2 y la mutación Abl1IsoAR47G fue la única mutante que no cayó en regiones conservadas.

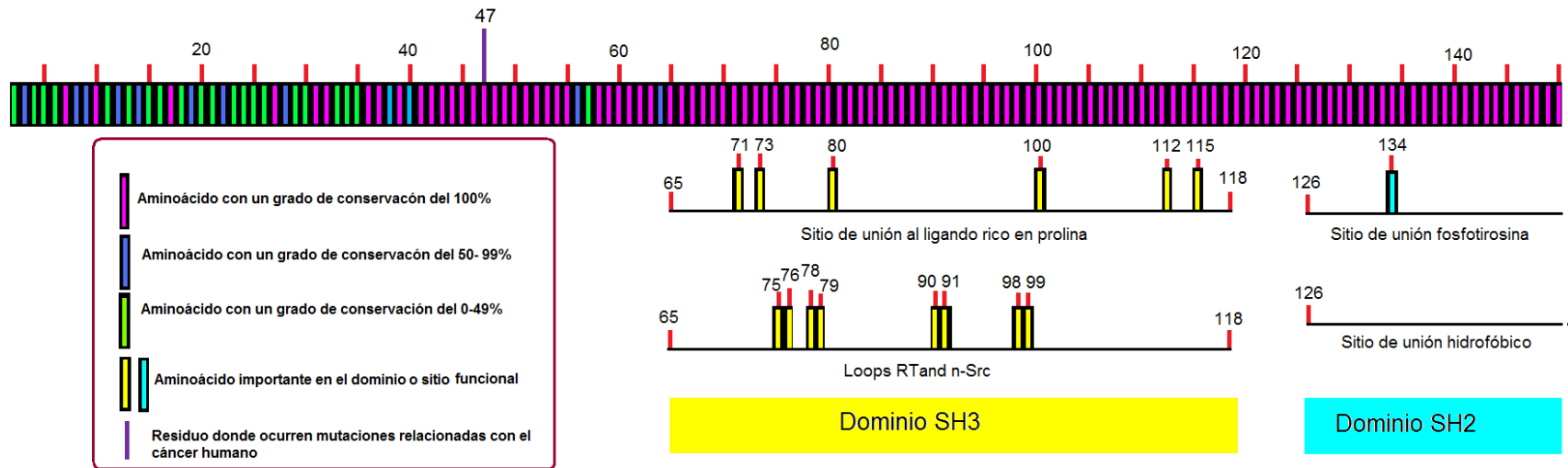


Figura 28. Patrones mutagénicos de Abl1 isoforma A (Fragmento 1): En esta figura se muestra una relación entre las regiones conservadas de la proteína Abl1 isoforma A y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína, destacando aquellos residuos que son importantes para dichos subdominios y cuales residuos que están relacionados con el cáncer humano se encuentran en residuos importantes para estos subdominios o aquellos residuos que solamente se encuentran en los subdominios, sin ser residuos esenciales.

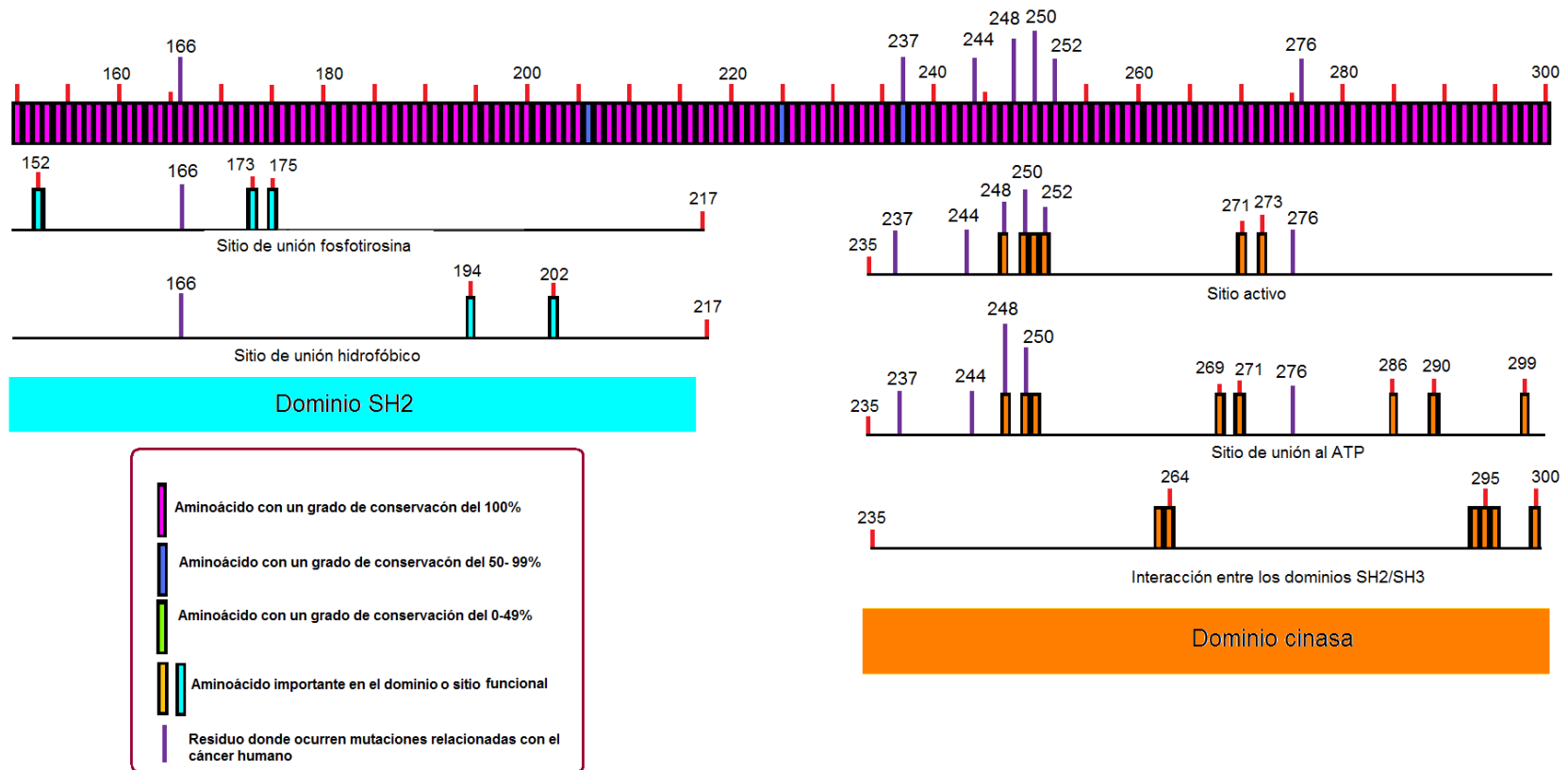


Figura 29. Patrones mutagénicos de de Abl1 isoforma A (Fragmento 2): En esta figura se muestra una relación entre las regiones conservadas de la proteína Abl1 isoforma A y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína, destacando aquellos residuos que son importantes para dichos subdominios y cuales residuos que están relacionados con el cáncer humano se encuentran en residuos importantes para estos subdominios o aquellos residuos que solamente se encuentran en los subdominios, sin ser residuos esenciales.

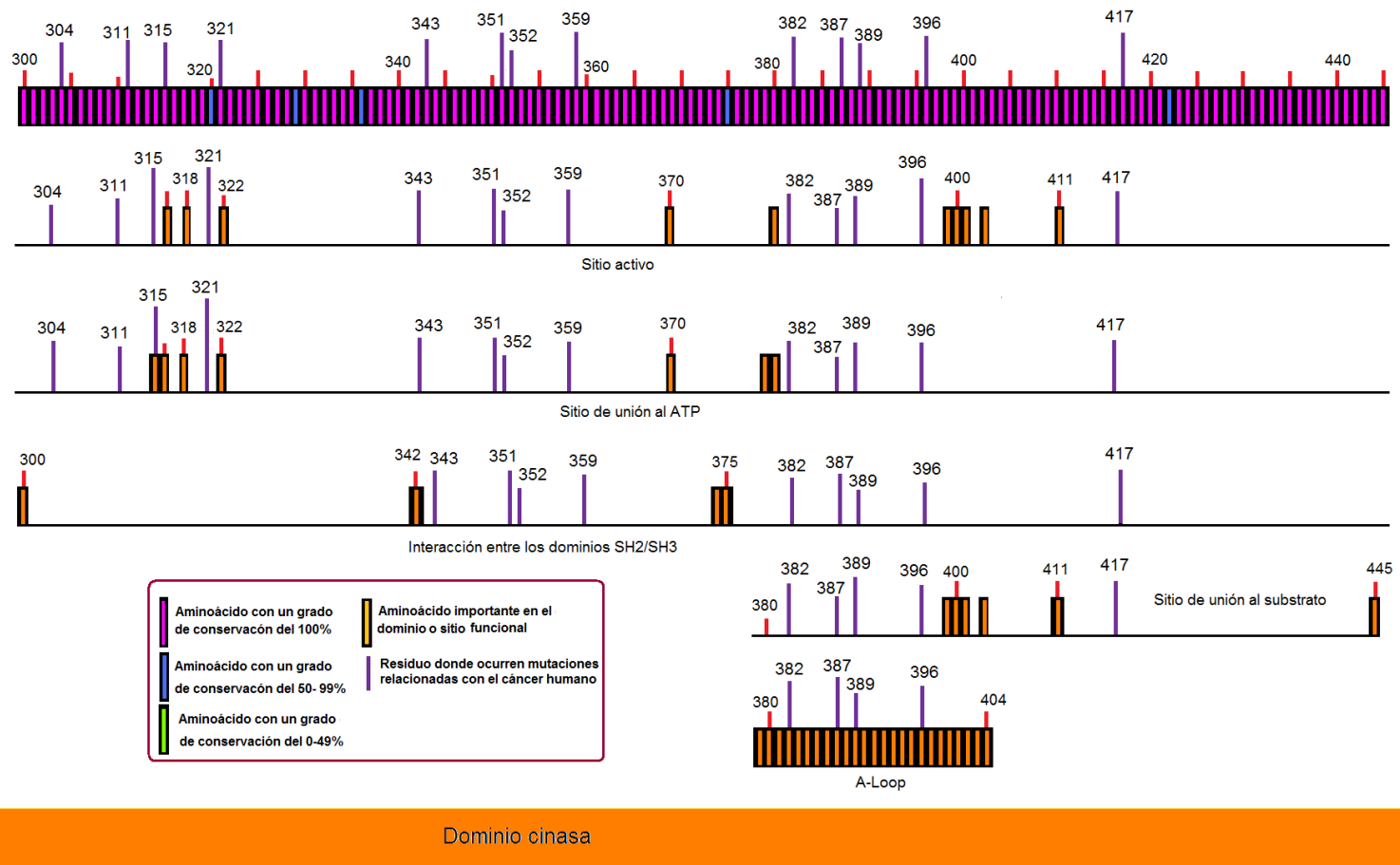


Figura 30. Patrones mutagénicos de de Abl1 isoforma A (Fragmento 3): En esta figura se muestra una relación entre las regiones conservadas de la proteína Abl1 isoforma A y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína, destacando aquellos residuos que son importantes para dichos subdominios y cuales residuos que están relacionados con el cáncer humano se encuentran en residuos importantes para estos subdominios o aquellos residuos que solamente se encuentran en los subdominios, sin ser residuos esenciales.

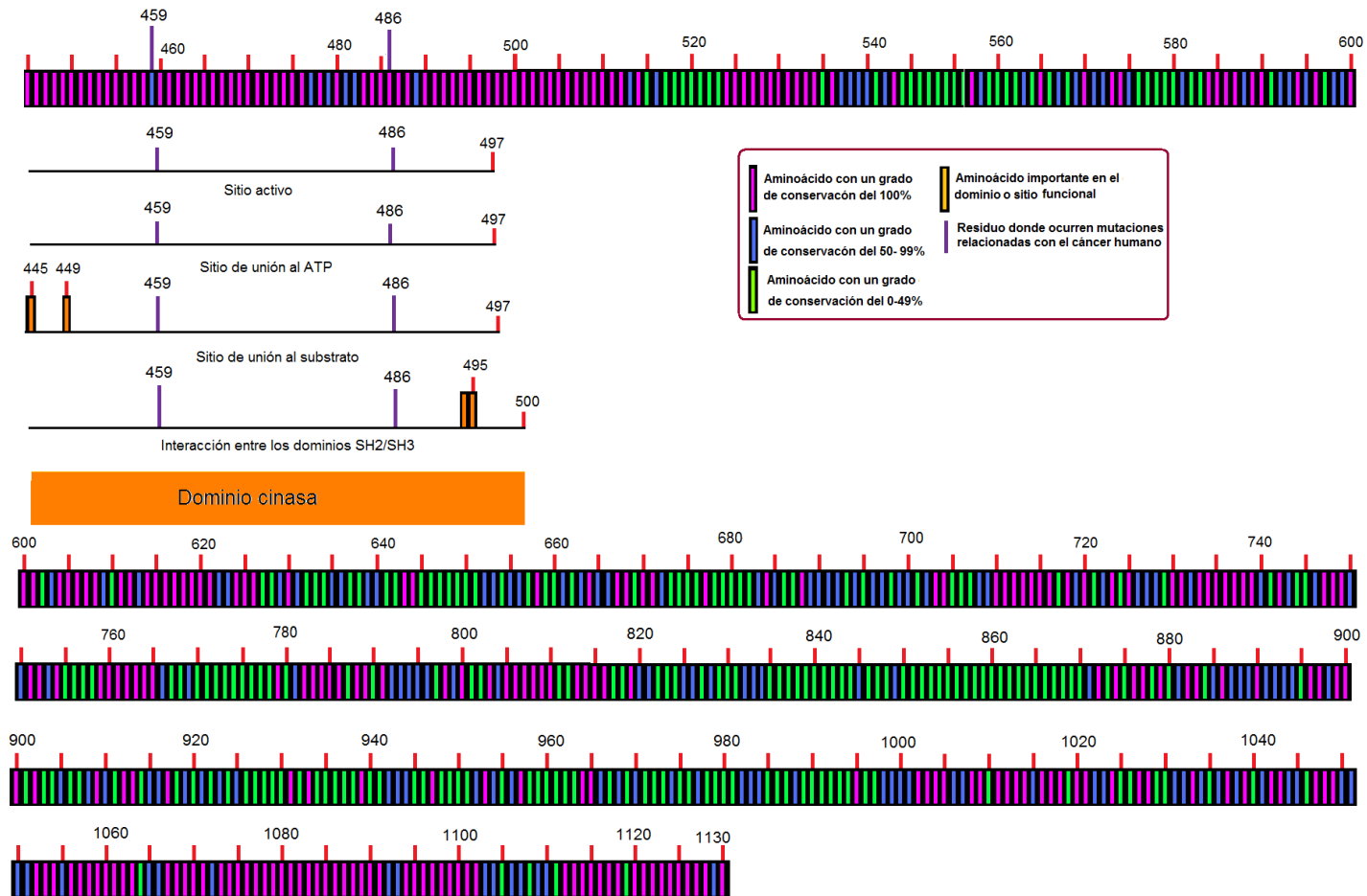


Figura 31. Patrones mutagénicos de de Abl1 isoforma A (Fragmento 4): En esta figura se muestra una relación entre las regiones conservadas de la proteína Abl1 isoforma A y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína, destacando aquellos residuos que son importantes para dichos subdominios y cuales residuos que están relacionados con el cáncer humano se encuentran en residuos importantes para estos subdominios o aquellos residuos que solamente se encuentran en los subdominios, sin ser residuos esenciales.

3.2.2.3. Patrones mutagénicos de p16

La proteína p16 tiene solamente una región conservada que se denomina ANK, la cual tiene como función primordial participar en las interacciones de proteínas con proteínas (ver la figura 32). Esta zona posee una característica estructural muy particular, ya que el dominio tiene una serie de repeticiones en la secuencia de aminoácidos que se denota a nivel estructural, ya que esto provoca 4 repeticiones consecutivas de dos hélices antiparalelas y una horquilla beta. Las repeticiones son apiladas en un arreglo (Batrukova *et al* 2000 y Li J *et al* 2006). Diversos estudios han asociado la participación de esta región en el cáncer y otras enfermedades. El 90% de las mutaciones puntuales estudiadas para estas proteínas se encuentran en la región ANK. Estas substituciones de aminoácidos se encuentran ampliamente dispersas en esta región, las posiciones en donde hay mayor frecuencia de mutaciones son las siguientes 30, 40, 43, 51, 52, 57 59, 63, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 80, 81, 82, 83, 84, 87, 88, 89, 93, 95, 97, 98, 99, 102, 109, 112, 113, 114, 115, 125, 126, 127, y 129. Las mutaciones que no caen dentro de las regiones conservadas se encuentran cercanas a estas regiones funcionales. Sin embargo puede afectar a la proteína a nivel estructural y no funcional.

La secuencia de p16 está constituida por 156 aminoácidos de los cuales 119 sufren mutaciones que están relacionadas con el cáncer, el 67% de estos residuos tienen una conservación estructural del 100% (ver la figura 32), siendo los siguientes aminoácidos M9, D14, L16, A17, A20, A21, G23, R24, V28, A30, L31, E33, A34, G35, P38, N39, A40, N42, G45, P48, I49, Q50, V51, M52, M53, G55, V59, A60, L63, L64, G67, A68, N71, C72, D74, P75, T77, R80, P81, V82, H83, D84, A85, A86, R87, E88, G89, F90, L91, D92, T93, V95, L97, H98, G101, A102, R103, D105, A106, R107, D108, A109, G111, R112, L113, P114, D116, A118, E120, G122, H123, D125, R128, Y129, L130, R131, A134, G135, S140 y G150. Destacando que 13 puntos calientes se encuentran es este porcentaje (A20, G35, P48, V51, M52, D74, P81, H83, A102, D108, R112, P114 y D116), solamente un aminoácido con mayor frecuencia no tiene el 100% de conservación estructural, pero el aminoácido A100 tiene un porcentaje de conservación del 73%.

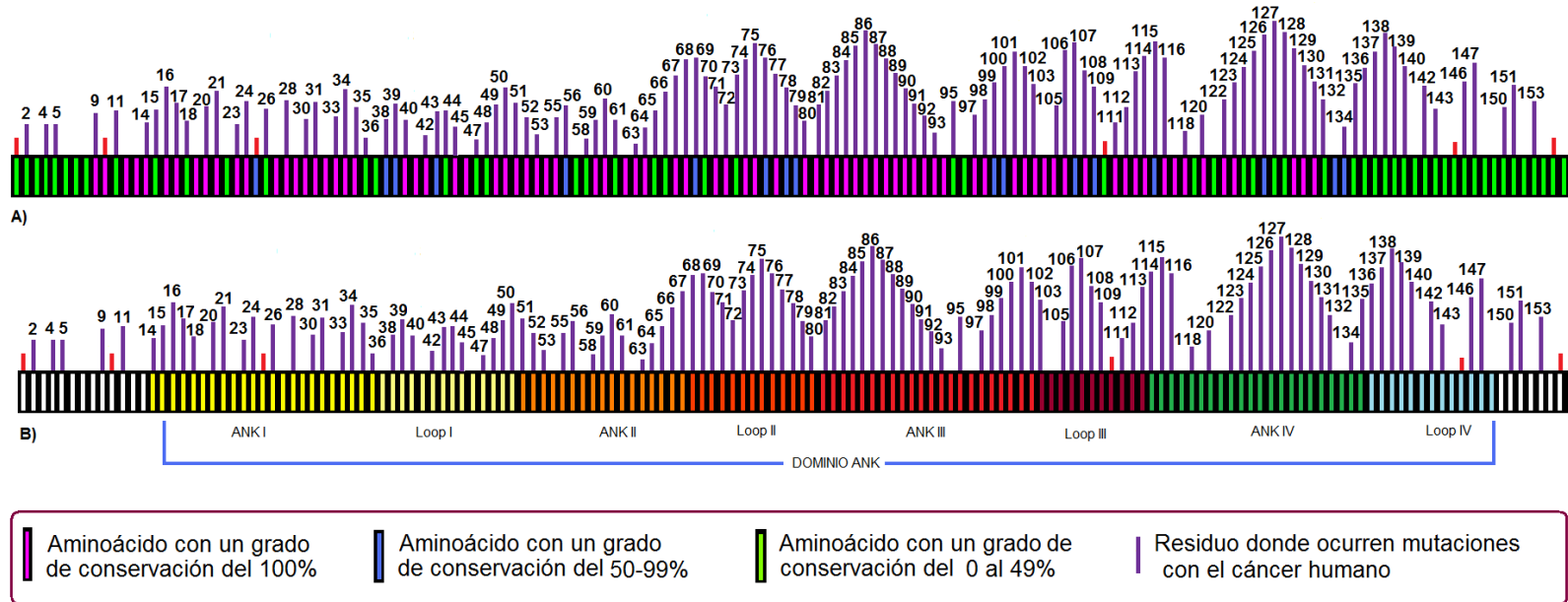


Figura 32. Patrones mutagénicos de p16: En esta figura se muestra una relación entre las regiones conservadas de la proteína p16 y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína.

3.2.2.4. Patrones mutagénicos de p53

Para p53 (ver la figura 33) los residuos que presentan el 100% de conservación son los siguientes P98, F113, K120, L130, K132, C135, V143, V147, P151, P152, V157, R158, A159, A161, I162, Y163, K164, V173, R175, H179, R181, H193, L194, I195, Y205, R213, S215, V216, Y220, P223, M237, C238, S241, C242, G244, G245, M246, R248, R249, P250, I251, I254, I255, E258, D259, G262, L265, G266, R267, E271, V272, R273, V274, C275, C277, P278, P280, D281, R282, E285 y E286, destacando que los 6 residuos denominados puntos calientes, están entre ellos (R175, G245, R248, R249, R273 y R282).

Todas las mutaciones puntuales estudiadas de la proteína p53 (ver la figura 33) se localizaron en el dominio central. La mayoría de todas las mutaciones se concentran en 6 puntos calientes o "hotspots" en los codones R175, G245, R248, R249, R273 y R282 del dominio central. Esta región es muy importante para la interacción y el reconocimiento del DNA, que es donde recae la función principal, que tiene como objetivo el reconocimiento de mutaciones en el DNA, y la activación de procesos de reparación del DNA, o activar la apoptosis de la célula, cuando el daño es irreparable (Marchler-Bauer A *et al.* 2009 y Marchler-Bauer A, Bryant SH *et al* 2004).

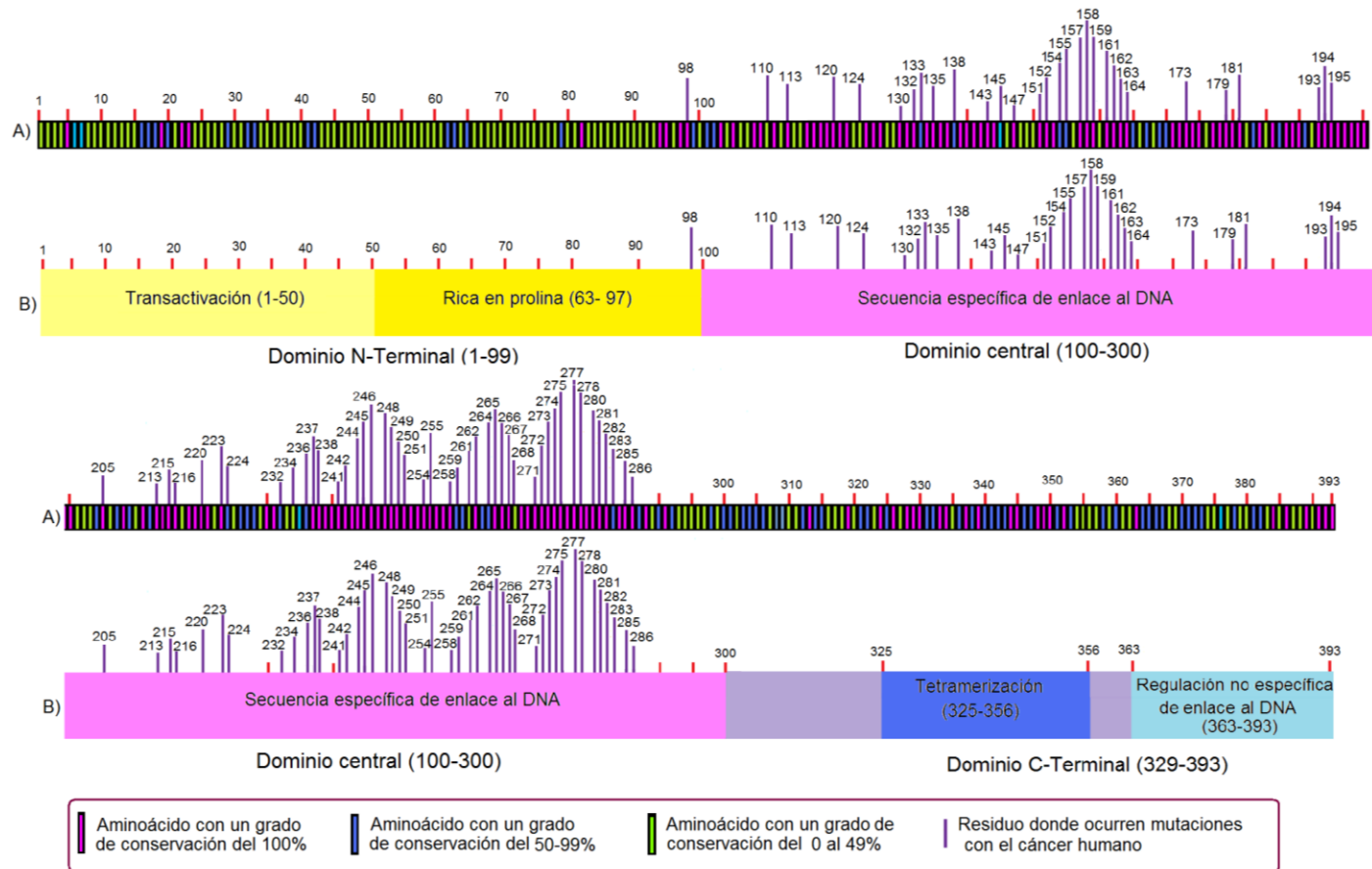


Figura 33. Patrones mutagénicos de p53: En esta figura se muestra una relación entre las regiones conservadas de la proteína p53 y los residuos en donde ocurren mutaciones relacionadas con cáncer y además los subdominios funcionales de la proteína.

Para poder definir el tipo de cambio se clasificaron los aminoácidos por su tamaño y su afinidad por el agua, formándose 7 grupos de aminoácidos, la cual está descrita en el anexo I. Se encontró la siguiente distribución en los distintos cambios que se dieron en las proteínas silvestres y sus mutantes.

3.2.3. Identificación del tipo del cambio

Por ejemplo (Ver el anexo I) los cambios de la proteína Hras y sus mutantes fue la siguiente, el 28% de los aminoácidos fueron reemplazados de aminoácidos pequeños y no polares a aminoácidos pequeños y polares, el 12% fue de aminoácidos pequeños y no polares a voluminosos y polares ácidos, el otro 12% fue de aminoácidos voluminosos y polar a voluminosos polar básico. El 8% de las mutaciones fue de pequeño y no polar a voluminoso y polar básico. En general se observa cambios significativos de polaridad. Se observan (Ver los anexos I, II, III, IV, V y VI o en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/localizacion_mutaciones.html) que en todas las proteínas de estudio la mayoría de los cambios de los aminoácidos se produjeron con aminoácidos que tienen baja similitud fisicoquímica, hay cambios muy frecuentes de aminoácidos hidrofóbicos a hidrofílicos, y de hidrofílicos a hidrofóbicos. Las diferencias de propiedades fisicoquímicas con lleva a cambios en las interacciones con sus residuos vecinos en la estructura tridimensional de la proteína o incluso con otras moléculas.

3.4. Análisis estructural de la proteína silvestre con sus mutantes

En la base de datos PDB se realizó la búsqueda de las estructuras tridimensionales de las proteínas de estudio, de las cuales se encontraron la proteína Hras, p53, p16 y Abl1 isoforma A. Las estructuras 3D de las proteínas Kras y Nras no fueron encontradas en la base de datos por lo tanto se realizó la búsqueda de un homólogo cercano mediante BLAST ocupando la base de datos PDB, para realizar la modelación por homología con Swiss-Model. De la base de datos de COSMI del instituto de Sanger se descargó la tabla de las mutaciones por sustitución de un nucleótido que está relacionada con el cáncer de cada una de las proteínas silvestres (Tabla 6). A partir de estas mutaciones se

escogieron y modelaron las estructuras tridimensionales de las proteínas mutantes mediante el programa Modeller 9v7.

TABLA 6. PROTEÍNAS QUE SE MODELARON	
PROTEÍNA SILVESTRE	MUTANTES
Hras	25
Kras	53
Nras	33
Abl1	30
p53	148
p16	207

Tabla 6. Proteínas modeladas: Se menciona cuantas mutantes fueron modeladas mediante el programa Modeller v97.

La última etapa consintió en comparar las proteínas silvestres, con cada una de sus mutantes, para ello se realizaron alineamientos estructurales globales y locales.

Comparaciones globales:

- RMSD

Comparaciones locales:

- RMS posicional.
- Potencial DOPE
- Potencial de ángulos de torsión.

Comparaciones globales:

La primera comparación global de la estructura proteica que se realizo fue el calcular el RMSD (Root Mean Square Deviation), que es la desviación media de la raíz cuadrada, la cual se obtiene de la superposición de los conjuntos de coordenadas atómicas de la cadena principal (N, Ca, C, O) pertenecientes a parejas de residuos que quedan juntos en un alineamiento estructural de las dos proteínas. Siendo este valor una distancia en amstrongs (Å). El RMSD en la comparación de una proteína consigo misma es de 0 Å y

entre más se incremente el valor del RMSD al comparar dos proteínas, significa que estas tienen mayor diferencia conformacional.

Se realizaron alineamientos estructurales entre la proteína silvestre y cada una de sus respectivas mutantes, tanto optimizadas como sin optimizar, mediante el programa VMD 1.8.7 para obtener el RMSD. Al revisar los valores obtenidos, se observó que todas las mutantes sin optimizar dieron valores a 0.0 Å, lo cual significa que las proteínas sin optimizar, no se realizaron cambios en la cadena principal de la proteína por lo tanto el valor es de cero, en cambio las mutantes optimizadas ya tienen modificaciones en la cadena principal, por lo tanto nos dan valores mayores a 0.0 Å, pero como son mutaciones puntuales los incrementos son pequeños.

En general los valores de RMSD de la mutante con su proteína silvestre no se obtuvieron diferencias altas debido a que son mutaciones de sustitución de un solo aminoácido, lo cual el 99% de la estructura proteica es igual, por ello no se esperaban grandes diferencias, estos valores son menores de 0.5 Å.

Comparaciones locales:

Dado que el RMSD, es una superposición de la molécula entera puede no proporcionar datos suficientes al nivel de la perturbación de la estructura al nivel local. Por este motivo también se calculó el RMS posicional. En esta comparación se obtienen las distancias por cada pareja de aminoácido a partir del alineamiento estructural de las dos proteínas comparadas.

En todas las mutantes de las proteínas Nras, Hras y Abl1 isoforma A (ver la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/analisis_estructural.html o en las figuras 34, 35 y 36), se observó que el único aminoácido donde fue mayor a cero fue en el residuo mutado.

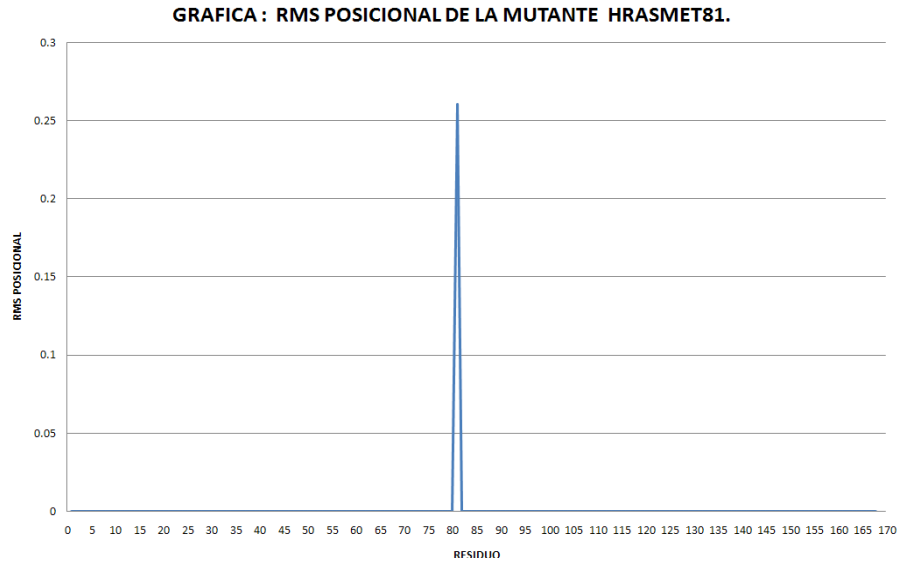


Figura 34. RMS Posicional de la mutante HrasV81M. Al realizar el alineamiento estructural entre la proteína silvestre y la mutante HrasV81M se observa que solamente en el residuo donde se ocasionó la mutación, cambio la distancia en este aminoácido, sin afectar los demás residuos (Se pueden observar todos los resultados de las comparaciones de RMS posicional de Hras silvestre con cada una de sus mutaciones estudiadas en la página http://bioinformatica.homelinux.org/cancer_proteinas/rmsp_hras.html).

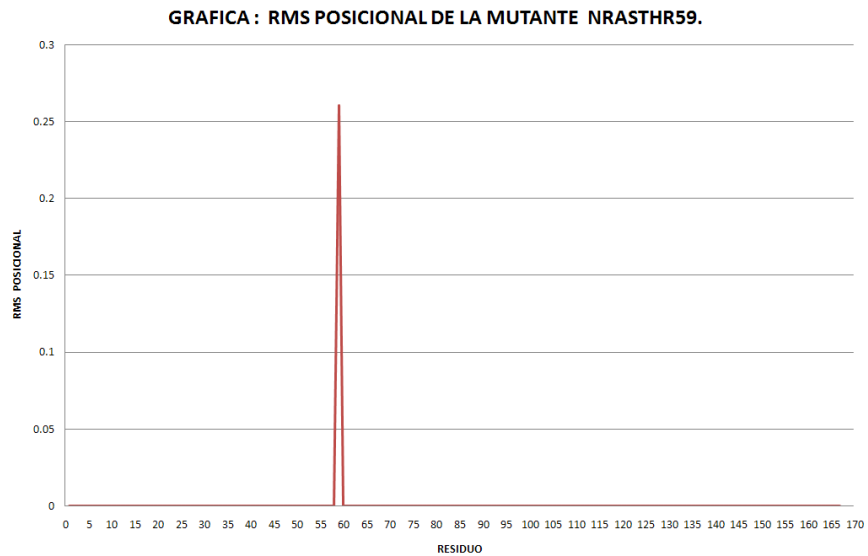


Figura 35. RMS posicional de Nras y NrasA59T. Al realizar el alineamiento estructural entre la proteína silvestre y la mutante NrasA59T se observa que solamente en el residuo donde se ocasionó la mutación, cambio la distancia en este aminoácido, sin afectar los demás residuos (Se pueden observar todos los resultados de las comparaciones de RMS posicional de Nras silvestre con cada una de sus mutaciones estudiadas en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/rmsp_nras.html).

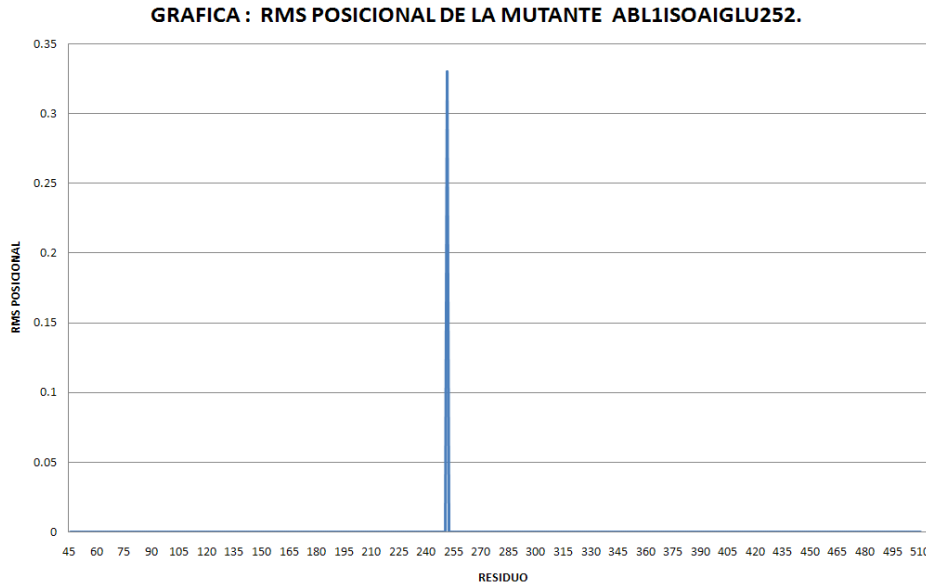


Figura 36. RMS posicional Abl1 y Abl1IsoAQ252E: Se observa las distancias de cada residuo al ser comparado con la proteína silvestre. Al realizar el alineamiento estructural entre la proteína silvestre y la mutante Abl1IsoAQ252E se observa que solamente en el residuo donde se ocasionó la mutación, cambio la distancia en este aminoácido, sin afectar los demás residuos (Se pueden observar todos los resultados de las comparaciones de RMS posicional de Nras silvestre con cada una de sus mutaciones estudiadas en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/rmsp_abl1.html).

En las únicas mutantes donde se observaron valores mayores a cero no solamente en el residuo mutado, sino también en algunos aminoácidos que se encuentran cerca del sitio de cambio, fue en algunas mutantes de Kras isoforma B, p16 y p53 (ver en la siguiente pagina http://bioinformatica.homelinux.org/cancer_proteinas/analisis_estructural.html o en las figuras 37, 38 y 99). Por ejemplo en la mutante KrasIsoBT35I tiene un valor mayor a cero en los aminoácidos 9 al 17, 29 al 38, 59 al 102, 115 al 138 con un valor de 0.001 Å excepto el residuo 35 con un valor de 0.73 Å.

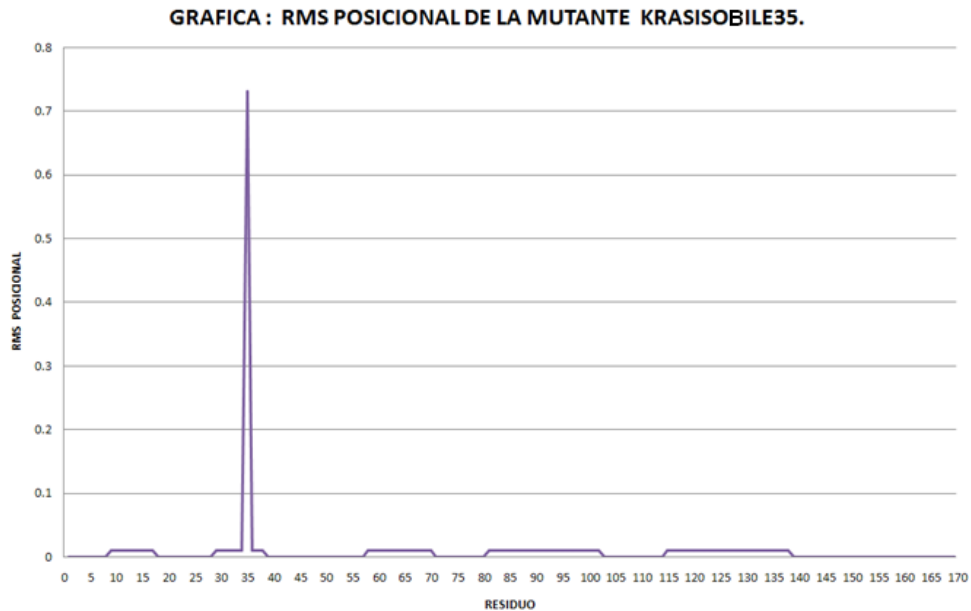


Figura 37. RMS posicional de la mutante KraslsoBT35I: Al realizar el alineamiento estructural entre la proteína silvestre y la mutante KraslsoBT35I se observa que esta mutación ocasionó cambios no solamente en el residuo 35 donde ocurrió la sustitución, sino también en los residuos 9 al 17, del 29 al 38, del 58 al 70, del 80 al 103, del 115 al 138 (Se pueden observar todos los resultados de las comparaciones de RMS posicional de Nras silvestre con cada una de sus mutaciones estudiadas en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/rmsp_kras.html).

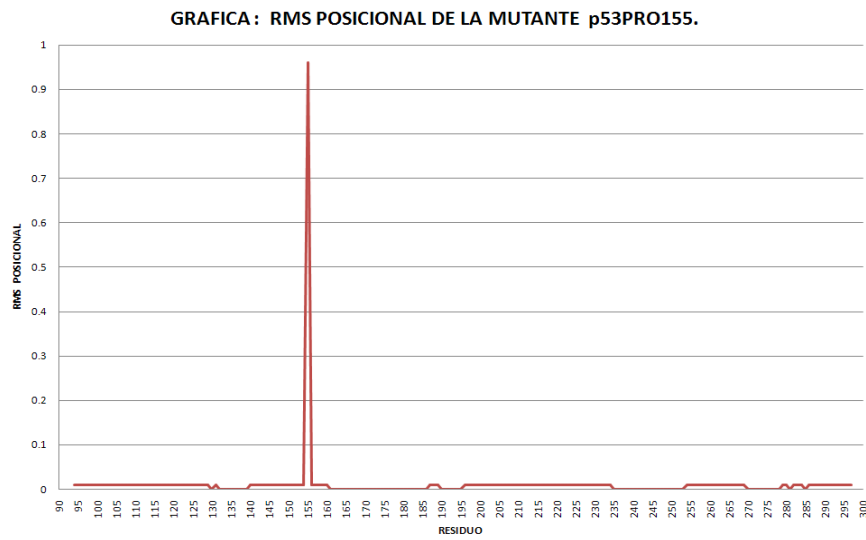


Figura 38. RMS posicional de p53 y p53T155P. Al realizar el alineamiento estructural entre la proteína silvestre y la mutante p53T155P se observa que esta mutación ocasionó cambios no solamente en el residuo 155 donde ocurrió la sustitución, sino también en los residuos 94 al 129, el 131, del 140 al 160, del 187 al 190, del 196 al 234, del 254 al 269 y del 279 al 297 (Se pueden observar todos los resultados de las comparaciones de RMS posicional de la proteína p53 silvestre con cada una de sus mutaciones estudiadas en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/rmsp_p53.html).

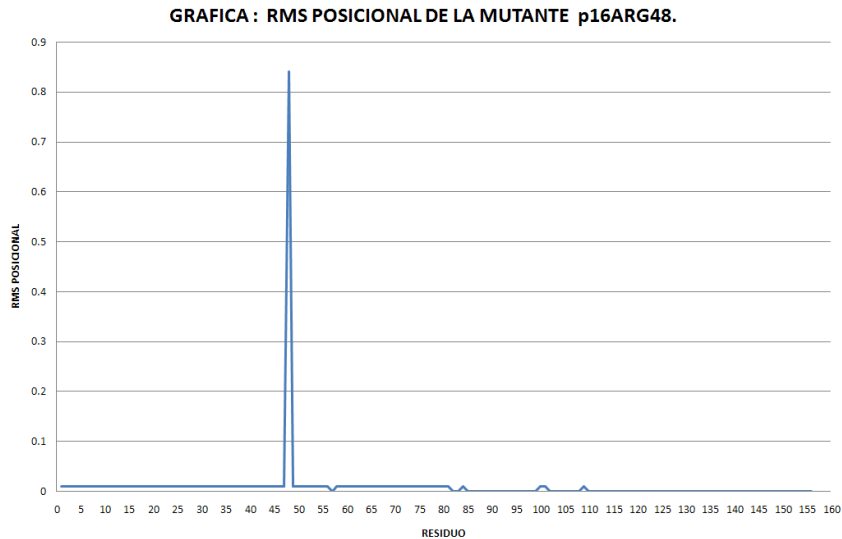


Figura 39. RMS posicional de p16 y p16P48R. Al realizar el alineamiento estructural entre la proteína silvestre y la mutante p16P48R se observa que esta mutación ocasionó cambios no solamente en el residuo 48 donde ocurrió la sustitución, sino también en los residuos 1 al 56, el 58 al 81, el 84, el 100, el 101 y el 109 (Se pueden observar todos los resultados de las comparaciones de RMS posicional de la proteína p16 silvestre con cada una de sus mutaciones estudiadas en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/rmsp_p16.html).

Las perturbaciones locales en la estructura de la proteína pueden originarse por mutaciones puntuales, modificaciones químicas, interacciones entre moléculas, etc. Para evaluar tales perturbaciones estructurales de la proteína, usualmente deben ser comparadas de un modo detallado, para observar la posición y orientación de átomos individuales, residuos o de estructura secundaria.

Para observar esos cambios locales se calculo la energía DOPE (Discrete Optimized Protein Energy) de cada una de las mutantes y se comparo con la energía DOPE de la proteína silvestre. Obteniéndose información que aquí se trata de manera muy general:

En el análisis estructural del impacto mutacional de la familia Ras (ver ver en la siguiente pagina http://bioinformatica.homelinux.org/cancer_proteinas/analisis_estructural.html o en las figuras 40, 41 y 42) se observo que las mutaciones relacionadas con el cáncer humano afectaban principalmente los residuos 12, 13, 60, 61, las perturbaciones encontradas no fueron tan altas como las demás proteínas estudiadas, estos cambios de energía y de residuos mutado ocasiona que la proteína este continuamente activada, aumentando la proliferación celular en aquellas células con estas mutaciones.

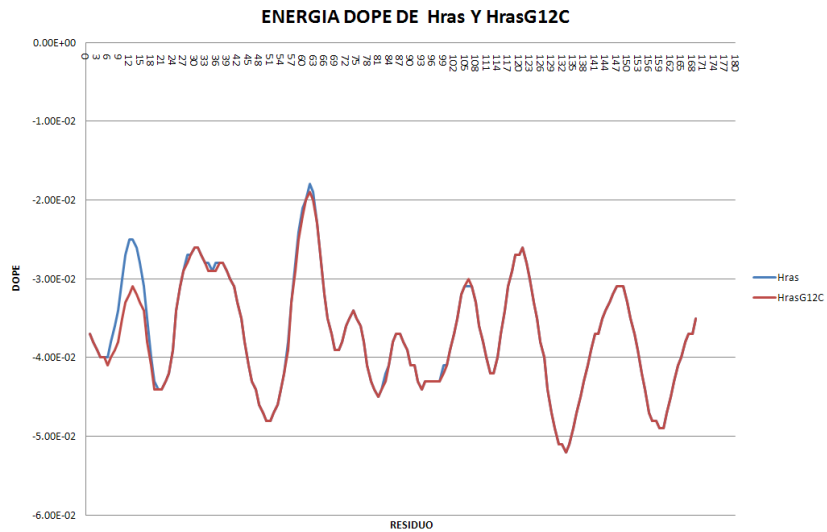


Figura 40. Comparación del potencial dope entre Hras y HrasG12C. En la mutación G12C se observan cambios locales en la energía proteica la región más importante se observa desde el residuo 6 al residuo 19, en donde la perturbación más alta se encuentra en el aminoácido mutado. Los residuos 28, 34, 36, 56, 58, 59, 60, 62, 63, 83, 99 y 106 tienen perturbaciones mínimas (Se pueden observar todos los resultados de las comparaciones del potencial DOPE de la proteína Hras silvestre con cada una de sus mutaciones estudiadas en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/dope_hras.html).

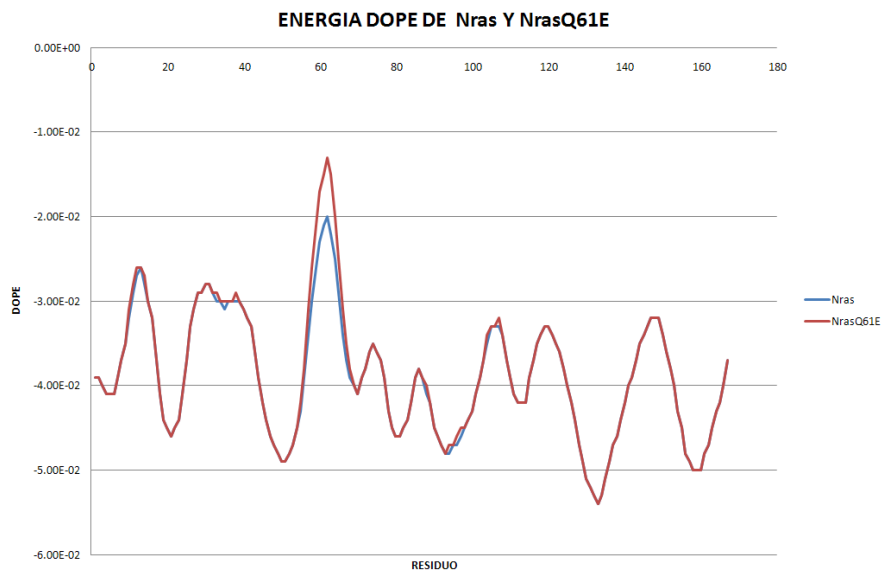


Figura 41. Comparación del potencial dope entre Nras y NrasQ61E: La mutación Q61E se produce perturbaciones en la energía varias regiones de la estructura proteica, pero la región principal está localizada desde el residuo 55 al 68, en donde el cambio de Q a E afectó a los residuos 62 E y 63 E, ocasionado el mayor cambio de energía DOPE en la estructura. Los otros residuos donde se observan diferencias de energía están en las siguientes posiciones 10, 11, 12, 14, 33, 35, 38, 88, 94, 96, 97, 104 y 107 (Se pueden observar todos los resultados de las comparaciones del potencial DOPE de la proteína Nras silvestre con cada una de sus mutaciones estudiadas en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/dope_nras.html).

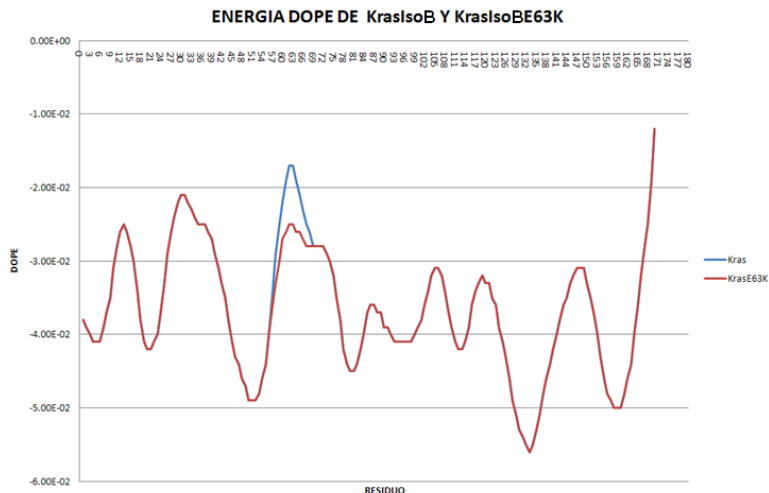


Figura 42. Comparación del potencial dope entre Kras y KraslsoBE63K: La mutante KraslsoBE63K es aquella proteína donde se produjo la diferencia más alta de energía Dope de todas las mutantes analizadas. Se observaron cambios en una sola región local que está comprendida desde el residuo 57 al 68, los residuos con mayor diferencia de energía son los residuos 62 y 63 (Se pueden observar todos los resultados de las comparaciones del potencial DOPE de la proteína Kras silvestre con cada una de sus mutaciones estudiadas en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/dope_kras.html).

En la proteína Abl1 isoforma A (ver la figura 43) las alteraciones con mayor impacto se encuentran en regiones funcionales importantes como en el dominio PTKc_Abl, las mutaciones son diversas y no se encontró un residuo en específico donde allá una predicción en estas mutaciones, los cambios de energía fueron menos localizadas que en la familia Ras, pero menos que las proteínas p16 y p53. En el caso de las proteínas p53 y p16 (ver las figuras 44 y 45 o en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/analisis_estructural.html).

Las perturbaciones son muy elevadas en comparación a la familia Ras, las alteraciones con mayor impacto se encuentran en zonas funcionales y estructurales de las mutaciones, sin embargo las perturbaciones no se encuentran en regiones pequeñas, sino están en la mayor parte de la estructura, esto puede desestabilizar la proteína, ocasionando que dejen de cumplir sus funciones como proteínas supresoras de tumores.

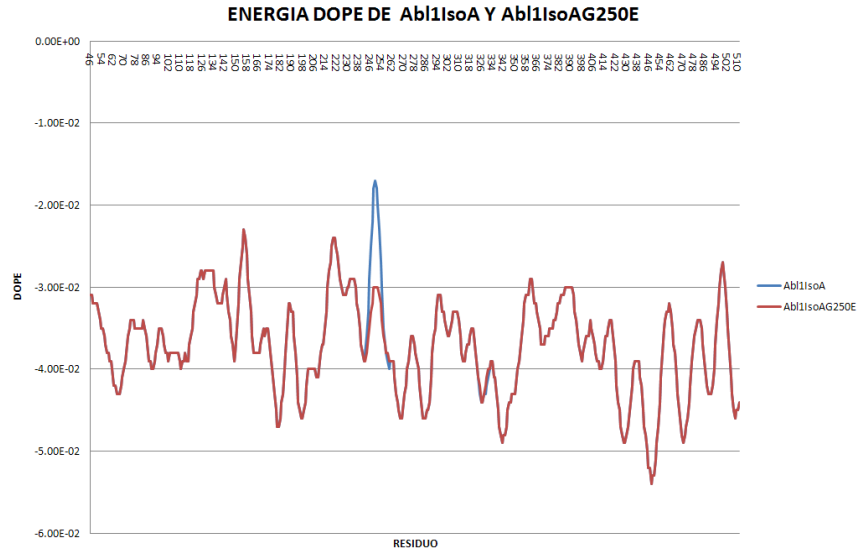


Figura 43. Comparación del potencial dope entre Abl1 y Abl1IsoAG250E: En la mutante Abl1IsoAG250E se originaron perturbaciones en la energía DOPE la principal zona en donde se observan estos cambios se sitúa en la zona 243 a la zona 260, los aminoácidos más afectados son 247, 248 al 253, en la cual en donde se produjo la mutación recae el mayor impacto con una diferencia de 0.013. También hubo afectación en los residuos 325, 329 al 331, pero la diferencia de energías fue mínima (Se pueden observar todos los resultados de las comparaciones del potencial DOPE de la proteína Abl1 silvestre con cada una de sus mutaciones estudiadas en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/dope_abl1.html).

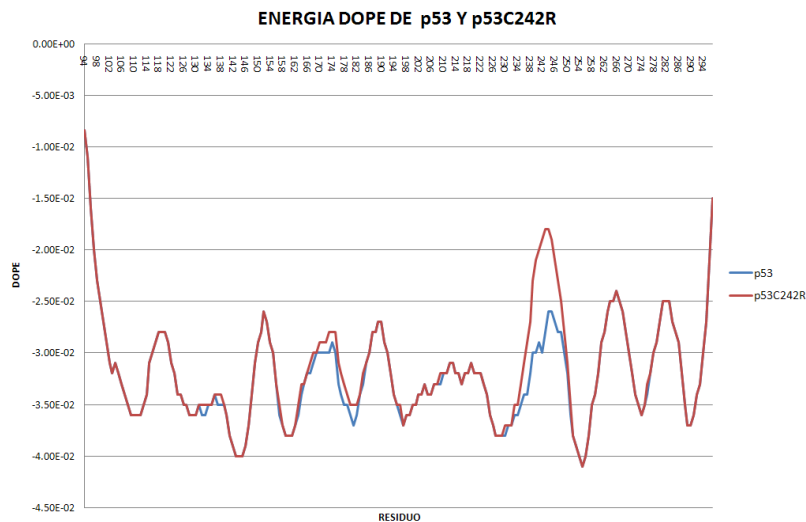


Figura 44. Comparación del potencial dope entre p53 y p53C242R: En la mutante p53C242R se observaron las siguientes perturbaciones que se encuentran una región principal que va desde el residuo 233 al 251, la mayor modificación se encuentra en el residuo mutado, esta mutante tiene varias modificaciones a lo largo de la estructura situadas en las siguientes posiciones 132, 133, 137, 138, 157, 163, 164, 167, 168, 170 -182, 184, 196, 209, 230 y 276 (Se pueden observar todos los resultados de las comparaciones del potencial DOPE de la proteína p53 silvestre con cada una de sus mutaciones estudiadas en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/dope_p53_2.html).

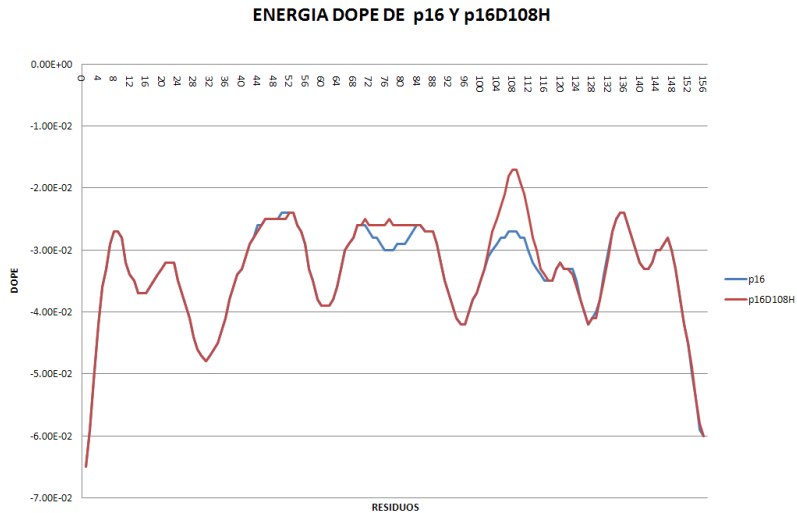


Figura 45. Comparación del potencial dope entre p16 y p16D108H: En la mutante p16D108H se percataron modificaciones en la energía DOPE, la zona donde hay más alteraciones está comprendida desde el aminoácido 102 al 124, en la cual los residuos 108 y 109 tienen una diferencia de 0.01, la segunda zona afectada se encuentra desde el aminoácido 71 al 83 en esta región el residuo 77 es el que tiene la mayor perturbación la cual la diferencia es de 0.005, por último existen otros aminoácidos que también están afectados pero con una diferencia mínima los cuales son el 44, 50, 51, 131, 132, 153 y 154 (Se pueden observar todos los resultados de las comparaciones del potencial DOPE de p16 silvestre con cada una de sus mutantes en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/dope_p16.html).

También se calculó el potencial de ángulos de torsión encontrando las siguientes observaciones: En las diferentes mutantes de la proteína Hras (ver la figura 46 o en la siguiente página http://bioinformatica.homelinux.org/cancer_proteinas/torsion_hras.html) se encontraron cambios de rotación en los aminoácidos donde ocurrió la mutación se encuentra la mayor diferencia de energía con respecto a la silvestre, esta sustitución puntual afecta a los 4 residuos vecinos de cada lado de la secuencia, es decir si la mutación ocurre en el residuo 12, los aminoácidos afectados serán los del residuo 8 al 16, en todas las estructuras mutantes tiene un cambio de energía de torsión del aminoácido 163 al 168.

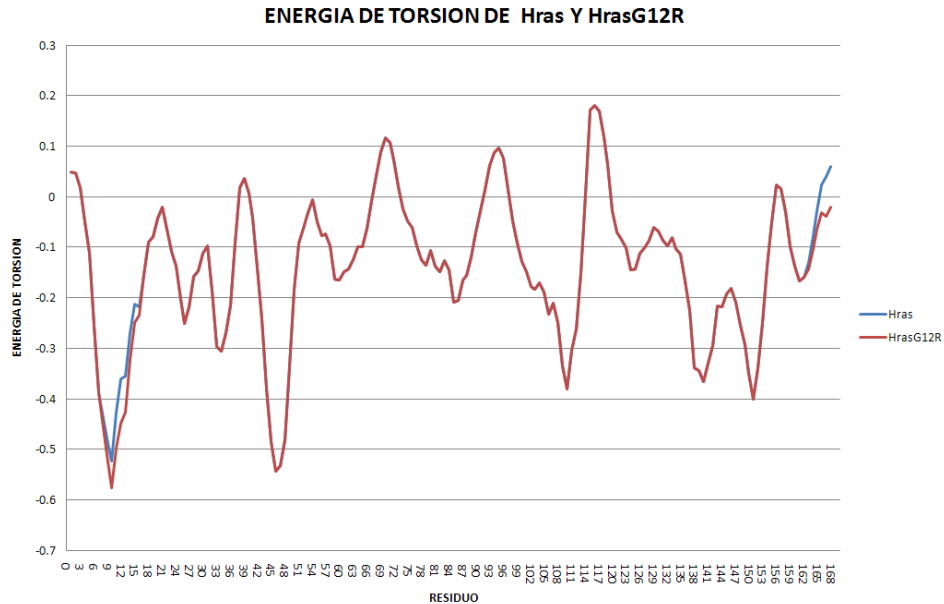


Figura 46. Comparación de energía de torsión entre Hras y HrasG12R: En esta grafica se puede observar los cambios de energía de los ángulos de torsión entre una de las mutantes con la proteína silvestre (Se pueden observar todos los resultados de las comparaciones de energía de torsión de Hras silvestre con cada una de sus mutantes en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/torsion_hras.html).

La afectación en la energía de torsión para las mutantes Nras, Kras, p16, p53 con respecto a la proteína silvestre se encuentra en una zona local, en donde se observa claramente que el residuo mutado obliga a la proteína a rotar, afectando a los residuos vecinos de la secuencia primaria, las mutaciones que tienen mayor afectación son los que se encuentran el residuo 13 para el caso particular de Nras y Kras (ver la figura 47), además afecta un residuo mas, es decir estos cambios afectan a los 5 aminoácidos que se encuentran a lado del cambio, tanto a la derecha como a la izquierda de la secuencia; y en las otras mutaciones solamente afecta cuatro residuos de cada lado. En todas las mutantes de p16 los residuos afectados son 4 de cada lado. En la proteína p53 el cambio del residuo obliga que 5 aminoácidos de cada lado de la posición mutante en la secuencia de la proteína rotaran.

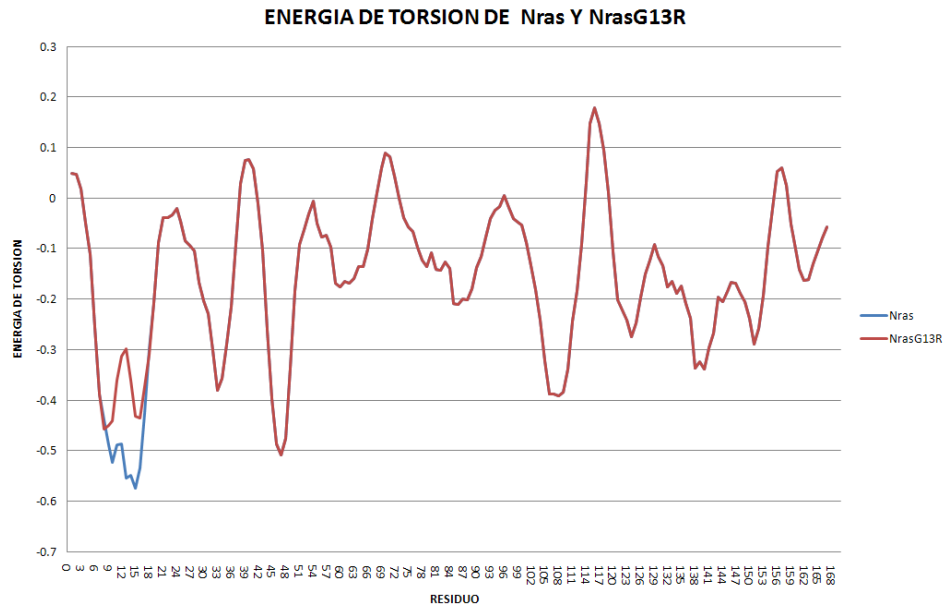


Figura 47. Comparación de energía de torsión entre Nras y NrasG13R: En esta grafica se puede observar los cambios de energía de los ángulos de torsión entre una de las mutantes con la proteína silvestre (Se pueden observar todos los resultados de las comparaciones de energía de torsión de Hras silvestre con cada una de sus mutantes en la siguiente página electrónica http://bioinformatica.homelinux.org/cancer_proteinas/torsion_nras.html).

4. DISCUSION

El estudio que se llevo a cabo en el presente trabajo se puede dividir en tres grandes etapas: 1) El análisis evolutivo y funcional de la proteína silvestre, 2) la localización de las mutaciones puntuales relacionadas con el cáncer humano y 3) el análisis estructural de las proteínas silvestres y sus mutantes. Los cuales serán explicados a continuación:

1) Análisis evolutivo de la proteína silvestre:

La historia evolutiva de las proteínas estudiadas nos confiere mucha información sobre la importancia de dichas secuencias proteicas, ya que al ser comparadas las secuencias de aminoácidos que realizan una misma función y que procede de organismos distintos, observamos que son parecidas y que se puede establecer una correspondencia entre las posiciones de las secuencias de unas y otras, existiendo posiciones conservadas y otras variables. Dichas secuencias tienen un origen común y las diferencias en la secuencia se deben a que a lo largo del tiempo los genes han divergido, mediante la acumulación de mutaciones. La comparación de proteínas homólogas nos puede dar mucha información acerca de cómo se ha producido la evolución de los organismos y también acerca de las propias proteínas, ya que el hecho de que unos cambios o mutaciones se mantengan está relacionado de cómo afecta a la estructura o de la función de la estructura, y también con la importancia que tenga la función para el organismo. De este modo, los residuos más directamente implicados en la realización de la función suele presentar menor variación. Ya que los cambios en estos pueden provocar pérdida de función. La importancia de estos estudios comparativos de proteínas homólogas fue destacada por primera vez por Zuckerkandl y Pauling, 1965

Como podemos observar la evolución de las proteínas Hras, Nras y Kras demuestra la importancia de la función biológica que realiza en el ciclo celular, ya que al encontrar secuencias homólogas remotas, que incluyen solamente a la gran familia de Ras, que tienen funciones similares, al observar los alineamientos con los que se obtuvieron los árbol filogenético encontramos regiones con un grado de conservación alto, los cuales coinciden con las posiciones de los aminoácidos donde ocurren mutaciones relacionadas con el cáncer humano. Al observar cómo se distribuyeron las proteínas Hras, Kras y Nras en el árbol filogénico nos percatamos que la mayoría de las otras secuencias de las mismas proteínas pero encontradas en otras especies, se agruparon

en sus respectivos subnodos, lo cual demuestra que a pesar de que las secuencias de proteínas de Hras, de Kras y de Nras humano presentan una gran similitud, en nuestro árbol filogenético se agrupan en diferentes regiones. Por lo tanto demuestra que estas proteínas se originaron a partir de la duplicación de genes, ya que cada una de ellas está codificada por distintos genes.

2) Localización de las mutaciones puntuales relacionadas con el Cáncer humano.

Un aspecto importante para evaluar el impacto de las mutaciones asociadas con el cáncer es el que está relacionado con la evolución de las proteínas. Durante el proceso evolutivo, las proteínas sufren una serie de mutaciones de diversos tipos, desde simples sustituciones de aminoácidos, hasta procesos más complejos como eliminaciones e inserciones de residuos. Sin embargo, dependiendo de la importancia de la función de las proteínas, algunas regiones de las proteínas sufren pocas variaciones ya que éstas son importantes para la función y mantenimiento de la estructura de la proteína. De esta forma un alineamiento múltiple de secuencias homólogas puede revelar zonas conservadas que estén íntimamente relacionadas con la actividad y la estructura de la proteína. A medida que transcurre el proceso evolutivo, muchas proteínas pueden incluso modificar su función, sin embargo la estructura suele conservarse. Por este motivo un alineamiento de secuencias que incluyera homólogos distantes de proteínas, puede revelar la presencia de residuos que juegan un papel muy importante en el mantenimiento de la estructura.

Por este motivo, para este estudio se buscó exhaustivamente los homólogos de cada una de las secuencias de proteínas en estudio, incluyendo posibles homólogos remotos en diversas especies. Posteriormente se realizó para cada proteína sus alineamientos múltiples locales. Dichos alineamientos son una fuente de información importante ya que permitió identificar aquellas regiones altamente conservadas y las regiones con menor conservación.

Las regiones que están conservadas en los diferentes alineamientos de Kras, Hras, Nras, Abl1 isoforma A, p16 y p53 son cruciales en la estructura o en la función; estas regiones son dominios como en las proteínas Hras, Kras y Nras tiene el dominio de unión a los nucleótidos de guanina y el dominio catalítico; La Abl1 isoforma A posee los

dominios SH2, SH3 y PTKc_Abl; en p16 se encuentra el dominio ANK y en p53 se observa el dominio central, el dominio C-terminal y el dominio N-terminal. La mayoría de las mutaciones puntuales analizadas, que están relacionadas con el cáncer de cada una de las proteínas estudiadas en este proyecto se encuentran en dominios funcionales o estructurales. En el caso de Hras, Kras, Nras, p53 el 100%; en Abl1 isoforma A el 97% y en p16 el 90% de las sustituciones están en dominios funcionales.

La mayoría de los residuos en donde se producen mutaciones relacionadas con cáncer humano tienen un grado alto o intermedio de conservación. En Hras y Nras el 100% los residuos en donde ocurren mutaciones puntuales que están relacionadas con cáncer se ubican en regiones con un grado de conservación del 100%. En la proteína Kras isoforma B el 97% de los residuos en donde ocurren mutaciones puntuales relacionadas con cáncer tiene un alto grado de conservación y el 1.5% de los residuos en donde hay mutaciones puntuales relacionadas con el cáncer tienen un intermedio grado de conservación, en la proteína Abl1 isoforma A el 92% de los aminoácidos donde ocurren sustituciones relacionadas con cáncer se encuentran conservadas al 100% y el 8% de los residuos donde ocurren mutaciones relacionadas con cáncer tienen un grado de conservación intermedio, para la proteína p53 el 80% de los residuos donde ocurren sustituciones de un solo aminoácido y que están relacionadas con el cáncer tienen un grado de conservación del 100% y el 13% de los residuos en donde ocurren mutaciones relacionadas con el cáncer tiene un grado de conservación intermedio, y en la proteína p16 el 74% de los residuos en donde ocurren mutaciones relacionadas con el cáncer tienen un 100% de conservación. Por lo tanto esto demuestra que las regiones que se encuentran con un grado alto o intermedio de conservación son primordiales para la función biológica, por lo que las mutaciones relacionadas con cáncer tienden a localizarse en dichos residuos. Se puede sugerir que las mutaciones que ocurran en residuos que son muy conservadas, en este tipo de proteínas (oncoproteínas y proteínas supresoras de tumores) pueden propiciar que la proteína confiera propiedades cancerígenas.

Los patrones mutagénicos que se observa acentúan más que las mutaciones relacionadas con el cáncer humano que tienen estas proteínas estudiadas caen en residuos importantes en la función biológica o en la estructura de dichas proteínas.

Como se observa en la proteína Hras la mayoría de los residuos que están relacionados con mutaciones que participan en la cancerogénesis, son aminoácidos que son esenciales y que participan directamente en la función biológica, los residuos que son primordiales en el sitio de unión GTP/Mg²⁺ son los que se encuentran en las siguientes posiciones 12,13, 13,18 y 117. Los aminoácidos que actúan directamente en el sitio de interacción GEF y que están involucradas con el cáncer son los que se encuentran localizados en las posiciones 17, 18, 59 y 61 de la estructura primaria de Hras. En el sitio de interacción GDI también se observan residuos que son importantes para la función de este sitio y que además sus mutaciones están relacionadas con el cáncer los cuales se encuentran ubicados en las posiciones 11, 12, 13 y 59 en la secuencia de la proteína. En la caja G1 también hay residuos que sus mutaciones están relacionadas con el cáncer humano y son esenciales para que dicha región lleve a cabo su función, los cuales son los aminoácidos que se encuentran en las posiciones 11, 12, 13 y 17, y por último el residuo que se encuentra en la posición 117 es un aminoácido importante en la Caja G4 y además ocurren mutaciones relacionadas con el cáncer humano. También el residuo 59 es esencial en la función de la caja G3 y es donde ocurren sustituciones de un solo aminoácido que están relacionadas con el cáncer. El aminoácido que se encuentra en la posición 20 no es un residuo que interactúan directamente en algún sitio o caja, pero sin embargo se encuentran en las siguientes regiones funcionales: el sitio de unión GTP/Mg²⁺, en el sitio de interacción GEF y en el sitio de interacción GDI. Y por último el residuo localizado en la posición 83 no es un residuo que interactúan directamente en algún sitio o caja, pero sin embargo se encuentran en las siguientes regiones funcionales: el sitio de unión GTP/Mg²⁺ y en el sitio de interacción GEF. Este patrón mutagénico es muy similar al que ocurre a sus dos proteínas homólogas estudiadas (Kras y Nras), por lo tanto para estas proteínas Ras se puede deducir que los residuos que están implicados directamente en las funciones biológicas que llevan a cabo dichas proteínas, son aminoácidos críticos para convertir proteínas promotoras al cáncer, y sobre todo se hace evidente que una sustitución de un solo aminoácido en residuos que actúan directamente en sitios funcionales es suficiente para modificar la función de las proteínas.

El patrón mutagénico de la proteína ABI1 isoforma A podemos recalcar la importancia de la localización de los residuos en regiones funcionales de las proteínas, y el grado de conservación que tienen estas posiciones son determinantes para que cuando ocurran

mutaciones en dichos aminoácidos pueden ocasionar la proteínas que propicien al cáncer. Ya que tienen un alto o intermedio grado de conservación aquellos residuos donde se ocurren mutaciones relacionadas con cáncer, además el 94% de las mutaciones analizadas de la proteína Abl1 isoforma A se encuentran localizadas en el dominio catalítico cinasa. El 3% de las mutaciones se encontró en el dominio SH2 y la mutación Abl1IsoAR47G fue la única mutante que no cayó en regiones funcionales, pero sin embargo tiene un grado de conservación estructural del 100%. Otro aspecto que se puede observar en este patrón fue que las mutaciones estudiadas caen en regiones con un alto grado de conservación y los residuos que al ocurrir sustituciones simples ocasionan mutaciones relacionadas con el cáncer no ocurren en las regiones con un bajo grado de conservación.

El patrón mutagénico de las proteína p53 es muy similar a las anteriores proteínas, ya que recalca la influencia que tiene que ver el grado de conservación y la función de los residuos en los cuales ocurren mutaciones para que ocasionen cambios que impliquen la pérdida de función, en dicha proteína los lugares donde ocurren las mutaciones estudiadas y que están relacionadas con cáncer tienen una gran tendencia a localizarse en el dominio central, en donde cumplen la función de unión específica al DNA, se puede observar que el dominio central es la región más conservada de la proteína p53.

Las mutaciones que están implicadas en cáncer humano en la proteína p16 se encuentran distribuidas en toda la estructura de la proteína, la mayoría de las mutaciones se encuentran en el dominio ANK el cual está integrado por cuatro regiones que son importantes para mantener la estructura de la proteína, siendo esa estructura para que dicha proteína interactúe con otras biomoléculas, es por ello que todos los residuos del dominio ANK son propensos a la pérdida de una función o un cambio en la estructura cuando ocurren mutaciones puntuales y por lo tanto puede inducir al cáncer.

3) Análisis estructural de las proteínas silvestres y sus mutantes.

Aún cuando la presencia mutaciones en regiones conservadas sugiere un efecto sobre la estructura-función de la proteína, es necesario un estudio al nivel de la estructura tridimensional de la proteína para comprender su posible efecto. Para esto es necesario llevar a cabo la comparación de las estructuras proteicas silvestres y mutantes. En este

estudio las proteínas mutantes se generaron mediante el programa Modeller 9v7, mediante un script que permite, sustituir el aminoácido mutado en la estructura de la proteína silvestre. El modelo de la proteína mutante se puede generar con o sin su optimización posterior por dinámica molecular, la cual permite calcular el cambio en la conformación de la proteína ocasionado por la mutación. Los estudios conformacionales en este estudio son limitados ya que por un lado no se está calculando el cambio conformacional que puede sufrir la proteína en forma natural debida al “respiración molecular”. Solo los cambios conformacionales que sufren las proteínas mutantes, que superen la variación natural de la conformación de la proteína”, se puede catalogar que afectan significativamente la estructura de la misma. El cálculo de estas variaciones conformacionales es de altos requerimientos de cómputo y el tiempo necesario para llevarlos a cabo en este estudio es prohibitivo. Sin embargo, se puede argumentar que las regiones más conservadas en una proteína participan en estructuras relativamente rígidas, por lo que los cambios conformacionales detectados por la técnica en este estudio, pueden proporcionar indicios del efecto de las mutaciones al nivel estructural.

Después de modelar las proteínas mutantes se lleva a cabo un alineamiento estructural entre las proteínas que permite realiza una comparación para identificar diferencias estructurales o energéticas entre las proteínas mutantes y la silvestre. Estas comparaciones se pueden dividir en dos tipos: *globales* en las cuales se compara toda la proteína y *locales* en donde las comparaciones se realizan a nivel del alineamiento de residuos.

Los estudios bioinformáticos estructurales muestran que las mutaciones puntuales desestabilizan la estructura disminuyendo o inactivando la función de las proteínas supresoras de tumores. Y en el caso de las proteínas Nras, Kras y Hras las mutaciones puntuales modifican la estructura de la misma o la afinidad con otras moléculas, manteniendo a la las proteínas Nras, Kras y Hras activada constitutivamente.

Las comparaciones globales de estructuras completas (mutante/silvestre), no proporcionan suficientes datos respecto a cambios conformacionales, ya que en esta comparación al existir un solo cambio de aminoácido en dicha secuencia proteica, los cambios conformacionales, no son tan evidentes. Sin embargo, examinando las estructuras a nivel local, todas estas mutaciones causaron cambios importantes en la conformación de las proteínas mutantes respecto a la silvestre por lo que es importante

resaltar que para estudiar el impacto de las mutaciones estudiadas en este proyecto es necesario realizar comparaciones locales.

En las tres comparaciones locales se encontraron cambios significativos, identificando aquellos residuos donde ocurrieron cambios conformacionales ocasionados por la mutación. En el cálculo del RMS posicional se encontraron diferencias conformacionales, en el caso de Nras, Hras y Abl1 isoforma A la mutación solamente afectó a los residuos donde ocurrió la mutación, en cambio para Kras isoforma B, p16 y p53 ocurrieron cambios conformacionales en el sitio de la mutación y además en otros residuos. Fue interesante observar que estos aminoácidos adicionales se ubican cerca del aminoácido mutado cuando la proteína adquiere estructura terciaria. En todos los casos el mayor cambio ocurrió en el residuo mutado.

En el potencial DOPE se encontraron diferencias entre la proteína silvestre y la mutante, para el caso de Hras, Kras y Nras las mutaciones afectan principalmente los residuos 12, 13, 60 y 61. Estas mutaciones provocan la activación constitutiva del efecto de estas proteínas. En p53 y p16 las diferencias entre el potencial DOPE, no se encuentran en regiones pequeñas como en los casos de Hras, Kras y Nras, sino que afectan a gran parte de la estructura de la proteína. Por ello se puede sugerir que en el caso de las mutaciones que activan una oncoproteína provoca cambios en la energía DOPE en regiones pequeñas ocasionando un cambio conformacional, que provoca la activación. Para las proteínas codificadas por genes supresores de tumores los cambios en el potencial DOPE están situados en casi toda la estructura de la proteína, ocasionando perturbaciones que conducen a la pérdida de función.

En el cálculo de la energía de torsión se puede observar que el cambio de un solo aminoácido ocasiona modificaciones en los ángulos de torsión tanto en el residuo mutado así como de 4 a 5 aminoácidos a la derecha y de 4 a 5 aminoácidos a la izquierda de la mutación en la secuencia primaria.

5. CONCLUSIONES

1. Después de realizar una búsqueda exhaustiva en las bases de datos especializadas en mutaciones relacionadas con cáncer en el humano, se escogieron para este estudio las proteínas por ser aquellas para las cuales existió mayor información relativa a mutaciones que las asocian con la presencia de este padecimiento.
2. Los residuos donde ocurren mutaciones relacionadas con cáncer humano son importantes en la función biológica o en la estructura de la proteína. Para la proteína Hras, Kras y Nras caen en residuos que actúan directamente en las siguientes regiones funcionales: En el sitio de unión GTP/Mg²⁺, en el sitio de interacción GEF, en el sitio de interacción GDI, en la caja G1 y caja G4. Para la proteína p53 el 100% de las mutaciones implicadas en cáncer se encuentran en el dominio central; en Abl1 isoforma A el 97% se encuentra en el dominio cinasa las mutaciones y en p16 el 90% de las sustituciones están en el dominio ANK.
3. Las mutaciones puntuales que se estudiaron y que están relacionada con el cáncer humano de las oncoproteínas Hras, Nras, Kras y Abl1 isoforma A afectan a la estructura tridimensional en regiones localizadas.
4. Las mutaciones puntuales que se estudiaron y que están relacionadas con el cáncer humano de las proteínas supresoras de tumores p53 y p16 afecta toda la estructura tridimensional.

6. PERSPECTIVAS

Realizar el estudio evolutivo, estructural y funcional en otras proteínas que están relacionadas con el cáncer humano, para poder entender el impacto mutacional de dichas mutantes.

Se comprobó que las mutaciones que caen en regiones con alto o intermedio grado de conservación tienden a ser sitios muy importantes en la transformación de proteínas silvestres con funciones biológicas adecuada a proteínas mutantes que propicien la formación de células cancerígenas.

Este estudio proporciona herramientas bioinformáticas con las cuales pueden identificarse mutaciones que son potencialmente peligrosas para provocar cáncer. Estos estudios involucran la identificación de regiones conservadas y cambios tanto conformacionales como energéticos de las proteínas involucradas.

Estos estudios muestran, que aún cuando no se recurra a estudios computacionales más complejos como la dinámica molecular, el empleo de estas herramientas informáticas puede ayudar mutacionales que pueden tener gran impacto en la función de las proteínas en estudio.

Esto podría permitir utilizar este tipo de estudios para desarrollar técnicas de diagnóstico encaminadas a la detección de mutaciones especialmente peligrosas para padecimientos como el cáncer u otras enfermedades, por ejemplo el diseño de microarreglos que permitan detectar estas mutaciones.

7. REFERENCIA

1. Giuriato S, Felsher DW. How cancers escape their oncogene habit. *Cell Cycle* 2003;2:329-32.
2. Heeg S, Doebele M, von Werder A, Opitz OG. In vitro transformation models: modeling human cancer. *Cell Cycle* 2006;5:630-4.
3. Cooper GM. *The cell : a molecular approach*. 2nd ed. Washington, D.C. Sunderland, Mass.: ASM Press ; Sinauer Associates, 2000.
4. Osborne C, Wilson P, Tripathy D. Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. *Oncologist* 2004;9:361-77.
5. Kufe DW, Bast RC. *Cancer medicine*. 6th ed. Hamilton, Ont. ; London: B C Decker, 2003.
6. Alberts B. *Molecular biology of the cell*. 5th ed. New York: Garland Science, 2008.
7. Lodish HF. *Molecular cell biology*. 6th ed. New York: W.H. Freeman, 2008.
8. Weinberg RA. How cancer arises. *Sci Am* 1996;275:62-70.
9. McManus EJ, Alessi DR. Cancer, oncogenes and signal transduction. *Genome Biol* 2004;5:332.
10. Baba AI, C*atoi C, National Center for Biotechnology Information (U.S.), National Institutes of Health (U.S.). *Comparative oncology*. Bucharest: The Publishing House of the Romanian Academy, 2007.
11. Croce CM. Oncogenes and cancer. *N Engl J Med* 2008;358:502-11.
12. Wee KB, Aguda BD. Akt versus p53 in a network of oncogenes and tumor suppressor genes regulating cell survival and death. *Biophys J* 2006;91:857-65.
13. Martin AC, Facchiano AM, Cuff AL et al. Integrating mutation data and structural analysis of the TP53 tumor-suppressor protein. *Hum Mutat* 2002;19:149-64.
14. Soussi T, Caron de Fromental C, Mechali M et al. Cloning and characterization of a cDNA from *Xenopus laevis* coding for a protein homologous to human and murine p53. *Oncogene* 1987;1:71-8.
15. Picksley SM, Vojtesek B, Sparks A, Lane DP. Immunochemical analysis of the interaction of p53 with MDM2;--fine mapping of the MDM2 binding site on p53 using synthetic peptides. *Oncogene* 1994;9:2523-9.
16. Thut CJ, Chen JL, Klemm R, Tjian R. p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science* 1995;267:100-4.

17. Harris N, Brill E, Shohat O et al. Molecular basis for heterogeneity of the human p53 protein. *Mol Cell Biol* 1986;6:4650-6.
18. Soussi T, Beroud C. Assessing TP53 status in human tumours to evaluate clinical outcome. *Nat Rev Cancer* 2001;1:233-40.
19. Jiao YF, Nakamura S, Sugai T et al. p53 gene mutation and MDM2 overexpression in a case of primary malignant fibrous histiocytoma of the jejunum. *Apmis* 2002;110:165-71.
20. Verhagen AM, Lock P. Revealing the intricacies of cancer. *Genome Biol* 2002;3:REPORTS4015.
21. Soussi T, Asselain B, Hamroun D et al. Meta-analysis of the p53 mutation database for mutant p53 biological activity reveals a methodologic bias in mutation detection. *Clin Cancer Res* 2006;12:62-9.
22. Herrera PJC, Vásquez, Palacio. Gonzalo, Ramírez Castro José Luis. Papel del gen TP53 en la oncogénesis. *Salud UIS* 2004.
23. Hansen M, Rusyn EV, Hughes PE et al. R-Ras C-terminal sequences are sufficient to confer R-Ras specificity to H-Ras. *Oncogene* 2002;21:4448-61.
24. Abankwa D, Gorfe AA, Inder K, Hancock JF. Ras membrane orientation and nanodomain localization generate isoform diversity. *Proc Natl Acad Sci U S A* 2010;107:1130-5.
25. Prior IA, Hancock JF. Compartmentalization of Ras proteins. *J Cell Sci* 2001;114:1603-8.
26. Henis YI, Hancock JF, Prior IA. Ras acylation, compartmentalization and signaling nanoclusters (Review). *Mol Membr Biol* 2009;26:80-92.
27. Quinlan MP, Quatela SE, Philips MR, Settleman J. Activated Kras, but not Hras or Nras, may initiate tumors of endodermal origin via stem cell expansion. *Mol Cell Biol* 2008;28:2659-74.
28. Omerovic J, Prior IA. Compartmentalized signalling: Ras proteins and signalling nanoclusters. *Febs J* 2009;276:1817-25.
29. Kalas W, Yu JL, Milsom C et al. Oncogenes and Angiogenesis: down-regulation of thrombospondin-1 in normal fibroblasts exposed to factors from cancer cells harboring mutant ras. *Cancer Res* 2005;65:8878-86.
30. Stone S, Jiang P, Dayananth P et al. Complex structure and regulation of the P16 (MTS1) locus. *Cancer Res* 1995;55:2988-94.
31. Robertson KD, Jones PA. Tissue-specific alternative splicing in the human INK4a/ARF cell cycle regulatory locus. *Oncogene* 1999;18:3810-20.

32. Lin YC, Diccianni MB, Kim Y et al. Human p16gamma, a novel transcriptional variant of p16(INK4A), coexpresses with p16(INK4A) in cancer cells and inhibits cell-cycle progression. *Oncogene* 2007;26:7017-27.
33. Forbes SA, Tang G, Bindal N et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* 2008;38:D652-7.
34. Tamm I, Schumacher A, Karawajew L et al. Adenovirus-mediated gene transfer of P16INK4/CDKN2 into bax-negative colon cancer cells induces apoptosis and tumor regression in vivo. *Cancer Gene Ther* 2002;9:641-50.
35. Ghazizadeh M, Jin E, Shimizu H et al. Role of cdk4, p16INK4, and Rb expression in the prognosis of bronchioloalveolar carcinomas. *Respiration* 2005;72:68-73.
36. Han XL, Wu FG, Zhang ZY, Tong TJ. Posttranscriptional induction of p21Waf1 mediated by ectopic p16INK4 in human diploid fibroblast. *Chin Med J (Engl)* 2007;120:405-9.
37. Okamoto A, Demetrick DJ, Spillare EA et al. Mutations and altered expression of p16INK4 in human cancer. *Proc Natl Acad Sci U S A* 1994;91:11045-9.
38. Campbell I, Magliocco A, Moyana T et al. Adenovirus-mediated p16INK4 gene transfer significantly suppresses human breast cancer growth. *Cancer Gene Ther* 2000;7:1270-8.
39. Sharpless NE, Alson S, Chan S et al. p16(INK4a) and p53 deficiency cooperate in tumorigenesis. *Cancer Res* 2002;62:2761-5.
40. Bartkova J, Lukas J, Guldborg P et al. The p16-cyclin D/Cdk4-pRb pathway as a functional unit frequently altered in melanoma pathogenesis. *Cancer Res* 1996;56:5475-83.
41. Lukas J, Sorensen CS, Lukas C et al. p16INK4a, but not constitutively active pRb, can impose a sustained G1 arrest: molecular mechanisms and implications for oncogenesis. *Oncogene* 1999;18:3930-5.
42. Marchini S, Codegoni AM, Bonazzi C et al. Absence of deletions but frequent loss of expression of p16INK4 in human ovarian tumours. *Br J Cancer* 1997;76:146-9.
43. Liggett WH, Jr., Sidransky D. Role of the p16 tumor suppressor gene in cancer. *J Clin Oncol* 1998;16:1197-206.
44. Greenblatt MS, Beaudet JG, Gump JR et al. Detailed computational study of p53 and p16: using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene* 2003;22:1150-63.
45. Tang KS, Fersht AR, Itzhaki LS. Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure* 2003;11:67-73.

46. Zheng X, Guller S, Beissert T et al. BCR and its mutants, the reciprocal t(9;22)-associated ABL/BCR fusion proteins, differentially regulate the cytoskeleton and cell motility. *BMC Cancer* 2006;6:262.
47. Smith DL, Burthem J, Whetton AD. Molecular pathogenesis of chronic myeloid leukaemia. *Expert Rev Mol Med* 2003;5:1-27.
48. Mughal TI, Goldman JM. Targeting cancers with tyrosine kinase inhibitors: lessons learned from chronic myeloid leukaemia. *Clin Med* 2006;6:526-8.
49. Gu JJ, Ryu JR, Pendergast AM. Abl tyrosine kinases in T-cell signaling. *Immunol Rev* 2009;228:170-83.
50. Filippakopoulos P, Muller S, Knapp S. SH2 domains: modulators of nonreceptor tyrosine kinase activity. *Curr Opin Struct Biol* 2009;19:643-9.
51. Nachman J, Gish G, Virag C et al. Conformational determinants of phosphotyrosine peptides complexed with the Src SH2 domain. *PLoS One* 2010;5:e11215.
52. Moore JH. *Bioinformatics*. *J Cell Physiol* 2007;213:365-9.
53. Xiong J. *Essential Bioinformatics* Cambridge, UK: Cambridge university press, 2006.
54. Forman MR, Greene SM, Avis NE et al. *Bioinformatics: Tools to accelerate population science and disease control research*. *Am J Prev Med* 2010;38:646-51.
55. Kim JH. *Bioinformatics and genomic medicine*. *Genet Med* 2002;4:62S-65S.
56. Chou KC. *Structural bioinformatics and its impact to biomedical science*. *Curr Med Chem* 2004;11:2105-34.
57. Coulson A. High-performance searching of biosequence databases. *Trends Biotechnol* 1994;12:76-80.
58. Altschul SF, Boguski MS, Gish W, Wootton JC. Issues in searching molecular sequence databases. *Nat Genet* 1994;6:119-29.
59. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673-80.
60. Dopazo J, Zanders E, Dragoni I et al. Methods and approaches in the analysis of gene expression data. *J Immunol Methods* 2001;250:93-112.
61. Scherf U, Ross DT, Waltham M et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24:236-44.
62. Bhagwat M, Aravind L. PSI-BLAST tutorial. *Methods Mol Biol* 2007;395:177-86.

63. Lee MM, Chan MK, Bundschuh R. Simple is beautiful: a straightforward approach to improve the delineation of true and false positives in PSI-BLAST searches. *Bioinformatics* 2008;24:1339-43.
64. Przybylski D, Rost B. Consensus sequences improve PSI-BLAST through mimicking profile-profile alignments. *Nucleic Acids Res* 2007;35:2238-46.
65. Agrawal A, Huang X. PSIBLAST_PairwiseStatSig: reordering PSI-BLAST hits using pairwise statistical significance. *Bioinformatics* 2009;25:1082-3.
66. Gondro C, Kinghorn BP. A simple genetic algorithm for multiple sequence alignment. *Genet Mol Res* 2007;6:964-82.
67. Rausch T, Emde AK, Weese D et al. Segment-based multiple sequence alignment. *Bioinformatics* 2008;24:i187-92.
68. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792-7.
69. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205-17.
70. Larkin MA, Blackshields G, Brown NP et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;23:2947-8.
71. Keith J. *Bioinformatics: Structure, Function and Applications* Brisbane, Queensland, Australia: Humana Press, 2008.
72. Waterhouse AM, Procter JB, Martin DM et al. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189-91.
73. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406-25.
74. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596-9.
75. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 2008;9:299-306.
76. Berman HM. The Protein Data Bank: a historical perspective. *Acta Crystallogr A* 2008;64:88-95.
77. Arnold K, Kiefer F, Kopp J et al. The Protein Model Portal. *J Struct Funct Genomics* 2009;10:1-8.
78. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 2003;31:3381-5.

79. Bordoli L, Kiefer F, Arnold K et al. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 2009;4:1-13.
80. Bourne PE, Weissing H. *Structural Bioinformatics* New Jersey: Wiley-Liss, 2003.
81. Jefferson ER, Walsh TP, Barton GJ. Biological units and their effect upon the properties and prediction of protein-protein interactions. *J Mol Biol* 2006;364:1118-29.
82. Marchler-Bauer A, Anderson JB, Chitsaz F et al. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 2009;37:D205-10.
83. Finn RD, Mistry J, Tate J et al. The Pfam protein families database. *Nucleic Acids Res* 2010;38:D211-22.
84. Olivier M, Eeles R, Hollstein M et al. The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum Mutat* 2002;19:607-14.
85. Eswar N, Eramian D, Webb B et al. Protein structure modeling with MODELLER. *Methods Mol Biol* 2008;426:145-59.
86. Lema MA, Echave J. Assessing local structural perturbations in proteins. *BMC Bioinformatics* 2005;6:226.
87. Mayorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol* 1994;235:625-34.
88. Bruschweiler R. Efficient RMSD measures for the comparison of two molecular ensembles. Root-mean-square deviation. *Proteins* 2003;50:26-34.
89. Yu H, Kohl A, Binz HK et al. Molecular dynamics study of the stabilities of consensus designed ankyrin repeat proteins. *Proteins* 2006;65:285-95.
90. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 2006;15:2507-24.
91. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* 2008;71:261-77.
92. Benkert P, Kunzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res* 2009;37:W510-4.

8. ANEXOS

8.1. Anexo I

Clasificación de los aminoácidos basado en su tamaño y su afinidad con el agua.		
Grupo de aminoácidos	Nombre de los aminoácidos	Código de los aminoácidos
Pequeño y no polar	Glicina Alanina Prolina	Gly, G Ala, A Pro, P
Pequeño y polar	Cisteína Serina Threonina	Cys, C Ser, S Thr, T
Voluminoso y polar	Glutamina Asparigina	Gln, Q Asn, N
Voluminoso y polar (básico)	Arginina Lisina Histidina	Arg, R Lys, K His, H
Voluminoso y polar (ácido)	Acido glutámico Acido aspártico	Glu, E Asp, D
Voluminoso y no polar (alifático)	Isoleucina Leucina Metionina Valina	Ile, I Leu, L Met, M Val
Voluminoso y no polar (aromático)	Fenilalanina Tirosina Triptófano	Phe, F Tyr, Y Trp, w

8.2. Anexo II

NATURALEZA DEL CAMBIO DE LA PROTEÍNA HRAS			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutantes
Pequeño y no polar	Pequeño y polar	28	HrasA11S, HrasA18T, HrasA59T, HrasG12C, HrasG12S, HrasG13C, HrasG13S
Pequeño y no polar	Voluminoso y polar ácido	12	HrasA83D, HrasG12D, HrasG13D
Voluminoso y polar	Voluminoso y polar básico	12	HrasQ61H, HrasQ61K, HrasQ61R
Pequeño y no polar	Voluminoso y polar básico	8	HrasG12R, HrasG13R
Pequeño y no polar	Voluminoso y no polar alifático	8	HrasG12V, HrasG13V
Pequeño y no polar	Pequeño y no polar	4	HrasG12A
Pequeño y polar	Pequeño y no polar	4	HrasS17G
Pequeño y polar	Voluminoso y no polar alifático	4	HrasT20I
Voluminoso polar básico	Voluminoso polar ácido	4	HrasK117E
Voluminoso y polar	Pequeño y no polar	4	HrasQ61P
Voluminoso y polar	Voluminoso y no polar alifático	4	HrasQ61L
Voluminoso y polar básico	Voluminoso y polar ácido	4	HrasQ61E
Voluminoso y no polar alifático	Voluminoso y no polar alifático	4	HrasV81M

NATURALEZA DEL CAMBIO DE LA PROTEÍNA NRAS			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutantes
Pequeño y no polar	Pequeño y polar	23.53	NrasA11T, NrasA18T, NrasA59T, NrasA146T, NrasG12C, NrasG12S, NrasG13C, NrasG13S
Pequeño y no polar	Voluminoso y polar ácido	11.7	NrasG10E, NrasG12D, NrasG13D, NrasG60E
Pequeño y no polar	Pequeño y no polar	8.8	NrasG12A, NrasG12P, NrasG13A
Voluminoso polar	Voluminoso y polar básico	8.8	NrasQ61H, NrasQ61K, NrasQ61R
Pequeño y no polar	Voluminoso y polar	5.8	NrasG12N, NrasG13N
Pequeño y no polar	Voluminoso y polar básico	5.8	NrasG12R, NrasG13R
Pequeño y no polar	Voluminoso y no polar alifático	5.8	NrasG12V, NrasG13V
Pequeño y no polar	Voluminoso y no polar aromático	5.8	NrasG12Y, NrasG13Y
Pequeño y polar	Pequeño y polar	2.95	NrasS65C
Pequeño y polar	Voluminoso y polar básico	2.95	NrasS65R
Voluminoso y polar	Pequeño y no polar	2.95	NrasY64N
Voluminoso y polar	Voluminoso y polar ácido	2.95	NrasQ61E
Voluminoso y polar	Voluminoso y no polar alifático	2.95	NrasQ61L
Voluminoso y polar ácido	Voluminoso y polar básico	2.95	NrasE49K
Voluminoso y no polar alifático	Pequeño y no polar	2.95	NrasV45A
Voluminoso y no polar aromático	Voluminoso y polar	2.95	NrasY64N

8.3. Anexo III

NATURALEZA DEL CAMBIO DE LA PROTEÍNA KRAS			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutantes
Pequeño y no polar	Pequeño y polar	15.6	KraslsoBA18T, KraslsoBA59T, KraslsoBA146T, KraslsoBG12C, KraslsoBG12S, KraslsoBG13C, KraslsoBG13S, KraslsoBG15S
Pequeño y no polar	Voluminoso y polar ácido	13.7	KraslsoBA18D, KraslsoBA59E, KraslsoBG12D, KraslsoBG12E, KraslsoBG13D, KraslsoBG13E, KraslsoBG15D
Pequeño y no polar	Voluminoso y no polar alifático	11.7	KraslsoBA11V, KraslsoBA146V, KraslsoBG12L, KraslsoBG12V, KraslsoBG13I, KraslsoBG13V
Pequeño y no polar	Pequeño y no polar	9.8	KraslsoBA11P, KraslsoBA146P, KraslsoBG12A, KraslsoBG13A, KraslsoBG60A
Voluminoso y polar	Voluminoso y polar básico	5.8	KraslsoBQ22K, KraslsoBQ22R, KraslsoBQ61H, KraslsoBQ61K, KraslsoBQ61R,
Voluminoso y no polar alifático	Voluminoso y no polar alifático	5.8	KraslsoBL23I, KraslsoBM72I, KraslsoBV14I
Pequeño y no polar	Voluminoso y polar	3.9	KraslsoBG12N, KraslsoBG13N
Pequeño y no polar	Voluminoso y polar básico	3.9	KraslsoBG12R, KraslsoBG13R
Pequeño y no polar	Voluminoso y no polar aromático	3.9	KraslsoBG12F, KraslsoBG12Y
Voluminoso y polar básico	Voluminoso y polar	3.9	KraslsoBK5N, KraslsoBK117N
Pequeño y polar	Pequeño y no polar	1.96	KraslsoBS17G
Pequeño y polar	Voluminoso y no polar alifático	1.96	KraslsoBT35I
Voluminoso y polar	Pequeño y no polar	1.96	KraslsoBH27L
Voluminoso y polar	Voluminoso y polar ácido	1.96	KraslsoBQ61P
Voluminoso y polar	Voluminoso y no polar alifático	1.96	KraslsoBQ61E
Voluminoso y polar básico	Voluminoso y no polar alifático	1.96	KraslsoBQ61L
Voluminoso y polar ácido	Voluminoso y polar	1.96	KraslsoBE31Q
Voluminoso y polar ácido	Voluminoso y polar básico	1.96	KraslsoBE63K
Voluminoso y polar ácido	Voluminoso y polar ácido	1.96	KraslsoBE62D
Voluminoso y polar ácido	Voluminoso y no polar aromático	1.96	KraslsoBD92Y
Voluminoso y no polar alifático	Voluminoso y no polar aromático	1.96	KraslsoBL19F

8.4. Anexo IV

NATURALEZA DEL CAMBIO DE LA PROTEÍNA ABL1			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutantes
Voluminoso y no polar alifático	Voluminoso y no polar alifático	13.33	Abl1IsoAL248V, Abl1IsoAL387M, Abl1IsoAM237I, Abl1IsoAM244V
Voluminoso y no polar aromático	Voluminoso y no polar alifático	10	Abl1IsoAF311L, Abl1IsoAF359V, Abl1IsoAF382L
Pequeño y no polar	Voluminoso y polar ácido	6.6	Abl1IsoAH396R, Abl1IsoAR166K
Voluminoso y polar	Voluminoso y polar básico	6.6	Abl1IsoAQ252H, Abl1IsoAQ252R
Voluminoso y polar básico	Pequeño y no polar	6.6	Abl1IsoAH396P, Abl1IsoAR47G
Voluminoso y polar básico	Voluminoso y polar básico	6.6	Abl1IsoAH396R, Abl1IsoAR166K
Voluminoso y polar ácido	Pequeño y no polar	6.6	Abl1IsoAD276G, Abl1IsoAE352G
Voluminoso y polar ácido	Voluminoso y polar básico	6.6	Abl1IsoAE255K, Abl1IsoAE459K
Voluminoso y no polar alifático	Pequeño y polar	6.6	Abl1IsoAM343T, Abl1IsoAM351T
Pequeño y polar	Pequeño y no polar	3.3	Abl1IsoAT389A
Pequeño y polar	Voluminoso y polar	3.3	Abl1IsoAT315N
Pequeño y polar	Voluminoso y no polar alifático	3.3	Abl1IsoAT315I
Pequeño y no polar	Voluminoso y no polar aromático	3.3	Abl1IsoAS417Y
Voluminoso y polar	Voluminoso y polar ácido	3.3	Abl1IsoAQ252E
Voluminoso y polar ácido	Voluminoso y no polar alifático	3.3	Abl1IsoAE255V
Voluminoso y no polar alifático	Pequeño y no polar	3.3	Abl1IsoAV304G
Voluminoso y no polar aromático	Pequeño y no polar	3.3	Abl1IsoAF359A
Voluminoso y no polar aromático	Pequeño y polar	3.3	Abl1IsoAF486S

8.5. Anexo V

NATURALEZA DEL CAMBIO DE LA PROTEÍNA P16			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutante
Pequeño y no polar	Pequeño y polar	12	p16G67S, p16G89S, p16G122S, p16P11S, p16P70S, p16P75S, p16P114S, p16P151S
Pequeño y no polar	Voluminoso y no polar alifático	10.6	p16A30V, p16A34V, p16A40V, p16A57V, p16A60V, p16A68V, p16A76V, p16A86V, p16A100V, p16A102V, p16A118V, p16A127V, p16A134V, p16G35V, p16G67V, p16G89V, p16G150V, p16P38L, p16P70L, p16P75L, p16P81L, p16P114L
Voluminoso y no polar alifático	Voluminoso y no polar alifático	6.2	p16L63V, p16L113M, p16M9I, p16M52I, p16V28M, p16V51I, p16V59M, p16V82L, p16V82M, p16V95L, p16V106M, p16V115L, p16V126I
Pequeño y no polar	Voluminoso y polar ácido	5.3	p16A20E, p16A21D, p16A68E, p16A102E, p16G23D, p16G35E, p16G111D, p16G122D, p16G135E, p16G136D, p16G139D
Voluminoso y no polar alifático	Pequeño y no polar	5.3	p16L16P, p16L31P, p16L63P, p16L64P, p16L65P, p16L97P, p16L113P, p16V28G, p16V51A, p16V59G, p16V95A
Voluminoso y polar básico	Voluminoso y polar	4.8	p16H83N, p16H83Q, p16H123Q, p16H142Q, p16R24Q, p16R58Q, p16R80Q, p16R87Q, p16R99Q, p16R103Q
Voluminoso y polar ácido	Voluminoso y polar	4.3	p16D74N, p16D84N, p16D92N, p16D105N, p16D108N, p16D116N, p16D125N, p16E69Q, p16E119Q
Voluminoso y polar ácido	Voluminoso y polar ácido	3.8	p16D14E, p16D74E, p16D105E, p16D125E, p16D146E, p16E26D, p16E33D, p16E88D,
Pequeño y no polar	Pequeño y no polar	3.3	p16A20P, p16A30P, p16A57P, p16A73G, p16A132P, p16G135A, p16P70A
Voluminoso y polar básico	Voluminoso no polar aromático	3.3	p16H66Y, p16H83Y, p16H98Y, p16H142Y, p16R87W, p16R99W, p16R128W
Voluminoso y polar ácido	Voluminoso y polar básico	3.3	p16D84H, p16D108H, p16E2K, p16E26K, p16E69K, p16E88K, p16E120K
Pequeño no polar	Voluminoso y polar básico	2.9	p16G35R, p16G55R, p16P48R, p16P75H, p16P81H, p16P114H
Voluminoso y polar ácido	Voluminoso y no polar alifático	2.9	p16D14V, p16D74V, p16D84V, p16D116V, p16E69V, p16E88V
Pequeño polar	Voluminoso y no polar alifático	2.4	p16S43I, p16T18M, p16T79I, p16T93M, p16T137I
Voluminoso y polar básico	Voluminoso y polar básico	2.4	p16H66R, p16H83R, p16H98R, p16R124H, p16R131H
Voluminoso y polar ácido	Pequeño y no polar	2.4	p16D84G, p16E61G, p16E88A, p16E88G, p16E120A
Voluminoso y polar ácido	Voluminoso no polar aromático	2.4	p16D74Y, p16D84Y, p16D108Y, p16D116Y, p16D153Y
Voluminoso y polar básico	Pequeño y polar	1.9	p16R107C, p16R112S, p16R131C, p16R138T
Voluminoso y polar básico	Pequeño y no polar	1.9	p16H83P, p16R24P, p16R112G, p16R112P
Voluminoso y polar básico	Voluminoso y no polar alifático	1.9	p16R80L, p16R87L, p16R131L, p16R138I
Pequeño no polar	Pequeño no polar	1.5	p16C72S, p16S140C, p16T77S

NATURALEZA DEL CAMBIO DE LA PROTEÍNA P16			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutante
Voluminoso y polar	Voluminoso y polar básico	1.5	p16L16R, p16L78H, p16M52K
Pequeño polar	Voluminoso y polar básico	0.9	p16S43R, p16T93R
Voluminoso y no polar alifático	Pequeño polar	0.9	p16I49S, p16M53T
Voluminoso y no polar alifático	Voluminoso y polar	0.9	p16L91Q, p16L130Q
Voluminoso y no polar alifático	Pequeño y polar	0.9	p16Y44S, p16Y129C
Voluminoso y no polar aromático	Voluminoso y no polar alifático	0.9	p16F90L, p16W15L
Pequeño no polar	Voluminoso no polar aromático	0.5	p16G101W
Pequeño y polar	Voluminoso y polar	0.5	p16S56N
Pequeño polar	Voluminoso y no polar aromático	0.5	p16C72Y
Voluminoso polar	Pequeño polar	0.5	p16N42T
Voluminoso polar	Voluminoso polar alifático	0.5	p16N42D
Voluminoso no polar	Voluminoso polar acido	0.5	p16N71I
Voluminoso no polar alifático	Voluminoso polar acido	0.5	p16V115E
Voluminoso no polar alifático	Voluminoso no polar aromático	0.5	p16V126F
Voluminoso no polar aromático	Voluminoso polar básico	0.5	p16Y129H
Voluminoso aromático no	Voluminoso no aromático	0.5	p16F90F

8.6. Anexo VI

NATURALEZA DEL CAMBIO DE LA PROTEÍNA P53			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutantes
Pequeño y polar	Voluminoso y no polar aromático	9.5	p53C135F, p53C135Y, p53C176F, p53C176W, p53C176Y, p53C238Y, p53C242F, p53C242W, p53C242Y, p53C275F, p53C275Y, p53C277F, p53S241F, p53S241Y
Pequeño y no polar	Voluminoso y no polar alifático	6	p53A159V, p53G154V, p53G245V, p53G262V, p53G266V, p53P152L, p53P223L, p53P250L, p53P278L
Voluminoso y polar básico	Pequeño y no polar	6	p53R110P, p53R156P, p53R158G, p53R248G, p53R248P, p53R249G, p53R267P, p53R280G, p53R283P
Voluminoso y polar básico	Voluminoso y polar básico	6	p53H179R, p53H193R, p53K132R, p53R158H, p53R175H, p53R181H, p53R273H, p53R280K, p53R283H
Voluminoso y no polar alifático	Voluminoso y no polar alifático	6	p53I251L, p53L130V, p53M237I, p53M246I, p53V173M, p53V216L, p53V216M, p53V272L, p53V272M
Pequeño y no polar	Pequeño y polar	5.7	p53A161T, p53G244C, p53G244S, p53G245C, p53G245S, p53P98S, p53P151S, p53P278S
Voluminoso y polar básico	Voluminoso y no polar alifático	5.7	p53H193L, p53K120M, p53R110L, p53R158L, p53R175L, p53R248L, p53R267L, p53R273L
Voluminoso y polar básico	Voluminoso y polar	4.7	p53H179Q, p53K120N, p53K132N, p53K132Q, p53K164N, p53R213Q, p53R248Q
Voluminoso y polar ácido	Voluminoso y polar básico	4.08	p53D281H, p53E224K, p53E258K, p53E271K, p53E285K, p53E286K
Voluminoso y no polar aromático	Pequeño y polar	4.08	p53F113C, p53F113S, p53Y163C, p53Y220C, p53Y236C, p53Y234C
Voluminoso y polar básico	Pequeño y polar	3.4	p53R181C, p53R249S, p53R273C, p53R280T, p53R283C
Pequeño y no polar	Voluminoso y polar ácido	3.4	p53A159D, p53A161D, p53G244D, p53G262D, p53G266E
Voluminoso y no polar alifático	Pequeño y polar	3.4	p53I195T, p53I251T, p53I255S, p53I255T, p53M246T
Pequeño y no polar	Voluminoso y polar básico	2.7	p53G245R, p53G266R, p53P151H, p53P278R
Voluminoso y polar básico	Voluminoso y no polar aromático	2.7	p53H193Y, p53R248W, p53R267W, p53R282W
Voluminoso y no polar alifático	Voluminoso y polar	2.7	p53I162N, p53I232N, p53I251N, p53I255
Voluminoso y no polar alifático	Voluminoso y polar básico	2.7	p53L145R, p53L264R, p53M133K, p53M246R
Pequeño y no polar	Pequeño y no polar	2.04	p53A138P, p53G244A, p53P278A
Pequeño y polar	Voluminoso y polar básico	2.04	p53C124R, p53C135R, p53C242R
Pequeño y polar	Pequeño y polar	2.04	p53C242S, p53S241C, p53S261T
Voluminoso y no polar alifático	Voluminoso y no polar aromático	2.04	p53L194F, p53V157F, p53V274F
Pequeño y polar	Pequeño y no polar	1.3	p53C275G, p53T155P
Voluminoso y polar básico	Voluminoso y polar ácido	1.3	p53K132E, p53K164E
Voluminoso y polar ácido	Voluminoso y no polar aromático	1.3	p53D259Y, p53D281Y

NATURALEZA DEL CAMBIO DE LA PROTEÍNA P53			
Propiedad del aa normal	Propiedad del aa mutado	Porcentaje	Mutantes
Voluminoso y no polar alifático	Pequeño y no polar	1.3	p53L265P, p53V143A
Voluminoso y no polar alifático	Voluminoso y polar ácido	1.3	p53I254D, p53V147D
Pequeño y polar	Voluminoso y polar	0.7	p53T155N
Pequeño y polar	Voluminoso y no polar alifático	0.7	p53S215I
Voluminoso y polar	Voluminoso y no polar alifático	0.7	p53N268I
Voluminoso y polar ácido	Pequeño y no polar	0.7	p53D281G
Voluminoso y polar	Voluminoso y polar ácido	0.7	p53N239D
Voluminoso y no polar aromático	Voluminoso y polar básico	0.7	p53Y234H
Voluminoso y no polar aromático	Voluminoso y polar ácido	0.7	p53Y220D
Voluminoso y no polar aromático	Voluminoso y no polar alifático	0.7	p53F113V
Voluminoso y no polar aromático	Voluminoso y no polar aromático	0.7	p53Y205F
Voluminoso y polar ácido	Voluminoso y polar ácido	0.7	p53E224D

8.7. Anexo VII

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de Hras

Posición del aminoácido	Mutaciones	Porcentaje de conservación
11	A11S	100
12	G12A, G12C, G12D, G12R, G12S, G12V	100
13	G13C, G13D, G13R, G13S, G13V	100
17	S17G	100
18	A18T	100
20	T20I	100
59	A59T	100
61	Q61E, Q61H, Q61K, Q61L, Q61P, Q61R	100
81	V81M	100
83	A83D	100
117	K117E	100

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de Nras

Posición del aminoácido	Mutaciones	Porcentaje de conservación
10	G10E	100
11	A11T	100
12	G12A, G12C, G12D, G12N, G12P, G12R, G12S, G12V, G12Y	100
13	G13A, G13C, G13D, G13N, G13R, G13S., G13V, G13Y	100
18	A18T	100
45	V45A	100
49	E49K	100
59	A59T	100
60	G60E	100
61	Q61E, Q61K, Q61L, Q61P, Q61R	100
64	Y64N	100
65	S65C, S65R	100
146	A146T	100

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de Kras

Posición del aminoácido	Mutaciones	Porcentaje de conservación
5	K5N	100
11	A11P, A11V	100
12	G12A, G12C, G12D, G12E, G12F, G12L, G12N, G12R, G12S, G12V, G12Y	100
13	G13A, G13C, G13D, G13E, G13I, G13N, G13R, G13S, G13V	100
14	V14I	100
15	G15D, G15S	100
17	S17G	100
18	A18D, A18T	100
19	L19F	100
22	Q22R	100
23	L23I	100
27	H27L	100
59	A59E, A59T	100
60	G60A	100
61	Q61E, Q61H, Q61L, Q61P, Q61R	100
62	E62D	100
63	E63K	100
72	M72I	100
92	D92Y	100
117	K117N	100
146	A146P, A146T, A146V	100
31	E31Q	58
35	T35I	46

8.8. Anexo VIII

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de p53		
Posición del aminoácido	Mutaciones	Porcentaje de conservación
98	P98S	100
113	F113C, F113S, F113V	100
120	K120M, K120N	100
130	L130V	100
132	K132E, K132N, K132Q, K132R	100
135	C135F, C135R, C135Y	100
143	V143A	100
147	V147D	100
151	P151H, P151S	100
152	P152L	100
157	V157F	100
158	R158G, R158H, R158L	100
159	A159D, A159V	100
161	A161D, A161T	100
162	I162N	100
163	Y163C	100
164	K164E, K164N	100
173	V173M	100
175	R175H, R175L	100
179	H179Q, H179R	100
181	R181C, R181H	100
193	H193L, H193R, H193Y	100
194	L194F	100
195	I195T	100
205	Y205F	100
213	R213Q	100
215	S215I	100
216	V216L, V216M	100
220	Y220C, Y220D	100
223	P223L	100
237	M237I	100
238	C238Y	100
241	S241C, S241F, S241Y	100
242	C242F, C242R, C242S, C242W, C242Y	100
244	G244A, G244C, G244S	100
245	G245C, G245R, G245S, G245V	100
246	M246I, M246R, M246T	100
248	R248G, R248L, R248P, R248Q, R248W	100
249	R249G, R249S	100
250	P250L	100
251	I251L, I251N, I251T	100
254	I254D	100
255	I255N, I255S, I255T	100
258	E258K	100
259	D259Y	100
262	G262D, G262V	100
265	L265P	100
266	G266E, G266R, G266V	100
267	R267L, R267P, R267W	100
271	E271K	100

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de p53

Posición del aminoácido	Mutaciones	Porcentaje de conservación
272	V272L, V272M	100
273	R273C, R273H, R273L	100
274	V274F	100
275	C275F, C275G, C275Y	100
277	C277F	100
278	P278A, P278L, P278R, P278S	100
280	R280G, R280K, R280T	100
281	D281G, D281H, D281Y	100
282	R282W	100
285	E285K	100
286	E286K	100
145	L145R	82
283	R283C, R283H, R283P	82
232	I232N	81
224	E224D, E224K	76
155	T155N, T155P	73
133	M133K	72
264	L264R	72
236	Y236C	67
154	G154V	55
138	A138P	54
268	N268I	47
234	Y234C, Y234H	45
110	R110L, R110P	40
261	S261T	35
124	C124R	25

8.9. Anexo IX

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de p16

Posición del aminoácido	Mutaciones	Porcentaje de conservación
9	M9I	100
14	D14E, D14V	100
16	L16P, L16R	100
17	A17T	100
20	A20E, A20P, A20S, A20T	100
21	A21D	100
23	G23D	100
24	R24P, R24Q	100
28	V28G, V28M	100
30	A30P, A30V	100
31	L31P	100
33	E33D	100
34	A34V	100
35	G35E, G35R, G35V	100
38	P38L	100
39	N39K	100
40	A40S, A40V	100
42	N42D, N42T	100
45	G45S	100
48	P48R	100
49	I49S	100
50	Q50R	100
51	V51A, V51I	100
52	M52I, M52K	100
53	M53T	100
55	G55R	100
59	V59G, V59M	100
60	A60V	100
63	L63P, L63V	100
64	L64P	100
67	G67S, G67V	100
68	A68E, A68T, A68V	100
71	N71I, N71K	100
72	C72G, C72S, C72Y	100
74	D74E, D74N, D74V, D74Y	100
75	P75H, P75L, P75S	100
77	T77S	100
80	R80L, R80Q	100
81	P81H, P81L	100
82	V82L, V82M	100
83	H83N, H83P, H83Q, H83R, H83Y	100
84	D84G, D84H, D84N, D84V, D84Y	100
85	A85T	100

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de p16

Posición del aminoácido	Mutaciones	Porcentaje de conservación
86	A86V	100
87	R87L, R87Q, R87W	100
88	E88A, E88D, E88G, E88K, E88V	100
89	G89S, G89V	100
90	F90L	100
91	L91Q	100
92	D92N	100
93	T93A, T93M, T93R	100
95	V95A, V95L	100
97	L97P	100
98	H98R, H98Y	100
101	G101W	100
102	A102E, A102T, A102V	100
103	R103Q	100
105	D105E, D105N	100
106	V106M	100
107	R107C	100
108	D108H, D108N, D108Y	100
109	A109T	100
111	G111D	100
112	R112G, R112P, R112S	100
113	L113M, L113P	100
114	P114H, P114L, P114S	100
116	D116N, D116V, D116Y	100
118	A118V	100
120	E120A, E120K	100
122	G122D, G122S	100
123	H123Q	100
125	D125E, D125N	100
128	R128W	100
129	Y129C, Y129H	100
130	L130Q	100
131	R131C, R131H, R131L	100
134	A134V	100
135	G135A, G135E	100
140	S140C	100
150	G150V	100
43	S43I, S43R	75
100	A100S, A100V	73
126	V126F, V126I	72
79	T79I	68
76	A76T, A76V	64
69	E69K, E69Q, E69V	61
56	S56N	51
11	V115E, V115L	51
44	Y44S	44
58	R58Q	40

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de p16

Posición del aminoácido	Mutaciones	Porcentaje de conservación
78	L78H	40
36	A36T	39
124	R124H	39
18	T18M	36
61	E61G	35
73	A73G, A73S	35
138	R138I, R138T	34
11	P11S	31
70	P70A, P70L, P70S	31
143	A143T	30
136	G136D	29
142	H142Q, H142Y	29
57	A57P, A57V	28
99	R99Q, R99W	28
127	A127S, A127V	28
137	T137A, T137I, T137P	28
26	E26D, E26K	27
151	P151S	27
147	A147T	26
153	D153Y	26
132	A132P	25
139	G139D	21
66	H66R, H66Y	19
65	L65P	12
146	D146E	7
15	W15L	5
2	E2K	2
4	A4S	2
5	A5T	2

8.10 Anexo X

Grado de conservación evolutiva de los residuos donde ocurren mutaciones puntuales relacionadas con el cáncer de Abl1

Posición del aminoácido	Mutaciones	Porcentaje de conservación
47	R47G	100
166	R166K	100
244	M244V	100
248	L248V	100
250	G250E	100
252	Q252E, Q252H, Q252R	100
255	E255K, E255V	100
276	D276G	100
304	V304G	100
311	F311L	100
315	T315I, T315N	100
321	G321E	100
343	M343T	100
351	M351T	100
352	E352G	100
359	F359A, F359V	100
382	F382L	100
387	L387M	100
389	T389A	100
396	H396P, H396R	100
417	S417Y	100
486	F486S	100
237	M237I	82
459	E459K	82