



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE MEDICINA

**IDENTIFICACIÓN Y ANÁLISIS DE UN GEN PUTATIVO
PARA UNA CINASA DEPENDIENTE DE CICLINAS
EN *TRICHOMONAS VAGINALIS***

T E S I S

**QUE PARA OBTENER EL TÍTULO DE:
LICENCIADA EN INVESTIGACIÓN
BIOMÉDICA BÁSICA**

**P R E S E N T A :
OLIVIA ARIZMENDI PÉREZ**

**DIRECTOR DE TESIS:
DRA. MARÍA IMELDA LÓPEZ VILLASEÑOR**

MÉXICO, D.F.

2010



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

“Identificación y análisis de un gen putativo para una Kinasa Dependiente de Ciclinas en *Trichomonas vaginalis*”

Olivia Arizmendi Pérez

Asesor: Dra. María Imelda López Villaseñor

Departamento de Biología Molecular y Biotecnología

Instituto de Investigaciones Biomédicas, UNAM

Agradecimientos

A la Dra. Imelda López Villaseñor, por enseñarme cómo hacer ciencia, por la paciencia y la oportunidad de desarrollarme en su laboratorio. Gracias a su comprensión y su calidad humana, he logrado cerrar esta etapa.

A todos mis compañeros de laboratorio por su gran ayuda; en especial a Vanessa Fuentes por las largas horas compartidas en el laboratorio y a Pablo Hernández por siempre enseñarme nuevos métodos para este trabajo.

A la Lic. en Investigación Biomédica Básica, su coordinador Dr. Rafael Camacho y a la UNAM, porque la educación recibida en esta máxima casa de estudios siempre me acompañará.

Al Dr. Roberto Hernández y la Dra. Ana María Cevallos del Instituto de Investigaciones Biomédicas, por las valiosas discusiones de este proyecto en cada una de sus etapas.

Al Dr. Guillermo Mendoza Hernández, del Depto. de Bioquímica de la Facultad de Medicina, por su valiosa contribución a este trabajo con la secuenciación de péptidos realizada para la comprobación experimental del mismo.

A mi esposo, Dr. Francisco Martínez Becerra, por su constante recordatorio de que debía cumplir esta meta, por ser mi ánimo, mi familia y mi cómplice.

A mi mamá, por enseñarme que podía lograr cualquier cosa que me propusiera.

A mi hermano Pedro, por haberme cuidado siempre, por haber sido mi luz en mi niñez.

A Lupita, Paco, Guillermo y Rodrigo por hacerme parte de su casa y su felicidad.

I. ÍNDICE	
I. ÍNDICE	4
III. INTRODUCCIÓN	7
1. Protozoarios	7
2. <i>Trichomonas vaginalis</i>	9
4. Proteínas Cdk	13
IV. ANTECEDENTES	17
1. CDK en Levadura	17
2. CDK en Mamíferos	17
3. CDK en Protozoarios	20
4. Clona genómica <i>pgactin10</i>	22
V. PLANTEAMIENTO DEL PROBLEMA	24
VI. OBJETIVOS	25
VII. METODOLOGÍA	26
VIII. RESULTADOS	31
1. Análisis de la secuencia nucleotídica de <i>pgactin10</i>	31
2. Expresión y número de genes de <i>TvCDK</i>	32
3. Análisis <i>in silico</i> de <i>TvCDK</i>	35
4. Producción de una proteína recombinante	40
5. Unión de <i>TvCDK</i> a <i>p13suc⁺¹</i>	43
6. Análisis del genoma liberado por TIGR	43
7. Método de análisis teórico	48
8. Elaboración de modelos tridimensionales	49
IX. DISCUSIÓN	56
X. CONCLUSIONES	59
XI. PERSPECTIVAS	60
XII. BIBLIOGRAFÍA	61

II. RESUMEN

Trichomonas vaginalis (*T. vaginalis*) es un parásito protozooario flagelado hospedero del humano, causando la enfermedad conocida como tricomonosis. La tricomonosis ha sido catalogada por la Organización Mundial de la Salud como la enfermedad de transmisión sexual no viral de mayor incidencia en el mundo, siendo los más afectados los países en vías de desarrollo.

El estudio de *T. vaginalis* también resulta relevante en investigación básica pues los mecanismos de un gran número de procesos biológicos aún no están definidos ni estudiados a profundidad en organismos protozoarios. Es el caso del ciclo celular de *T. vaginalis*, cuyo estudio ha permanecido sin avances durante un largo periodo y que podría llegar a esclarecer principios básicos de ese proceso biológico dentro de un contexto filogenético.

En este proyecto a partir de una clona genómica de *T. vaginalis* que contenía un gen de actina en su región 3', se localizó un gen putativo para una Cinasa Dependiente de Ciclinas (CDK, por sus siglas en inglés) a 2 Kb río arriba del gen de actina. El trabajo consta de la caracterización de este gen a través de análisis experimentales y teóricos; los resultados nos sugieren que este es el primer gen de una proteína Cdk en ser descrito en *T. vaginalis*.

Como estudio práctico se realizaron análisis por medio de Southern y northern blot, producción de la proteína recombinante con un extremo de histidinas y ensayo de unión de la proteína recombinante a un sustrato específico para proteínas del tipo Cdk. Como estudio teórico se realizó el análisis de la secuencia nucleotídica y de posibles regiones consenso para transcripción; análisis de la secuencia peptídica y localización de motivos conservados de la familia de proteínas cinasa así como modelaje *in silico* de la posible estructura terciaria.

Como aportación final se elaboró un método teórico para realizar análisis *in silico* de un genoma anotado y así proponer en un organismo las posibles proteínas con actividad como CDK. Este método está basado en un puntaje asignado a secuencias de aminoácidos conservados. Se examinó el genoma de *T. vaginalis* con el método arriba descrito y se obtuvieron cinco secuencias con puntaje alto. A estas secuencias también se les realizó modelaje *in silico* y alineamiento de sus estructuras tridimensionales con una CDK convencional.

III. INTRODUCCIÓN

1. Protozoarios

Los protozoarios son seres unicelulares llamados comúnmente protistas y con un tamaño microscópico en el rango de 0.01 a 0.5 mm. La mayoría son de vida libre; sin embargo, algunos viven como parásitos y se sabe que todos los animales superiores pueden ser parasitados con una o más especies protozoarias provocando así infecciones que pueden ser asintomáticas o llevar a complicaciones severas².

Estos organismos se encuentran casi en cualquier hábitat: el plancton está conformado por organismos de los grupos taxonómicos dinoflagelados, foraminíferas y radiolarios y es parte del hábitat marino contribuyendo al balance del nicho ambiental; los rizópodos, flagelados y ciliados se encuentran comúnmente en agua dulce. También se encuentran diversos tipos de protozoarios en manantiales termales, lagos salados, vías de drenaje y en los suelos pueden ayudar a enriquecer el estrato forestal. Así, son componentes básicos y necesarios para una gran variedad de ecosistemas. Como parte de la cadena evolutiva tienen más diversidad que plantas y animales².

Los protozoarios constan de una anatomía compleja con componentes atípicos como peroxisomas, fotorreceptores, hidrogenosomas y estructuras móviles como cilios y flagelos. Tienen una nutrición holozoica (requieren sustancias orgánicas preformadas) y una gran parte de ellos se alimenta por medio de fagocitosis².

De acuerdo a algunas propuestas en sistemática, los protozoarios no son organismos ancestrales ni primitivos y la falta de algunos componentes y características podría deberse a una pérdida secundaria². Existe evidencia que sugiere que todos los eucariontes tuvieron mitocondrias en algún punto de su historia evolutiva³. Es por estos estudios filogenéticos que los protozoarios resultan de gran interés científico, así como

por las implicaciones del estudio de las especies parásitas en las ramas de la biomedicina y veterinaria.

Como agentes infecciosos representan un reto para encontrar medicamentos seguros, eficientes y de baja toxicidad ya que su carácter eucarionte hace que sea más difícil tratarlos sin afectar al hospedero a la vez. Este es el caso de la terapia con compuestos de la familia de los imidazoles, los cuales pueden tener una eficacia limitada y presentar riesgos de toxicidad importantes⁴. La Organización Mundial de la Salud (WHO, por sus siglas en inglés) marca ciertas enfermedades resultantes de la infección por protozoarios como críticas para la inversión en investigación y el entrenamiento de los servicios de salud. Se detalla en su plan de 'Control de Enfermedades Tropicales Relegadas' que más de mil millones de personas (un sexto de la población total mundial) sufre una ó más de las enfermedades tropicales relegadas y que éstas solo persisten en condiciones de pobreza y marginación. Dentro de tales enfermedades se encuentran la tripanosomiasis africana, leishmaniasis y la enfermedad de Chagas, las tres causadas por parásitos protozoarios¹².

Asimismo, WHO indica como un problema de salud pública a las enfermedades de transmisión sexual (ETS), con una marcada diferencia en su prevalencia dependiendo del progreso del país, siendo que la mortalidad para el año 2002 en los países en vías de desarrollo fue de 179,673 hombres y 88,728 mujeres; mientras que en países desarrollados fue de sólo 1,214 hombres y 832 mujeres¹².

En los países en vías de desarrollo las ETS y sus complicaciones son una de las cinco causas más comunes de consulta médica y pueden desembocar en síntomas agudos, infecciones crónicas y secuelas a largo plazo como incapacidad, infertilidad, embarazos extrauterinos, cáncer cervical y muerte. WHO estimó en 1999 que 340 millones de nuevos casos de sífilis, gonorrea, clamidia y tricomonosis habían ocurrido en personas

en el rango de 15 a 49 años de edad, aún cuando todas estas enfermedades son tratables¹².

2. *Trichomonas vaginalis*

Trichomonas vaginalis (*T. vaginalis*) es un protozoo que pertenece al grupo Parabasalia. Los parabasálidos son un linaje eucariótico altamente divergente y han sido divididos de acuerdo a sus características morfológicas y su filogenia molecular en hipermastigotes y tricomonádidos. Los hipermastigotes comprenden la mayoría de los organismos dentro del grupo Parabasalia y viven como comensales dentro del tracto intestinal de vertebrados e insectos².

El segundo grupo son los tricomonádidos, protozoarios flagelados anaeróbicos que se encuentran en vida libre o como parásitos y que incluyen a patógenos que causan enfermedades venéreas: En el hombre *T. vaginalis* y en el ganado *Tritrichomonas foetus*. Éste último puede causar infertilidad, lo que resulta en pérdidas económicas considerables para el sector ganadero⁹⁻¹¹.

T. vaginalis es el agente causante de la enfermedad de transmisión sexual (ETS, mencionadas con anterioridad) no viral de mayor prevalencia en el humano, pues se estima que anualmente origina de 174 a 180 millones de infecciones nuevas en todo el mundo. Su importancia médica se basa en que, al igual que las demás ETS, su prevalencia es extremadamente alta en algunos países en vías de desarrollo¹¹⁻¹³.

Este protozoo se transmite a través del tracto urogenital y las infecciones agudas están asociadas a un riesgo elevado de inflamación pélvica, infertilidad y problemas perinatales⁹. Causa síntomas en aproximadamente el 50% de las mujeres infectadas y a pesar de que en los hombres causa una infección uretral y de corta duración, ellos actúan como transmisores del parásito a la mujer. El cuadro clínico de la infección aguda

en mujeres incluye descarga vaginal, hedor, edema y eritema. También se ha documentado que *T. vaginalis* puede causar neumonía, bronquitis e infecciones orales y en algunos casos las infecciones respiratorias se adquieren perinatalmente de madres infectadas¹¹.

La infección por *T. vaginalis* ha sido asociada con un incremento del riesgo (1.52 veces) de infección con el Virus de Inmunodeficiencia Humana (VIH)¹⁴. Se cree que este incremento es debido a la inflamación local que causa *T. vaginalis* y la presencia de lesiones en las mucosas. También se ha observado que la co-infección con *T. vaginalis* predispone a individuos VIH positivos a desarrollar el Síndrome de Inmunodeficiencia Adquirida (SIDA)^{11,12}.

El metronidazol es el medicamento comúnmente utilizado para combatir la tricomonosis pero la resistencia a este compuesto es frecuente así como la resistencia cruzada con otros fármacos de la familia de los 5-nitroimidazoles, lo que representa un problema para el tratamiento de esta enfermedad¹⁵.

Se considera que el ciclo de vida de *T. vaginalis* consiste en una etapa única de trofozoito pues la etapa de quiste nunca ha sido documentada. El trofozoito varía en tamaño y forma, con una medida promedio de 10 micras de largo y 7 micras de ancho (**Fig. 1**). En cultivos axénicos su forma es ovalada y se torna ameboide cuando se adhiere a la pared del epitelio vaginal¹⁶.

El trofozoito tiene cuatro flagelos anteriores y un flagelo recurrente incorporado a una membrana ondulante que es sostenida por una costa no contráctil. Posee organelos denominados hidrogenosomas, similares anaeróbicos de las mitocondrias¹³.

Su división celular se lleva a cabo por fisión binaria (la forma más común de reproducción asexual) en la que una célula se divide transversalmente para dar lugar a

dos células hijas. En los tricomonádidos la fisión binaria ocurre como una división longitudinal cuya polaridad se determina por la duplicación inicial de los cuerpos basales de los flagelos¹⁷.

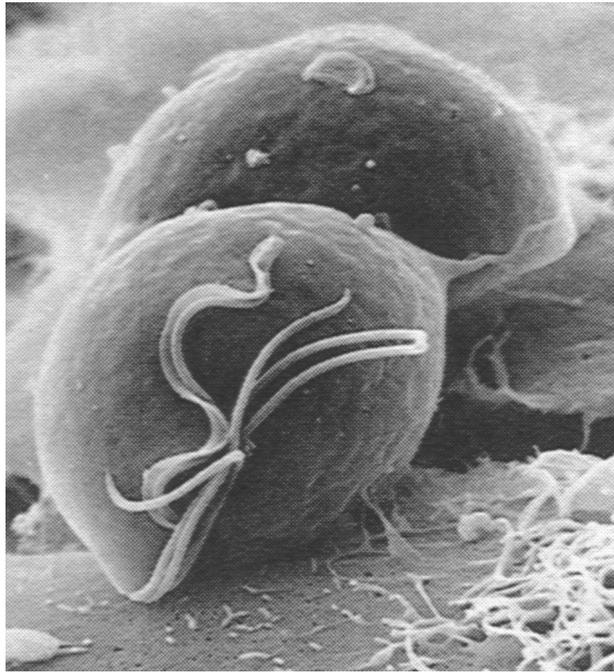


Figura 1. El trofozoito de *T. vaginalis* tiene una medida promedio de 10 micras de largo y 7 micras de ancho. En esta micrografía se encuentra antes de tomar una forma ameboide en la superficie del epitelio vaginal y se aprecian los cuatro flagelos anteriores y un flagelo recurrente incorporado a una membrana ondulante¹.

3. Ciclo celular y Cdk (Cinasa Dependiente de Ciclinas)

Las células eucariontes realizan su división celular a través de una serie de eventos coordinados que en conjunto se conocen como ciclo celular. El ciclo celular asegura la segregación correcta de la información genética por medio de etapas sucesivas de

duplicación y división del genoma y comprende todos los procesos bioquímicos y morfológicos necesarios para dar lugar a dos células hijas que mantengan la misma carga genética y la misma composición fundamental que su progenitora^{18,19}.

El ciclo celular se ha estudiado y definido como una sucesión de cuatro fases reguladas de manera precisa y denominadas G1 (*Gap 1* o postmitótica), S (Síntesis), G2 (*Gap 2* o postsintética) y M (Mitosis). Cada una de estas etapas comprende diferentes estados fisiológicos de la célula. En los eucariontes superiores el ciclo celular determina el crecimiento del organismo por medio de señales extracelulares mitogénicas y antimitogénicas; asimismo, define la diferenciación y el desarrollo dependiendo de los estímulos recibidos¹⁸. En eucariontes unicelulares como las levaduras y los protozoarios, el control del ciclo celular está íntimamente relacionado con el control de su ciclo de vida²⁰.

Para asegurar el control adecuado del ciclo celular, éste se vale de los '*checkpoints*' o puntos de control ejercidos por complejos proteicos que previenen que el ciclo celular avance hasta que se hayan completado eventos críticos de cada fase; por ejemplo, el *checkpoint* de replicación evita que la mitosis prosiga si la replicación del DNA aún no se ha completado y el *checkpoint* de ensamblaje del huso cromático detiene la separación de las cromátidas hermanas hasta que todos los cinetocoros se encuentren unidos al huso cromático.

Algunos de estos *checkpoints* funcionan inhibiendo la actividad de los complejos proteicos CDK para así evitar la progresión del ciclo celular¹⁸. Los complejos CDK son heterodímeros conformados por una subunidad catalítica conocida como Cdk (se utiliza el término en mayúsculas CDK para el heterodímero y el término con minúsculas Cdk para la subunidad catalítica)²¹. Las proteínas Cdk han sido ampliamente estudiadas en modelos que van desde las levaduras hasta los mamíferos y fueron conocidas anteriormente como proteínas Cdc (por las siglas en inglés de Cell division control).

El heterodímero funcional CDK se forma con una proteína del grupo heterogéneo de las Ciclinas, que reciben su nombre debido a la expresión oscilatoria que muestran durante las diferentes etapas del ciclo celular^{22,23}. Las ciclinas fungen como subunidades reguladoras del heterodímero, ya que permiten la activación de la proteína Cdk, ayudan a su estabilización, su localización celular y coadyuvan a la especificidad de sustrato²⁴.

Los complejos CDK se regulan de diferentes maneras: El número de genes presente en cada organismo, la disponibilidad de ciclinas y su especificidad para cada etapa del ciclo celular, la fosforilación y defosforilación regulatoria, la unión de inhibidores, la unión de otras moléculas reguladoras como Suc1 (llamada Cks1 en mamíferos), ubiquitinación y regulación de niveles transcripcionales¹⁸.

4. Proteínas Cdk

Las proteínas Cdk son importantes para mantener la progresión sincronizada del ciclo celular, ya que la activación secuencial y periódica de los complejos CDK determina el estado de fosforilación de sus proteínas blanco, las cuales comprenden a proteínas relacionadas al ciclo celular como son Whi5 (inhibidor de la fase S), Sic1 (inhibidor de la CDK de mitosis), Rb (represor transcripcional antiproliferativo) y otras asociadas a diferentes eventos como la condensación y dinámica cromosomal, transcripción de genes de histonas, proteólisis, replicación y reparación de DNA^{21,25}, etc.

Las proteínas Cdk son proteínas de la familia de las cinasas y fosforilan a sus proteínas blanco al transferirles un grupo fosfato proveniente de la molécula ATP²⁶. Sus características proteicas las incluyen en la superfamilia de las proteínas cinasas del tipo serina/treonina dirigidas a prolina y su peso molecular se encuentra entre los 30 a 36 kDa^{27,26}.

La regulación de las proteínas Cdk influye directamente en la progresión del ciclo celular. Estas proteínas no se regulan por sus niveles de concentración, ya que éstos se mantienen estables a lo largo de todo el ciclo celular. En cambio, las proteínas Cdk están sujetas a diferentes modificaciones postraduccionales y a eventos de fosforilación y defosforilación en algunos de sus aminoácidos significativos. Existen varios motivos de aminoácidos y estructuras características que han sido descritos como necesarios para la función de Cdk, su unión a ciclina y a Suc1 (otra proteína reguladora)^{23,28-33}.

Siguiendo el patrón de las proteínas cinasas, la conformación de las proteínas Cdk es bilobulada con una parte amino terminal (lóbulo N) rica en hojas β -plegadas y una parte carboxilo terminal (lóbulo C) más grande y conformada por α -hélices. Entre estos dos lóbulos existe una hendidura, que corresponde al sitio catalítico de unión a ATP^{26-28,32} **(Fig. 2)**.

Las proteínas Cdk conservan los subdominios descritos para todas las proteínas cinasas²⁸ y en su conformación proteica además tienen dos estructuras características: Una de ellas es llamada hélice *PSTAIRE* (por los aminoácidos que comúnmente la conforman) y se ha sugerido que funciona como un sitio de reconocimiento para la ciclina. La segunda estructura es llamada *T-loop* y se conforma por la región que va desde un motivo *DFG* (Aspartato-Fenilalanina-Glicina) hasta un motivo *APE* (Alanina-Prolina-Glutamato), ambos motivos son característicos de las proteínas cinasas.

En su estructura también se han definido como importantes una región helicoidal conocida como $\alpha L12$ y un sitio de fosforilación activadora T₁₆₁ (los números de aminoácidos se nombran así de acuerdo a sus equivalentes en Cdc2)^{27,29,31}.

Las ciclinas interactúan con ambos lóbulos de las proteínas Cdk y las activan al causarles cambios conformacionales: En un estado inicial (sin unión a una ciclina) Cdk tiene residuos de aminoácidos que son necesarios para la unión de ATP, pero que no

son funcionales debido a que la estructura $\alpha L12$ interfiere con la posición de la hélice *PSTAIRE*²⁶. Las ciclinas al unirse a las proteínas Cdk inducen la rotación de la hélice *PSTAIRE* y el contacto de ésta con el sitio catalítico.

También en un estado inicial (sin unión a una ciclina), la estructura *T-loop* está bloqueando al sitio catalítico. Una vez que la ciclina se une, ocasiona la apertura de la estructura *T-loop*.

La unión de ciclina también expone el sitio de fosforilación activadora T₁₆₁^{27,33}. Este sitio es fosforilado por la proteína llamada *Cinasa Activadora de CDK* (CAK, por sus siglas en inglés)^{18,26}.

La inhibición de los complejos CDK es ocasionada generalmente por fosforilación de las subunidades Cdk en sus aminoácidos T₁₄ y Y₁₅ (los números de aminoácidos se nombran así de acuerdo a sus equivalentes en Cdc2). Estos aminoácidos forman parte del sitio catalítico (en el motivo GXGTYG) y si son modificados irrumpen en la disposición de la hendidura del sitio catalítico y otros sitios de reconocimiento peptídico³³. Esta fosforilación la ocasionan las cinasas *Wee1* y *Myt1*¹⁸ y es removida por la fosfatasa *Cdc25*.

Los complejos CDK son regulados también por otras proteínas, subunidades inhibitorias conocidas como *CKIs* (por las siglas en inglés de Cyclin-dependent Kinase Inhibitors) que se unen a la proteína tipo Cdk o a ambas partes del complejo para detener el ciclo celular^{27,35}.

Otra forma de regulación de los complejos CDK es conferida por la proteína *Suc1* (también conocida como *Cks1* en levaduras), que es necesaria para la terminación completa de la mitosis y se considera un componente integral de la maquinaria que regula la función de los complejos CDK. *Suc1* se une a la proteína Cdk y actúa al inhibir la

defosforilación del aminoácido Y₁₅, asegurando así que los complejos CDK sólo actúen en la etapa del ciclo celular en que son requeridos. La unión a esta proteína reguladora ha sido utilizada como un criterio de análisis para la identificación de una proteína Cdk.

Esta inhibición en la defosforilación del aminoácido Y₁₅ se ha postulado que puede ser el resultado de que *Suc1* ejerza una vía de menor activación de la fosfatasa Cdc25 o porque permita una falta de inhibición de las cinasas *Wee1* y *Myt1*³⁶. Estudios cristalográficos y bioquímicos han sugerido que esa función dual puede ser debido a que *Suc1* esté actuando como una plataforma de acoplamiento para reguladores positivos y negativos de Cdk³⁷. También se ha estudiado la posibilidad de que *Suc1* adicionalmente regule a los complejos CDK al modular la degradación de algunas ciclinas³⁸.

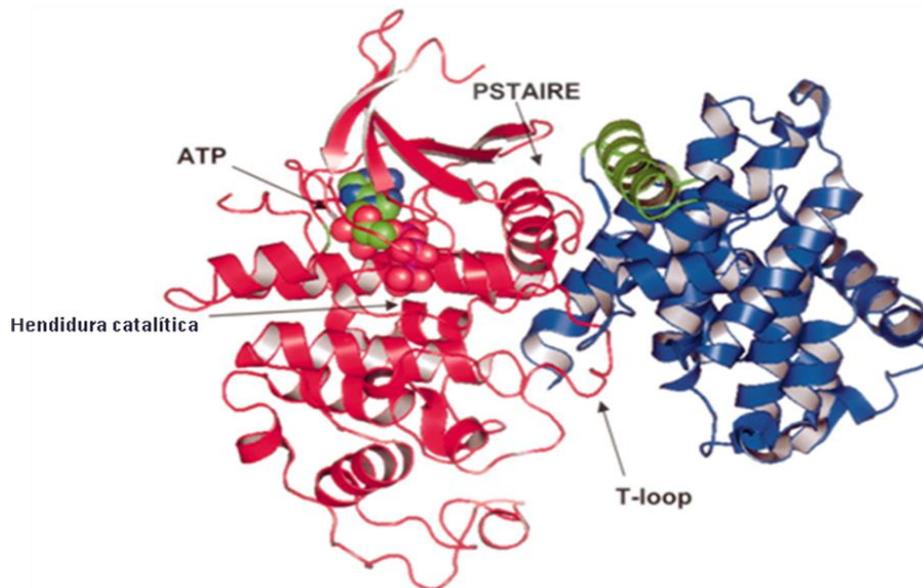


Figura 2. Representación en listones del complejo formado por una de las proteínas Cdk presentes en humano, llamada Cdk2 (en rojo) y su subunidad reguladora Ciclina A (en azul). Se indican las estructuras más representativas de las proteínas de tipo Cdk, como la hélice PSTAIRE, la hendidura catalítica y la estructura conocida como *T-loop*; así como una molécula de ATP unida al sitio catalítico^{34(modificada)}.

IV. ANTECEDENTES

1. CDK en Levadura

En *Saccharomyces cerevisiae*, la progresión del ciclo celular se encuentra regulada por una sola proteína Cdk llamada *Cdc28* y cuya especificidad para cada etapa del ciclo celular es dada por la ciclina a la que se ha asociado ¹⁹. Existen tres ciclinas del tipo *CLN* para el inicio del ciclo celular (en las etapas G1 y el inicio de S) y seis ciclinas del tipo *B* para la progresión (a partir del inicio de la replicación del DNA en S) **(Fig. 3)**.

En *Schizosaccharomyces pombe* también se requiere una sólo proteína Cdk llamada *Cdc2* y existen tres ciclinas del tipo *B*: *cdc13* y *cig1* se requieren para las fases G2 y M y *cig2* para las fases G1 y S.

2. CDK en Mamíferos

En los mamíferos la familia de proteínas Cdk comprende a 20 proteínas, de las cuales 11 son homólogas a *Cdc2* (*CDK1* a *CDK11*). Sólo cinco de estas proteínas Cdk tienen actividad durante el ciclo celular: *CDK4/6* que se unen a ciclinas tipo *D*; *CDK2* que se une a ciclinas tipo *A* y *E*; *CDK1* que se une a ciclinas tipo *A* y *B*; y *CDK3* que se une a ciclinas de los tipos *A*, *C* y *E* ²⁵ **(Fig. 4)**.

Las proteínas *CDK7* a *CDK11* se han encontrado involucradas en eventos de regulación transcripcional. El complejo conformado por *CDK7* y ciclina *H* está presente en el ciclo celular al actuar como Cinasa Activadora de CDK (*CAK*, por sus siglas en inglés) pero además actúa en la activación de la transcripción al ser un componente del factor general de transcripción IIH (*TFIIH*).

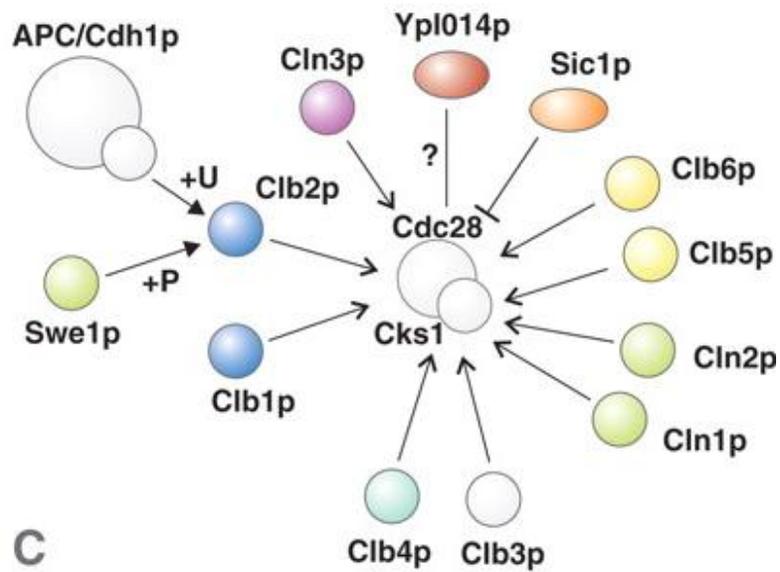


Figura 3. En levaduras, una sola proteína tipo Cdk se encuentra involucrada en el ciclo celular. En este modelo se muestra la proteína Cdk única de *S. cerevisiae*, llamada Cdc28. También se aprecia la posible unión de Cdc28 a las diferentes ciclinas presentes en este organismo, al homólogo en levaduras de la proteína reguladora *Suc1*, llamada *Cks1* y la subunidad inhibitoria *Sic1p*, que en el ciclo celular de levaduras representa una *CKI*, proteínas que se unen a la proteína tipo Cdk y a las ciclinas para detener el ciclo celular³⁹.

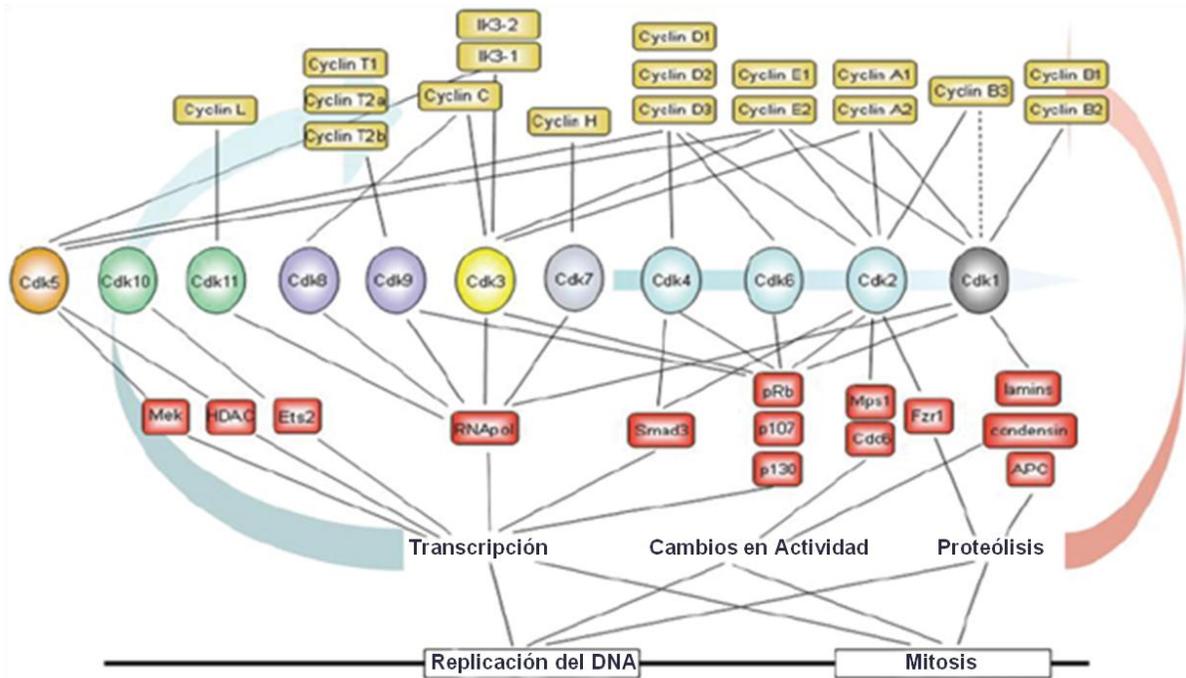


Figura 4. En mamíferos, varias proteínas del tipo Cdk se encuentran involucradas en el ciclo celular. También existen otras proteínas del tipo Cdk que se encuentran asociadas a otros procesos; por ejemplo, Cdk7 se encuentra relacionada con la progresión de la transcripción y además actúa como una CAK, que son las proteínas que se encargan de efectuar la fosforilación activadora a las proteínas Cdk. En esta imagen se aprecian las diferentes proteínas Cdk descritas en mamíferos, así como las ciclinas a que se unen, las interacciones proteicas que establecen y las funciones interdependientes que ejercen en el ciclo celular, la transcripción y otros procesos biológicos^{25(modificada)}.

El complejo *CDK8*-ciclina *C* fosforila a la RNA polimerasa II, el complejo *CDK9*-ciclina *T/K* regula el proceso de elongación y el complejo *CDK11*-ciclina *L* está involucrado en el proceso de splicing⁴⁰.

CDK5 es un factor ligado a eventos del sistema nervioso y que se ha descrito como activo durante senescencia y regulador de neurofilamentos junto con la proteína *p35*²⁵.

3. CDK en Protozoarios

Las proteínas Cdk descritas en diversos protozoarios a veces han sido llamadas CRK (por las siglas en inglés de Cdc2-Related Kinase), debido a que presentan cambios en la conformación típica de las proteínas Cdk de otros tipos de organismos o cambios en los motivos proteicos generalmente encontrados en éstas⁴¹.

Por ejemplo, algunas de las *CRK* conocidas pueden presentar cambios a la región regulatoria *PSTAIRE* y se ha descrito una secuencia mínima conservada de los aminoácidos equivalentes a P₄₅, R₅₀ y E₅₁^{20,41,42}. Es importante señalar que existen trabajos realizados en *Schizosaccharomyces pombe* que demostraron que se le podían modificar algunos aminoácidos a la región *PSTAIRE* de *Cdc2* y ésta retenía su función²⁰.

Las *CRK* comúnmente presentan extensiones en sus extremos amino y carboxilo, aunque se desconoce la utilidad que estas modificaciones les confieren²⁰. También se han descrito *CRK* que contienen extensiones o inserciones en sus dominios catalíticos, sin que esto les impida ser funcionales⁴¹. Un caso particular para estos hallazgos es la *CRK* descrita en el organismo *Crithidia fasciculata*, la cual tiene un peso molecular de 53 kDa y tiene dos largas inserciones de 66 y 79 aminoácidos; los estudios de esta *CRK* han demostrado que a pesar de estas extensiones, es una proteína funcional⁴³.

Protozooario	Número CRKs	Nombre (s)
<i>Trypanosoma brucei</i>	6	TbCRK1-4 y TbCRK6-7
<i>Trypanosoma cruzi</i>	2	TzCRK1 y TzCRK3
<i>Leishmania mexicana</i>	2	LmmCRK1 y LmmCRK3
<i>Leishmania major</i>	1	LmajCRK3
<i>Trypanosoma congolense</i>	1	TcCRK1
<i>Plasmodium falciparum</i>	2	PfPK5-6 y Pfmrk
<i>Crithidia fasciculata</i>	1	Cfa CRK
<i>Toxoplasma gondii</i>	1	TgCRK2
<i>Theileria annulata</i>	1	TaCRK2
<i>Dictyostelium discoideum</i>	1	Cdc2Dd
<i>Plasmodium vivax</i>	1	PvCRK2
<i>Entamoeba histolytica</i>	1	EhCDK2
<i>Giardia intestinalis</i>	1	CRK1

Tabla 1. CRKs (CDK Related Kinases) que han sido descritas en organismos protozoarios y cuya función en el ciclo celular ha sido comprobada por medio de análisis experimentales. En estos protozoarios y otros no mostrados en esta tabla, existen proteínas adicionales que se encuentran bajo estudio para identificar si ejercen una función en el ciclo celular ^{43,42,44-46,20,41,47}.

4. Clona genómica *pgactin10*

De acuerdo a lo anteriormente publicado por nuestro grupo de trabajo⁵¹, una alta proporción de guaninas y citosinas (G+C) dentro de una secuencia genómica de *T. vaginalis*, sugiere la presencia de una posible región codificadora.

Así, se analizó una clona genómica de *T. vaginalis* clonada en el vector pBluescript para *E. coli*, llamada *pgactin10* y que contenía una copia del gen de actina cerca de su extremo 3'. Se encontró que esta clona exhibía un mayor contenido de G+C aproximadamente 2 kb río arriba del gen de actina. La clona genómica *pgactin10* fue secuenciada y esta secuencia se verificó a través de un análisis manual de los nucleótidos completos para el inserto genómico. De esta manera, se confirmó que el inserto genómico tiene una longitud de 7,337 pb (**Fig. 5**).

El inserto genómico fue analizado con la herramienta *in silico* BLAST ('Basic Local Alignment Search Tool'), la cual consiste en un programa al que el usuario suministra secuencias nucleotídicas y/o peptídicas para ser comparadas con bases de datos de secuencias ya reportadas. El programa localiza regiones de similitud entre ellas y calcula la relevancia estadística de las correlaciones, lo que permite identificar de manera teórica miembros de familias de genes así como inferir relaciones de carácter funcional o evolutivo entre secuencias.

Este análisis permitió identificar que en la región del inserto de la clona que había sido marcada con un aumento en el contenido G+C, existía una región con homología a genes codificantes para proteínas del tipo Cdk en organismos filogenéticamente diversos como son levaduras, otros protozoarios, anfibios, aves y mamíferos. Esta región de homología comprende una sección mayor a 1 kb de longitud, con las características de abarcar en su totalidad la extensión de los genes homólogos a proteínas Cdk de la base de datos de la herramienta BLAST.

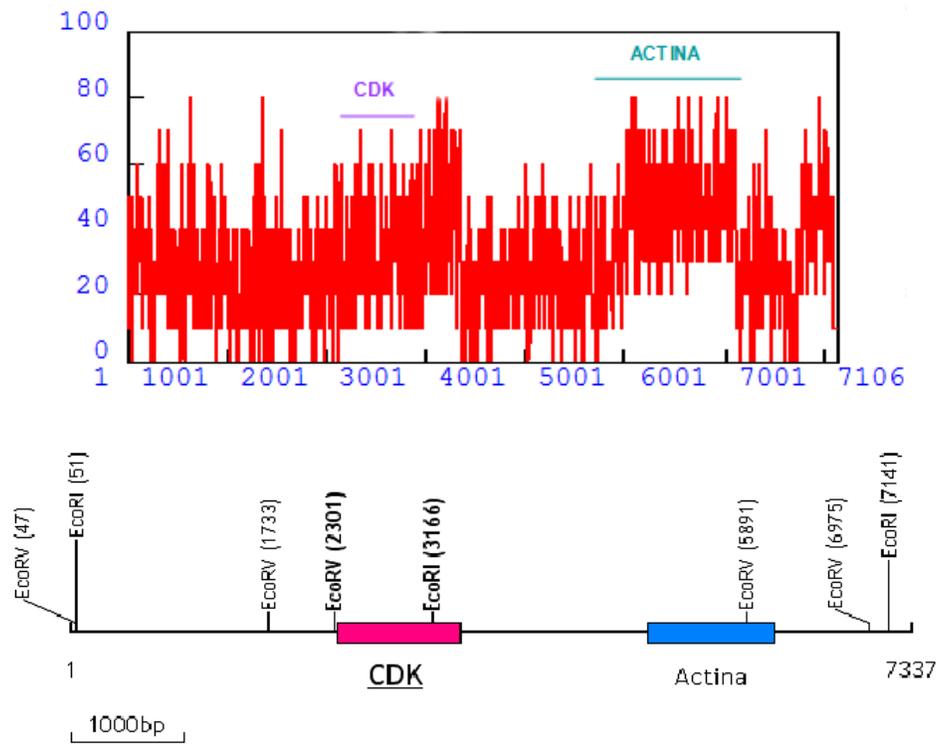


Figura 5. Mapa del contenido de G+C en la clona genómica *pgactin10* y esquema del inserto genómico presente en la clona *pgactin10* de *T. vaginalis*. Una alta proporción de guaninas y citosinas (G+C) dentro de una secuencia genómica de *T. vaginalis*, sugiere la presencia de una posible región codificadora. En este mapa, las regiones con alto contenido de G+C corresponden a: Una región de homología con proteínas del tipo Cdk (identificada mediante análisis tipo BLAST) y una copia del gen de actina.

V. PLANTEAMIENTO DEL PROBLEMA

Anteriormente, se realizó una biblioteca genómica de *T. vaginalis* y se obtuvieron ocho clonas que contenían copias del gen constitutivo para actina. Al realizar un análisis *in silico* de las regiones asociadas al gen de actina, se describió que la clona denominada *pgactin10* contenía aproximadamente a 2 kb río arriba del gen de actina, una región de alta homología a genes codificantes para Cinasas Dependientes de Ciclinas (Cdk) de organismos muy diversos.

Al tomar en cuenta que el estudio del ciclo celular de *T. vaginalis* es escaso y aún no se han descrito las proteínas que forman parte de su regulación, se consideró relevante caracterizar la región con homología descrita y determinar si ésta codifica para una proteína del tipo Cdk.

Este trabajo se divide en dos partes: Por una parte está la comprobación experimental de que esta secuencia homóloga a Cdk se trata de un gen que se expresa en *T. vaginalis* y su posible funcionalidad a nivel proteico; por otra parte se realizó un estudio *in silico* de las secuencias nucleotídica y peptídica; asimismo se creó un sistema teórico de evaluación para proteínas que se proponen como Cdk.

VI. OBJETIVOS

1. Analizar la secuencia nucleotídica de *pgactin10* y describir los elementos de un gen en la región con homología a genes codificantes para proteínas Cdk.
2. Identificar el número de copias de este gen en el genoma de *T. vaginalis*.
3. Analizar la secuencia peptídica y elaborar un modelo estructural por homología.
4. Comprobar la expresión del gen a nivel de transcripción de RNA mensajero.
5. Clonar la secuencia codificante en un vector de expresión y purificar la proteína recombinante.
6. Determinar la capacidad de la proteína recombinante para unirse al sustrato específico $p13^{suc+1}$.
7. Elaborar un método de evaluación teórico para determinar la identidad de una proteína como CDK.
8. Analizar el genoma de *T. vaginalis* con el método de evaluación teórico desarrollado.

VII. METODOLOGÍA

1. Cultivos

El aislado CNC147⁵¹ de *T. vaginalis* fue utilizado para todos los experimentos. Estos protozoarios fueron cultivados en medio TYM (1% triptona, 0.5% extracto de levadura y 0.25% maltosa) suplementado al 10% con suero de caballo inactivado.

2. Extracción de DNA genómico y RNA total

Cultivos de *T. vaginalis* fueron tratados para la extracción de DNA por el método de fenol-cloroformo⁴⁸ con una modificación del amortiguador de lisis (0.2% SDS, 100mM NaCl, 10 mM EDTA y 10 mM Tris-HCl pH 8) y para la extracción de RNA por el método de TRIzol-cloroformo, una modificación al método de un solo paso⁴⁹.

3. Southern blot

Para el ensayo de Southern blot con amortiguador de acetato de amonio, DNA genómico de *T. vaginalis* (5 µg) fue digerido con la enzima de restricción EcoRI y separado por electroforesis en un gel de agarosa al 0.8%. El DNA genómico fue digerido con esta enzima pues la clona genómica *pgactin10* fue construida con un inserto de extremos EcoRI. El gel resultante fue transferido a un filtro de nylon que posteriormente fue fijado por entrecruzamiento con luz UV, hibridado a 42°C con sondas elaboradas por random-priming con α [³²P]dCTP, lavado con la solución 2X SSC/0.5% SDS y expuesto a autoradiografía.

4. Northern blot

Para los ensayos de northern blot se utilizó RNA total de *T. vaginalis* (5 µg), que fue separado por electroforesis en un gel de agarosa-formaldehído al 1.75%. Este gel fue transferido a una membrana de nitrocelulosa, la cual fue procesada de la misma manera que los filtros de Southern blot.

5. Sondas

Las sondas utilizadas en este trabajo fueron realizadas por el método de random-priming⁵⁰ con α [³²P]dCTP.

En el caso de la sonda para la secuencia de *TvCDK* el fragmento marcado fue de 865 pb, que fue purificado de una digestión de la clona genómica *pgactin10* de *T. vaginalis* con las enzimas de restricción EcoRV y EcoRI.

6. Análisis de secuencias

La secuencia de la clona genómica de *T. vaginalis* llamada *pgactin10* fue analizada con la herramienta *in silico* BLAST de NCBI para análisis de secuencias nucleotídicas o peptídicas (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Se utilizaron las variantes *blastn* para comparación entre secuencias nucleotídicas, *blastp* para comparación de secuencias peptídicas y *blastx*. Esta última herramienta traduce la secuencia nucleotídica provista a una secuencia de aminoácidos; y la secuencia de aminoácidos resultante la compara con una base de datos de secuencias de proteínas anteriormente reportadas.

7. Modelo de la estructura terciaria de *TvCDK*

El modelo de la conformación que tendría la proteína *TvCDK* putativa se realizó por medio de la herramienta 3D PSSM Protein Fold Recognition (<http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html>). Esta herramienta realiza una búsqueda de identidad entre una base de datos de secuencias pertenecientes a proteínas cuya estructura tridimensional ya se conoce y una secuencia de aminoácidos que el usuario provee. Al encontrar similitud con algunas de las secuencias reportadas, la herramienta toma como patrón el modelo terciario correspondiente para proporcionar un modelo hipotético de la estructura terciaria que tendría la proteína resultante de la secuencia de interés. Los archivos PDB fueron visualizados con el

programa PyMOL de DeLano Scientific para obtener imágenes en representación de listones y hélices con regiones y residuos de aminoácidos resaltados en colores únicos.

8. Construcción de un plásmido recombinante

El marco abierto de lectura del gen putativo para *TvCDK* fue fusionado a una cola de seis histidinas usando el vector de expresión pQE30 de QIAgen en los sitios BamHI y PstI. Estos sitios se insertaron en la secuencia de cDNA por medio de PCR de la clona genómica *pgactin10* con los oligos presentes en la **Tabla 2**.

Nombre	Secuencia	Sitio de restricción
CDK30B	5'CATCGGATCCAT GGATTCCAGCGATGATTTT 3'	BamHI
CDK30R	5' CATCCTGCAGTTAAATTT CAGCTT TCTCAGG3'	PstI

Tabla 2. Oligos utilizados para la obtención por medio de PCR del ORF de *TvCDK* con extremos BamHI y PstI desde la clona *pgactin10*. Con letra resaltada se marca la región de cada oligo que empalma con la secuencia de *pgactin10* y subrayado el sitio de corte de la enzima de restricción señalada.

El PCR fue llevado a cabo con la polimerasa Pfu y su producto fue clonado en el vector pCR-BluntII-TOPO. Este nuevo plásmido fue digerido con ambas enzimas (BamHI/PstI) y el fragmento con el gen para *TvCDK* fue purificado de esta doble digestión. Este fragmento fue clonado en el vector pQE30 previamente digerido por las mismas enzimas, utilizando la cepa M15 [pREP4] de *E. coli*.

9. Secuenciación

El plásmido resultante de la clonación de la región codificante de *TvCDK* en el vector pQE30 fue nombrado pCDK30. Este plásmido recombinante se purificó con el Estuche QIAprep de Qiagen y posteriormente se secuenció la unión 5' del vector con *TvCDK* para asegurar que se encontraran en fase las secuencias de la región codificante de *TvCDK* y la región del vector que codifica para las seis histidinas. La secuenciación se realizó con

un plásmido templado pCDK30 a concentración 0.3 µg/µl y el oligo '10gactin14' (5'TCTGCTTGCTTGCAATTGAAG3') a concentración 4 pmol/µl por el método de Sanger.

10. Ensayo de inducción de *His-CDK*

Cultivos de *E. coli* transfectados con el plásmido recombinante pCDK30 fueron inducidos con dos concentraciones de IPTG (0.1 mM y 1 mM) al alcanzar una D.O. de 0.6 a 600 nm de luz. Se tomaron alícuotas cada 30 minutos durante 5 horas, y los extractos proteicos totales se obtuvieron por ebullición de las muestras del cultivo en amortiguador de Laemmli. Las muestras se sometieron a electroforesis en gel de poliacrilamida (12% gel separador, 5% gel concentrador) con amortiguador Tris-glicina (25 mM Tris base, 250 mM glicina pH 8.3, 0.1% SDS).

11. Ensayo de unión a *p13suc⁺¹*

Cultivos de *E. coli* transfectados con el plásmido recombinante CDK30 fueron inducidos con 0.1 mM de IPTG por 3 horas. Las células se lavaron con PBS, se lisaron con 250 µl de Amortiguador de lisis K (50 mM Tris-HCl pH 7.4, 250 mM NaCl, 5 mM EDTA, 1 mM DTT, 0.1% NP-40, 2 µg/ml Leupeptina y 50 µg/ml PMSF). El lisado se incubó 15 minutos en hielo y finalmente fue centrifugado. El sobrenadante con el lisado proteico total se mantuvo a -20°C.

Proteína *p13suc⁺¹* conjugada a agarosa a una concentración 2.5 mg/500 µl en suspensión del 50% fue adquirida de *Upstate Cell Signaling Solutions*. De estas perlas, se lavaron 100 µg por muestra con 1 ml de Amortiguador *Beads* (50 mM Tris pH 7.4, 250 mM NaCl, 5 mM EDTA, 5 mM EGTA y 0.1% NP-40), se centrifugaron y se añadió 250 µl de Amortiguador *Beads* más 240 µg del lisado proteico a estudiar. Esta suspensión se mantuvo en rotación constante durante 1 hora a 4°C. Posteriormente se centrifugó y el botón de perlas fue lavado en 5 ocasiones con 250 µl de Amortiguador *Beads*. Las perlas resultantes fueron resuspendidas en 10 µl de Amortiguador Laemmli y sometidas a ebullición por 5 minutos. *Esta es una variación al método utilizado por Zhang⁵².*

Las muestras se sometieron a electroforesis en gel de poliacrilamida (12% gel separador, 5% gel concentrador) con amortiguador Tris-glicina (25 mM Tris base, 250 mM glicina pH 8.3, 0.1% SDS). La proteína resultante de la unión con *p13suc⁺¹* fue purificada y secuenciada por medio de espectrometría de masas en el laboratorio a cargo del Dr. Guillermo Mendoza Hernández de la Facultad de Medicina, UNAM.

VIII. RESULTADOS

1. Análisis de la secuencia nucleotídica de pgactin10

Con esta información, se analizó si esta región de homología podría contener un gen completo y funcional para una proteína del tipo Cdk de *T. vaginalis*. Se procedió a estudiar la región para localizar los elementos de transcripción a nivel secuencia nucleotídica que podrían identificarla como un posible marco abierto de lectura (ORF, *open reading frame*).

En el análisis, se verificó en primera instancia la existencia dentro de la secuencia de dos codones: uno de inicio de la traducción (ATG) y otro de término (TAA o 'codón ocre', el cual es el codón preferentemente utilizado por *T. vaginalis* para término de la traducción). Se consideró que ambos codones estuvieran en fase y se encontró un ORF que codificaría una proteína de 360 aminoácidos con un peso aproximado de 41.4 KDa.

También se examinó la presencia de otros elementos que le podrían conferir características de un gen apto para la transcripción. Se encontró una secuencia en el extremo 5' que concordaba con el consenso YCA₊₁NW reportado por Liston y Johnson para un promotor del tipo Inr específico de *T. vaginalis*, donde el nucleótido A señalado con un +1 marca el sitio de inicio de la transcripción⁵³. Tal promotor putativo en esta región de homología de la clona genómica, corresponde a la secuencia TCA₊₁AA y se encuentra a 14 pb río arriba del codón de inicio de traducción ATG.

Se consideraron también los trabajos realizados con anterioridad en nuestro grupo de investigación que describen una secuencia que fungiría como señal de poliadenilación en la transcripción y otra secuencia consenso que equivaldría a un sitio de corte y procesamiento post-transcripcional⁵¹. Así, dentro de esta región de homología de la

clona *pgactin10* se encontraron en su extremo 3' la probable señal de poliadenilación TAAA que contiene al codón TAA terminador (consenso UAAA) y dos probables sitios de procesamiento TAAATT (consenso $Y_{\downarrow}(A)_{0-3}AAUU$).

Con este análisis se identificó una región putativa para un transcrito de RNA de aproximadamente 1.1 Kb y un posible ORF codificante para una proteína de 360 aminoácidos; este ORF lo denominaremos en lo subsecuente '*TvCDK*' (**Fig. 6**).

2. Expresión y número de genes de *TvCDK*

Para analizar el número de genes probables similares a *TvCDK* dentro del genoma de *T. vaginalis* y comprobar si existía la expresión de un RNA mensajero (mRNA), se realizaron los ensayos de Southern y northern blot.

En el ensayo de Southern blot se analizó cuántas eran las regiones en el DNA genómico de *T. vaginalis* con similitud a *TvCDK*. Este ensayo se realizó empleando DNA genómico digerido con EcoRI y la sonda descrita en la sección de métodos y comprende los nucleótidos de la posición -22 a la posición +886 del ORF mostrado en la Figura 8, incluyendo así la mayor parte del ORF.

El número de segmentos con los que se halla hibridación en este tipo de ensayos de Southern blot son un indicativo del número de posibles regiones con similitud a la sonda presentes en el genoma de *T. vaginalis*. En este estudio, se encontró sólo una banda de hibridación de aproximadamente 7.1 Kb que corresponde al tamaño del inserto que contiene la clona *pgactin10* (**Fig. 7 Izq**).

En el ensayo de northern blot se comprobó la expresión de un RNA mensajero (mRNA) de *TvCDK*, al obtener una señal resultante de la hibridación del segmento de DNA de la sonda con la población de mRNA presente en un extracto de RNA total de *T. vaginalis*.

```

aataaaacag aaaatcgaga tatcgaaatt gagagcaatc aaaatctttt tATGGATTCC 60
AGCGATGATT TTAGCATAGA AACAAAGGAT CTGAGTACGC TCTTTACCAA GGTAGAAAAG 120
ATCGGTCTGG GAACCTATGG CAAGGTTTAT AAGTGCTTG ATATAAGAAA GAAACAACT 180
GTTGCTCTAA AGAAGATTAA AATCCTAAAG CCAAACGACG GCTTTCCTGT AAATACTGTT 240
AGAGAAATCA AGTGCCTAAA AGAACTAAAA CATGACAATA TTATCAGATT AAGAAATGTA 300
ATTACATTAT ACAAACCAA TAAGCCGTCA ACAGTATGGT TATCATTTGA TTACTGTGAA 360
TTTGACTTAT TTGGATTACT CCACAAAACA GATTTTCCAC AATTATCAAC AAAACAGGTT 420
TTATGCTACT CCCGTCAATT AATTCTTGCA ATGCTTCAAT GCAAGCAAGC AGAAATACTT 480
CATCGTGATC TAAAACCTGC AAATTTATTC ATTACTCGTG ATAACGTTTT GAAGATCGGC 540
GATTTTGGCT TAGCTCGCAA ATTTAAGTCT GACGATATTA AGTATACATA TAATGTTATC 600
ACTTTATACT ACAGAGCACC AGAGTTAATC TTAGGCTGTC AAAATTATCA ATATGAAGTT 660
GATGTTTGGA GTGTTGGATG CATTTTATAT GAGCTTTGCA CCAACAAATA CTTTTTCAAG 720
GCCCTATAG GTAAAGAAAT AGATCAACTT ACTGCAATTT TCAAGATTAC AGGTACACCT 780
GACAATGAAG AGTGGCCAGA GTTCAAAGAG CTCGACAAAG GTGGTCTTTT TACTCAGAAG 840
ATAGAAGGCA ACCTCCTTGA ATACTTGGA AAGACAATTC CACCTGAATT CGAAGGTGCC 900
GCAGATTTAA TCTCAAAAAT GTGCAGGTTA ACACCATCAA AGAGAATTTT CATGCAGGAG 960
GCATTTATGC ATCCTTTCAT TTCACGTAAT GGCCAACCAA TTGAAGCTTC TCAGCTTCCT 1020
CCGATCAGTT TAATGGAAAT GCACCAAATG CAAGCATCAG CTGAGCTCGG CAAGAATAAG 1080
AAAGATAAAT CTGCTCACGA TTCCCCACCA AGACCTGAGA AAGCTGAAAT TAAgaatt 1140
atataatatt gtttttgtaa attattgtaa attttatatt gtgatacaca tttccaacac 1200

```

Figura 6. Secuencia nucleotídica de la zona de homología a proteínas tipo Cdk obtenida con el análisis por BLAST. El marco abierto de lectura identificado (mayúsculas) cuenta con 1080 pb, comprendidas entre un codón de inicio ATG (morado y subrayado) y un codón ocre de término TAA (rojo y subrayado). También se identificaron las secuencias consenso para elementos de transcripción: Promotor tipo Inr (azul y subrayado, sitio +1 de inicio de la transcripción marcado con una flecha superior); señal de poliadenilación (rojo y subrayado); dos posibles sitios de procesamiento (verde y subrayado). La secuencia no codificadora se muestra en minúsculas.

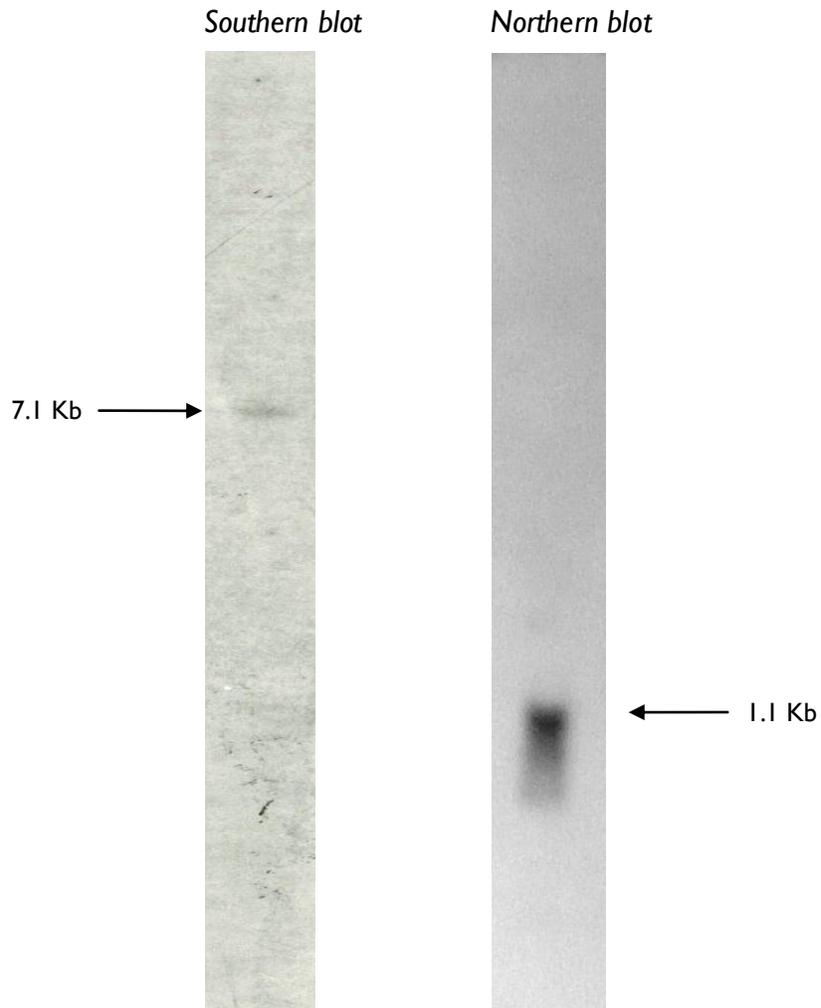


Figura 7. Se analizó experimentalmente el número de copias presentes en el genoma de *T. vaginalis* para el gen identificado, así como la expresión de su mRNA. *Izq.* Southern blot de 5 μ g de DNA genómico de *T. vaginalis* digerido con la enzima EcoRI. Se encontró una banda única de aproximadamente 7.1 Kb. *Der.* Northern blot de 5 μ g de RNA total de *T. vaginalis*. Se encontró una banda única de aproximadamente 1.1 Kb. (En ambos ensayos se utilizó una sonda construida con el fragmento de 865 pb descrito en el texto).

La señal que se encontró fue estimada como un mRNA de aproximadamente 1.1 Kb, lo que correlaciona con el tamaño esperado para el transcrito de *TvCDK* (**Fig. 7 Der**).

3. Análisis *in silico* de *TvCDK*

Nuestro análisis *in silico* comenzó en tres fases que serían determinantes para el método teórico desarrollado posteriormente:

1. Comparación de la secuencia peptídica de *TvCDK* y secuencias peptídicas de proteínas tipo Cdk reportadas en las bases de datos de NCBI. Se realizó la búsqueda por medio de la variante blastp de la herramienta *in silico* BLAST (Ver métodos). Se realizó un alineamiento de la secuencia peptídica de *TvCDK* con proteínas Cdk provenientes de diversos organismos (**Fig. 8**).

Se encontraron secciones de similitud entre *TvCDK* y las proteínas comparadas, encontrando diversas regiones que corresponden a sitios canónicos de protein cinasas o proteínas del tipo Cdk, por lo que se procedió a revisar los motivos conservados en *TvCDK* respecto a otras protein cinasas.

2. Búsqueda manual en la secuencia peptídica de *TvCDK* de la conservación de motivos para regiones específicas de cinasas, regiones características del subtipo cinasa serina/treonina y regiones únicas de proteínas tipo Cdk, tomando como referencia lo publicado por Hanks, 1988 (**Fig. 9**).

Se encontraron diversos motivos con una identidad del 100% así como otros motivos con una alta homología (motivos con residuos conservados) a los motivos reportados en el estudio mencionado. Se muestra la secuencia con los motivos seleccionados (**Fig. 9**).

3. Elaboración de modelos teóricos con el programa 3D PSSM Protein Fold Recognition (ver métodos) de la estructura terciaria que tendría *TvCDK* usando como templado la proteína CDK2 de humano (**Fig. 10**).

Se encontró un plegamiento conservado con los lóbulos característicos para las proteínas de tipo Cdk con un lóbulo N (región amino terminal) formado por hojas β , un lóbulo C (región carboxilo terminal) formada por α hélices, presencia de la estructura denominada *T-loop*, presencia de la estructura conocida como hélice PSTAIRE.

También se encontró una disposición similar a la usual para los tres aminoácidos reguladores más importantes de las proteínas de tipo Cdk, cuyos símiles se determinaron de acuerdo al análisis de alineamiento (ver arriba): El aminoácido T₂₈ como equivalente a T₁₄ y el aminoácido Y₂₉ como equivalente a Y₁₅ (sitios de fosforilación inhibitoria); el aminoácido T₁₇₉ como equivalente a T₁₆₀ (sitio de fosforilación activadora) (**Fig. 10**).

Tv	MDSSDDFSIETKDLSTLFTTKVEKICLGGTYGKVMKCLDIRKKQ...TVALKKIKILKPNDEFFVNTVREIKCLKELK.HDN...IIRLRNVITLYKPNKP	92
Eh	MTRYEKKQQLGEGTYGVVCKAWDT...VCNRYVALKKIKQEREDDCIPVTSVREIAVLLLELK.HPN...VVDLYDIYLEDK...	74
Gi	MSVSAAPSATSIDRYRRIITKLGEGTYGEVYKAITD...VTNETVALKRIIRLEHEEEEGVPGTAIREVSLKELQ.HRN...IIEELMSVIHHH...	85
Dd	MESDGGLSRYQKLEKLGEGTYGKVMKAK...EKATGRMVALKKIRLE...DDGVPSTALREISLLKEVP.HPN...VVSFLDVLHCQN...	78
Tt	MNNNTFGSDRYEKLKLVGEGTYGEVYKAKDIQ...SSEIVALKKIKLENEDEGVPSTALREISILKELQPHPN...IVCMHEVIYQPQEK...	84
Ca	MVELSDYQRQKVCBEGTYGVVYKALDTRKHNN...RVVALKKIRLESEDEGVPSTALREISLLKEMK.DDN...IVRLYDIHSDSH...	79
Sc	MSGELANYRRLKVEGEGTYGVVYKALDLRPGQQRVVALKKIRLESEDEGVPSTALREISLLKELK.DDN...IVRLYDIVHSDAH...	82
Sp	MENYQKVEKIGEGTYGVVYKAR...HKLGRIVAMKKIRLESEDEGVPSTALREISLLKEVN.DENNRNSCNVRLDILHAES...	78
Zm	MEQYKVEKIGEGTYGVVYKALD...KATNETIALKKIRLEQDEDEGVPSTALREISLLKEMN.HGN...IVRLHDVVHSEK...	74
Xl	MENFQKVEKIGEGTYGVVYKARN...RETGEIVALKKIRLDTETEGVPSTALREISLLKELN.HPN...IVKLLDVIHTEN...	74
Mm	MENFQKVEKIGEGTYGVVYKARN...KLTGEVVALKKIRLDTETEGVPSTALREISLLKELN.HPN...IVKLLDVIHTEN...	74
Hs	MENFQKVEKIGEGTYGVVYKARN...KLTGEVVALKKIRXDTETEGVPSTALREISLLKELN.HPN...IVKLLDVIHTEN...	74
Tv	STVWLSFDYCEFDLFGLLHKTDFP...QLSTKQVLCYSRQLLILAMLQCKQAEILHRDLKPAFLFITRDN...VLKIGDFGLARKPKSDDIKYTYNVIT	184
Eh	FLYLWFEEFCDEDLQYFMSRSSKIP...INETRISIVYQILQGLAFCHYHQLHRDMKPNILINKG...TKLGEFGLARLTTINDRKYTSEVVT	163
Gi	RLHLHFEYAENDLKKYMDKNPDVS...MRVKSFLYQLINGVNFCHSRRLHRDLKPNLLSVSDASETFVLRIGDFGLARAFGPIRQTHEIIT	179
Dd	RLYLWVEFLDQDLKKYMDSVPALC...PQLIKSYLYQLLGLAYSHGHRILHRDLKPNLLIDRQG...ALKLADFGLARAVSIPIVRYVTHEIVT	167
Tt	KLYLWVEFLVDODLKKFLDQYRKDKKLQLRPYQIKLMMYQILNGLNFCHSRRIIHRDLKPNLLIDAKG...NKIADFGLARAFGVPKITTTHEVET	178
Ca	KLYLWVEFLDLDLKKYMESIPQG...VGLGANMIKRFMNQLIRGIKCHSHRVLHRDLKPNLLIDKEG...NLKLADFGLARAFGVPPLRAYTHEVVT	171
Sc	KLYLWVEFLDLDLKKRYMEGIPKD...QPLGADIVKRFMMQLCKGIAYCHSHRILHRDLKPNLLINKDG...NLKLADFGLARAFGVPPLRAYTHEIVT	174
Sp	KLYLWVEFLDMDLKKYMDRISETGATSLDPRLVQKFTYQLVNGVNFCHSRRIIHRDLKPNLLIDKEG...NLKLADFGLARAFGVPPLRNYTHEIVT	172
Zm	RIYLWVEFLDLDLKKFMDSCPEFA...KNPTLIKSYLYQILHGVAYCHSHRVLHRDLKPNLLIDRRTN...ALKLADFGLARAFGVPVRTTHEVVT	166
Xl	KLYLWVEFLNODLKKFMDRSNISG...ISLALVKSYLEQLLQGLAFCHSHRVLHRDLKPNLLINSDG...ALKLADFGLARAFGVPVRTTHEVVT	165
Mm	KLYLWVEFLHODLKKFMDASALTG...IPLPLIKSYLYQLLQGLAFCHSHRVLHRDLKPNLLINAEG...SLKLADFGLARAFGVPVRTTHEVVT	165
Hs	KLYLWVEFLHODLKKFMDASALTG...IPLPLIKSYLYQLLQGLAFCHSHRVLHRDLKPNLLINTEG...ALKLADFGLARAFGVPVRTTHEVVT	165
Tv	LWYRAPEIILGCGNQYQYEVVWVSVGCILYELCTN...KYFFKAEIKKEIDQLTAIFK	238
Eh	LWYRAPEIILGATQYGGADINWSTAIFGELINKE...ELFKGRCKIDQLFKIFS	215
Gi	LWYRAPEIILGSRHYSTVSDIWSIACIWAEMLMKT...PLFPDSEIDQLFKIFE	231
Dd	LWYRAPEVLLGSKSYVVPVDMVSVGCIFGEMLN...KKPLFSGDCEIDQIFRIFR	219
Tt	LWYRAPEIILGQKAYSIGVDIWSLGCIFHEMVEKR...ALFMDGSEIDQIFKIFQ	230
Ca	LWYRAPEIILGKQYSTGVDMVSVGCIFAEMCN...RKPLFPDSEIDQIFRIFR	223
Sc	LWYRAPEVLLGKQYSTGVDTWSIGCIFAEMCN...RKPIFSGDSEIDQIFKIFR	226
Sp	LWYRAPEVLLGSRHYSTGVDIWSVGCIFAEMIR...RSLFPDSEIDEIFKIFQ	224
Zm	LWYRAPEIILGARQYSTPVDVWVSVGCIFAEM...VNQKPLFPDSEIDELFKIFR	218
Xl	LWYRAPEIILGCKFYSTAVDIWSLGCIFAEM...ITRRALFPDSEIDQLFRIFR	217
Mm	LWYRAPEIILGCKYYSTAVDIWSLGCIFAEMHVLVCTQHAKCGEHRNRGRHSLCPLCSYLEVAASQGGGMTAVSAPHVTRRALFPDSEIDQLFRIFR	265
Hs	LWYRAPEIILGCKYYSTAVDIWSLGCIFAEM...VTRRALFPDSEIDQLFRIFR	217
Tv	ITGTPDNEBWPEFKELDKGGLFTQKIEGNLLEYLEKTIPEPEFEG.AADLISKMCRLTTPSKRISMQBTFMHPFISRNGQPIEASQLPFIISLMEMHQMQASA	337
Eh	QIGTPTEDIWNGVTKLFFYLSTFPKWA...KDLHTIFHT.DER.AVDLLQKMFITPEKRISAADALKHPPFDPLNKPNN	291
Gi	VILGLPDDTTPGVTALPDWKQSFPPKFRG...KTLKRVLGALLDDEGLDLTAMLEMDFVKRISAKNALEHPYFSHNDFDP	308
Dd	VILGTPDDSWPGVTKLPEYVSTFPNWPFG...QPYNKIFRCEPL.ALDLTAKMLQYEPSKRISAKBALLHPYFGDLDTSSF	296
Tt	YHGTPTTEQWPAKCEPYFKPIYPRKFT...ADPKTYFRN.FCDKGFDTIQMIALDPAKRISVKDALRHPYFEDLSREDIAKFEPNQVHMY	318
Ca	ILGTPNEEITWPDVNYLPDFKSSFPQWKK...KPLSEAVPSLDAN.GIDLQMLVYDPSRISAKRALIHPYFNDDRRHNNYEDNIGIDKHQNMQ	317
Sc	VILGTPNEAIVPDIIVLPDFKPSFPQWRR...KDLQVVPVSLDPR.GIDLQMLVYDPSRISAKRALIHPYFNDDRRHNNYEDNIGIDKHQNMQ	298
Sp	VILGTPNEEIVWPGVTLQDVKSTFPWRKR...MDLHKVVPNGEED.AIELLSAMLVYDPAHRISAKRALQNYLRDFH	297
Zm	ITGTPNEQSWPGVSLPDFKTAFFPRQA...QDLATVVPNLDP.AGLDLSKMLRYEPSKRITARQALEHEYFKDLEVVQ	294
Xl	TILGTPDEVSWPGVTPMPDYKSTFPKWIR...QDFSKVVPLEDED.GRDLQAQMLQVDSNKRISAKVALTHFFFRDVSPTPHLI	297
Mm	TILGTPDEVWPGVTSMPDYKSPFKWAR...QDFSKVVPLEDED.GRSLLSQMLHYDPNKRISAKAALAHPPFQDVTKVPVPHLRL	346
Hs	TILGTPDEVWPGVTSMPDYKSPFKWAR...QDFSKVVPLEDED.GRSLLSQMLHYDPNKRISAKAALAHPPFQDVTKVPVPHLRL	298
Tv	ELGKNKKDKSAHDSPPRPEKAEI	360

Figura 8. Alineamiento de la secuencia peptídica de *TvCDK* y secuencias peptídicas de proteínas tipo Cdk reportadas en la base de datos de NCBI. El alineamiento fue generado por el programa DNAMAN (Lynnon BioSoft). El sombreado negro indica una identidad del 100% entre las secuencias, el rosa indica una similitud del $\geq 75\%$ y el azul una similitud del $\geq 50\%$.

Las proteínas utilizadas para este alineamiento son indicadas por medio de las siglas a la izquierda de cada renglón: *Tv*, indica la secuencia de estudio *TvCDK*. Los organismos a los que pertenecen las demás secuencias alineadas son (las referencias de NCBI se muestran entre paréntesis): *Eh*, *Entamoeba histolytica* (AAA51480); *Gi*, *Giardia intestinalis* (AAM14635); *Dd*, *Dictyostelium discoideum* (S24386); *Tt*, *Tetrahymena thermophila* (AAD46564); *Ca*, *Candida albicans* (CAA56338); *Sc*, *Saccharomyces cerevisiae* (TVB8); *Sp*, *Schizosaccharomyces pombe* (TVZP2); *Zm*, *Zea mays* (AAA33479); *Xl*, *Xenopus laevis* (P23437); *Mm*, *Mus musculus* (AAH05654); *Hs*, *Homo sapiens* (CAA43985).

```

MDSSDDFSIE TKDLSTLFTK VEKIGLGTYG KVYKCLDIRK KQTVALKKIK ILKPNDGFPV
NTVREIKCLK ELKHDNI IRL RNVITLYKPN KPSTVWLSFD YCEFDLFGLL HKTDFPQLST
KQVLCYSRQL ILAMLQCKQA EILHRDLKPA NLFITRDNVL KIGDFGLARK FKSDDIKYTY
NVITLYRAP ELILGCQNYQ YEVDVWSVGC ILYELCTNKY FFKAPIGKEI DQLTAIFKIT
GTPDNEEWPE FKELDKGGLF TQKIEGNLLE YLEKTIPPEF EGAADLISKM CRLTPSKRIS
MQEAFMHPFI SRNGQPIEAS QLPPISLMEM HQMQASAELG KNKKDKSAHD SPPRPEKAEI

```

Subdominio I	GXGXXGXV T₁₄ / Y₁₅	Sección de unión a ATP de protein cinasas Sitios de fosforilación inhibitoria
Subdominio III	EGVPSTAIRESILLKE PSTAIRES	Común en proteínas Cdk Región característica de proteínas Cdk
Subdominio VI	HRDLKPQNL D / N LKPQ	Sección de loop catalítico de proteínas Cdk Especificidad cinasa tipo serina/treonina Unión a ATP
Subdominio VII	DFG	Conservado en todas las protein cinasas
Subdominio VII	T₁₆₀	Sitio de fosforilación activadora
Subdominio VIII	G-T/S-X-X-Y/F-X-APE APE	Conservado en protein cinasas Indicador de dominio catalítico
Subdominio IX	DXWSXG	Conservado en proteínas Cdk
Subdominio X	GDSEID	Sección de contacto con CAK
Subdominio XI	R	Conservado en todas las protein cinasas

Figura 9. Análisis de la secuencia de aminoácidos correspondiente a *TvCDK*, por búsqueda manual de motivos conservados a través de los subdominios de una protein cinasa (de acuerdo al estudio realizado por Hanks). Los motivos idénticos u homólogos encontrados en *TvCDK* son marcados de acuerdo a la clave de colores mostrada en el anexo inferior de la imagen.

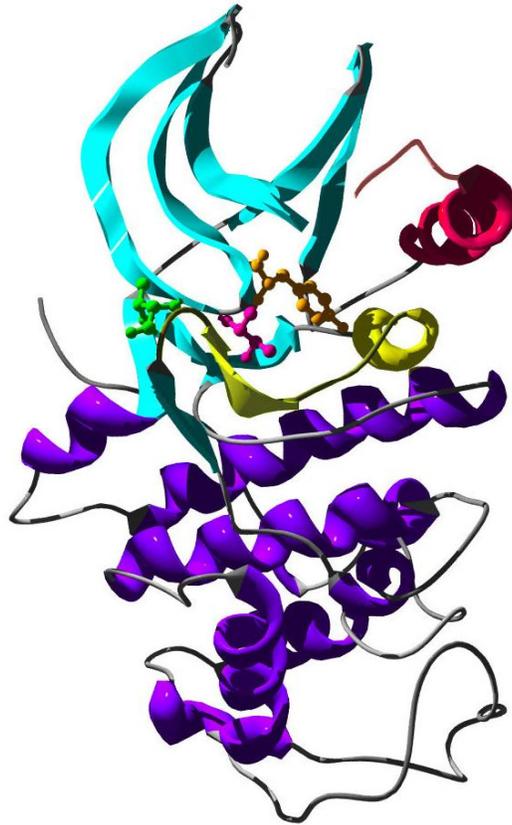


Figura 10. Modelo teórico de la proteína *TvCDK* elaborado por el programa *3D PSSM Protein Fold Recognition* tomando como referencia la configuración de la proteína CDK2 de humano. En este se muestran el lóbulo N formado por hojas β (azul), el lóbulo C formado por α hélices (morado), *T-loop* (amarillo), hélice *PSTAIRE* (rojo) y los equivalentes a los aminoácidos reguladores: T₂₈, Y₂₉ y T₁₇₉ (rosa, naranja y verde respectivamente).

4. Producción de una proteína recombinante

Los datos obtenidos de los primeros análisis *in silico* propusieron que la proteína codificada por este gen podría ser funcional, por lo que se propuso construir una proteína recombinante que permitiera evaluar su funcionalidad como proteína del tipo Cdk.

Se generó un plásmido recombinante por clonación del marco abierto de lectura (ORF) del gen *TvCDK* en el vector *pQE30* (Ver sección de métodos). Este plásmido recombinante expresa a *TvCDK* con una cola de histidinas anexada en el extremo amino terminal. El resultado es un inicio de la serie peptídica MRGSHHHHHHGSM donde la M subrayada marca la metionina iniciadora del ORF de *TvCDK*. La proteína resultante fue nombrada 'His-CDK' (Fig. 11).

La secuencia de la clona fue corroborada por secuenciación, indicando que los nucleótidos codificadores para la metionina iniciadora de la proteína recombinante, las seis histidinas y la metionina iniciadora de la secuencia original de *TvCDK* se encuentran clonados en fase.

Se realizó un ensayo de inducción de la proteína recombinante, exponiendo el cultivo de *E. coli* a dos diferentes concentraciones de IPTG durante intervalos de tiempo establecidos. Este ensayo mostró la inducción progresiva de una banda con un peso molecular de aproximadamente 42 KDa (esperado) iniciando desde el punto de medición de 0.5 horas y hasta el punto final de 5 horas, sin observarse diferencias significativas entre la inducción a cada concentración de IPTG.

Asimismo, la proteína recombinante *His-CDK* no resultó tóxica para la bacteria en ningún punto a lo largo del tiempo de inducción que se realizó, ya que los cultivos se mantuvieron en crecimiento a lo largo de los ensayos (**Fig. 12**).

El peso molecular de 42 KDa concuerda con el calculado para *His-CDK*, pues *TvCDK* tendría un peso aproximado de 41.4 KDa lo que ascendería a un estimado de 42.8 KDa al añadir la secuencia amino anexada.

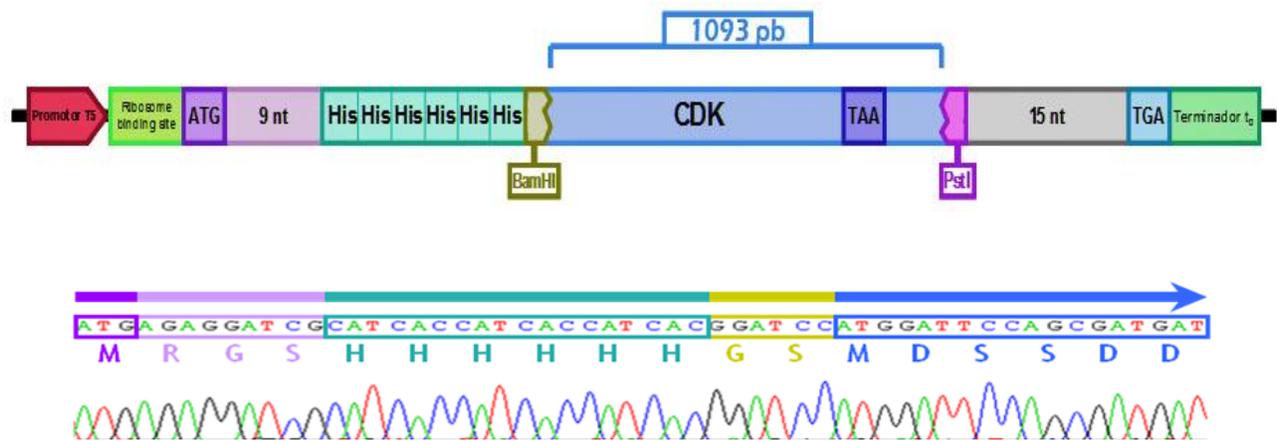


Figura 11. Diagrama de la construcción del plásmido recombinante y secuenciación. Se generó un plásmido recombinante por clonación del marco abierto de lectura (ORF) del gen *TvCDK* en el vector *pQE30*. Este plásmido recombinante expresa a *TvCDK* con una cola de histidinas anexada en el extremo amino terminal. *Arriba*. En este diagrama se aprecian: codón de inicio de la traducción para la proteína recombinante; sección codificante para la cola de 6 histidinas; así como el inserto con extremos BamHI/PstI de *TvCDK*, que incluye al codón original de término de la traducción de *TvCDK*. *Abajo*. La secuencia de la clona fue corroborada por secuenciación, indicando que los nucleótidos codificadores para la metionina iniciadora de la proteína recombinante, las seis histidinas y la metionina iniciadora de la secuencia original de *TvCDK* se encuentran clonados en fase.

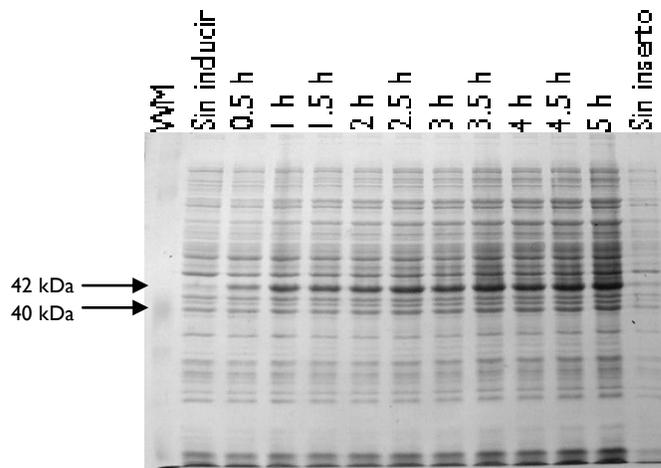
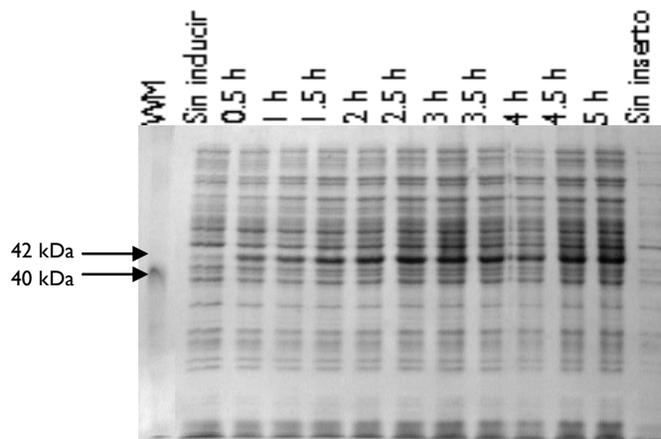


Figura 12. Geles de poliacrilamida con extractos proteicos totales de cultivos de *E. coli* con el plásmido recombinante para *His-CDK*. *Arriba*. Curva de inducción a concentración de 0.1 mM de IPTG. *Abajo*. Curva de inducción a concentración de 1 mM de IPTG. La banda que se intensifica a lo largo del tiempo de los ensayos es la correspondiente a *His-CDK* con un peso estimado de 42 KDa.

5. Unión de TvCDK a $p13suc^{+1}$

De acuerdo a lo estudiado como mecanismos regulatorios de las proteínas Cdk, la unión a una molécula pequeña denominada *Suc1* (*Cks1* en *S. cerevisiae*) es importante para la correcta progresión del ciclo celular y específicamente, para la entrada a mitosis. Por esta razón, se ha descrito un método experimental que utiliza esta característica para determinar la funcionalidad de una proteína hipotética de tipo Cdk.

El método experimental consiste en poner en interacción un extracto proteico total o un concentrado de proteína purificada y proteína $p13suc^{+1}$ conjugada a agarosa. La suspensión resultante es centrifugada y sólo aquellas proteínas capaces de unirse a $p13suc^{+1}$ son recuperadas. En este caso, se realizó este ensayo utilizando extractos proteicos totales procedentes de cultivos de *E. coli* transfectados con *pTvCDK*. Como control se utilizó un cultivo sin inducir con IPTG y como experimental un cultivo inducido por 0.1 mM de IPTG. Este ensayo reveló la unión de una proteína con un peso molecular de aproximadamente 42 kDa a la proteína $p13suc^{+1}$ (**Fig. 13**).

Para comprobar la identidad de la proteína que se une a $p13suc^{+1}$ en este ensayo, se determinó la secuencia proteica de la banda de 42 kDa por medio de un análisis de espectrometría de masas elaborada en el Laboratorio a cargo del Dr. Guillermo Mendoza en la Facultad de Medicina. En este análisis la secuencia obtenida fue idéntica a la secuencia de la proteína de interés *TvCDK*.

6. Análisis del genoma liberado por TIGR

En Enero de 2007 The Institute for Genomic Research (TIGR), que ahora forma parte del J. Craig Venter Institute, publicó un borrador inicial del genoma de *T. vaginalis*⁵⁴. Se realizó una búsqueda en este borrador a través de su sitio de Internet utilizando la

herramienta BLAST en su variante *blastp* (ver métodos) para determinar las secuencias contenidas en el genoma de *T. vaginalis* que representaran proteínas putativas con homología a *TvCDK*.

Las secuencias de aminoácidos resultantes de esta búsqueda por *blastp* fueron 813, cuya similitud variaba significativamente entre sí. Por esta razón, se eligió a las primeras 32 secuencias que mostraban valores de homología con una probabilidad hasta $1e^{-35}$ (este parámetro indica la probabilidad de que la homología encontrada sea explicada por el azar, por lo que a menor valor el resultado se considera más significativo).

De estas 32 secuencias, la primera (número identificador 88601.m00465) corresponde al fragmento genómico de *TvCDK* ya que muestra una identidad del 100%. La secuencia correspondiente a *TvCDK* y las 31 secuencias restantes fueron analizadas utilizando PROSITE (<http://expasy.org/prosite/>), una base de datos de dominios proteicos, sitios funcionales y motivos de aminoácidos que provee una herramienta de detección de motivos de proteínas. En estas secuencias se encontraron cuatro tipos de motivos:

- a) Dominio de unión a ATP de protein cinasas (Número de motivo PROSITE PS00107).
- b) Sitio activo de protein cinasas tipo serina/treonina (Número de motivo PROSITE PS00108).
- c) Sitio característico de MAP cinasas (Número de motivo PROSITE PS01351).
- d) Sitio característico de proteína ribosomal S2 (Número de motivo PROSITE PS00962).

Se realizó una selección de aquellas secuencias proteicas que mostraran un patrón coincidente de los motivos a) y b) mostrados arriba, por ser éstos motivos encontrados comúnmente en las proteínas del tipo Cdk; también se consideró que las secuencias

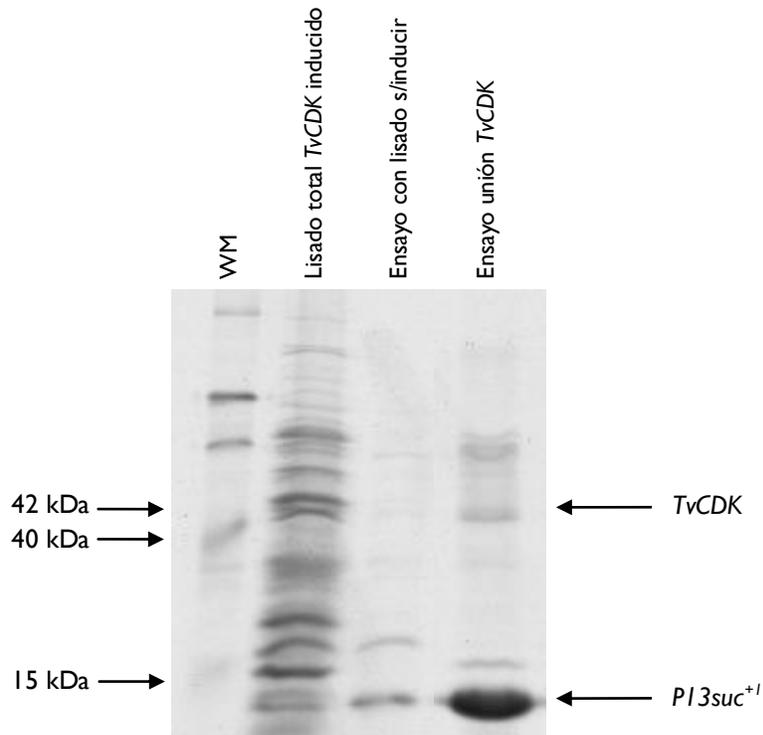


Figura 13. Gel de poliacrilamida del ensayo de unión a proteína $p13suc^{+1}$ conjugada a perlas de agarosa. El segundo carril muestra el extracto proteico total de un cultivo de *E. coli* transfectado con el plásmido recombinante para *His-CDK* e inducido con 0.1 mM de IPTG por 1 hora. El tercer carril corresponde al ensayo de unión utilizando un cultivo sin inducir con IPTG. El último carril muestra el ensayo de unión utilizando el cultivo inducido con IPTG; se resalta la banda de aproximadamente 42 kDa correspondiente a *TvCDK* y la banda que corresponde a $p13suc^{+1}$.

Tabla 3. Selección de secuencias proteicas por los motivos conservados que se encontraron por medio del análisis con PROSITE

	ID secuencia	ATP	S/T	MAPK	RIBO S2	Selección
1	88601.m00465	IGLGTYGKVKCLD IRKKQTVALK	ILHRDLKPANLFI			<i>TvCDK</i>
2	89123.m00090	IGDGTYSVQKCKD ERTNEIVALK	VVHRDLKPANMFI			✓
3	88200.m00230	IGKGTYATVSKCID KATNKPVALKAIKK	VIHRDLKPANLFI			✓
4	82761.m00463	VGKGYAVSKAKM KGTNQIVALKKVSR DQIAMK	IYHRDIKPSNIYV			✓
5	92757.m00246		ILHRDLKPANLLV			
6	87854.m00199	LGEGTYGVVFKAID KRTNQVVVALK	IIHRDMKPQNLLI			✓
7	86244.m00029		IVHRDLKPANIFI			
8	87414.m00372		IVHRDLKPANIFV			
9	87835.m00290		IIHRDLKPSNIFV			
10	95940.m00220	LGEGTYGVVYRAVD NRTGEIVALK	IMHRDMKPQNLLI			✓
11	91895.m00068	IGTGTYSVYKAVD LDQKQIVALK	FVHRDVKPANILI			✓
12	82761.m00542	LGEGTYGAVYKARN KTTGEILAMK	IIHRDMKPQNLLL			✓
13	97324.m00300	LGEGSYGVVYRATD TRDKSTVALK	IIHRDIKPDNLLL			✓
14	97466.m00043	IGVGAYGSVFKAFD LETNEVVALK	YAHRLKSSNILI			✓
15	86498.m00569	IGKGSNSTVFKVIE KRTKKVLALK	VLHRDIKPENIMI			✓
16	85889.m00444	IGQGTYSVYKAQN TQNKEIVAIK	IIHRDVKPQNILL			✓
17	82736.m00203	IGQGAYGSVYKALD LLNKKIVALK	FAHRDIKPANILL			✓
18	88273.m00284	LGAGTYGMVFKAKD KTTGKIVAIK	VIHRDIKPDNLLI			✓
19	87150.m00250	LGEGTYSVFRATH IPTDQPVVLK	IVHRDIKPENILI			✓
20	81864.m00296	LGRGGFGIVYIAED TVTKETLALK	IMHRDIKPDNVLF			✓
21	84685.m00003					

22	92043.m00144		YLHRDIKPENLLI			
23	84511.m00298	IGDGGFGVVTKMRN IATGQIVAMK	FFHRDLKPENLVF			✓
24	96281.m00300					
25	96252.m00271	IGKGTYGTVYSIAIW TSTNEVVAIKK	IIHRDIKPQNLLI			✓
26	91562.m00103	LGEGSFSEVFKVKS MKNQQFYAIK	LFHRDIKPENCFI			✓
27	88914.m00152	VGKGYGIVVRAQD RVTGQLVALK	VIHRDLKPSNMLI	FQNVDAQRTYREITLLRQLKSHPFIV GLLQVHRAENDNDIYLAFEFMDTDVHT VIHAGILLDVHQRYIFWQLLCALKFIH SAGVIHRDLKPSNMLIKSDS		
28	94102.m00175	VGDGAFGTVLKCKN RETGQLVAIKK				
29	108748.m00003	LGDGAFGVVTKCRD KETGEIVAIKK			ASACIMA ELYMG	
30	81864.m00329	IGKGYGVVWRCIN KQNGQVVALK	VIHRDLKPSNLLI			✓
31	86673.m00415	IGKGYGVVWKAMD KVTHQMVALK	LIHRDLKPSNLLI	FRNSTDAQRTFREITFLRQLRGHPHIV DLLAVYKAVNNLDLYVVFELLES DVHS VIRANILMDVHKRFIVWQVLVALKYLH SRRLIHRDLKPSNLLINS DS		
32	94734.m00162	IGQGYGVVWRVLN KQTQQVVALK	IIHRDLKPSNLLI			✓

Tabla 3. Motivos presentes en las secuencias seleccionadas del genoma de TIGR. En esta tabla se muestran las 32 secuencias analizadas con PROSITE. En la columna ‘ID secuencia’ se muestra el número identificador de la secuencia de acuerdo a TIGR. En las columnas adyacentes se muestran las regiones que tienen homología a los motivos encontrados por PROSITE: ‘ATP’, homología con el motivo PS00107 de dominio de unión a ATP de protein cinasas; ‘S/T’, homología con el motivo PS00108 de sitio activo de protein cinasas tipo serina/treonina; ‘MAPK’, homología con el motivo PS01351 de sitio característico de MAP cinasas; y ‘RIBO S2’, homología con el motivo PS00962 de sitio característico de proteína ribosomal S2. La columna ‘Selección’ indica si esa secuencia fue seleccionada para los análisis *in silico* posteriores.

seleccionadas no contuvieran los motivos c) y d), por considerarse que dichos motivos son encontrados en proteínas de otros tipos no relacionadas con Cdk (**Tabla 3**).

La selección final fue de 20 secuencias más la secuencia correspondiente a *TvCDK*.

7. Método de análisis teórico

Para analizar las secuencias seleccionadas por medio del análisis descrito arriba, se revisó la literatura en búsqueda de motivos proteicos que se encuentren conservados a través de la familia de proteínas de tipo Cdk y en los estudios de caracterización de los dominios necesarios para su funcionalidad, éstos últimos consisten en estudios de mutación dirigida y cuantificación de la unión de la proteína mutada a proteínas relevantes como ciclinas y *p13^{suc+1}*.

Se creó una tabla con la información y se le asignó un valor numérico a la presencia de los sitios incluidos en esa tabla, de acuerdo a la relevancia reportada en los artículos que se tomaron como base para la tabla. Esta tabla de valorización toma en cuenta además que los motivos hayan sido reportados como sitios característicos a través de los estudios de revisión de proteínas del tipo Cdk (**Tabla 4**).

Como control positivo para comprobar la funcionalidad del método de valorización, se analizaron con éste secuencias correspondientes a proteínas Cdk que ya han sido reportadas en diversos organismos, obteniendo así un rango de valores de 51 a 72. Como control negativo se analizaron las secuencias de la proteína MAPK de ratón, la cual no tiene actividad alguna como Cdk (valor obtenido de 46) y la proteína CDK7 de humano, la cual no tiene una actividad en el ciclo celular, sino forma parte de un complejo proteico de la maquinaria transcripcional (valor obtenido de 43) (**Tabla 5**).

Con este método de valorización se analizaron las 21 secuencias resultantes de la selección explicada arriba. Es importante mencionar que el valor de los residuos presentes se contabilizó tomando en cuenta si el residuo era idéntico al reportado, en cuyo caso se contabilizaba con el valor mostrado en la tabla de valorización; o si el residuo era solamente conservado (presencia de residuos de aminoácidos con características bioquímicas similares), en cuyo caso se contabilizaba con la mitad del valor mostrado en la tabla de valorización (**Tabla 6**).

Se realizó una segunda selección de secuencias, tomando sólo aquellas que mostraran un valor asignado mayor a 50 de acuerdo al método de valorización de motivos conservados. Adicionalmente, se eliminó de la selección a una de las secuencias que sí mostraban un valor mayor a 50 (número identificador 88273.m00284), debido a que ésta carecía de un sitio con similitud al motivo característico PSTAIRE. Esta selección consideró así a 7 secuencias además de la secuencia correspondiente a *TvCDK*. Un alineamiento de las 8 secuencias proteicas del genoma de *T. vaginalis* que fueron seleccionadas muestra las regiones encontradas de homología entre ellas (**Fig. 14**).

Se encontró que los controles negativos mostraban algunos de los motivos conservados para protein cinasas, sin embargo, su puntuación es menor a 50 ya que carecen de los sitios característicos de una proteína del tipo Cdk. También se encontró que en todos los casos las secuencias de proteínas Cdk utilizadas como control positivo mostraban un valor mayor al umbral establecido de 50.

8. Elaboración de modelos tridimensionales

Como análisis complementarios a las secuencias seleccionadas del genoma de TIGR por medio de los análisis *in silico* descritos y el método teórico desarrollado de valorización de aminoácidos, se elaboraron modelos teóricos con el programa 3D PSSM Protein Fold Recognition (ver métodos) de la estructura terciaria que tendrían estas secuencias si se

comprobara su existencia a nivel proteico, similar al modelo desarrollado anteriormente para *TvCDK* usando como templado la proteína CDK2 de humano (**Fig. 15**).

Con estos análisis, se pudo observar que la posible conformación que tendrían las proteínas seleccionadas sigue un patrón de plegamiento como el esperado para las proteínas tipo Cdk y que fue anteriormente detallado en el análisis a *TvCDK*.

Domínios funcionales	Motivos	Ducommun MolCellBiol 91 p13suc1 binding (%)	Ducommun EMBO J 91 cyclin A binding (%)	Marcote MolCellBiol 93 p13suc1 & cyclin (clusters or alone)	Marcote MolCellBiol 93 PK conserved	Hanks Science 88 PK family	Harper ChemRev 01 CDK review	Sitios característicos Cdk
	E ₁₁ K ₆	A - 104	A - 18	cyc - cluster	+			2+1 3
	G ₁₁					+		1+1 2
Domínio de unión a ATP	E ₁₂	A - 104	A - 18	cyc - cluster				2+1 3
	G ₁₃				+	+		1+1 2
[GXGXGXV]	Y ₁₅	F - 117, Q.E - 7	F - 97	phospho	+	+		1+1+1+2 6 5
	G ₁₆				+	-		1+1 2 2
	K ₂₀	A - 7 cluster R22H23K	A - 134	suc - alone				1 1 1
	R ₂₂ H ₂₃ K ₂₄	A - 7	A - 134	suc - cluster				1 1 1
	K ₃₃	A - 68 cluster K33K34F	A - 13	cyc - alone	+	+		2+2+1+1+2 8 4
	K ₃₄ R ₃₆	A - 68	A - 13	cyc - cluster				2+1 3 3
	E ₃₅ E ₄₀ E ₄₁	A - 104	A - 13	cyc - cluster				2+1 3 3
	E ₄₂ G ₄₃	A - 104	A - 13	cyc - alone				2+2 4 4
Hélice PSTAIRE [EGV PSTAIRESLIKE]	PSTAR ₅₉ E	A - 95	A - 0	cyc - R alone, RE cluster	E+	E+	+	Halmark de CDK 2+2+1+1+2 8 5
	K ₆₇ E ₆₇ R ₆₉	A - 5	A - 75	suc - cluster				1+1 2 2
	H ₆₉	A - 5	A - 75	suc - alone	+			1+1+1+1 4 4
	D ₇₈				+	+		1+1 2 2
	N ₁₃₃				+	+		1+1 2 2
	D ₁₄₆ FG				D+	all 3	+	1+1+2 4 4
	R ₁₅₁	A - 74		cyc - alone				2+1 3 3
T-loop (con dt.12)	T ₁₆₁	V - 87, D - 99, E - 101	E - 9, V - 5	phospho		+	+	Fosforilación activadora CAK 2+1+1+2 6 5
	R ₇₀	A - 10	A - 110	suc - cluster				1 1 1
	APE ₁₇₃	A - 10	A - 110	suc - E cluster	E+	A-E+		1+1+2 4 4
	R ₁₈₀	A - 10	A - 105	suc - cluster				1 1 1
	D ₁₈₆	A - 10	A - 105	suc - alone	+			1+1+1 3 3
	R ₂₁₈	A - 3	A - 78	suc - alone, cluster				1+1+1 3 3
	D ₂₇₁ K ₂₇₄ R ₂₇₅	A - 10	A - 101	suc - cluster	R+			1+1 2 2

Tabla 4. Esquema de valorización de aminoácidos relevantes para la funcionalidad de una proteína del tipo Cdk. Se tomaron en cuenta estudios de mutación dirigida (primeras tres columnas en gris), en las que se valoró que la unión de la CDK estudiada a p13suc⁺, a ciclina o a ambas hubiera disminuido significativamente al ser mutado el residuo o motivo en cuestión.

En las siguientes tres columnas se valoró que estos motivos se encontraran reportados en los estudios de revisión de los dominios conservados a través de la superfamilia de protein cinasas y de la familia de proteínas tipo Cdk. Se contabilizó también la presencia de tres sitios canónicos para las proteínas tipo Cdk: Los residuos de fosforilación regulatoria y el motivo PSTAIRE.

En las columnas del lado derecho se muestra consecutivamente el valor que se asignó a cada uno de estos parámetros y finalmente, el valor asignado a la conservación de un motivo en una secuencia de interés.

K ₂₀	R ₂₀ H ₁₂ K ₂₄	R ₁₇₀	R ₁₈₀	4(2)	G ₁₁	G ₁₃	G ₁₆	K ₁₆ E ₁₇ R ₁₈	D ₂₂₈	N ₁₅₅	D ₂₇₁ K ₂₇₂ R ₂₇₅	7	E ₁ K ₂ E ₂	K ₁₉ R ₁₈	E ₁₈ E ₁₉ E ₁₁	R ₁₅₁	D ₁₉₈	R ₂₁₈ R ₂₁₈	7	K ₁₉	E ₁ G ₁₃ H ₁₆	D ₁₄₈ FG	APE ₁₇₃	5	Y ₁₅	P ₅₁ T ₁₉₁ R ₁₉₁ E	T ₁₉₁	3	Val ₁₀₁												
88601.m00465	K	LDI	R	N	2	2.0	G	G	KEK	D	N	TKR	5	12	12.0	EK	L	KK	LPN	R	D	AK	3	1	10.5	K	DG	H	DFG	APE	4	11	18.0	Y	PONTI	T	2	11	12.5	55.0	SI
89123.m00990	K	KDE	R	E	2	2.0	G	G	KQR	D	N	NKR	5	11	11.0	EK	D	KK	ION	R	D	EK	3	2	12.0	K	DG	H	DFG	PPE	3	2	16.0	Y	PONTI	S	1	2	10.0	51.0	SI
89200.m00230	K	IDK	R	E	2	2.0	G	G	QKN	D	N	NKR	4	12	10.0	EN	K	AK	VDE	R	D	RK	2	1	7.5	K	DG	H	DFG	PPE	3	2	16.0	Y	TKTIL	T	2	10.0	45.5	NO	
82761.m00463	K	KWK	R	H	2	2.0	G	G	RAS	D	N	NER	5	10	10.0	ET	K	KS	DIA	R	D	SS	2	2	6.0	MG	N	DFG	APE	3	12.0	Y	PENAW	P	1	1	7.5	37.5	NO		
87854.m0189	K	IDK	R	F	2	2.0	G	G	KEH	D	N	DKR	6	11	13.0	EK	E	RR	DEE	R	D	SK	4	2	15.0	K	EG	H	DFG	PPE	4	11	18.0	Y	PPTSI	T	2	11	12.5	60.5	SI
95940.m00220	R	VDN	R	L	1	1.5	G	G	RNN	D	N	NKR	5	11	11.0	EK	E	QR	EEE	R	D	KR	5	1	16.5	K	EG	H	DFG	PVE	4	11	18.0	Y	PVTAL	T	2	11	12.5	59.5	SI
91895.m0088	K	MDL	R	N	2	2.0	G	G	QTS	D	N	NKR	5	11	11.0	EK	T	VK	FOK	R	D	KR	3	1	10.5	K	ES	H	DFG	APE	4	16.0	Y	PLSPY	T	2	11	12.5	52.0	SI	
82761.m00542	K	RNK	R	A	2	2.0	G	G	SES	D	N	DKR	6	12	12.0	EK	E	VH	EEE	R	D	TK	5	15.0	K	EG	H	DFG	APE	5	20.0	Y	PPTSV	T	2	11	12.5	61.5	SI		
97324.m00300	R	TDT	K	-	2	1.0	G	G	KTN	D	N	DKR	6	12	12.0	EK	E	IK	DLE	R	D	KS	4	1	13.5	K	DG	H	DFG	APE	4	11	18.0	Y	PSTLR	T	2	11	12.5	57.0	SI
97466.m00043	K	FDL	R	D	2	2.0	G	G	QNK	D	N	NKR	5	10	10.0	EN	V	IK	YQK	R	D	KR	2	1	7.5	K	ES	H	DFG	APE	4	16.0	Y	PLSPY	T	2	11	12.5	48.0	NO	
86498.m00569	K	IEK	R	Y	2	2.0	G	G	S	KDN	D	N	NKR	4	8	8.0	EK	K	KK	RKT	F	D	KS	2	1	7.5	K	OS	H	DFG	PPE	2	11	10.0	N	PDFEIA	S	1	2.5	30.0	NO
85889.m00444	K	QNT	R	N	2	2.0	G	G	KEK	D	N	DKR	6	11	13.0	NR	Q	RK	EOE	R	D	KV	2	1	7.5	K	EG	H	DFG	PPE	4	11	18.0	Y	PS7AR	S	2	11	12.5	53.0	SI
82736.m00203	K	LDL	R	D	2	2.0	G	G	QSK	D	N	NKR	5	11	11.0	EN	Q	TK	FHK	K	D	KQ	1	1	4.5	K	OS	D	DLG	APE	3	12.0	Y	PVSFF	T	2	11	12.5	42.0	NO	
88273.m00284	K	KDK	R	I	2	2.0	G	G	KKN	D	N	DKR	6	12	12.0	EK	A	EI	DEE	R	D	RS	3	1	10.5	K	EG	H	DFG	PPE	4	11	18.0	Y	SSITM	T	2	10.0	52.5	SI	
87150.m00250	R	THI	L	A	1	0.5	G	G	KSN	D	N	NHR	5	10	10.0	MK	E	LR	DEE	A	D	KE	2	1	7.5	K	DG	H	DFG	APE	4	11	18.0	Y	PPSSV	I	1	1	7.5	43.5	NO
81864.m00296	I	EDT	R	N	1	1.0	G	G	NEQ	D	N	DKR	6	12	12.0	KK	R	KM	RPK	C	D	OK	1	3.0	K	DG	H	DFG	APE	4	11	18.0	F	EYHAE	I	1	0.0	34.0	NO		
84511.m00298	K	RNI	R	F	2	2.0	G	G	RKK	D	N	DKR	6	12	12.0	CS	D	RK	KQS	R	D	KA	2	2	9.0	K	-F	H	DFG	APE	4	16.0	Y	ECLLE	T	1	5.0	44.0	NO		
96252.m00271	S	IMT	R	K	1	1.0	G	G	NNK	D	N	NR	5	10	10.0	TK	K	KK	ESN	R	D	EF	2	1	7.5	K	EG	H	DFG	PPE	4	11	18.0	Y	PCTSL	S	1	2	10.0	46.5	NO
91562.m0103	K	KSM	R	S	2	2.0	G	G	SAE	D	N	DKR	4	8	8.0	NK	E	IMK	RRS	R	D	RN	3	9.0	K	-V	N	DFG	APE	3	12.0	F	EV7RL	T	1	5.0	36.0	NO			
81864.m00329	R	INK	R	S	1	1.5	G	G	RQK	D	N	DKR	5	10	10.0	QK	K	KF	AGN	R	D	RA	2	6.0	K	LT	C	DFG	APE	3	12.0	Y	AO7TY	T	2	10.0	39.5	NO			
94734.m001627	R	LNK	R	S	1	1.5	G	G	RQQ	D	N	DER	5	10	10.0	KK	Q	KF	AQN	R	D	RS	2	6.0	K	VT	H	DFG	APE	4	16.0	Y	AL7TY	T	2	10.0	43.5	NO			

Tabla 6. Asignación de valores a las secuencias previamente seleccionadas del genoma de *T. vaginalis*, por medio del método desarrollado para valorización de aminoácidos relevantes en una proteína del tipo Cdk.

Como encabezado se muestran los motivos que se analizaron y como columna izquierda y derecha se muestran los identificadores asignados por TIGR de las secuencias analizadas (en azul se recalcan las secuencias que obtuvieron un valor mayor a 50).

Los motivos fueron organizados de acuerdo a la relevancia asignada, donde el color amarillo claro muestra una relevancia asignada de 1, el color amarillo oscuro una relevancia asignada de 2, el color naranja claro una relevancia asignada de 3, el color naranja oscuro una relevancia asignada de 4 y por último, el color rojo una relevancia asignada de 5.

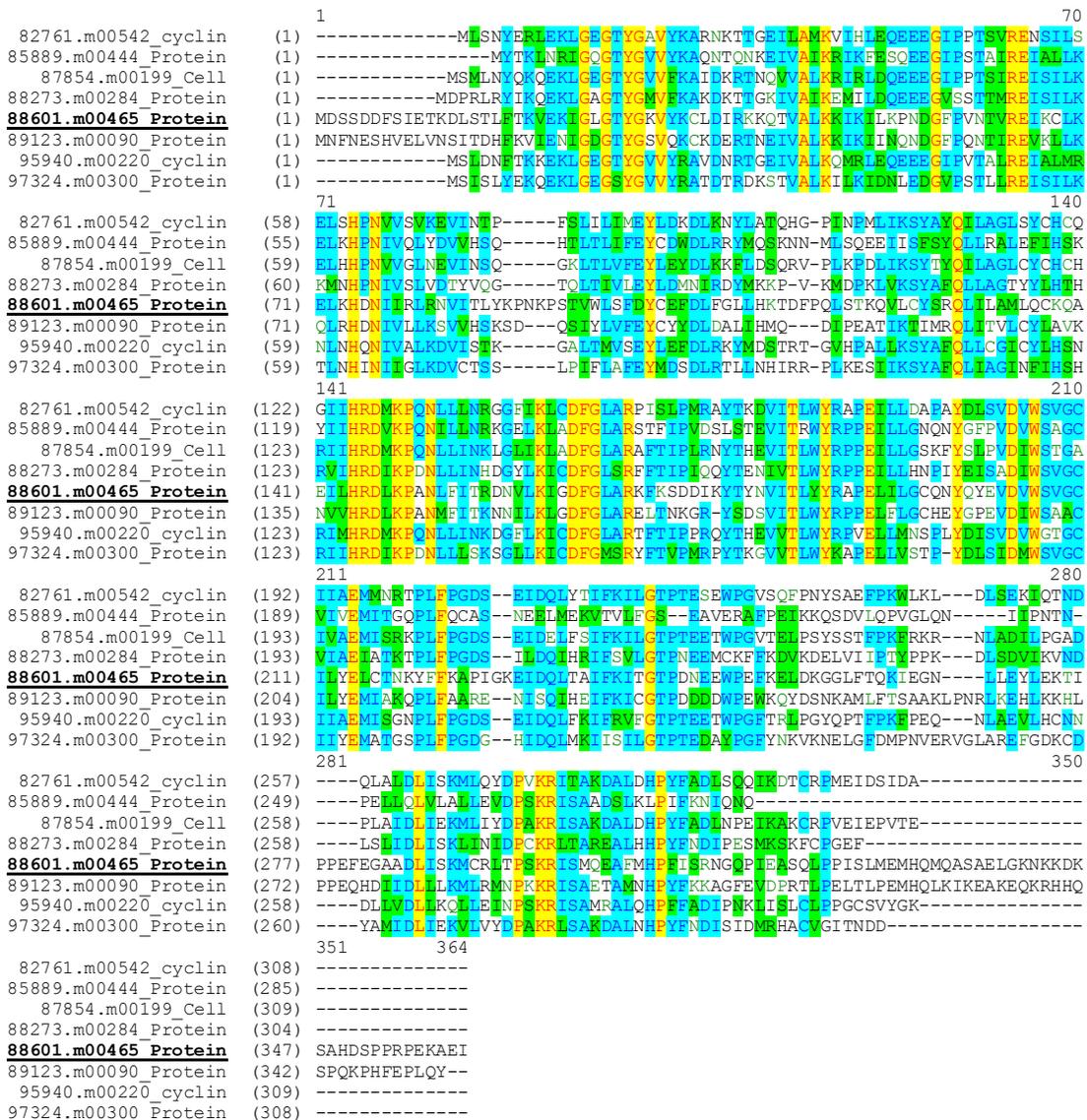


Figura 14. Alineamiento de la secuencia peptídica de *TvCDK* (secuencia 88601.m00465, se muestra subrayada y en negritas) y 7 secuencias peptídicas del genoma de TIGR seleccionadas por el método teórico desarrollado. El alineamiento fue generado por el programa AlignX (Invitrogen Corporation). El sombreado amarillo con letras rojas indica una identidad del 100%, el azul indica un cambio conservado y el verde una similitud de la región alineada.

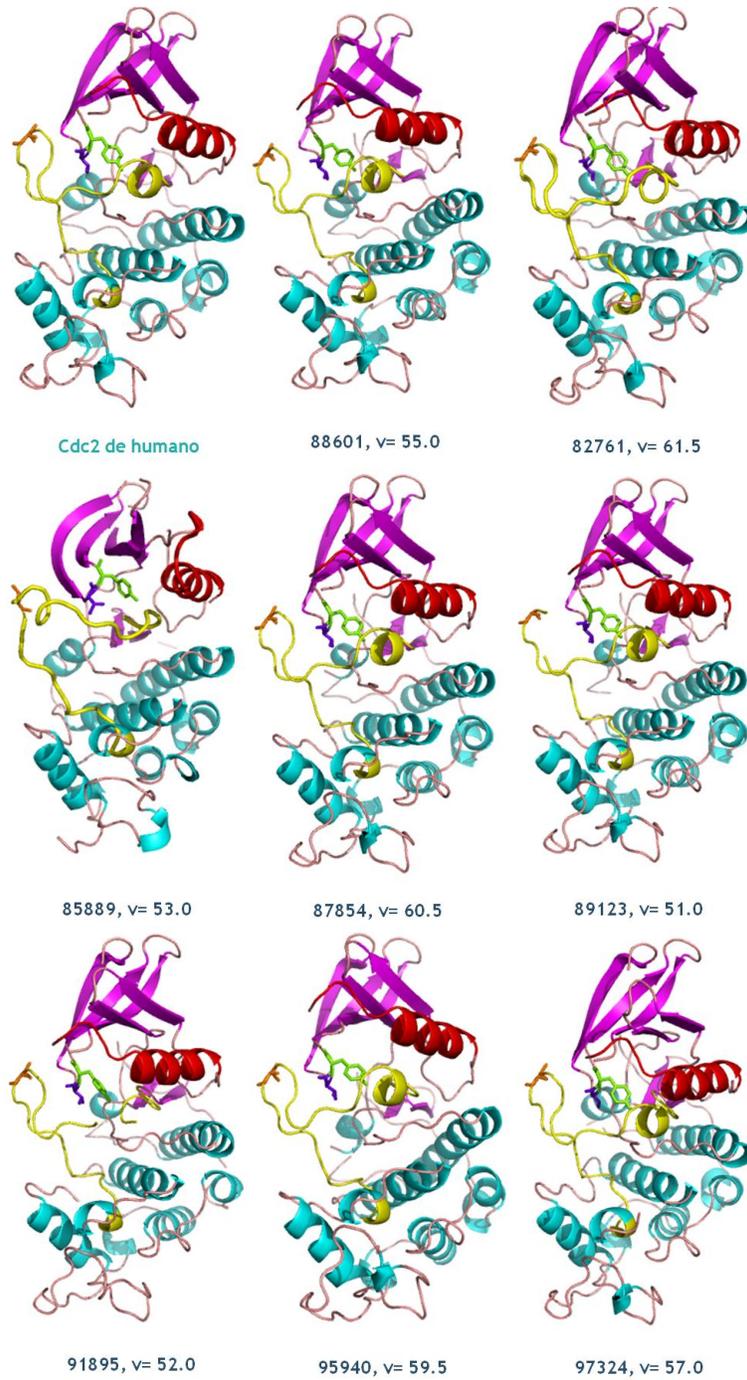


Figura 15. Modelos proteicos de las secuencias seleccionadas del genoma de *T. vaginalis* por medio del método teórico desarrollado. También aparecen las estructuras de Cdc2 de humano y el modelo para *TvCDK*. Se muestran el lóbulo N (violeta), el lóbulo C (azul), *T-loop* (amarillo), hélice *PSTAIRE* (rojo) y los equivalentes a los aminoácidos reguladores: T₂₈, Y₂₉ y T₁₇₉ (morado, verde y naranja respectivamente).

IX. DISCUSIÓN

El análisis de la secuencia nucleotídica del gen contenido en la clona genómica *pgactin10* mostró elementos conservados que podrían ser funcionales para la transcripción de un RNA mensajero de aproximadamente 1.1 Kb (tomando en cuenta el tamaño desde el sitio +1 hasta alguno de los sitios posibles de corte del transcrito).

Posteriormente, la existencia de este transcrito fue comprobada a través de un ensayo de northern blot, lo que indica que éste es un gen que se expresa a nivel de mRNA. Asimismo, por medio de un ensayo de Southern blot se observó que este gen no guarda amplias similitudes con otras regiones genómicas de *T. vaginalis* ya que se encontró una sola banda de aproximadamente 7.1 Kb, tamaño correspondiente a la clona genómica *pgactin10*. Estos hallazgos nos permiten pensar que el gen codificante para la proteína putativa *TvCDK* es un gen de copia única el cual se expresa a nivel transcripcional.

Por medio del análisis de la secuencia peptídica de *TvCDK* con la herramienta BLAST de NCBI, se encontró una similitud con otras CDK de alrededor del 50%. Esto se considera un valor alto de homología ya que las protein cinasas de protozoarios generalmente son parecidas en un 40% a sus homólogas en otros organismos⁴².

De acuerdo a los tres primeros análisis *in silico* que se describen en la sección 5 del apartado de resultados, se observó que la secuencia peptídica de *TvCDK* contenía múltiples motivos conservados que corresponden a una protein cinasa y específicamente, a proteínas del tipo Cdk; también, el modelo tridimensional mostró que esta proteína podría tener una conformación similar a la reportada para proteínas del tipo Cdk, conservación del plegamiento de hélices funcionales y posición equivalente de los aminoácidos involucrados en la regulación por fosforilación de proteínas del tipo Cdk.

Es importante remarcar que entre las estructuras en *TvCDK* que conservan una disposición estándar para proteínas del tipo Cdk, se encuentra la hélice PSTAIRE, que sin tener una identidad exacta con los aminoácidos consenso (el cambio es a la secuencia PVNTVRE), muestra un plegamiento hipotético que le brindaría la funcionalidad descrita para esta estructura. El motivo de 16 residuos que contiene a la hélice PSTAIRE, se considera característico de las proteínas de tipo Cdk y algunos estudios muestran que podría ser necesario para la unión a ciclinas ³⁰.

La secuencia peptídica de *TvCDK* posee extensiones de aminoácidos no relacionados a las proteínas de tipo Cdk convencionales en sus extremos amino y carboxilo terminales, lo que se ha documentado en diversas proteínas CRK de protozoarios como extensiones de las que se desconoce su función pero que no se ha observado que comprometan la funcionalidad de la proteína ⁴². Para el organismo *C. fasciculata* se ha descrito una CRK que contiene largas inserciones de aminoácidos, pero se postula que dado que esas inserciones ocurren entre dominios proteicos bien conservados para una protein cinasa, es poco probable que afecten la funcionalidad de una CRK⁴³.

La expresión de la proteína recombinante *His-CDK* es inducida por IPTG aún a bajas concentraciones, asimismo, la expresión de esta proteína recombinante no es tóxica para las bacterias transformadas ya que aún a un largo tiempo de inducción (5 horas), no se observaron cambios en la tasa de crecimiento de los cultivos. Por medio de electroforesis se encontró que *His-CDK* pesa un poco más de 40 kDa, lo que coincide con el peso esperado de acuerdo a los aminoácidos extra que se tuvieron que insertar para fusionar la cola de histidinas.

El ensayo de unión a perlas de agarosa asociadas a la proteína $p13^{suc+1}$ mostró que la proteína recombinante desarrollada es capaz de unirse a este ligando, lo que sugiere que *TvCDK* contiene las regiones necesarias para la unión de esta proteína reguladora y es capaz de obtener una conformación suficiente para esta unión. De acuerdo a los

análisis de mutación realizados por ³⁰, la capacidad de unión a *p13^{suc+1}* es indicador de la funcionalidad de una proteína del tipo Cdk; por lo que este ensayo nos proporciona evidencia experimental indirecta de la posible funcionalidad de *TvCDK*.

Con diversos análisis *in silico* y utilizando el método teórico desarrollado, se encontraron 7 secuencias presentes en el genoma anotado de *T. vaginalis* que poseen motivos y residuos de aminoácidos conservados que sugieren su posible funcionalidad como una proteína tipo Cdk en este organismo. Asimismo, los modelos teóricos de la conformación que tendrían las proteínas derivadas de estas secuencias muestran una disposición similar a la que se detalló para el modelo tridimensional de *TvCDK*.

El método teórico desarrollado para la valorización de los residuos de aminoácidos o motivos presentes en una secuencia peptídica determinada y su validación por medio de controles positivos y negativos nos muestra que este método es una aproximación que podría ser aplicada a gran escala para detectar proteínas en el genoma anotado de un organismo particular o ser utilizado en búsquedas metagenómicas de bases de datos con genomas anotados para diversos organismos.

Adicionalmente, este método puede ser modificado para incorporar nuevos parámetros de análisis, según la identificación posterior de secuencias o motivos importantes para proteínas del tipo Cdk.

X. CONCLUSIONES

1. La secuencia nucleotídica de *pgactin10* contiene elementos específicos de un gen en la región descrita para *TvCDK*.
2. El gen correspondiente a *TvCDK* se expresa a nivel de transcripción de RNA mensajero.
3. La proteína recombinante desarrollada se expresa por medio de inducción por IPTG a bajas concentraciones.
4. La proteína recombinante para *TvCDK* es capaz de unirse al ligando *p13^{suc+1}*, que es un factor regulador de las proteínas del tipo Cdk.
5. La secuencia peptídica de *TvCDK* muestra una alta homología a otras proteínas del tipo Cdk y conserva motivos necesarios para este tipo de proteínas.
6. El modelo teórico de la estructura terciaria de *TvCDK* muestra una conformación similar a la reportada en cristales de proteínas del tipo Cdk cuya funcionalidad ha sido comprobada.
7. El genoma de *T. vaginalis* contiene otras 7 secuencias además de *TvCDK* que podrían codificar para una proteína del tipo Cdk, de acuerdo a los análisis *in silico* realizados.

XI. PERSPECTIVAS

Se propone realizar estudios para la caracterización y expresión de las 7 secuencias del genoma reportado por TIGR que se analizaron con estrategias *in silico* en este trabajo y que podrían tener relevancia en el ciclo celular de *T. vaginalis*.

Además, se propone la purificación de la proteína recombinante *TvCDK* por cromatografía de afinidad en columnas de Ni-NTA para la realización de un ensayo de fosforilación de la histona H1 *in vitro* y la caracterización funcional de la misma a través de la producción de anticuerpos.

XII. BIBLIOGRAFÍA

- (1) Arroyo, R.; Gonzalez-Robles A.; Martinez-Palomo A.; Alderete, J.F. Mol. Microbiol. 1993, 7, 299-309
- (2) Dacks, J. B.; Walker, G.; Field, M. C. Parasitol Int 2008, 57, 97-104.
- (3) Roberts; Natural History Museum, UK, 1998 (<http://www.nhm.ac.uk/research-curation/research/projects/euk-extreme/>).
- (4) Nanda, N.; Michel, R. G.; Kurdgelashvili, G.; Wendel, K. A. Expert Rev Anti Infect Ther 2006, 4, 125-35.
- (5) Baron, S. Medical microbiology; 4th ed.; University of Texas Medical Branch at Galveston: Galveston, Tex., 1996.
- (6) Alberts, B. Molecular biology of the cell; 4th ed.; Garland Science: New York, 2002.
- (7) Lodish, H. F. Molecular cell biology; 5th ed.; W.H. Freeman and Company: New York, 2003.
- (8) Lehninger, A. L.; Nelson, D. L.; Cox, M. M. Lehninger principles of biochemistry; 4th ed.; W.H. Freeman: New York, 2005.
- (9) Malik SB, P. A., Stefaniak LM, Schurko AM, Logsdon JM Jr. PlosOne 2007, 3.
- (10) Zubacova, Z.; Cimburek, Z.; Tachezy, J. Mol Biochem Parasitol 2008, 161, 49-54.
- (11) Ali, V.; Nozaki, T. Clin Microbiol Rev 2007, 20, 164-87.
- (12) WHO, Global prevalence and incidence of selected curable sexually transmitted infections. Overview and estimates; World Health Organization, 2001.
- (13) Miller, M.; Liao, Y.; Gomez, A. M.; Gaydos, C. A.; D'Mellow, D. J Infect Dis 2008, 197, 503-9.
- (14) McClelland, R. S.; Sangare, L.; Hassan, W. M.; Lavreys, L.; Mandaliya, K.; Kiarie, J.; Ndinya-Achola, J.; Jaoko, W.; Baeten, J. M. J Infect Dis 2007, 195, 698-702.
- (15) Dunne, R. L.; Dunn, L. A.; Upcroft, P.; O'Donoghue, P. J.; Upcroft, J. A. Cell Res 2003, 13, 239-49.
- (16) Petrin, D.; Delgaty, K.; Bhatt, R.; Garber, G. Clin Microbiol Rev 1998, 11, 300-17.

- (17) Mehlhorn, H.; Armstrong, P. M. Encyclopedic reference of parasitology: diseases, treatment, therapy; 2nd ed.; Springer: Berlin ; New York, 2001.
- (18) Stein, G. S. The molecular basis of cell cycle and growth control; J. Wiley: New York, 1998.
- (19) Feldmann, H. Yeast molecular biology: A short compendium on basic features and novel aspects., 2005.
- (20) Wong, J. T. Biol Signals 1996, 5, 301-8.
- (21) Harper, J. W.; Adams, P. D. Chem Rev 2001, 101, 2511-26.
- (22) Nurse, P. Mol Biol Cell 1994, 5, 613-6.
- (23) Ducommun, B.; Brambilla, P.; Felix, M. A.; Franza, B. R., Jr.; Karsenti, E.; Draetta, G. EMBO J 1991, 10, 3311-9.
- (24) Pines, J. Nature 1995, 376, 294-5.
- (25) Malumbres, M. Cell Cycle 2005, 4, 206-10.
- (26) Morgan, D. O. Annu Rev Cell Dev Biol 1997, 13, 261-91.
- (27) Pavletich, N. P. J Mol Biol 1999, 287, 821-8.
- (28) Hanks, S. K.; Quinn, A. M.; Hunter, T. Science 1988, 241, 42-52.
- (29) Ohta, T.; Okamoto, K.; Isohashi, F.; Shibata, K.; Fukuda, M.; Yamaguchi, S.; Xiong, Y. Cancer Res 1998, 58, 1095-8.
- (30) Ducommun, B.; Brambilla, P.; Draetta, G. Mol Cell Biol 1991, 11, 6177-84.
- (31) Brown, N. R.; Noble, M. E.; Endicott, J. A.; Johnson, L. N. Nat Cell Biol 1999, 1, 438-43.
- (32) Hanks, S. K.; Hunter, T. FASEB J 1995, 9, 576-96.
- (33) Marcote, M. J.; Knighton, D. R.; Basi, G.; Sowadski, J. M.; Brambilla, P.; Draetta, G.; Taylor, S. S. Mol Cell Biol 1993, 13, 5122-31.
- (34) Gondeau, C.; Gerbal-Chaloin, S.; Bello, P.; Aldrian-Herrada, G.; Morris, M. C.; Divita, G. J Biol Chem 2005, 280, 13793-800.
- (35) McGowan, C. H.; Russell, P. EMBO J 1993, 12, 75-85.
- (36) Patra, D.; Dunphy, W. G. Genes Dev 1996, 10, 1503-15.
- (37) Pines, J. Curr Biol 1996, 6, 1399-402.

- (38) Hayles, J.; Aves, S.; Nurse, P. *EMBO J* 1986, 5, 3373-9.
- (39) de Lichtenberg, U.; Jensen, L. J.; Brunak, S.; Bork, P. *Science* 2005, 307, 724-7.
- (40) Loyer, P.; Trembley, J. H.; Grenet, J. A.; Busson, A.; Corlu, A.; Zhao, W.; Kocak, M.; Kidd, V. J.; Lahti, J. M. *J Biol Chem* 2008, 283, 7721-32.
- (41) Hammarton, T. C.; Mottram, J. C.; Doerig, C. *Prog Cell Cycle Res* 2003, 5, 91-101.
- (42) Doerig, C.; Endicott, J.; Chakrabarti, D. *Int J Parasitol* 2002, 32, 1575-85.
- (43) Brown, L.; Hines, J. C.; Ray, D. S. *Nucleic Acids Res* 1992, 20, 5451-6.
- (44) Mottram, J. C.; Kinnaird, J. H.; Shiels, B. R.; Tait, A.; Barry, J. D. *J Biol Chem* 1993, 268, 21044-52.
- (45) McKean, P. G. *Curr Opin Microbiol* 2003, 6, 600-7.
- (46) Speranca, M. A.; Vinkenoog, R.; Ocampos, M.; Fischer, K.; Janse, C. J.; Waters, A. P.; del Portillo, H. A. *Exp Parasitol* 2001, 97, 119-28.
- (47) Vinkenoog, R.; Speranca, M. A.; Ramesar, J.; Thomas, A. W.; del Portillo, H. A.; Janse, C. J.; Waters, A. P. *Mol Biochem Parasitol* 1998, 95, 229-40.
- (48) Chou, C. F.; Tai, J. H. *Biotechniques* 1996, 20, 790-1.
- (49) Chomczynski, P.; Sacchi, N. *Anal Biochem* 1987, 162, 156-9.
- (50) Suganuma, A.; Gupta, K. C. *Anal Biochem* 1995, 224, 605-8.
- (51) Espinosa, N.; Hernandez, R.; Lopez-Griego, L.; Arroyo, R.; Lopez-Villaseñor, I. *Acta Trop* 2001, 78, 147-54.
- (52) Zhang, H., *Involvement of CDK/Cyclin motif in ciliate cell cycle regulation*; The University of British Columbia, 2000.
- (53) Liston, D. R.; Johnson, P. J. *Mol Cell Biol* 1999, 19, 2380-8.
- (54) Carlton, J. M.; Hirt, R. P.; Silva, J. C.; Delcher, A. L.; Schatz, M.; Zhao, Q.; Wortman, J. R.; Bidwell, S. L.; Alsmark, U. C.; Besteiro, S.; Sicheritz-Ponten, T.; Noel, C. J.; Dacks, J. B.; Foster, P. G.; Simillion, C.; Van de Peer, Y.; Miranda-Saavedra, D.; Barton, G. J.; Westrop, G. D.; Muller, S.; Dessi, D.; Fiori, P. L.; Ren, Q.; Paulsen, I.; Zhang, H.; Bastida-Corcuera, F. D.; Simoes-Barbosa, A.; Brown, M. T.; Hayes, R. D.; Mukherjee, M.; Okumura, C. Y.; Schneider, R.; Smith, A. J.; Vanacova, S.; Villalvazo, M.; Haas, B. J.; Perteza, M.; Feldblyum, T. V.; Utterback,

T. R.; Shu, C. L.; Osoegawa, K.; de Jong, P. J.; Hrdy, I.; Horvathova, L.; Zubacova, Z.; Dolezal, P.; Malik, S. B.; Logsdon, J. M., Jr.; Henze, K.; Gupta, A.; Wang, C. C.; Dunne, R. L.; Upcroft, J. A.; Upcroft, P.; White, O.; Salzberg, S. L.; Tang, P.; Chiu, C. H.; Lee, Y. S.; Embley, T. M.; Coombs, G. H.; Mottram, J. C.; Tachezy, J.; Fraser-Liggett, C. M.; Johnson, P. J. *Science* 2007, 315, 207-12.