Pontifícia Universidade Católica do Rio Grande do Sul
Faculdade de Biociências
Programa de Pós-Graduação em Biologia Celular e Molecular

Modelagem molecular de grupos funcionais dos domínios Ets envolvidos na ligação ao DNA humano

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Celular e Molecular como requisito para a obtenção do grau de Mestre em Biologia Molecular e Celular.

Autor
Rafael Andrade Caceres

Orientador
Prof. Dr. Walter Filgueira de Azevedo Junior

Porto Alegre, RS
Março, 2008

Dedico este trabalho à minha mãe, Vera Maria Andrade, exército de um homem só, mulher de coragem e determinação, que fez os meus sonhos os dela. Exemplo de coragem, perseverança e bondade. A você mãe, por todas dificuldades, sofrimentos, privações e certamente alegrias que vivenciamos juntos, dedico-te esta conquista como uma ínfima parcela de minha gratidão.

Agradecimentos

Inicialmente gostaria de agradecer ao Prof. Dr. Hermes L. N. de Amorim, meu primeiro mestre, grande amigo, pessoa ao qual devo o início da minha carreira acadêmica.

Ao Prof. Dr. Diógenes Santiago Santos e Prof. Dr. Luiz Augusto Basso pois, sem eles certamente não estaria cursando a minha Pós-Graduação na PUCRS e a eles devo a indicação para ser orientado pelo Prof. Dr. Walter F. de Azevedo Jr.

Ao Prof. Dr. Walter Filgueira de Azevedo Junior por ter aceito ser meu tutor e realizar meu treinamento científico, meu professor, grande amigo e por vezes pai. Através dele aprendi a tirar o máximo de mim, aprendi que ciência pode ser feita com alegria e informalidade sem perder de vista a produção científica. Por mais que tentasse me expressar através de palavras (e todos sabem que não sou bom em expressar meus sentimentos), estas não seriam suficientes para demonstrar o quanto sou grato por ter me acolhido e dia a dia me ensinado o valor do trabalho, dedicação e ética na ciência.

Aos meus amigos do Laboratório de Bioquímica Estrutural (LaBioQuest), Ana Luiza Vivan, Guy, Ivani, Luis Fernando, Patrícia Perez, Raquel Dias pelo apoio, companhia, amizade e carinho em todas as horas.

À minha mãe, Vera Maria Andrade, pelo amor incondicional, apoio em todas as horas difíceis, e aqueles que me conhecem sabem que foram muitas, fez dos meus sonhos os dela, e não mediu esforços para que pudesse realizá-los. Seria impossível transcrever o quanto sou agradecido a ela.

À todas as pessoas que direta ou indiretamente contribuíram para a execução deste trabalho.

INDICE

LISTA DE BREVIATURAS

ACE: energia atômica de contato

DNA: Ácido desoxiribonucleico

ED: Domínio ETS

ETS: E26 twenty six

HTH: Hélice volta hélice

MD: Dinâmica molecular

ns: nanosegundos (10^{-9} segundos)

NERF: Novo fator Ets relacionado

PDB: Banco de dados de proteínas

RMSD: Desvio médio quadrático

RMSF: Flutuação média quadrática

1. RESUMO

Considerando que alguns fatores de transcrição da família ETS regulam, crescimento, apoptose, angiogênese e genes relacionados a metástase em células tumorais, nós propomos modelos tridimensionais de domínios ETS (ETV-2, SPI-C and NERF) aplicando técnicas de modelagem molecular por homologia e utilizando estruturas de cristais de Ets humana-DNA como template. Estes modelos serão úteis para estabelecer semelhanças estruturais entre duas ou mais moléculas relacionadas.

A obtenção destes detalhes estruturais das interações entre proteína e DNA fornecerá pistas sobre suas consformações dos complexos binários Ets-DNA, que podem a vir tornar alvos moleculares para o desenvolvimento de fármacos seletivos para o tratamento de câncer no futuro.

Abstract

Considering that some ETS transcriptions factors regulate growth, apoptosis, angiogenesis, invasions and metastasis-related genes in tumor cells, these transcription factors we proposed and describes three-dimensional models of ETS (ETV-2, SPI-C and NERF) by applying computational homology-modeling techniques and utilizing the high-resolution crystals structures of Ets-human DNA complexes as templates. These models should be useful to establish overall structural similarities between two or more related molecules, and establish the structural key features between members of related subfamilies and their relationship with their diverse biological roles. Obtaining structural details of macromolecular interactions provide insights into static conformation of DNA-Ets domain binary complexes, which may become a molecular targeting therapy against ETS transcription factors that could be a novel approach to a selective cancer treatment in the future.

2. INTRODUÇÃO

Câncer é uma doença que envolve mudanças dinâmicas no genoma. Baseadas no fato de que todas as células de mamíferos carregam uma maquinaria celular semelhante que regula sua proliferação, diferenciação e morte, inúmeros traços moleculares, bioquímicos e celulares são compartilhados talvez por todos os tipos de células cancerosas humanas. Tumorigênese é um processo constituído de múltiplas etapas que reflete alterações genéticas guiando uma transformação progressiva de células sadias em derivantes altamente malignas [1]. Fatores de transcrição são moléculas chave, envolvidas em vários processos biológicos, incluindo proliferação, diferenciação, desenvolvimento, transformação e apoptose. A detecção do alvo dos genes que são modulados por um fator de transcrição específico é crítica para compreender os genes envolvidos no crescimento de células sadias e para fornecer informações diretas e indiretas no desenvolvimento do câncer [2]. A perda de homeostase celular é regulada por fatores de transcrição específicos que ativam ou reprimem a expressão dos genes que controlam o balanço celular. A identificação de moléculas envolvidas em etapas biológicas é necessária para a descoberta de novos potenciais marcadores de diagnósticos/prognósticos tanto quanto alvos para o desenvolvimento de fármacos contra o câncer [2].

Os fatores de transcrição que coordenam programas celulares apresentam um conjunto de motivos estruturais de reconhecimento e ligação em seqüências específicas de DNA. Uma das maiores famílias de reguladores

transcricionais é a família ETS[1]. A superexpressão ou mesmo a perda de fatores de transcrição Ets estão relacionados ao desenvolvimento de câncer [3].

A identificação da proteína Ets-1 no vírus da eritroblastose aviária E26, portador do oncogene *v-ets,* em 1983, levou a descoberta da família Ets de fatores de transcrição. Tanto a proteína Ets-1, quanto a família Ets foram denominadas em decorrência do vírus aviário E26 (*E twenty six*). As demais proteínas que seguidamente tornaram-se membros desta família foram identificados e agrupados com base nas identidades dos seus domínios de ligação ao DNA referente a Ets-1, chamado de domínio ETS [4].

[1] ETS = E26 (*E twenty six*)

3. APRESENTAÇÃO DO TEMA

O domínio ETS é um motivo de reconhecimento da cadeia de DNA, denominado *winged helix-turn-helix*, wHTH, pois, até meados dos anos 90 acreditava-se pertencer à superfamília *helix-turn-helix* (HTH) de reconhecimento de DNA. Este domínio é constituído de aproximadamente 85 resíduos de aminoácidos, composto por duas hélices-α conectadas por uma volta e por duas folhas β conectadas por uma alça (winged). Estudos estruturais mostraram uma grande conservação no arranjo deste domínio, com as três hélices-α e quatro fitas β arranjadas na seguinte ordem: α1–β1 –β2–α2–α3–β3–β4 [5], sendo que, as hélices-α2 e 3 correspondem ao motivo hélice-volta-hélice (hélix-turn-hélix HTH), a asa (winged) é composto por resíduos entre as fitas β3 e 4 (**Figura 1**). Estudos prévios de RMN e difração de raios X demonstraram que a hélice α3 é o principal ponto de interação com o sulco maior da cadeia de DNA da seqüência consenso 5'-GGAA/T-3' [4].

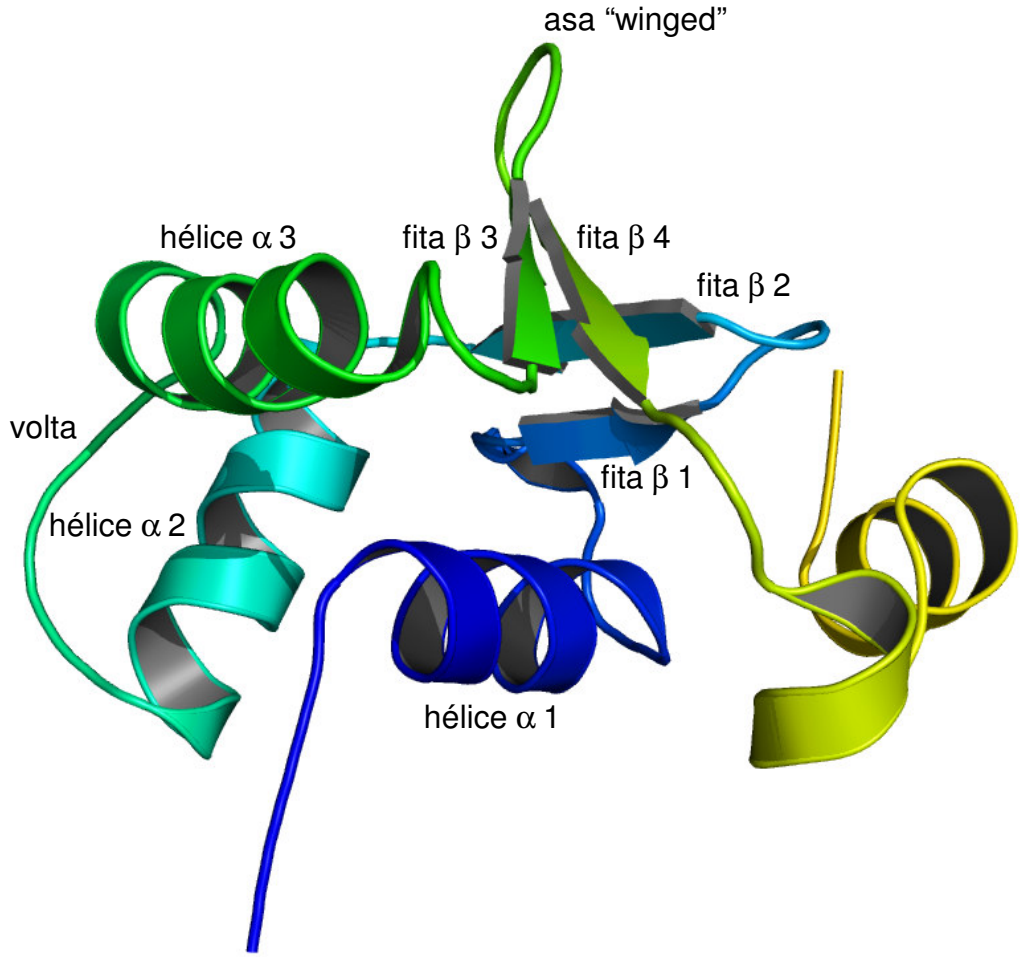


Figura 1 – Estrutura do domínio ETS da proteína ETS-1 (PDB ID: 1K79 [5]), a cadeia de DNA não está representada. A proteína é mostrada no modelo de fitas e colorida do azul para o amarelo, da extremidade amino terminal (N-terminal) para C-terminal. Os pontos de contato com a cadeia de DNA são a hélice alfa 3, volta e a asa (winged), características do motivo de reconhecimento wHTH.

Todas proteínas de domínio ETS reconhecem e se ligam à seqüência de nucleotídeos 5’-GGA(A/T)-3’. Estes quatro nucleotídeos localizados nas posições 5’e 3’ desta seqüência são responsáveis pela especificidade de ligação de cada um dos membros da família ETS.

Apenas os organismos metazoários apresentam fatores de transcrição da família ETS, cuja função parece estar relacionada a cascatas regulatórias exclusivas a estes organismos [4].

Fatores de transcrição atuam como receptores nucleares de cascatas de transdução de sinais, geralmente funcionando em colaboração com outros fatores de transcrição, ligados ou não à cadeia de DNA, com co-reguladores, e também com histona-acetilases [6,7,8]. Os mecanismos de sinalização variam significativamente entre os membros da família, havendo tanto ativadores quanto repressores do processo de transcrição, com funções tão diversas quanto o desenvolvimento hematopoiético e neuronal em mamíferos, desenvolvimento vascular e angiogênese [9].

O padrão de distribuição temporal e espacial dos fatores de transcrição e de seus co-reguladores em embriões e em organismos adultos contribui para a especificidade de seus efeitos de controle transcricional [10].

As proteínas desta superfamília de fatores de transcrição são classificadas em subfamílias de acordo com a identidade de suas seqüências e outros domínios conservados em suas estruturas terciárias, o que se reflete nas funções biológicas e moleculares de cada uma das subfamílias.

As proteínas ETS estão relacionadas a processos biológicos distintos, sem ainda haver uma função específica que seja compartilhada por todos os membros da sua família. Funções tão diversas como crescimento, indução e inibição de apoptose, desenvolvimento, diferenciação, regulação da senescência, hematopoiése, angiogênese, transformação oncogênica entre

outras são decorrência da divergência das seqüências destas proteínas nos seus demais domínios estruturais, o que contribui para diferentes padrões de expressão, propriedades de controle da transcrição e por fim, diferentes papéis biológicos [8]. Diversas proteínas ETS são tecido-específicas, e estão envolvidas no crescimento e diferenciação das células do tecido em que são expressas através da ativação da transcrição de genes que codificam fatores de crescimento característicos de cada tecido [8]. A sinalização mediada por fatores de crescimento e seus receptores modulam a atividade de diversas proteínas ETS pela via MAP cinase (MAPK) através de fosforilações que podem afetar a sua capacidade de ligação à cadeia de DNA, e a sua capacidade de associação com outras proteínas regulatórias ou co-regulatórias, bem como a sua localização celular [8,11].

A subfamília de proteínas ETS está envolvida na regulação da expressão de genes relacionados com o controle do ciclo celular, ontogênese e genes supressores de tumor [8]. A desregulação ou expressão inapropriada destas proteínas parecem ter o papel crucial no desencadeamento de alguns tipos de câncer [5], estando co-relacionadas com altas taxas de invasividade e metástase [8,12,13].

Conhecer e caracterizar cada um dos membros da família ETS, seus mecanismos moleculares e funções biológicas têm tido relevância maior que a procura pelo fator comum entre seus membros [12]. Uma vez que estas proteínas estão envolvidas em processos de crescimento celular, desenvolvimento, diferenciação e apoptose, o entendimento preciso do seu

funcionamento pode auxiliar no controle de várias doenças humanas incluindo desordens imunológicas, doenças vasculares e o desenvolvimento de tumores malignos.

A partir dos dados obtidos através da análise estrutural destas proteínas a etapa seguinte deverá ser a integração deste conhecimento na análise funcional de múltiplas proteínas ETS expressas em um mesmo tipo celular e sua capacidade de regular a expressão dos genes em comum, com diferentes magnitudes e propriedades (ativador/repressor) e ainda padrões de expressões diferentes (constitutiva/transitória), assim como a relação com outros fatores de transcrição e seu sinergismo no controle transcricional [12].

Para tanto, faz-se necessário conhecer que resíduos são responsáveis pela interação proteína-DNA. Cada proteína possui uma especificidade de ligação, e este fenômeno parece estar relacionado com a diferença dos resíduos de aminoácidos observados nos membros destas subfamílias Ets e de propriedades estruturais seqüência-dependente da cadeia de DNA enquanto substrato de ligação [14], sendo os mecanismos de reconhecimento intermolecular determinantes na formação dos complexos proteína-DNA para a família ETS [15,16].

4. OBJETIVOS

As diferenças na afinidade dos domínios ETS das proteínas são decorrentes das substituições nas suas estruturas primárias e de diferentes propriedades estruturais da cadeia de DNA conferidas pelas seqüências de nucleotídeos 5'-GGAA-3' de alta afinidade e 5'-GGAT-3'de baixa afinidade.

Características estruturais das proteínas e da cadeia de DNA são responsáveis pela complementaridade dos sítios de ligação e formação do complexo proteína-DNA.

Através da proposição de modelos estruturais das proteínas ETS, ainda não determinadas por métodos empíricos, será possível identificar os resíduos envolvidos na formação do complexo e a influência dos fatores intrínsecos ao complexo proteína-DNA

4.1. OBJETIVOS GERAIS

Obter as estruturas dos domínios ETS, mais especificamente da NERF, SpiC, e ETV2, através do método de modelagem molecular por homologia e posteriormente a obtenção dos complexos ETS-DNA através de simulações de *docking rígido.*

4.2. OBJETIVOS ESPECÍFICOS

Identificar quais os resíduos que participam e quais são importantes para a interação molecular entre proteína e DNA.

5. ARTIGO CIENTÍFICO 1

Esta revisão foi um convite feito pelos editores da Current Bioinformatics (CBIO) no segundo semestre de 2006. A revisão contribuiu de forma a consolidar os conhecimentos necessários para a execução deste projeto sem qual certamente o trabalho de revisão biblioráfica se tornaria um trabalho laborioso.

A revisão abrenge toda pesquisa feita nos últimos 24 anos em relação as ETS, focando principalmente estudos estruturais através dos métodos de modelagem molecular por homologia, difração de raios X, ressonância magnética nuclear (RMN) e dinâmica molecular. A revisão ainda expressa a visão dos autores sobre o futuro nas pesquisas sobre ETS utilizando ferrementas de bioinformática estrutural e de quão importante é esta área da pesquisa ainda pouco explorada, mas, que vem conquistando espaço devido a sua capacidade de propor estruturas e modelos de sistemas complexos (proteína-ligante), permitindo assim a obtenção de dados moleculares que, senão impossíveis, são muito difíceis de se obter em laboratório por meio das técnicas atualmente utilizadas.

Molecular Dynamics and Structural Studies of the Ets Domain-DNA Complexes

Rafael A. Caceres[1], Cláudia P. Nunes[1], Walter F. de Azevedo Jr.[1], Luiz A. Basso[*,1,2] and Diógenes S. Santos[*,2]

[1]*Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga, 6681, Porto Alegre-RS CEP 90619-900, Brazil*

[2]*Pontifícia Universidade Católica do Rio Grande do Sul, Centro de Pesquisa em Biologia Molecular e Funcional, Porto Alegre, RS, Brazil*

Abstract: Ets family transcription factors, characterized by an evolutionary conserved Ets domain, play important roles in cell development, cell differentiation, apoptosis and tissue remodeling. All members of this family have an activation or a repression domain for DNA binding. The Ets domain has been shown to bind 5'-GGAA/T-3' core motif of DNA by its wHTH (winged helix-turn-helix) structure, as determined by NMR or X-Ray crystallography analysis. A number of Ets transcription factors have been shown to be involved or directly implicated in the pathogenesis of a wide spectrum of human cancers. They are thus potential molecular targets for selective cancer therapy. Computer simulations using molecular dynamics allows monitoring the dynamics of individual atoms, thereby giving a unique insight into the structural flexibility of binary complexes between Ets-domain and a specific DNA sequence that cannot be easily extracted from laboratory experiments. Moreover, an understanding of structural motifs that influence the specificity of Ets domains is essential for antitumor drug design. Here we describe molecular and structural studies that provide a detailed description of the direct and indirect readout mechanisms of DNA recognition by Ets domains, a conserved mechanism observed in other DNA-protein complexes.

Keywords: Molecular dynamics, Ets domain, DNA binding site, structure.

INTRODUCTION

Cancer is a disease involving dynamic changes in the genoma. Based upon the fact that all mammalian cells carry similar molecular machinery regulating their proliferation, differentiation and death, a number of molecular, biochemical and cellular traits are shared by perhaps all types of human cancer cells. Tumorigenesis is a multistep process that reflects genetic alterations driving progressive transformation of normal cells into highly malignant derivatives [1]. Transcription factors are key molecules involved in various biological processes, including cellular proliferation, differentiation, development, transformation and apoptosis. Detection of target genes that are modulated by a specific transcription factor is critical to understanding the genes involved in control of normal cell growth and to provide direct and indirect information on cancer development [2]. Loss of cellular homeostasis is regulated by specific transcription factors that activate and repress the expression of genes that control cellular balance. The identification of molecules involved in biological steps is required to unveil novel potential diagnostic/prognostic markers as well as targets for drug development against cancer [2].

The transcription factors that coordinate cellular programs display a set of DNA binding motifs. One of the largest families of transcriptional regulators is the Ets family. Ets factors have been found to be overexpressed or lost during cancer development [3]. This family is characterized by an evolutionary conserved binding DNA domain, termed the Ets domain. Members of this large family are divided into subfamilies, according to the extent of homology in the Ets domains. All members have much higher homology in the Ets domain within an individual subfamily as compared with members in other subfamilies. The pattern of tissue-specific expression of *ets* genes varies among family members, and a number of them are associated with human malignancies [4, 5] (Table **1**).

The expression of the entire mouse Ets family was analyzed in different tissues and 24 of the 25 Ets factors were expressed in normal mammary tissue. A comparison of Ets expression shows increased expression of 5 Ets in tumoral tissue and other 4 in normal tissue. However, it has been pointed out that large increases in Ets factor expression do not necessarily reveal a causal role in tumorigenesis [2]. Accordingly, a complementary approach to identifying Ets factors that may mediate breast cancer signaling by determining which Ets factor mRNAs are abundant in all mammary tissue contexts showed that Elk4, Elf1, Ets2, and Ets1 are as abundant in normal mammary tissue as in tumors. These results prompted the authors to propose that Ets factors abundantly expressed in normal mammary tissues are also important candidates for mediating tumor signaling [2]. Similarity in DNA binding specificities makes it difficult to infer from overexpression studies which Ets factor(s) regulate(s) the expression of most target genes *in vivo* [6].

Recent molecular dynamics (MD) simulations of Ets1-DNA complex indicate that helix-3 in the winged helix-turn-

*Address correspondence to either author at the Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF)-Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Av. Ipiranga, 6681-Tecnopuc-Prédio 92A, ZIP CODE 90619-900, Porto Alegre, RS, Brazil; Tel/Fax: +55 51 33203629; E-mails: diogenes@pucrs.br or luiz.basso@pucrs.br

1574-8936/07 $50.00+.00

Table 1. Ets Family Members

Family Members	Biological Roles	Cancer and Disease Related
Ets-1, Ets-2	T-cell survival, NK cell differentiation, ECM remodeling	Thyroid, colon, liver, pancreas, prostate, breast, lung, gastric
Erg-2, Fli-1, FEV	Megakaryocyte differentiation	Ewing's sarcoma
GABPÎ±	Muscular sinapse function	
TEL (Etv-6), TEL2, Yan	Yolk sac angiogenesis, adult hematopoiesis	Fusions with leukaemias genes
PEA3, ERM, ETV1, ER81, ETV5	Neuronal pathfinding, muscle differentiation	Breast, Ewing's sarcoma, lung, gastric
Elk-1, SAP1, NET/ERP/SAP2	Immediate-early gene differentiation, neuronal differentiation, lymphatic development	Cerebral ischemia, prostate, breast
Elf-1, NERF 1b, MEF	Cellular growth and differentiation	Prostate, endometrial, ovarian
PU.1, Spi-B, Spi-C	Myeloid/lymphoid differentiation, B cell function	Leukaemias
ERF, PE-1	Related to macrophage differentiation	Not known
ESX/ESE-1, ESE-2, ESE-3, PDEF	Ephitelial specific expression	Breast, prostate, Ewing's sarcoma

Adapted from [2] and [5].

helix (wHTH) motif binds in the major groove of the consensus DNA sequence. Here we describe recent results on this interaction and the relationship between the core sequences of DNA and Ets domain-DNA binary complex formation affinity.

THE ETS BINDING DOMAIN

Members of Ets family were originally identified by a region of primary sequence homology with the protein product of *v-ets* oncogene encoded by E26 (Ets = **E** t**w**enty-**s**ix) avian erythroblastosis virus. The conserved region, named the Ets-domain, corresponds to the DNA-binding domain. Ets-domain transcription factor can be further subclassified primarily because of high amino acid conservation in their Ets-domain [7]. Moreover, based on observed organization of secondary structural elements, Ets proteins are proposed to be part of a superfamily of winged helix-turn-helix DNA-binding proteins [8] consisting of three alpha-helices and four beta-strands. The DNA-binding Ets-domain is approximately 85 amino acids long, and is usually at the N-terminal polypeptide sequence. The Ets domain is composed of three alpha-helices (H) and four beta-strands (S) arranged in the order H1-S1-S2-H2-H3-S3-S4 [8]. It has been proposed for the Ets family the nomenclature wHTH (winged helix-turn-helix) [8], in which helices 2 and 3 correspond to helix-turn-helix motif, the DNA recognition region. HTH motif makes specific contacts with the major groove of DNA. The winged helix is a helix flanked by a loop connecting the last two -strands on one side and a long loop at C-terminal on the other side as shown in Fig. **1**. Previous X-ray and NMR studies have shown that helix-3 in the winged helix-turn-helix motif binds in the major groove of the consensus DNA sequence. The third helix in the Ets-domain is responsible for making major grove contacts with the central GGAA/T motif, whilst additional DNA contacts are made by a loop between the β-strands 3 and 4 (the wing) and a loop between α-helix 2 and 3 as depicted in Fig. **1**.

The Ets-domain can bind DNA as a monomer and induces a smooth 8º bend in the DNA target site. Single substitutions on the hydrophobic core result in loss of DNA binding, probably because the changes disrupt the tight globular structure of the domain. Summarizing, the residues present in this domain (mainly in the helix-3) seem to be pivotal for the formation of the overall scaffold of DNA recognition by the Ets domain [9]. Interaction between DNA and protein occurs through two surfaces on which interacting groups are located, as observed for protein-protein interactions. The interaction surface of many transcription factors consists predominantly of an -helix and the major (M) groove of DNA [10] as shown in Fig. **2**. The Ets domains bind specifically to Ets binding sites (EBSs) composed of a purine-rich central core sequence 5´GGA(A/T)3´ surrounded by more variable sequences that contribute to the binding specificity of individual proteins [11-13].

Interestingly, the Ets proteins found in metazoans revealed structural similarity with regions of *Trichomona vaginalis* IBD (Initiator Binding Domain). This region of similarity corresponds to the structural scaffold of the latter used for presenting the DNA binding elements, demonstrating that a fundamental mechanism of recognition is also present in primitive eukaryotes [14]. Conserved regions indicate sequences and structures likely to be important in the biological function of these proteins.

Structural analysis of the murine Ets-1 Ets domain indicates that a hydrophobic core of highly conserved residues of all three helices and four -strands are essential to maintain the structural integrity of the domain. On the other hand, highly variable sequences are on the loops or turns connecting the -helices and -sheets [15].

ETS-DNA COMPLEXES

One of the goals of structural studies on DNA-binding transcription factors is identify how particular sequences of DNA are recognized. The apparent DNA binding specificity appears to depend on the experimental approach. Multiplex analysis measures the contribution of a single base to complex stabilization without knowledge of the sequence context, whereas SELEX experiments involve the isolation of individual high affinity sequences from a randomized pool of oligonucleotides. Based on the former approach, a mecha

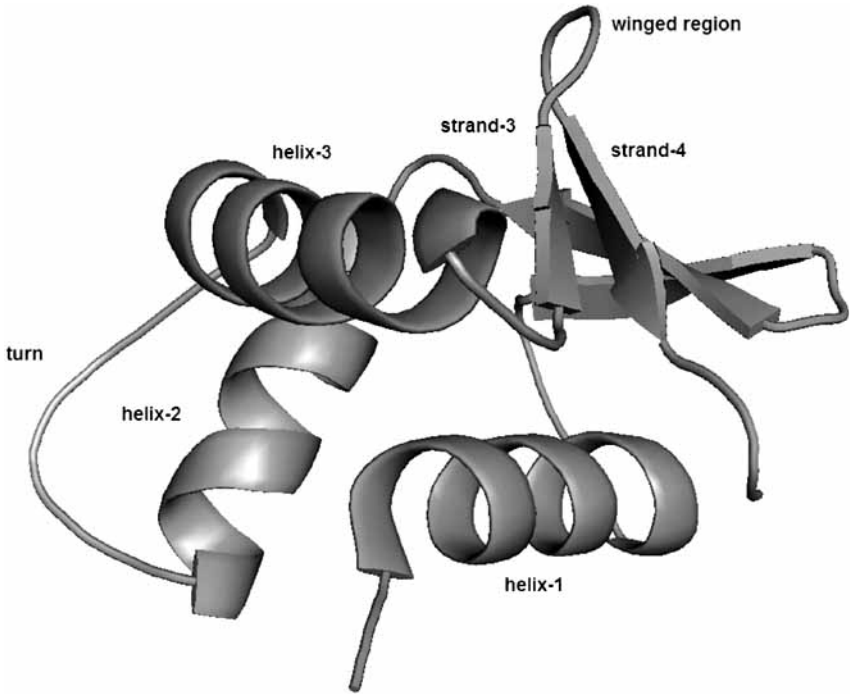


Fig. (1). Tertiary structure of the PU.1 ETS domain [38]. ETS domain is presented as ribbon diagrams. The PU.1 (PDB Code: 1PUE) structure was solved by crystallographic analysis. The DNA helix and crystallographic waters were present in the reported structure and removed in this figure. Three helices (helix-1-helix-3) form a helix-turn-helix (HTH) motif that packs against a four-stranded, antiparallel -sheet (strand-1-strand-4). The term "winged helix-turn-helix" is used to describe this domain and highlights the importance of both the HTH element and the -sheet (wing formed between strand-3 and strand-4) in the structural fold in contacting DNA. The image was generated using Pymol [34].

nism has been put forward in which cooperative effects among neighboring bases flanking the central purine-rich 5´GGA(A/T)3´ core sequence contribute to the formation of stable Ets-DNA complexes [11]. In this indirect readout mechanism, the Ets domain recognizes a sequence-dependent structure that is either induced or already present in DNA. The indirect readout mechanism for Ets domain is supported by X-ray crystallography and NMR. This model is in contrast to a direct readout mechanism of DNA binding specificity based on SELEX results, in which protein residues recognize and interact with unique base pairs within an Ets-binding site (EBS) [11]. In both mechanisms water molecules could potentially mediate contacts between the protein and the DNA sequence [11].

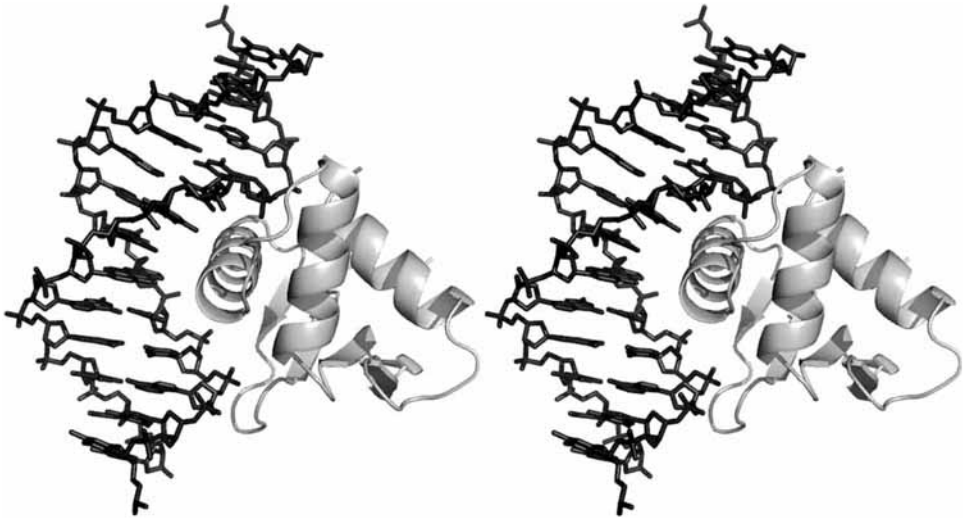


Fig. (2). Structure of Elk-1-DNA complex shown in stereoview. The three dimensional structure of the Elk-1 Ets domain bound to a 13 base-pair DNA duplex [35]. Ets domain is presented as ribbon diagram (light gray) and DNA duplex is presented as stick diagram (dark gray). The Elk-1 (PDB Code: 1DUX) structure was solved by crystallographic analysis. The crystallographic waters present in the reported structure were removed in this figure. The third helix (H3) in the Ets-domain is responsible for making major groove contacts with the central GGAA/T motif.

Structures of Ets domain-DNA complexes of some Ets members have been determined by nuclear magnetic resonance (NMR), which allow to distinguish unique DNA contacts from protein-DNA interactions. Different Ets domains demonstrate differences in DNA sequence specification, despite their overall structural similarities and a shared mechanism of recognition at the core and the flanking regions of their DNA sites. The combination of different amino acids at key locations in individual Ets domains results in different specificity and affinity for DNA sites for each member [9, 16]. Retardation of specific DNA-protein complexes were detected at the Ets site by EMSA (electrophoretic mobility-shift assay). Two Ets-1 isoforms bind and form stable complexes with optimal Ets binding site [17, 18]. The DNA-binding domain of Sp1 was determined by solution NMR, showing that two residues (Lys550 and His553 at positions -1 and 3 of the recognition helix) are likely to form several hydrogen-bonding partners at multiple base positions with consequently less sequence and site specificity [19]. This property allows Sp1 to bind various DNA sequences with high affinity [19].

Ets domain may have some degree of flexibility and diversity in its mode of interaction with DNA. Because Ets proteins have overlapping DNA binding specificities and because their expression is often tissue specific, an understanding of the sequence specificity of Ets proteins could be important for understanding the mechanisms of deregulation in Ets-related cancers [11].

MODULATION OF ETS FUNCTION

Ets functional activity is modulated at multiple levels, including interaction with other factors and binding to target DNA sequence for precise transcriptional regulation. Nucleotides adjacent to the 5´GGAA/T 3´core sequence were found to play a role in DNA binding. These nucleotides may either affect the interaction with other transcription factors or enhance transcriptional activity [17]. A correlation has been found between the immediate flanking nucleotides of the consensus PU.1 binding sequence (GAGGAA) and transcription factor PU.1 binding affinity and promoter activity [20], thereby providing further insight into the mechanism by which the nucleotides flanking the core sequence affect transcription.

Protein-protein interactions are critical for the function of Ets-related proteins and occur with transcription factors of various other families. Analysis of physical interaction of AML1 (a transcription factor critical for definitive hematopoiesis) with NERF-2 (an isoform of *n*ew *E*ts-*r*elated *f*actor) activates the *blk* promoter (Blk is a B cell-specific tyrosine kinase expressed in pre-B and mature B cells), whereas interaction of AML1 with NERF-1a inhibits transactivation [21]. Analysis of gene expression under the control of regulatory elements has detected Ets DNA-binding as indispensable regulatory component for Tlr4 (Toll-like receptor 4) gene expression [17]. *In vitro* and *in vivo* studies have been carried out to determine target sequences of Ets members as well as interactions with other transcription factors. The Ets full protein or the Ets domain have been used to study target sequence specificity and protein-protein interactions, since the Ets domain can be produced as a stable protein fragment and it is sufficient for DNA binding, being a structural and functional domain [12]. Many Ets proteins are end effector molecules of multiple signal transduction pathways, being controlled by phosphorylation that mediates effects on DNA-protein, protein-protein interaction, transcriptional activation and subcellular localization. More recently, acetylation was found to be involved in Ets1 modification. This post-translational modification of Ets factors is dependent on cell-type and appears to have important implications for studies on functionality [22].

STUDIES ON KINETICS OF INTERACTIONS

To understand the control of transcriptional machinery, data on DNA binding thermodynamic and kinetic properties displayed by each transcription factor are required. Surface Plasmon Resonance (SPR) technology, which is an optical label-free detection technique, makes it possible to obtain quantitative information in real-time on specificity of binding, active concentration of molecule in a sample, kinetics and affinity [23]. SPR is thus a powerful tool to study dynamics of both DNA-protein and RNA-protein interactions. A number of authors have used SPR to study interactions between transcription factors and target DNA sequences, obtaining measurements of kinetics and equilibrium binding constants of protein-DNA interactions with high sensitivity [24-28]. These experimental results should be reconciled with molecular dynamic data to help determine the key features to Ets-DNA complex formation. Experimental studies provide a large amount of information for understanding biological systems, including cancer. However, technologies and approaches must still be devised to connect this vast amount of data and to use it to elucidate cellular pathways involved in tumorigenesis. The most obvious immediate benefit of having access to large amounts of data besides sequence information is the ability to use bioinfomatics approaches, such as molecular modeling and molecular dynamics simulations of the complex Ets-DNA.

MOLECULAR MODELING

The obstacles for 3D structure determination by NMR or X-ray crystallography methods have been overcome by a recent area of bioinformatics named molecular modeling. The molecular modeling has become a reliable tool in structural biology because the increase of efficiency in use of comparative models that are detectably related to proteins of known structure. However, with regard to Ets studies the molecular modeling seems to be neglected. This powerful tool could be widely used for proteins that lack experimentally solved 3D structures. Obtaining structural details of macromolecular interactions could provide insights into static conformation of DNA-Ets domain binary complex that could be submitted to dynamic studies by molecular dynamics (MD) simulations. These models of Ets domains should be useful to establish overall structural similarities between two or more related molecules, and to establish the structural key features between members of related subfamilies and their relationship with their diverse biological roles. At present there are several servers for automated comparative modeling that are available for researchers, making this task more user friendly and fully automated [29-31].

MOLECULAR DYNAMICS OF ETS-DNA COMPLEX

Considerable progress in computer simulation has been made in recent years. Improvement in force fields, simulation techniques, protocols and increasing computer power are playing prominent roles in a wide area of research, including molecular dynamics (MD). The use of MD simulation allows monitoring the dynamics of individual atoms thereby giving a unique insight into the molecule behavior that cannot be easily extracted from laboratory experiments.

The rapid accumulation of X-ray and NMR data and increasing computer power have contributed to an increasing number of simulations of protein-DNA complexes. The CHARMM, AMBER and GROMACS are suites that can be employed to simulate the system, in which the complex is solvated within a water sphere and either shift/switch technology for long-range interactions or periodic boundary conditions and particle-mesh Edwald electrostatics is utilized [32]. The application of MD makes possible to predict structural flexibility features of binary complexes of Ets-domain DNA, which allow deepening our understanding of molecular recognition process involving a specific DNA sequence and Ets-domain structure in terms of dynamics motion of the complex. Computer simulations are also powerful tools for analyzing the components of recognition processes. This can involve looking at structural adaptation upon binding or the consequences of point of mutations. The latter has been used to analyze, at molecular level, the effects of point mutations in the DNA core sequence on Ets domain and DNA binary complex formation [33]. These results prompted the authors to propose that the GGAA core sequence (high affinity) is distinguished from the mutated GGAC sequence (low affinity) by a direct readout mechanism in the Ets-1 Ets domain-DNA complex [33]. Analysis of the direct contacts in the interface between the helix-3 region of Ets-1 and the major groove of the core DNA sequence showed that the highly conserved arginine residues (Arg391 and Arg394) play a critical role in binding the GGAA core sequence [33]. Although these arginine residues show flexibility in side-chain conformation, they make bidentate contacts with G_1 and G_2 to stabilize the complex (Fig. **3**) [34]. In Ets-GGAA complex the hydroxyl group of Tyr395 is hydrogen bonded to N7 nitrogen of A_3 (the third adenosine in the GGAA core), whereas the hydroxyl group makes a contact with N4 nitrogen of $C_{4'}$ (the complementary nucleotide of the fourth guanosine G_4 in GGAC sequence) in the Ets-GGAG complex [33]. These MD results showed that this change of Tyr395 interaction partner is a consequence of the relatively large motion of helix-3 in the Ets-GGAG complex, resulting in the collapse of bidentate contacts between Arg391/Arg394 and the G_1G_2 dinucleotide in the GGAG sequence.

The affinity of binary complex formation between transcription factors and their target DNA sequences is determined by both the network of interactions at the interface and the entropy change associated with the complex formation. Despite SAP-1 and Elk-1 proteins having identical amino acid sequence for the DNA recognition helix within their Ets domains, they exhibit different DNA binding properties. A comparison of the crystal structures of binary complexes formed between SAP-1 or Elk-1 and a high affinity

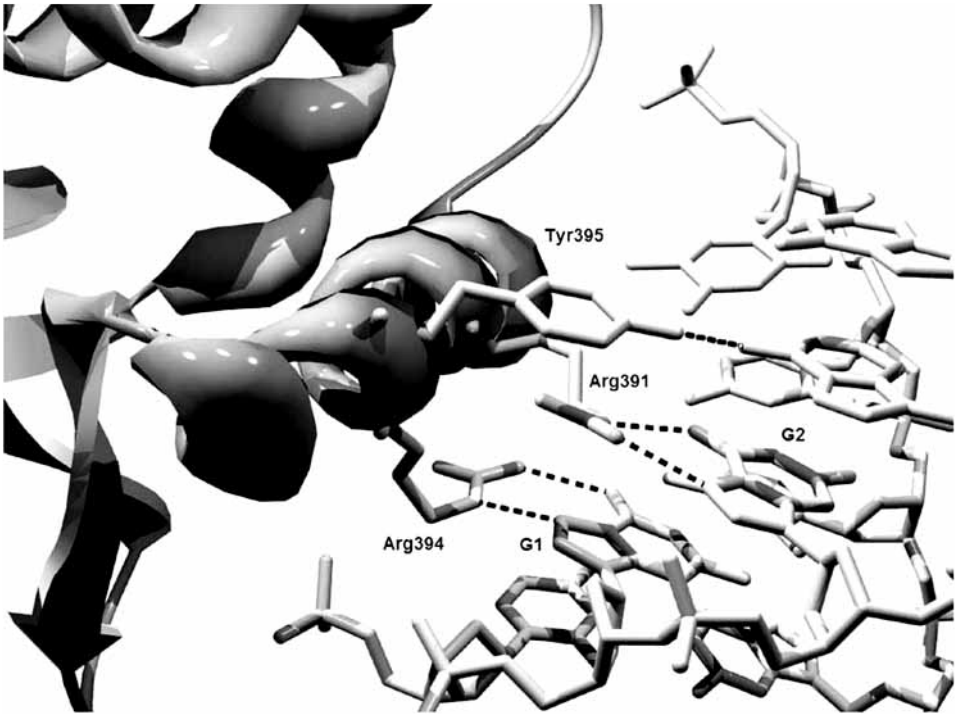


Fig. (3). Crystal structure of Ets-1 ES domain-DNA complex. (PDB Code: 1K79). The DNA double helix is represented as stick diagram (dark gray) and de ETS domain as ribbon diagram (light gray) The close-up of the Ets domain-DNA nucleobase contact site (helix-3 and GGAA core sequence) shows the interactions of the Arg391 and Ag394 with G_2 and G_1 nucleotide respectively and Tyr395 hydroxyl group with A_3. These two arginines residues, which are in the helix-3 region and conserved among the Ets family, make bidentate interactions. However the pattern of these interactions is not maintained in others structures. The image was generated using SwissPDB viewer [39] and rendering with POV-ray v. 3.5.

DNA sequence showed that the differential DNA binding properties of these proteins are mediated by non-conserved residues distal to the DNA binding surface that function to orient conserved residues in the DNA recognition helix [35]. More recently, MD dynamics simulations of Elk-1 and SAP-1 were carried out to study the role of structural fluctuation in fine-tuning DNA affinity [36]. These studies showed that several residues in Elk-1 have significantly higher main-chain root-mean-square deviations than their counterparts in SAP-1. In particular, a single residue, D69, may contribute to Elk's lower DNA affinity by structurally destabilizing the carboxy terminus of the recognition helix. While D69 does not contact DNA directly, the increased mobility in the region may contribute to its weaker binding. Measurements of the ability of single point mutants of Elk-1, in which D69 of wild-type Elk-1 was replaced with residues with higher helix propensity in order to stabilize the local conformation, to bind its target DNA sequence showed that gains in transcriptional activity and the free energy of binding correlate well with stability gains computed from helix propensity and charge-macrodipole interactions [36]. The lower helix propensity of an amino acid residue in the carboxy-terminal end of the recognition helix (H3) appears to destabilize H3 and increase the entropic cost of DNA binding [36]. Accordingly, this study suggests that residues that are distal to the binding interface may indirectly modulate the binding affinity by stabilizing the protein scaffold required for efficient DNA interaction. MD studies have been performed on the Ets domain of Ets-1 transcription factor bound to the 14-bp DNA, d(AGTGCCGGAAATGT), comprising the core sequence of high-affinity (GGAA), Ets-GGAA, and compared with the mutant low-affinity core sequence, GGAG (Ets-GGAG) [37]. This single mutation resulted in variations of helical parameters and solvent-accessible surface area around the major and minor grooves of the DNA. The presence of water contacts to the fourth base pair step of core DNA sequence appears to be a characteristic feature of Ets-GGAA. These results indicate water contacts are involved in stabilizing the hydrogen bond interactions between helix-3 residues of Ets-1 and DNA during the transcription process.

CONCLUDING REMARKS

Computer simulations are becoming a powerful tool for analyzing biological processes and proving to be a reliable way to understand the static and dynamic aspects in solution, avoiding difficulties that can occur in a crystalline environment. The results of computer simulations could help to explain how these complexes respond to their environment and recognitions system of protein in DNA core motif. In addition, owing to improvements of force fields, molecular dynamic techniques and protocols can offer more reliable model proposals with high predictive accuracy. As seen above, MD simulations can help us understand how Ets-domain proteins recognize the same DNA sequence with differing affinity by simulating the dynamics of these Ets and comparing their backbone flexibility.

The structural characterization of Ets-DNA complexes is inaccurate in various enthalpic components, including key hydrogen bonds and salt bridges. However most Ets-DNA structures present a single, or at most a few, static snapshots of the complexes, giving a limited structural basis for a detailed analysis. The possibility of investigating in a dynamic environment should enhance our understanding of binary complex motions, protein-DNA recognition and affinity. A detailed structural description of interactive networks among different amino acid residues of Ets-domain proteins and DNA target sequences, and their effect on interaction parameters should provide a framework on which to base the prediction of outcomes of these interactions. It is hoped that this understanding will provide new therapeutic options to treat cancer in which Ets-domain proteins play important roles.

ACKNOWLEDGEMENTS

This work was supported by grants from Millennium Institute, MCT/CNPq, PRONEX, FAPERGS, and Hewlett-Packard of Brazil to DSS and LAB. WFA also thanks Millennium Institute, MCT/CNPq for financial support. DSS, LAB and WFA are research career awardees from the National Research Council of Brazil (CNPq). RC and CPN thank Hewlett-Packard of Brazil for fellowships.

REFERENCES

[1] Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell* **2000**; 100: 57-70.

[2] Seth A, Watson DK. ETS transcription factors and their emerging roles in human cancer. *Eur J Cancer* **2005**; 41: 2462-78.

[3] Sementchenko VI, Watson DK. Ets target genes: past, present and future. *Oncogene* **2000**; 19: 6533-48.

[4] Mao X, Miesfeldt S, Yang H, Leiden JM, Thompson CB. The Fli-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *J Biol Chem* **1994**; 269: 18216-22.

[5] Singh S, Barrett J, Sakata K, Tozer RG, Singh G. ETS proteins and MMPs: partners in invasion and metastasis. *Med Chem Rev* **2004**; 1: 39-46.

[6] Galang CK, Muller WJ, Foos G, Oshima RG, Hauser CA. Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. *J Biol Chem* **2004**; 279: 11281-92.

[7] Sharrocks AD, Brown AL, Ling Y, Yates PR. The Ets-domain Transcription Factor Family. *J Biochem Cell Biol* **1997**; 29: 1371-87

[8] Donaldson LW, Petersen JM, Graves BJ, McIntosh LP. Secondary structure of the ETS domain places murine Ets-1 in the superfamily of winged helix-turn-helix DNA-binding proteins. *Biochemistry* **1994**; 33: 13509-16.

[9] Pio F, Kondandapani R, Ni C, *et al*. News Insights on DNA Recognition by ets Proteins from Crystal Structure of the PU.1 ETS Domain-DNA. *J Biol Chem* **1996**; 31: 23329-37

[10] Suzuki M. Common features in DNA recognition helices of eukaryotic transcription factors. *EMBO J* **1993**; 12: 3221-6.

[11] Szymzcyna BR, Arrowsmith CH. DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. *J Biol Chem* **2000**; 275: 28363-70.

[12] Graves BJ, Petersen JM, Specificity within the ets family of transcription factors. *Adv Cancer Res* **1998**; 75: 1-55.

[13] Hollenhorst PC, Jones DA, Graves BJ. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Res* **2004**; 32: 5693-702.

[14] Schumacher MA, Lau AOT, Johnson PJ. Structural basis of core promoter recognition in a primitive eukaryote. *Cell* **2003**; 115: 413-24.

[15] Donaldson LW, Petersen JM, Graves BJ, McIntosh LP. Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J* **1996**; 15: 125-34.

[16] Wang Y, Feng L, Said M, Balderman S, Fayazi Z, Liu Y, Ghosh D, Gulick AM. Analysis of the 2.0 Å Crystal Structure of the Protein-DNA Complex of the Human PDEF Ets Domain Bound to the Prostate Specific Antigen Regulatory Site. *Biochemistry* **2005**; 44: 7095-106.

[17] Roger T, Miconnet I, Schiesser AL, Kai H, Miyake K, Calandra T. Critical role for Ets, AP-1 and GATA-like transcription factors in regulating mouse Toll-like receptor 4(Tlr4) gene expression. *Biochem J* **2005**; 387: 355-65.

[18] Fisher RJ, Fivash M, Casas-Finet J, *et al*. Real-time DNA binding measurements of the ETS1 recombinant oncoproteins reveal significant kinetc differences between the p42 and p51 isoforms. *Protein Sci* **1994**; 3: 257-66.

[19] Oka S, Shiraishi Y, Yoshida T, Ohkubo T, Sugiura Y, Kobayashi Y. NMR Structure of Transcription Factor Sp1 DNA Binding Domain. *Biochemistry* **2004**; 43: 16027-35.

[20] Li S, Schlegel W, Valente AJ, Clark RA. Critical flanking sequences of PU.1 binding sites in myeloid-specific promoters. *J Biol Chem* **1999**; 274: 32453-60.

[21] Cho J, Akbarali Y, Zerbini LF, *et al.* Isoforms of the Ets transcription factos NERF/ELF-2 phisically interact with AML1 and mediate opposing effects on AML1-mediated transcription of the B cell-specific *blk* gene. *J Biol Chem* **2004**; 279: 19512-22.

[22] Hsu T, Trojanowska M, Watson DK. Ets proteins in biological control and cancer. *J Cell Biochem* **2004**; 91: 896-903.

[23] Basso LA, Silva LHP, Fett-Neto AG, *et al*. The use of biodiversity as source of new chemical entities against defined molecular targets for treatment of malaria, tuberculosis, and T-cell mediated diseases-a review. *Mem Inst Oswaldo Cruz* **2005**; 100: 475-506.

[24] Shumaker-Parry JS, Aebersold R, Campbell CT. Parallel, quantitative measurement of protein binding to a 120-element double-stranded DNA array in real time using surface plasmon resonance microscopy. *Anal Chem* **2004**; 76: 2071-82.

[25] Galio L, Briquet S, Cot S, Guillet J, Vaquero C. Analysis of interactions between huGATA-3 transcription factor and three GATA regulatory elements of HIV-1 long terminal repeat, by surface plasmon resonance. *Anal Biochem* **1997**; 253: 70-7.

[26] Brockman JM, Frutos AG, Corn RM. A multistep chemical modification procedure to create DNA arrays on gold surfaces for the study of protein-DNA interactions with surface plasmon resonance imaging. *J Am Chem Soc* **1999**; 121: 8044-51.

[27] Smith EA, Erickson MG, Ulijasz AT, Weisblum B, Corn RM. Surface plasmon resonance imaging of transcription factor proteins: interactions of bacterial response regulators with DNA arrays on gold films. *Lagmuir* **2003**; 19: 1486-92.

[28] Katsamba PS, Park S, Laird-Offringa IA. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* **2002**; 26: 95-104.

[29] Uchoa HB, Jorge GE, da Silveira NJ, Camera JC, Canduri F, de Azevedo WF. Parmodel: a web server for automated comparative modeling of proteins. *Biochem Biophys Res Commun* **2004**; 325:1481-6.

[30] Silveira NJF, Bonalumi CE, Uchôa HB, Pereira JH, Canduri F, de Azevedo WF. DBMODELING-A Database Applied to the Study of Protein Targets From Genome Projects. *Cell Biochem Biophys* **2006**; 44: 366-74.

[31] Silveira NJF, Bonalumi CE, Arcuri HA, de Azevedo WF. Molecular modeling databases: a new way in the search of protein targets for drug development. *Curr Bioinformatics* **2007**; 2: 1-10.

[32] Giudice E, Lavery R. Simulations of Nucleic Acids and Their Complexes. *Acc Chem Res* **2002**; 35: 350-7.

[33] Obika S, Reddy SY, Bruice TC. Sequence specific DNA binding of Ets-1 transcription factor: molecular dynamics study on the Ets domain-DNA complexes. *J Mol Biol* **2003**; 331: 345-59.

[34] DeLano Wl. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA, 2002 [http://www.pymol.org].

[35] Mo Y, Vaessen B, Johnston K, Marmorstein R. Structure of the elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA. *Nat Struct Biol* **2000**; 7: 292-7.

[36] Park S, Boder ET, Saven JG. Modulating the DNA affinity of Elk-1 with computationally selected mutations. *J Mol Biol* **2005**; 348: 75-83.

[37] Reddy SY, Obika S, Bruice TC. Conformations and dynamics of Ets-1 ETS domain-DNA complexes. *Proc Natl Acad Sci USA* **2003**; 100: 15475-80.

[38] Kodandapani R, Pio F, Ni CZ, *et al*. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* **1996**; 380: 456-60.

[39] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**; 18: 2714-23 [http://www.expasy.org/spdbv].

Received: March 30, 2007 Revised: May 11, 2007 Accepted: May 15, 2007

6. ARTIGO CIENTÍFICO 2

Este artigo encerra os dados obtidos ao longo dos 18 meses de pesquisa, o qual, constitui a minha dissertação de mestrado. A partir da seqüência primária dos domínios das proteínas ETS estudadas (NERF, EV-2, SPI-C) obtivemos suas estrtuturas 3D e desenvolvemos um protocolo viável, confiável e rápido para simulação de *docking* e análise dos mesmo. A partir dos complexos binários ETS-DNA, utilizanddo técnicas de dinâmica molecular analisamos a estabilidade e principalmente o comportamento destes sistemas em solução, permitindo um detalhamento dos aspectos moleculares ainda desconhecidos.

Molecular modeling and docking of functional groups of Ets involved in human DNA binding domain

Rafael Andrade Caceres[1], Luis Fernando Saraiva Macedo Timmers[1], Luiz Augusto Basso[1], Walter Filgueira de Azevedo Jr[1]., Diógenes Santiago Santos[*2]

[1] Faculdade de Biociências - Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga, 6681. Porto Alegre-RS CEP 90619-900, Brazil.
[2] Pontifícia Universidade Católica do Rio Grande do Sul, Centro de Pesquisa em Biologia Molecular e Funcional, Porto Alegre, RS, Brazil.

[*]Corresponding author (diogenes@pucrs.br)

Introduction

The E26 transformation-specific (ETS) family is restricted to metazoans and thus represents an evolutionarily recent class of transcription factors [1-3].It is one of the largest transcription factors families, consisting of 27 ETS genes in humans, 26 in mice, 10 in *Caenorhabtitis elegans* and 9 in *Drosophila*, and are divided into several subfamilies based on homology within ETS domain (ED), that can be categorized into 11 subfamilies (ETS, ERG, ELG, ELF, ESE, ERF, TEL, PEA3, SPI, TCF and PDEF)[4;5].

The ETS proteins share a homologous sequence in the ED and correspond to the DNA-binding domain, they differ in length and in the relative position of this domain. In some molecules, the ED is found at the carboxyl terminus (e.g. PU.1 [6], Ets-1 [7] and Ets-2 [8]), while in others the domains is located in the middle of the sequence (ERG [9]), or in the amino-terminal region (ELK-7 [10]). Ets proteins are proposed to be a part of the superfamily of winged helix-turn-helix DNA-binding proteins [11] and consist of three α-helices and four β-strands. This conserved region, binding domain, corresponds to ~85 amino acids, located at the N-terminus in most of proteins. The ED is composed of three alpha-helices (H) and four beta-strands (S) arranged in the order H1-S1-S2-H2-H3-S3-S4 [11]. Donaldson et al., proposed for Ets family, the nomenclature wHTH (winged helix-turn-helix) where helices 2 and 3 correspond to helix-turn-helix motif, the DNA recognition region. HTH motif makes specific contact to the major groove of DNA. The winged is a loop made by connecting the last two β-strands on one side and a long loop at C-terminus on the other side. Previous X-ray and NMR studies

have shown that helix-3 in the winged helix-turn-helix motif bind in the major groove of the consensus DNA sequence, or either, is responsible for making DNA recognition. The third helix in the ED is responsible for making major grove contacts with the central 5'-GGAA/T-3' motif, whilst additional DNA contacts are made by a loop between the β-strands 3 and 4 (the wing) and the loop between α-helix 2 and 3.

Considering that some ETS transcriptions factors regulate growth- apoptosis, angiogenesis-, invasions and metastasis-related genes in tumor cells, these transcription factors we proposed and describes three-dimensional models of ETS (ETV-2, SPI-C and NERF) by applying computational homology-modeling techniques and utilizing the high-resolution crystals structures of Ets-human DNA complexes as templates. These models should be useful to establish overall structural similarities between two or more related molecules, and establish the structural key features between members of related subfamilies and their relationship with their diverse biological roles. Obtaining structural details of macromolecular interactions provide insights into static conformation of DNA-Ets domain binary complexes [12], which may become a molecular targeting therapy against ETS transcription factors that could be a novel approach to a selective cancer treatment in the future.

Biological roles

ETS-domains transcription factors have been linked with diverse biological processes, but not clear unifying theme has emerged. This is consistent with the

observation that these proteins function differently at the molecular level [2]. Our study focus at three Ets protein as follow: SPI-C, NERF, and ETV-2. These proteins can be categorized into subfamilies as show Table 1.

NERF

The NERF was identified and characterized cDNA clones encoding three alternative splice products of a novel member of the Ets gene family, NERF/ELF-2, NERF-1a, NERF-1b, and NERF-2, which differ in their N-terminal [13]. NERF is most closely related to ELF-1 and MEF/ELF-4. It was demonstrated that both, NERF and the related protein ELF-1 are involved in regulating a set of genes in B cells and myeloid cells and are highly expressed in these cells [13,14]. The NERF-2 is expressed in endothelial cells and transactivates the regulatory regions of the *tie2* gene [15]. Interestingly, NERF-2 expression is also increased in endothelial cells in response to hypoxia and to angiopoietin-1, indicating functions for NERF in the immune system and vasculature [16].

SPI-B/SPI-C

PU.1 is a member of a subfamily of Ets transcription factors that includes two other closely related factors, Spi-B and Spi-C. Spi-B exhibits 43% overall amino acid sequence identity to PU.1, but the Ets domain is 68% identical to PU.1. Spi-C (also called Prf) shares 40% overall sequence identity with PU.1, and shares 59%identity in the DNA binding domain [17,18]. PU.1 and Spi-B function as

strong transcriptional activators, while Spi-C has weak transcription activation ability [18]. PU.1 is expressed in multiple hemopoietic lineages, including all stages of B cell development, while the expression of Spi-B is restricted to B cells and T cells [19-21]. Targeted null mutation of Spi-B does not affect B cell development, but results in defective BCR-mediated responses [22;23]. Both PU.1 and Spi-B are expressed in the B cell lineage beginning at the pro-B cell stage and appear to interact with identical DNA binding sites. Therefore in developing B cells, the available evidence indicates that PU.1 and Spi-B are functionally interchangeable [23;24].

ETV-2

The overexpression of certain ETS protein has been associated with various diseases. For instance, transgenic mouse studies have shown that Ets-2 overexpression leads to a phenotype that is reminiscent of the one observed in Down's syndrome individuals, consistent with the location of the human Ets-2 gene on chromosome 21 [25].

Furthermore, mice that overexpress the ETS factor Fli-1 are affected by a systemic lupus erythematosus-like disorder [26]. Despite these studies, the precise role for most ETS factors remains to be elucidated.

ETV-2 also known as ER71 is a member of the ETS factor family that, until now, has not been studied in detail. It is a 336 amino acid protein with a calculated molecular weight of 37.2 kDa whose expression appears to be testis-specific [27]. ETV-2 is a remarkably unique protein; the degree of identity between its

domain and those other factors is at most 67%, while the amino acid sequences that lie outside its ETS domain do not share homology with any other protein in current databases.

ETV-2 is a constitutively nuclear protein whose intracellular localization is dependent on a portion of ETS domain. Furthermore, the DNA binding activity is intramolecularly regulated, as the N-terminus of ETV-2 has a negative effect on DNA binding while the C-terminus dramatically enhances this activity. Were also demonstrated that ETV-2 possesses an extremely potent N-terminal transactivation domain comprised of amino acids 1-157 and that ETV-2 is capable of directly activating both an E74 site-driven and the matrix metalloproteinase-1 promoter [28].

Material and Methods

Molecular modeling

Homology modeling is usually the method of choice when there is a clear relationship of homology between the sequences of a target protein and at least one experimentally determined three-dimensional structure. This computational technique is based on the assumption that tertiary structures of two proteins will be similar if their sequences were related.

Model building of the complexes was carried out using the program Parmodel [29], which is a web server for automated modeling and protein structural assessment. Parmodel runs a parallelized version of MODELLER version 7 [30].

MODELLER is an implementation of an automated approach to comparative modeling by satisfaction of spatial restraints. The modeling procedure begins with alignment of the sequence to be modeled (target) with the sequence of related known three-dimensional structures in complex with the inhibitor or human DNA (templates). This alignment is usually the input to the program. The output is a three-dimensional model for the target sequence containing all main-chain and sidechain nonhydrogen atoms [30].

The homology models of ETS were based on the atomic coordinates of ETS member family complexed with DNA, using E value as one selection criterion, except of ETS-1 which structure is available in the PDB (PDB code 1GVJ). The ETS-DNA complexes were defined as another selection criterion, because the DNA duplex induces a conformational changes in the amino acids, so the deformation caused by DNA complexaton is transferred to model built. This important aspect was taking account because the built models will be subjected a rigid-body docking of the DNA consensus sequence 5'-GGAA/T-3' extract from 3D structure template.

A total of 1000 models were generated by the program Parmodel [29], the final model was select based on and the final models were selected based on the MODELLER objective function. The model with the lowest value for the MODELLER objective function was selected as the best model. All optimization process was performed on a Beowulf cluster with 16 nodes (Bio-Comp, AMD Athlon XP 2100+, Brazil).

Analysis of models

The overall stereochemical quality of the final models of ETS were assessed by the program PROCHECK [31]. The cutoff for hydrogen bonds and salt bridges was 3.5 Å. For superposition of Cα, we used the program Swiss-PDBViewer v3.7 [32]. The root mean square deviations (RMSD) from ideal geometries for bond lengths, bond angles, dihedrals and impropers were calculated with X-PLOR [33]. The program VERIFY-3D was used to assess structural quality and compatibility of the homology models [34, 35]. The G-factor value is essentially just log-odds score based on the observed distributions of the stereochemical parameters. It is computed for the following properties: torsion angles (the analysis provided the observed distributions of ϕ–ψ, χ1–χ2, χ-1, χ-3, χ-4, and ω values for each of the 20 amino acid types) and covalent geometry (for the main-chain bond lengths and bond angles). These values average was calculated using PROCHECK [31]. The intermolecular hydrogen bonds of ETS-DNA complexes were assessed by the program LIGPLOT [36].

Rigid-body docking simulations

Predicting protein–nucleic acid interactions by rigid body docking is more difficult compared to protein–protein interactions as DNA and RNA are highly charged and more flexible than proteins [38]. ZDOCK 2.3 [38;39] has used for the first time in its ability to predict protein–DNA contacts with high accuracy and good predictive ability [40].

The rigid-docking protocol developed for this study was validated using ELK-1 complexed with thirteenth DNA base pair under PDB code 1DUX. The ELK-1 ETS and DNA fragment were split in two files and the crystallographic waters were kept in the protein file using Swiss-PDBViewer v3.7 [32]. The ETS-protein was rotated 180° along z axis and translated 1(fractional coordinates) along the three axis. The DNA was parameterized by adding the missing atomic radii, atomic contact energies (ACE) and atomic charges, these parameters were used by Fanelli *et al*.[40]

The DNA fragment was kept fixed (i.e., receptor), whereas the protein ELK-1 was allowed to rotate and translate around the receptor. A rotational sampling interval of 15° was used, i.e, dense sampling, and the best 50 solutions were retained and ranked according to the lowest RMSD in relation of crystallographic structure. The best solution given was of 1.01 Å of RMSD, proving the high potential of protocol to predict ETS-DNA complex 3D structure. All the ETS models obtained were subjected to this protocol and the best solutions compared with lowest Cα RSMD of helix-alpha 3 in relation with their 3D structures modeling templates.

Molecular dynamics simulations protocol

In order to obtain an optimized complex ETS-DNA models and examines detailed informations on the structures and its dynamical properties molecular dynamics (MD) simulations have been performed.

The (MD) were performed with the GROMACS[41] vesion 3.3.1 package using the Gromos 96.1 (53A6) [42] force field. Manipulation of structures was performed with Swiss-PDBViewer v3.7 program [32]. The simulations of nine systems were performed by a time period of 4 ns. In all systems were added Na^+ counter ions using *Genion* Program of the GROMACS simulation suite to neutralize the negative charge density of the systems. For each Ets were performed two simulations, the proteins in apo form and other simulation with ETS-DNA complex obtained from rigid *docking* simulations.

Each structure was placed in the center of a truncated cubic box filled with Simple Point Charge Extended (SPC/E) water molecules [43].The initial simulation cell dimensions and had the ETS protein solvated by a layer of water molecules of at least 10 Å in all directions in all systems. During the simulations, bonds lengths within the proteins were constrained by using LINCS algorithm.[44] The SETTLE algorithm was used to constrain the geometry of water molecules.[45] In the MD protocol, the binary complex, ions, and water molecules were first subjected to 1500 steps of energy minimization by steepest descent followed 1500 steps of conjugate gradient to remove close van der Waals contacts. The systems were then submitted to a short molecular dynamic with position restrains for period of 20 ps and afterwards performed a full molecular dynamics without restrains. The temperature of the system was then increased from 50 K to 300 K in 5 steps (50 K to 100 K, 100 K to 150 K, 150 K to 200 K, 200 K to 250 K, 250 K to 300 K), and the velocities at each step were reassigned according to the Maxwell-Boltzmann distribution at that temperature and

equilibrated for 10 ps except the last part of termalization phase, which was for 40 ps. Energy minimization and MD were carried out under periodic boundary conditions. The simulation was computed in the NPT ensemble at 300 K with the Berendsen temperature coupling and constant pressure of 1 atm with isotropic molecule-based scaling.[46] The LINCS algorithm, with a 10^{-5} Å tolerance, was applied to fix all bonds containing a hydrogen atom, allowing the use of a time step of 2.0 fs in the integration of the equations of motion. No extra restraints were applied after the equilibration phase. The electrostatic interactions between atoms were evaluated by the particle-mesh Ewald method [47] with a charge grid spacing of 1.0 Å and a four-order spline, the charge grid was interpolated on a cubic grid with the direct sum tolerance set to 1.0 X 10^{-5}. The Lennard-Jones interactions were evaluated using a 10.0 Å atom-based cutoff.

All analysis were performed on the ensemble of system configurations extracted at 0.5-ps time intervals from the simulation and MD trajectory collection was initiated after 2 ns of dynamics to guarantee a completely equilibrated evolution. The MD simulation and results analysis were performed in a personal compute Intel DuoQuadiCore 2 Duo E6300 – 1.86 GHz and 4 Gb RAM.

The convergences of the different simulations were analyzed in terms of the root mean-square deviation (RMSD) from the initial models structures, order parameter *S2*, for ϕ/ψ torsion angles of the protein backbone of the MD-averaged structures (DNA bound and unbound), and root mean-square fluctuation (RMSF) to estimate flexibility of protein domains. For the S2 and RMSFs were calculated relative to the last 200 ps averaged backbone structures, and all coordinate

frames from the trajectories were first superimposed on the initial conformation to remove any effect of overall translation and rotation.

Results and Discussion

Primary sequence comparison

The alignment of ETS templates and their targets are shown in **Figure 1**. The sequences alignment ranged of 37% to 68% (**Table 1**) of identity, which was sufficiently reliable to generated models with high accuracy.

Quality of the models

Ramachandran plot for the five ETS-DNA complexes structures solved by crystallography were generated in order to compare the overall stereochemical quality of nine ETS modeled structures. Analysis of the Ramachandran ϕ – ψ plots for the models present a minimum of 89.3% (ETV-2) of the residues in the most favored regions, a maximum of 2.4% (ETV-2) of the residues in the generously allowed region and none in the disallowed regions, whereas structures solved by crystallography (templates) present a minimum of 86.8% (ETS1, PDB code 1K78) of the residues in the most favored regions also a maximum of 1.3% (PU-1, PDB code 1PUE) of the residues in the generously allowed region and none in the disallowed regions (**Table 2**). The Verify3D values, the average G-factor and RMSD values of bond lengths and bond angles are presented in **Table 3** and **Table 2**, respectively. From analysis of the overall

stereochemical quality of the molecular models, we feel that it is appropriate for structural studies.

Overall description and analysis of the interactions between molecular models and DNA

All models of ETS obtained maintained the HTH and winged motif which are the main characteristic of this superfamily, consisting at least of three α-helices and four β-strands. The complexes were superposed with their templates and ranked by lower RMSD between helix 3 which, as previously describe is responsible for making DNA recognition. The RMSD for each superposition is given Table 3.
In order to determine the amino acids which make contact with major grove of DNA with the central 5'-GGAA/T-3' the complexes were assessed by LIGPLOT [36] that gives intermolecular hydrogen bonds.

MD simulations

We performed molecular dynamics simulations of the ETS structures in the apo form (unbound DNA), and ETS-DNAA complexed, in order to elucidate the influence of the DNA on the overall structure of the ETS domains. The root-mean square deviation (RMSD) of the positions for all backbone Cα atoms from their initial configuration as a function of simulation time for all systems were calculated and are shown in Figure 2. Analysis of this figure indicates that the structure without DNA presents the high RMSD when compared with the complex structure.

The flexibilities of the proteins were assessed by the RMSF's from MD the trajectory which reflects the flexibility of each atom residue in a molecule **(Fig. 3).** The major backbone fluctuation occurs in the loop region and in the region surrounding the beta-alpha-beta fold, whereas regions with the low RMSF correspond exclusively to the rigid beta-alpha-beta fold. These facts indicate the stability of our model structures bounded in DNA sequence, the residues which compound the helix-3 in all system presents a lower RMSF as expect, since the DNA reduce its mobility.

Dihedral Order Parameters of ETS Proteins

The order parameter, *S2*, for N–Cα and Cα–C vectors (related to ϕ and ψ torsions, respectively), provides additional insights about correlated dynamics between different residues of protein bound to the DNA . S2 gives a measure of the system flexibility, being 1 in a completely rigid system or 0 in a system for total flexibility where all possible conformations are sample. **Figure 4** shows these order parameters for ETS domains. There is a loop composed by residues 14-20 which presents a pattern profile in apo form (unbound DNA). The SpiC ETS domain bound in the DNA show a higher mobility in the "winged" composed by residues ARG77-LYS79 **(Figure 4A)**, on the contrary of ETV2 "winged" (composed by residues GLY74-LYS77)**(Figure 4B)**, and was observed a loop (ARG48-LYS72) which presents a higher flexibility. The dihedral angles visiting a relatively limited conformational space in bound DNA sequence in the helix-3.

The NERF domain not presents abnormally results, except a high value for GLY34 which can not be explained.

Interaction between the helix-3 of ETS domain and DNA

In order to identify which residues interact with major groove with the central 5'-GGAA/T-3' motif were calculated the number of hydrogen bonds during the time evolution of MD simulation **(Table 4)**. In all of complexes simulated were observed the same patter of formation of bidentade complex composed by two ARG, in the NERF besides the ARG59 and ARG62 the presence of TYR64 it becomes its profile seemed of ETS-1 observed by Bruice *et al.* [48].

In order to understand part of DNA behavior the major grove widths **(Figure 5)** was analyzed during over simulation. Major groove widths are defined as distance between two appropriate phosphate atoms. The average major widths of seven base-pair DNA in homeodomain–DNA complex are summarized in Figure 5. The graphics showm a average of 18 Å of widths in those simulation demonstrating that widths is constant and stable independent of ETS complexed.

Conclusions

In this work the molecular modeling by homology was first time using for obtain a structural models of ETS proteins. The molecular models o ETS generated possess good stereochemical quality and appropriate for docking studies.

It was conducted 4ns MD simulations of ETS domain – ~13 base-pair DNA complexes with PME treatment of electrostatic interactions. These MD simulations have provided a good deal of information on the sequence specific interaction between ETS domain and the consensus DNA and which residue makes contacts between de helix-3 and DNA consensus.

References

[1] Seth, A. and Watson, D. K. *Eur. J. Cancer.* 2000; 41:2462-2478

[2] Sharrocks, A. D. *Nature.* 2001; 2:827.

[3] Wasylyk, B., Hangman, J. and Gutierrez-Hartmann, A. *Trends Biochem. Sci*. 1998; 279:213.

[4] Galang, C. K., Muller, W. J., Foos, G., Oshima, R. G. and Hauser, C. A. *J. Biol. Chem.* 2004; 279:11281.

[5] Hollenhorst, P. C., Jones, D. A. and Graves, B. J. *Nucleic Acids Res.* 2004; 32:5693.

[6] Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., and Maki, R. A. *Cell.* 1990; 61**:**113.

[7] Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J., and Papas, T. S. *Proc. Natl. Acad. Sci. U. S. A.* 1985; 82: 7294.

[8] Reddy, E. S., and Rao, V. N. *Oncogene Res.* 1988; 3:239.

[9] Reddy, E. S., Rao, V. N., and Papas, T. S. *Proc. Natl. Acad. Sci. U. S. A.* 1987; 84:6131.

[10] Rao, V. N., Huebner, K., Isobe, M., ar-Rushdi, A., Croce, C. M., and Reddy, E. S. *Science.* 1989; 244:66.

[11] Donaldson L.W., Petersen J.M., Graves B.J. and McIntosh L.P. *Biochemistry*, 1994; 46:13509.

[12] Caceres, R. A., Nunes, C. P., de Azevedo Jr., W. F., Basso, L. A. and Santos, D. S. *Current Bioinformatics*, 2007; 2:222.

[13] Oettgen, P., Akbarali, Y., Boltax, J., Best, J., Kunsch, C., and Libermann, T. A *Mol. Cell. Biol.* 1996; 16: 5091.

[14] Akbarali, Y., Oettgen, P., Boltax, J., and Libermann, T. A. *J. Biol. Chem.* 1996; 271: 26007.

[15] Dube, A., Akbarali, Y., Sato, T. N., Libermann, T. A., and Oettgen, *P. Circ. Res.* 1999; 84: 1177.

[16] Christensen, R. A., Fujikawa, K., Madore, R., Oettgen, P., and Varticovski, L.

J. Cell. Biochem. 2002; 85: 505.

[17] Bemark, M., A. Martensson, D. Liberg, and T. Leanderson. *J. Biol. Chem.* 1999; 274:10259.

[18] Hashimoto, S., H. Nishizumi, R. Hayashi, A. Tsuboi, F. Nagawa, T. Takemori, and H. Sakano. *Int. Immunol.* 1999; 11:1423.

[19] Klemsz, M. J., S. R. McKercher, A. Celada, C. Van Beveren, and R. A. Maki. *Cell.* 1990; 61:113.

[20] Hromas, R., A. Orazi, R. S. Neiman, R. Maki, C. Van Beveran, J. Moore, and M. Klemsz. *Blood.* 1993; 82:2998.

[21] Su, G. H., H. S. Ip, B. S. Cobb, M. M. Lu, H. M. Chen, and M. C. Simon. *J. Exp. Med.* 1996; 184:203.

[22] Su, G. H., H. M. Chen, N. Muthusamy, L. A. Garrett-Sinha, D. Baunoch, D. G. Tenen, and M. C. Simon. *EMBO J.* 1997; 16:7118.

[23] Garrett-Sinha, L. A., G. H. Su, S. Rao, S. Kabak, Z. Hao, M. R. Clark, and M. C. Simon. *Immunity.* 1999; 10:399.

[24] Rao, S., L. A. Garrett-Sinha, J. Yoon, and M. C. Simon. *J. Biol. Chem.* 1999; 274:34245.

[25] Sumarsono, S. H., Wilson, T.J., Tymms, M.J., Venter, D.J., Corrick, C.M., Kola, R., Lahoud, M.H., Papas, T.S., Seth, A. and Kola, I. *Nature*. 1996; 379:534.

[26] Zhang, L., Eddy, A., Teng, Y.T., Fritzler, M., luppel, M., Melet, F. and Bernstein, A. *Mol. Cell. Biol.* 1995; 15:6961.

[27] Brown, T. A. and McKnigth, S.L. *Science*. 1992; 6:2502.

[28] De Haro, L. and Janknecht, R. *Nucleic Acids Res.* 2002; 30:2972.

[29] Uchoa HB, Jorge GE, da Silveira NJ, Camera JC, Canduri F, De Azevedo W. F. *Biochem. Biophys. Res. Commun.* 2004; 325:1481.

[30] Sali A., Blundell T. L. *J. Mol. Biol.,* 1993; 234:779.

[31] Laskowski R. A., MacArthur M. W., Smith D. K., Jones D. T., Hutchinson E. G., Morris A. L., Naylor D., Moss D. S., Thorton J. M. *J. Appl. Cryst.* 1994. 26**:**283.

[32] Guex N., Peitsch M.C., *Electrophoresis.* 1997; 18:2714–2723 http://www.expasy.org/spdbv.

[33] A.T. Brünger, X-PLOR Version 3.1: A System for Crystallography and NMR, Yale University Press, New Haven, 1992.

[34] Bowie, J. U., Luthy, R., and Eisenberg, D. *Science.* 1991; 253:164.

[35] Luthy, R., Bowie, J., and Eisenberg D. *Nature* 1992; 356: 83.

[36] Wallace, A.C., Laskowski, R. A., and Thornton, J.M. *Protein Eng.* 1995; 2:127.

[37] Sternberg, M.J., Gabb, H.A., Jackson, R.M., *Curr. Opin. Struct. Biol.* 1998; 8: 250.

[38] Chen, R., Li, L., Weng, Z., *Proteins.* 2003;52: 80.

[39] Chen, R., Weng, Z., *Proteins* 2003;51: 397.

[40] Fanelli, F., Ferrari, S. *J. Struc. Biol.* 2006;153:278.

[41] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comp. Chem.* 2005**,** 26:1701.

[42] C. Ooestenbrik, T. A. Soares, N. F. A. van der Vegt, W. F. van Gunsteren, *Eur Biophys J.* 2005; 34: 273-284

[43] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman B., Ed.; Reidel D. Publishing Company: Dordrecht, The Netherlands, 1981; p 331.

[44] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. *J.Comput. Chem.* 1997; 18:1463.

[45] Miyamoto, S.; Kollman, P. A. J. *Comput. Chem.* 1992, 13:952.

[46] Chowdhuri, S.; Tan, Ming-Liang. and Ichiye, T.; J. *Chem. Phys.* 2006; 125: 144513.

[47] Darden, T; York, D.; Pedersen, L. A. *J. Chem. Phys*. 1993, 98:10089.

[49] Obika, S., Reddy, S. Y. and Bruice, T. C. *J. Mol. Biol.* 2003; 331:345.

Legends

Figure 1: Alignment of ETS domains of the templates and their targets.

Figure 2: Graphical representation of root-mean-square deviation (RMSD) of all Cα from starting structure of models as a function of time. The graphic A shows the RMSD of SpiC (black line) and SpiC-DNA complex (red line), B graphic shows the RMSD of ETV2 (black line) and ETV2-DNA complex (red line) and the graphic C shows the RMSD of NERF (black line) and NERF-DNA complex (red line).

Figure 3: Graphical representation of root-mean-square fluctuations (RMSF) of all Cα from starting structure of models as a function of time. The graphic A shows the RMSF of SpiC (black line) and SpiC-NA complex (red line), graphic B shows the RMSF of ETV2 (black line) and ETV2-NA complex (red line) and graphic C the RMSF of NERF (black line) and NERF-NA complex (red line).

Figure 4: Order parameter S^2, for the ϕ/ψ torsion of proteins residues of the MD-average structures of last 2 ns of simulation. The graphic A shows the S^2 parameter of SpiC (black line) and SpiC-NA complex (red line), graphic B shows the S^2 parameter of ETV2 (black line) and ETV2-NA complex (red line) and graphic C the S^2 parameter of NERF (black line) and NERF-NA complex (red line).

Figure 5: Major groove widths over 4 ns of MD. The graphic A shows DNA widths complexed with SpiC, B DNA widths complexed with ETV2 and C DNA widths complexed with NERF.

Table 1: Solved structures and their amino acid templates identities

Table 2: Analysis of Ramachandran plot and RMSD from ideal geometry for templates and models of Ets

Table 3: Analysis of the 3D profile and G-factor for the templates and models.

Table 4: Average number of hyrogen bonds of aminoacids residues which compose helix-alpha 3 over 4ns of MD simulation.

Table 1:

Target	PDB Acess code (template)	Ets (template)	Organism	Resolution (Å)	E value	Domain identity (%)
ETV-2	1K79	ETS-1	*Mus musculus*	2.40	1.2E-28	52.88
SPI-C	1PUE	PU-1	*Mus musculus*	2.10	1.6E-26	59.55
NERF	1DUX	ELK-1	*Homo sapiens*	2.10	4.4E-21	49.43

Table 2:

Protein	Region of Ramachandran plot				RMSD from ideal geometry	
	Most favorable (%)	Additional allowed (%)	Generously allowed (%)	Disallowed (%)	Bond lengths (Å)	Bond angles (°)
ETV-2	89.3 (86.8)	8.3 (13.2)	2.4 (0.0)	0.0 (0.0)	0.019	2.076
SPI-C	93.8 (91.1)	6.2 (7.6)	0.0 (1.3)	0.0 (0.0)	0.018	2.072
NERF	92.2 (87.0)	7.8 (13.0)	0.0 (0.0)	0.0 (0.0)	0.019	2.113

In parentheses are values obtained for the templates used for modeling

Table 3:

	3D profile			G-factor			Superposition of backbone
Protein	Total score	Ideal score	$S_{ideal\ score}$	Torsion angles	Covalent geometry	Global	RMSD (Å)
ETV-2	56.73 (53.42)	47.07 (46.15)	1.21 (1.16)	-0.13 (0.41)	-0.06 (0.63)	0.08 (0.32)	0.24
SPI-C	35.98 (34.79)	39.77 (39.77)	0.90 (0.87)	-0.04 (-0.04)	-0.04 (-0.03)	0.10 (0.14)	0.36
NERF	42.52 (40.77)	39.32 (38.86)	1.08 (1.05)	-0.14 (0.43)	-0.08 (0.67)	0.09 (0.33)	0.21

In parentheses are values obtained for the templates used for modeling

Table 4:

ETS protein	Residue	Atoms	NHL
NERF	GLU55	OE2---N4 (C1')	1.4±0.8
	ARG59	NH2---N7 (G1)	1.4±0.7
		NH2---O6 (G1)	
	ARG62	NE---N7 (G2)	0.9±0.6
	TYR64	OH---N6 (A3)	1.0±0.1
	TYR65	OH---O4 (T4')	0.8±0.5
SpiC	ARG62	NE---O6 (G1)	1.5±0.6
		NH2---O6 (G1)	
	ARG65	NE---N7 (G2)	0.6±0.7
ETV2	GLU55	OE1---N4(C1')	1.4±0.8
	LYS56	NZ---O2P (A3)	0.8±0.5
	ARG59	NE---N7 (G1)	1.3±0.7
		NH2---N7 (G1)	
	ARG62	NH2---O6 (G2)	0.4±0.6

Figure 1:

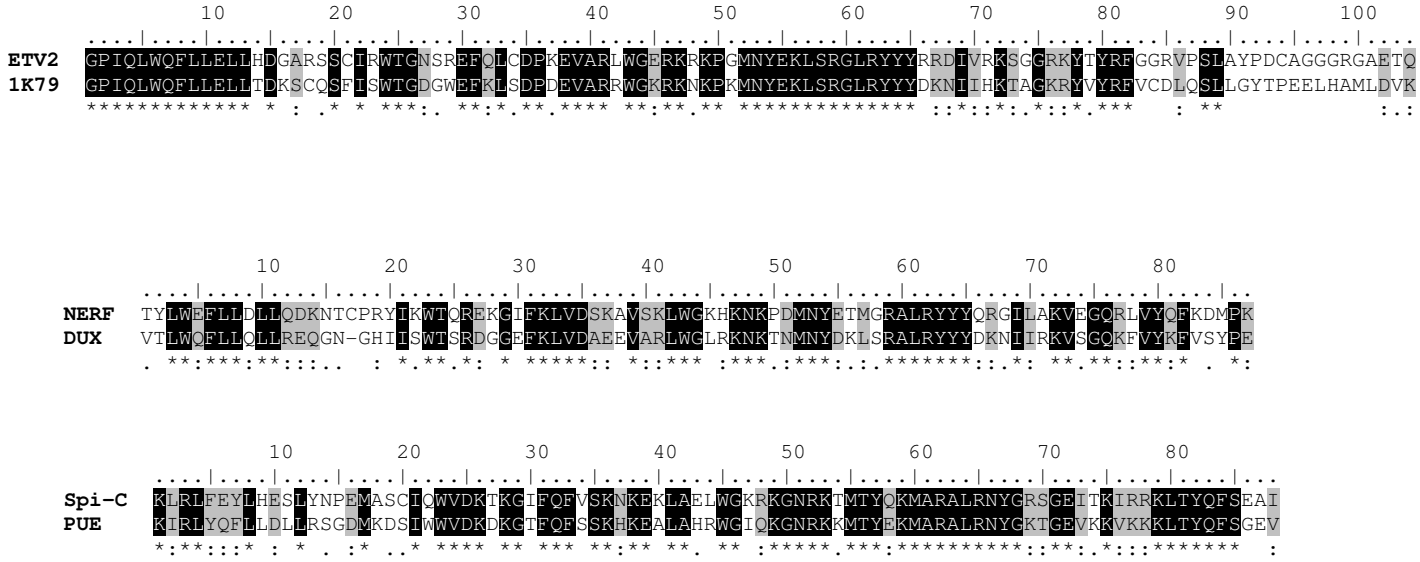


Figure 2:

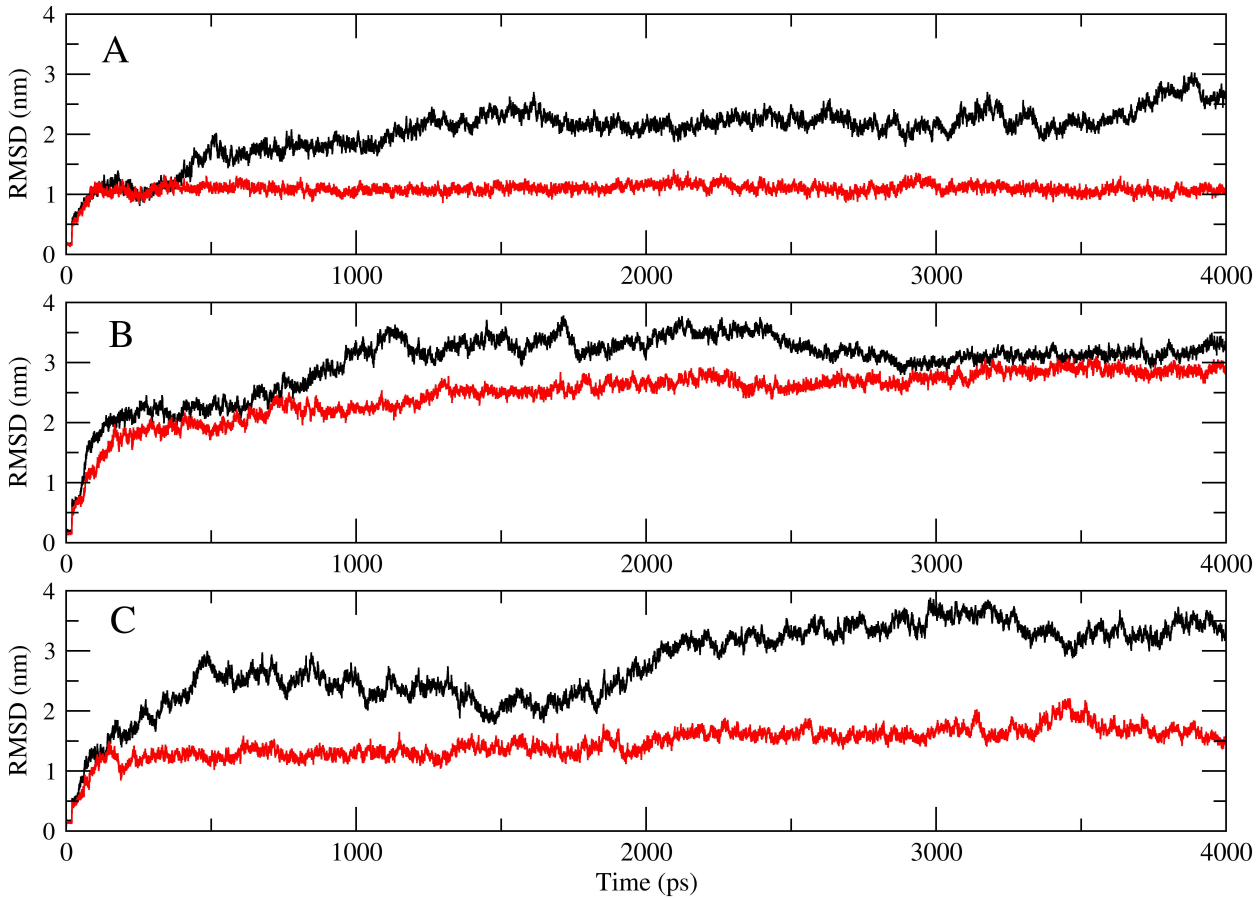


Figure 3:

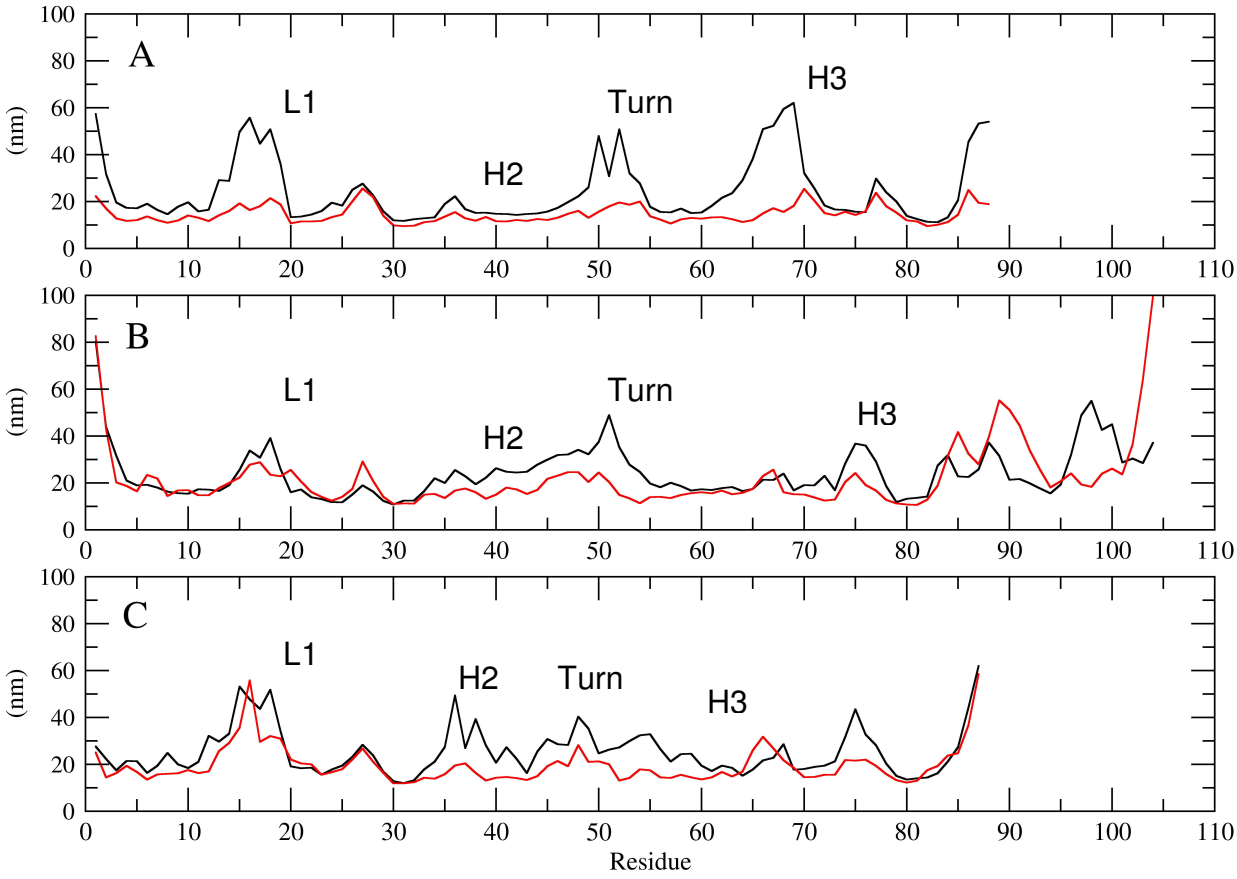


Figure 4:

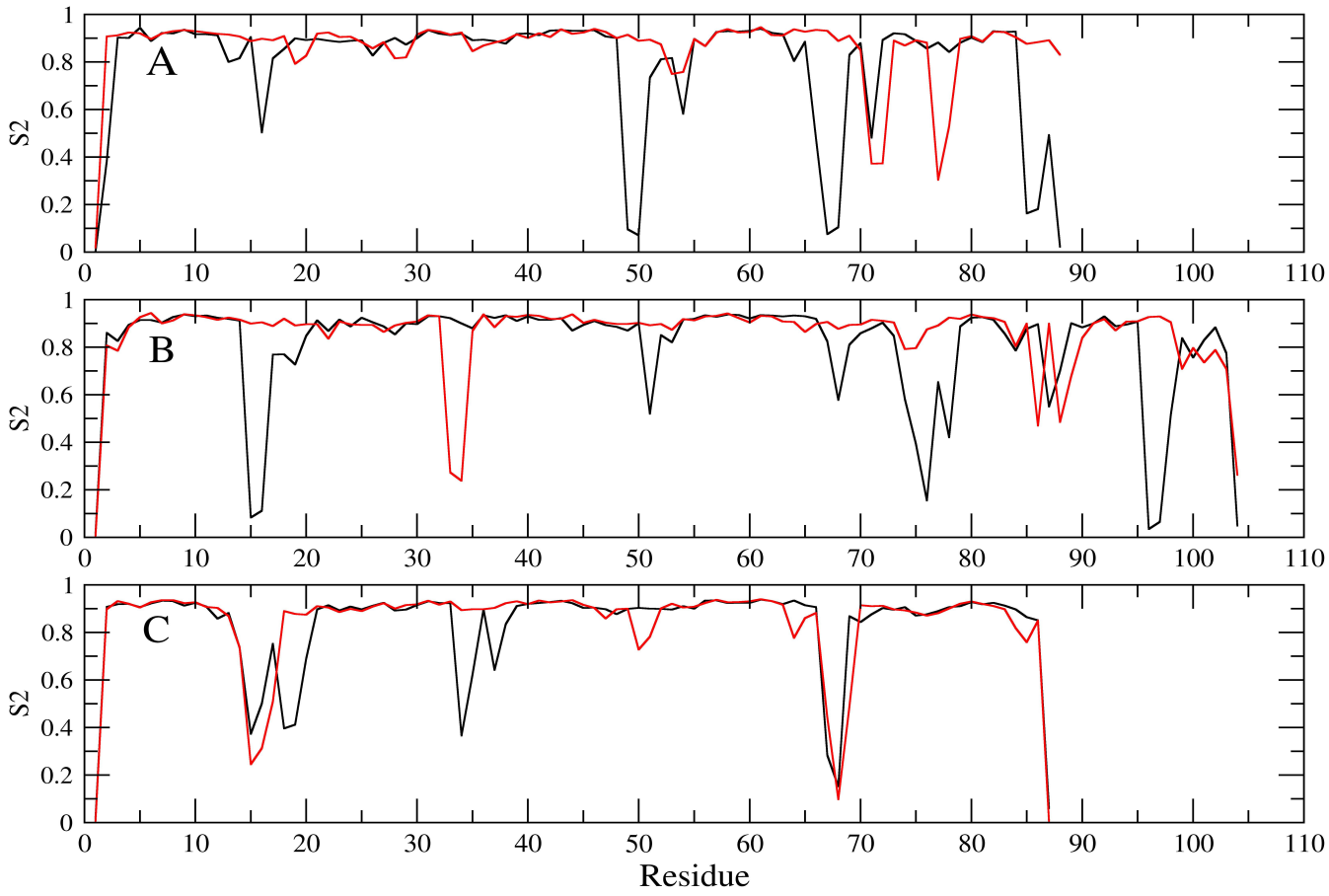
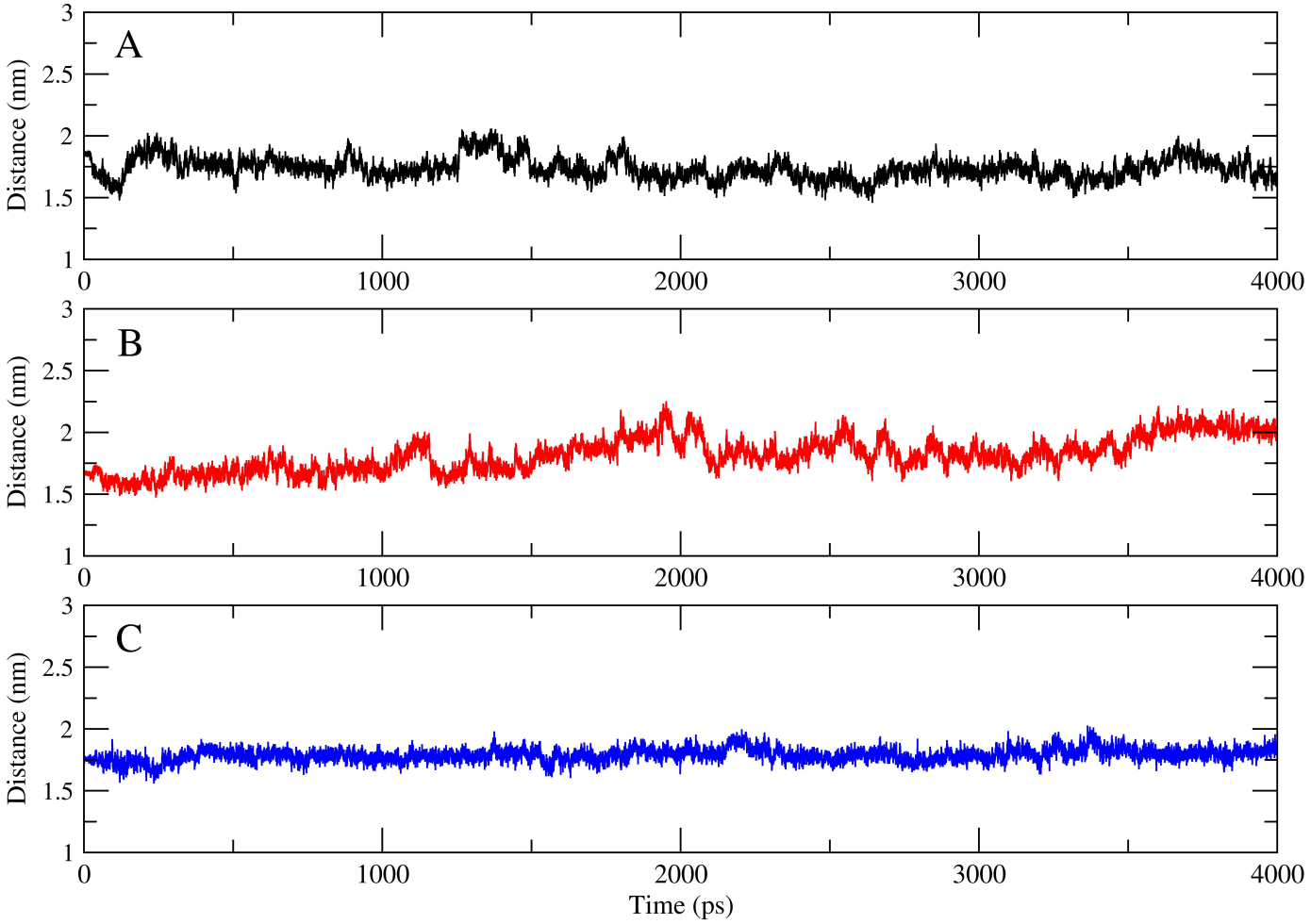


Figure 5:



7. ANEXOS

7.1. ARTIGOS DESENVOLVIDOS DURANTE O PERÍODO DE MESTRADO

Protein Kinases as Targets for Antiparasitic Chemotherapy Drugs

Fernanda Canduri[1,*], Patrícia Cardoso Perez[2], Rafael A. Caceres[2], and Walter F. de Azevedo Jr.[2]

[1]*Universidade Federal de Mato Grosso do Sul – UFMS. Centro de Ciências Biológicas e da Saúde – CCBS-Departamento de Morfofisiologia-Laboratório de Bioquímica, Caixa Postal 549- CEP 79070-900 Campo Grande –MS, Brazil and* [2]*Faculdade de Biociências-PUCRS. Av. Ipiranga, 6681. Porto Alegre, RS. 90619-900, Brazil*

Abstract: Parasitic protozoa infecting humans have a great impact on public health, especially in the developing countries. In many instances, the parasites have developed resistance against available chemotherapeutic agents, making the search for alternative drugs a priority. In line with the current interest in Protein Kinase (PK) inhibitors as potential drugs against a variety of diseases, the possibility that PKs may represent targets for novel anti-parasitic agents is being explored. Research into parasite PKs has benefited greatly from genome and EST sequencing projects, with the genomes from a few species fully sequenced (notably that from the malaria parasite *Plasmodium falciparum*) and several more under way, the structural features that are important to design specific inhibitors against these PKs will be reviewed in the present work.

Key Words: Kinases, parasites, drug design, crystallography, Apicomplexa

This paper is dedicated to Prof. Yvonne P. Mascarenhas on the occasion of her 75th birthday and 50th anniversary of her scientific work, as a tribute to her work training and inspiring a new generation of Brazilian scientist.

INTRODUCTION

Over the last few decades, progress in drug discovery has enabled increasingly sophisticated medicines to be developed to cure a wide variety of diseases. Some infectious diseases, however, have been progressively marginalized by research programs in both private and public sectors. These neglected diseases have a devastating impact on the world's poor population but because they affect only the poor, they do not constitute a lucrative market to attract investment in research and development of new drugs [1].

The Apicomplexa is a diverse group of intracellular protists [2]. They include species from the genus *Plasmodium*, *Toxoplasma*, and *Cryptosporidium*. These protozoan parasites cause some of the most serious, and in some cases, deadly diseases in humans [3]. Oocysts of these parasites can persist for long periods of time in the environment (i.e. in water, soil, on vegetation and other food resources), maintaining their infectivity even under harsh environmental conditions and therefore are important for dispersal and transmission to hosts [2].

Although chemo- and immuno-prophylactic strategies are available to control some of these parasites, they are inadequate. Currently, there is an urgent need to design new vaccines or chemotherapeutics for apicomplexan diseases. High-throughput global protein expression analyses using gel or non-gel based protein separation technologies coupled with mass spectrometry and structural bioinformatics provide means to identify new drug and vaccine targets in these pathogens. Protein identification based proteomic projects in apicomplexan parasites is currently underway, with the most significant progress made in the malaria parasite, *Plasmodium falciparum* [3].

Genes encoding CDK-related proteins have been identified in several apicomplexans [4-8]. PfPK5 and PfPK6 (protein kinase from *Plasmodium falciparum*), and Tgcrk2/TPK2 (cdc2-related kinase and protein kinase 2 from *Toxoplasma gondii*, respectively) share a high degree of homology and are most closely related to vertebrate CDK 1 and 2 which are involved in regulation of the G1/S phase and G2/M phases of the cell cycle. PfPK5, the best characterized of the apicomplexan crks, has been implicated in regulation of DNA synthesis [9] during parasite asexual development, but to date, expression of these apicomplexan crks has not been studied during the sexual cycle [10].

Cyclin-dependent kinases (CDKs) were first identified in the budding yeast *Saccharomyces cerevisiae*, and the fission yeast *Schizosaccharomyces pombe*, in genetic screens for conditional lethal mutants unable to progress through the cell cycle at an elevated temperature, although they can continue to grow. Such mutants are called cdc, for cell division cycle. Cdc mutants arrest at the particular point in the cell cycle where the mutant gene product normally functions. There are nearly 50 cdc mutants from each yeast whose of these correspond to CDKs. CDKs regulate progression through the eukaryotic cell cycle. They are also involved in the regulation of other processes such as gene expression and phosphate metabolism. CDKs are the catalytic subunits of heterodimeric complexes and they approximate the size of a minimal protein-kinase catalytic domain. Alone, they lack protein kinase activity. As their name implies, CDKs must associate with a member of the cyclin protein family to gain activity [11] and undergoes specific phosphorylation and desphosphorylation to be fully activated.

A key regulator in the cGMP signaling pathway, known as cGMP-dependent protein kinase (PKG), mediates a wide range of physiological effects in mammalian cells [12]. To date, two types of mammalian PKG have been reported, type I and type II. Both forms exist as homodimers: each polypeptide comprising a regulatory domain and a catalytic domain. The regulatory domain contains a dimerization domain, an auto-inhibitory domain, autophosphorylation sites, and two cyclin-nucleotide-binding sites. PKGs within the apicomplexan group of protozoa have been identified in *Toxoplasma gondii* (TgPKG) and the malaria parasite *Plasmodium falciparum* (PfPKG) [13-16]. TgPKG exist as monomers that also occur as two isoforms [13-14]. These apicomplexan PKGs uniquely contain three cGMP-binding sites, however, PfPKG contains an additional, degenerate cGMP-binding site [17].

Despite sharing identical amino-acid sequences within the cGMP-binding sites, mammalian PKG I and PKG I exhibit different kinetic properties and cyclin-nucleotide-analogue specificities [18]. The substitution of a conserved serine or threonine resi-

*Address correspondence to this author at the Universidade Federal de Mato Grosso do Sul – UFMS. Centro de Ciências Biológicas e da Saúde – CCBS-Departamento de Morfofisiologia-Laboratório de Bioquímica, Caixa Postal 549- CEP 79070-900 Campo Grande –MS, Brazil; Tel: ----------; Fax: ----------; E-mail: fcanduri@yahoo.com.br

1389-4501/07 $50.00+.00

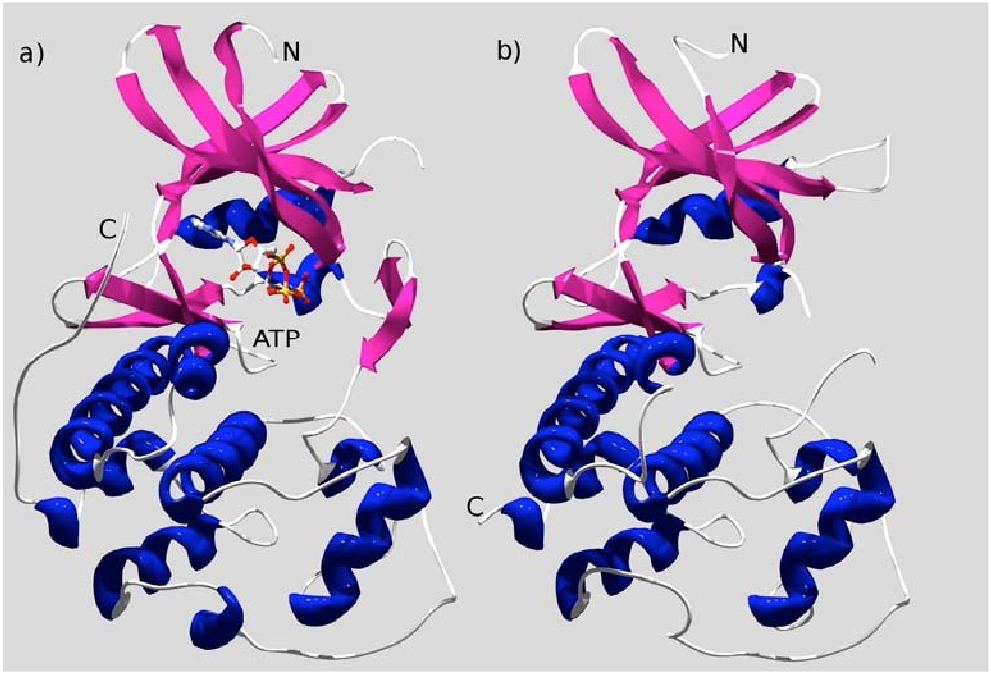


Fig. (1). Ribbon diagrams of (a) human CDK2 in complex with ATP (PDB access code: 1HCK), showing the binding pocket, and (b) PfPK5 (PDB access code: 1OB3).

due (with an alanine residue) within the cGMP-binding sites leads to distinct effects on PKG activation [19,20]. Apart from the predicted structural differences from the mammalian isoforms, a truncate form of PfPKG also possessed distinct cyclin-nucleotide-dependent kinase-activation properties [17]. Deng *et al.*, (2003) have shown that the C-terminal cGMP-binding site has the greatest effect on PfPKG activation. Detailed information on the distinct properties of PfPKG in terms of its cGMP-activation mechanism may prove useful for the design of a new anti-malarial drug [17].

This review will address the study about the proteins kinase from phylum Apicomplexa *Cryptosporidium*, *Plasmodium*, and *Toxoplasma* species, the most common causes of human parasitic diseases.

Plasmodium species

Plasmodium spp. are protozoan parasites with a complex life cycle, including alternating development phases in mosquito and vertebrate hosts. More than a hundred *Plasmodium* species have been described, of which four infect humans and cause malaria. The body of information available offers new opportunities in malaria research, and will hopefully accelerate rational design of novel intervention strategies, a pressing need in view of the rapidly deteriorating efficiency of existing control tools [21].

Plasmodium falciparum, a unicellular protozoan responsible for the most lethal form of human malaria, has re-emerged as a leading cause of mortality in developing countries, especially in the population of young children. The parasite accounts for up to 2.7 million deaths worldwide [22]. Drug-resistant malaria parasites have spread to the vast majority of malaria-endemic countries, and disease is therefore very difficult to control [17,23]. Current efforts to produce effective vaccines have not yet resulted in any significant success. Identification of drugs that interfere with parasite development could be a useful way to inhibit parasite growth in humans. Detailed knowledge of molecular mechanisms that control the life cycle of malaria parasite could provide crucial information needed to achieve this goal [24].

A recently published genome sequence [25] and earlier studies suggest that several homologues of eukaryotic signaling proteins, such as protein kinases and cyclins are conserved in *P. falciparum* [26] and these enzymes (including PfPK5, PfPK6, Pfmrk, Pfcrk-1, Pfcrk-3, Pfcrk-4) play an important role in the development of the parasite [27]. The major challenge is to understand how these enzymes integrate in cellular machinery of *Plasmodium* and what roles they play in parasite development. The answer to these questions could lead to identification of important signaling pathways involved in the development of this parasite [24].

The intraerythrocytic developmental cycle of the malarial parasite diverges from the typical eukaryotic cell cycle where the chromosomes replicate only once. During schizogony, the parasite undergoes multiple rounds of DNA replication to yield polyploidity of chromosomal DNA. Regardless of the unique characteristics of the *Plasmodium* cell cycle, it is expected that CDKs will be key molecular switches for cell cycle progression in the parasite [26].

In this context, *Plasmodium* cell cycle regulators show considerable promise, both because their activities are most probably essential to parasite survival and because the overall organization of the cell cycle in malaria parasites differs considerably from that in mammalian cells. Available data indicate that these differences reflect structural and functional peculiarities of the regulatory molecules involved in the cell cycle [28].

PfPK5

P. falciparum protein kinase 5 (PfPK5) is the best characterized member of the *P. falciparum* CDK family [29]. In a sequence comparison with members of the human CDK family, it shares the greatest degree of sequence identity (about 60%) with human CDK1 and CDK5 [30]. The stage the *P. falciparum* schizont undergoes multiple rounds of nuclear division [9]. Its localization within the cell and the timing of its activation both strongly suggest that it plays a role in the regulation of the *P. falciparum* nuclear division cycle [28].

The crystallographic structure of PfPK5 was recently solved at 1.9 Å resolution (PDB access code: 1OB3) [28], and the structure

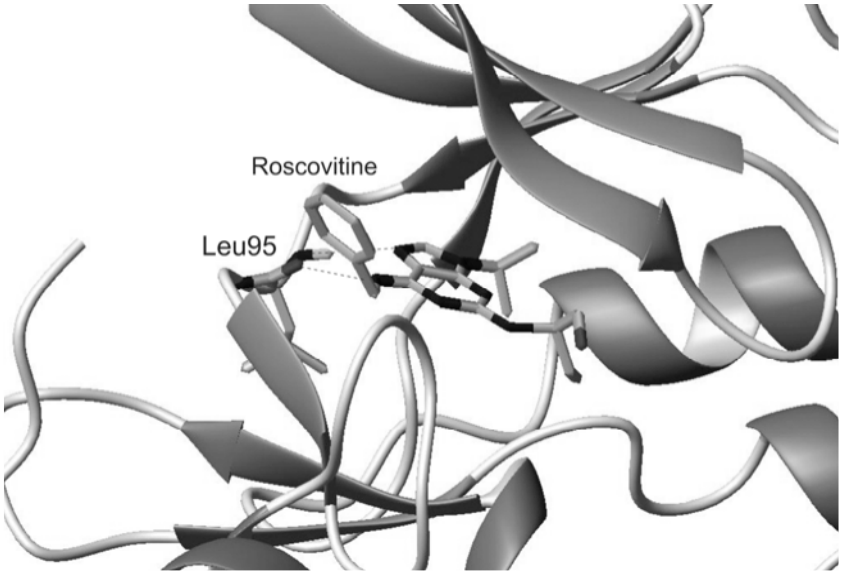


Fig. (2). Binding site region of PfPK6 showing the interaction with inhibitor R-roscovitine.

adopts the characteristic bilobate kinase fold (Fig. (**1**)) [31]. The smaller N-terminal lobe consists of five antiparallel strands (1–5) and a single prominent helix containing the sequence PSTTIRE that corresponds to the PSTAIRE cyclin binding motif of CDKs 1, 2, and 3. This lobe is connected through a hinge region to the larger predominantly -helical C-terminal lobe. The PfPK5 structure is the second determination (the first being that of CDK2 [32] of a structure of a monomeric CDK fold). The high degree of sequence identity between PfPK5 and CDK2 is reflected in their structures, which have an rmsd of 1.3 Å calculated over 235 equivalent C atoms [28]. All aspects of the monomeric CDK2 architecture are well conserved [32] (Fig. (**1**)). The two deletions in the PfPK5 sequence (equivalent to CDK2 residues Lys24 and Ala 93) are both accommodated at the apex of -turns, resulting in very little perturbation to the overall structure. The short CDK2 C-terminal tail (residues Val289–Leu298) that wraps toward the hinge region is absent from PfPK5 [28]. As observed in CDK2, PfPK5 is inactive as monomer. The C helix (residues 44–57), which is displaced from the body of the protein so that the side chain of Glu50 (the E of the PSTTIRE sequence) points towards the solvent and does not interact with Lys32. By analogy with other kinases, it is expected that these two residues will form a close interaction in the active state, which in turn allows Lys32 to correctly orientate the ATP and phosphate groups for phosphotransfer. A second member of the CDK family from a species phylogenetically remote from humans adopts this distinctive disposition of the C helix provides evidence that the same inappropriate C helix orientation is employed in the wider CDK family to inhibit CDK activity [28]. This fact can be exploited to design new inhibitors against protein kinases.

PfPK6

PfPK6 was recently isolated by differential display RT–PCR (DDRT–PCR) of mRNA obtained from different asexual erythrocytic stages of *Plasmodium falciparum*, which shows sequence similarity to cyclin-dependent kinase (CDK) [26]. The sites of regulatory phosphorylation, found in other CDKs (Thr14, Tyr15 and Thr160 in CDK2) are conserved in PfPK6 [26]. Over-expression of PfPK6 in *Escherichia coli* allowed studying the biochemical properties of the recombinant enzyme. In addition, immunolocalization and immunoblot analysis were performed to localize PfPK6 protein in various developmental stages [26].

A PfPK6 model has been constructed using as the template the structure of active CDK2 extracted from the structure of the phosphorylated CDK2/cyclin A complex. This homology model shows a typical bilobal structure, with the smaller N-terminal lobe consisting predominantly of -sheet structure and the larger C-terminal lobe consisting primarily of -helices. The N-terminal lobe of PfPK6, as observed for CDKs, consists of a sheet of five antiparallel -strands (1 5) and a single large helix (1). The C-terminal lobe contains a pseudo-4-helical bundle (2,3,4,6), a small -sheet ((6 8), and two additional helices (5,7) [33].

PfPK6 contained a 286-residue catalytic domain similar to that in eukaryotic protein kinases. The identity of PfPK6 ORF with that of human CDK2 and human p38 MAPK was 38% and 33%, respectively. The presence of DIKPEN and GTLWYRAPE motifs in subdomains VI and VIII respectively are consistent with the consensus sequence (DLKPEN and GT/SXXY/FXAPE) for serine/threonine kinases. PfPK6 contained 14 of 15 conserved residues found in the catalytic domains of eukaryotic kinases including the GXGXXG (glycine loop) ATP-binding subdomain I motif and the invariant lysine in the catalytic subdomain II [34]. Of these 15 signature residues, only the phenylalanine in the DFG motif in subdomain VII is replaced by a leucine, a conservative change. Several conserved residues whose phosphorylation regulates vertebrate CDK activity, such as Thr14, Tyr15, Thr161 and Ser277, are also conserved in PfPK6 (Thr19, Tyr20, Thr173 and Ser283).

Several lines of evidence suggests that PfPK6 is a cyclin-independent kinase: (a) the recombinant protein shows significant autophosphorylation and histone phosphorylation activity in the absence of cyclin and (b) unlike PfPK5, the PfPK6 activity is not stimulated by incubation with cyclins. By analogy with other well-studied eukaryotic CDKs it would be predicted that full activation of PfPK6 would require phosphorylation of Thr173, the residue homologous to Thr160 of human CDK2 [26]. The mechanism of PfPK6 cyclin-independent kinase activity is unclear, but presumably the major changes to vertebrate CDKs that are normally induced by cyclin binding must be constitutively present in PfPK6. Interestingly, PfPK6 contains a relatively large insertion (Asp80-Cys94) in the L6 loop which makes this loop about 10 amino acids longer than that of other CDKs. It is possible that this larger than usual L6 loop may contribute to the constitutive activity of PfPK6, although other protein kinases that are active as monomers do not have an equivalent insertion [26-34].

PfPK6 exhibits other unusual properties such as its preference for Mn^{2+} as the divalent cation, a property usually associated with kinases. PfPK6 also resembles the MAPKs in that its sensitivity to roscovitine, is in the micromolar range rather than the nanomolar range as reported for CDK1, CDK2 or CDK5 [34]. The increasing in the concentration of roscovitine causes the progressive inhibition of autophosphorylation of PfPK6 in a kinase assay *in vitro* with IC_{50} value of 30 M [34]. Fig. (**2**) shows the active site of the PfPK6 in complex with roscovitine. There are two intermolecular hydrogen bonds between the Leu95 of the PfPK6 and N6 and N7 of the inhibitor.

The TPT activation site that is conserved, and must be phosphorylated on both the Thr residues for the PfPK6 to become fully active. The CDKs possess only one residue that can be phosphorylated in the region, corresponding to Thr160 in CDK2. It would be informative to determine whether both residues must be phosphorylated for full activation of PfPK6. A number of critical domains known to be important for the binding of CDK-activating proteins are also conserved in PfPK6 [34]. Clearly, the determination of the crystallographic structure of PfPK6 will be important to elucidate the mechanism of its cyclin-independent activity and its relationship to MAPKs [26].

Pfmrk

A novel gene encoding a CDK-like protein, Pfmrk (for MO15-related kinase), has been isolated from the human malaria parasite *Plasmodium falciparum*. The gene has no introns and comprises an open reading frame encoding a protein of 324 amino acids with a predicted molecular mass of 38 kDa. Database searches revealed a striking similarity to the CDK subfamily with the highest similarity to human CDK7 (MO15). The overall sequence of Pfmrk shares 62% similarity and 46% identity with human MO15 (CDK7). Pfmrk contains two unique inserts: one consisting of 5 amino acids just before the cyclin-binding motif and the other composed of 13 amino acids within the T-loop equivalent region. [35]. Pfmrk-cyclin complexes have no detectable activity *in vitro* towards PfPK5 or other plasmodial CDK-related kinases. Whether this result reflects the fact that Pfmrk is not a CDK-activating kinase but has entirely different functions in the parasite (in addition to its CDK activating kinase activity, mammalian CDK7 is implicated in other functions such as transcriptional regulation), or whether no CDK-activating kinase activity is detectable because one or more accessory elements are lacking in the assays, is unclear [36, 37]. Pfmrk is a single-copy gene conserved between several parasite strains and is located on chromosome 10. A 2500-nucleotide transcript of this gene is expressed predominantly in the sexual blood stages (gametocytes), suggesting that Pfmrk may be involved in sexual stage development [35].

Pfcrk

There is a gene encoding another CDK-like protein in *Plasmodium falciparum*. It was identified using degenerate oligonucleotides designed to hybridise to regions that are conserved in members of the cdc2 gene family. This gene, called Pfcrk-1, for cdc2-related kinase, is located on chromosome 4. Pfcrk-1 is developmentally regulated, as indicated by stage-specific accumulation of mRNA in gametocytes [5,38]. No cyclin partner has yet been identified for these enzymes. The availability of genomic databases has allowed the identification of Pfcrk-3, a gene encoding a putative kinase with maximal homology to CDK1, and of Pfcrk-4, which, like PfPK6, displays similar levels of homology to both CDKs and MAPKs [26]. Both these ORFs are expressed in asexual and sexual blood stages. A threonine equivalent to CDK2 Thr160 appears to be present in all three ORFs, although in the case of Pfcrk-4 insertions within the activation loop make the alignment difficult. The residues corresponding to Thr14 and Tyr15 of CDK2 are conserved in Pfcrk-1, but in Pfcrk-3 and Pfcrk-4 Thr14 is substituted by an alanine and a valine, respectively. This finding may indicate that these enzymes have diverged in their ability to be regulated by inhibitory phosphorylation. Pfcrk-1, Pfcrk-3 and Pfcrk-4 all share the peculiarity of possessing a very large (several hundred amino-acids) extension at their N-terminus, which contains repeated motifs as found in many other types of *Plasmodium* proteins. It is unclear at this stage whether these extensions are involved in enzyme function, or whether they are processed during the infection. In addition to this extension, Pfcrk-3 and Pfcrk-4 also carry large insertions within their catalytic domains. No function has yet been demonstrated for Pfcrk-1, -3 and -4 [26].

The first three introns of Pfcrk2 have been compared in a total of seven *Plasmodium* species. The introns were located at conserved sites, suggesting an ancestral origin [39]. The complete sequence of the crk2 gene from *P. knowlesi* and from *P. berghei* was determined. In both species, the crk2 gene is closely linked to an elongation factor 1 alpha gene. The two crk2 proteins are highly homologous to the PfPK5 protein. The crk2 gene of both species is expressed at a low level during the asexual cell-cycle within the host erythrocytes. The *P. berghei* crk2 mRNA is also present in gametocytes and in stages during development in the mosquito, suggesting a role of this protein in different parts of the life cycle [26].

PKA

In *Plasmodium falciparum* the catalytic subunit gene of cAMP-dependent protein kinase exists as a single copy. Interestingly, its expression appears developmentally regulated, being at higher levels in the pathogenic asexual stages than in the sexual forms of parasite that are responsible for transmission to the mosquito vector. Within asexual parasites, PfPKA activity can be readily detected in schizonts. Similar to endogenous PKA activity of noninfected red blood cells, the parasite enzyme can be stimulated by cAMP and inhibited by protein kinase inhibitor. Importantly, *ex vivo* treatment of infected erythrocytes with the classical PKA-C inhibitor H89 leads to a block in parasite growth. This suggests that the PKA activities of infected red blood cells are essential for parasite multiplication [40].

PKB

Protein Kinase B (PKB) is an important member of the phosphatidylinositol 3-kinase-dependent signaling pathways in several eukaryotes, but the role of PKB in protozoan parasites is not known. A protein kinase B homologue in *Plasmodium falciparum* (PfPKB) is expressed mainly in the schizonts and merozoites. Even though PfPKB shares high sequence homology with PKB catalytic domain, it lacks a pleckstrin homology domain typically found at the N terminal of the mammalian enzyme. Biochemical studies performed to understand the mechanism of PfPKB catalytic activation suggested (i) its activation is dependent on autophosphorylation of a serine residue (Ser-271) in its activation loop region and (ii) PfPKB has an unusual N-terminal region that was found to negatively regulate its catalytic activity. Kumar *et al.*, (2004) identified an inhibitor of PfPKB activity that also inhibits *P. falciparum* growth, suggesting that this enzyme, as observed in PKA, may be important for the development of the parasite [24].

CDPKs

The *Plasmodium* genome [41] encodes a full complement of signaling molecules, allowing the parasite to interact with host cells and to adjust its complex life cycle to the changing environments of vertebrate and mosquito hosts. A notable expansion of proteins containing Ca^{2+} binding EF hand modules in *Plasmodium* relative to yeast [42] reflects the importance of Ca^{2+} as a secondary messenger in apicomplexan parasite-host interactions [43]. In response to host melatonin Ca^{2+} signaling synchronizes *Plasmodium* proliferation in the blood [44] and inhibitor studies show a role for

cellular Ca^{2+} in regulating gamete formation in response to conditions encountered in the mosquito midgut [45,46].

CDPKs combine an amino-terminal serine/threonine kinase domain and a carboxy-terminal calmodulin-like domain, composed of four EF hands, in the same molecule. Current evidence suggests that CDPKs may be limited to plants, green algae, ciliates, and apicomplexan parasites. The genome of *P. falciparum* encodes a family of six putative CDPKs and one CDPK-related kinase, in which the calmodulin domain is missing. Three CDPKs of unknown function have been characterized in *P. falciparum* [47, 48]. Their developmental regulation makes these kinases prime candidates for the molecular switches that translate ubiquitous Ca^{2+} signals into appropriate cellular responses at specific stages of the parasite's life cycle.

CDK Regulators

Several groups attempt to identify cyclins by PCR-based approaches using conserved sequences within the cyclin box, a malarial cyclin gene, but was not reported until the genomic databases became available. Database mining using cyclin sequences as queries led to the identification of an ORF called Pfcyc-1, which displayed maximal homology to cyclin H from a variety of organisms [49]. Pfcyc-1 presents approximately 17% homology with mammalian cyclin H. It came as a surprise that bacterially expressed Pfcyc-1 was *in vitro* able to efficiently activate PfPK5. This result suggested that PfPK5 may have a relaxed specificity in its cyclin requirements, a hypothesis that was confirmed by the demonstration that PfPK5 can be activated by other cyclins and cyclin-like proteins such as p25, a CDK5 activator protein structurally related to cyclins despite a lack of homology at the sequence level. Pfcyc-1 also activates Pfmrk, which was somewhat less surprising since Pfcyc-1 displays maximal homology to cyclin H and Pfmrk to CDK7 [26], cyclin H is the cyclin partner of CDK7.

Sequence-based searches of the *P. falciparum* genome have failed so far to identify any putative *P. falciparum* proteins with significant sequence homology to any CDK inhibitor homologues. Results previously presented suggest that *P. falciparum* CDK complexes may be subject to regulation by CDK inhibitors and by analogy with the roles of CDK inhibitors in higher eukaryotic cells the identification of bonafide CDK inhibitors in *P. falciparum* would suggest the existence of checkpoint pathways to regulate CDK activity. Hence, it cannot be excluded that the *Plasmodium* CDKs are regulated *in vivo* by CKI functional homologues [26].

Toxoplasma species

Ocular toxoplasmosis is a local manifestation of systemic infection in which *Toxoplasma* spreads into the eye, affecting mainly the posterior segment of the eye. Reactivation of the initial retinal condition presumably results from the rupture of quiescent parasitic cysts lying adjacent to pre-existing scars and may secondarily involve the choroid (leading to retinochoroiditis). Although the molecular mechanisms underlying host-parasite interaction are largely unknown, toxoplasmic retinochoroiditis usually remains a local event, and does not necessarily evoke a detectable systemic immune response. Local immunotolerance mechanisms may likewise confound attempts to confirm the clinical diagnosis by serology. Against this background, current diagnostic strategies need to be re-evaluated with a view to future improvements [50].

TPK2

The Apicomplexan parasite *Toxoplasma gondii* replicates by endodyogeny, an unusual form of binary fission. The role of TPK2 was tested, a homologue of the CDC2 cyclin-dependent kinases, in cell cycle regulation. TPK2 phosphorylates a peptide from Histone H1, proving that TPK2 is a functional kinase. Despite being a functional kinase, TPK2 did not rescue *Schizosaccharomyces pombe* cdc2 and *Saccharomyces cerevisiae* cdc28 mutant strains. Overexpression of a dominant-negative mutant of TPK2 (TPK2-dn) in *T. gondii* tachyzoites arrested replication. FACS analysis of tachyzoites expressing TPK2-dn revealed an increase in the fraction of cells in S-phase when compared with TPK2 transfected parasites. Expression of TPK2 did not arrest tachyzoite replication. No discernable G2 cell cycle block was evident suggesting that cell cycle checkpoints differ in *T. gondii* from most other eukaryotic cells. These data suggest that TPK2 executes an essential function in *T. gondii* cell cycle and is likely to be the *T. gondii* CDC2 orthologue [51].

PKG

cGMP-dependent protein kinase is an essential gene product for *T. gondii* and as such represents an attractive therapeutic target. *Toxoplasma* PKGs are encoded by single-copy genes, two serologically-related parasite PKG isoforms have been detected [52]. The identification of an ATP competitive active-site inhibitor selective for the parasite enzyme indicates that therapeutically exploitable structural differences exist between parasite and host enzymes [13]. The apicomplexan parasite enzymes are also 30% larger than their mammalian and insect counterparts. The size increase is due to an expanded N-terminal regulatory domain, which encompasses a putative third cGMP-binding site. Demonstration that the activity of *Toxoplasma* PKG is dependent on three putative cGMP binding sites distinguishes parasite PKG from the cGMP-dependent kinases of vertebrates and insects. The observation that the novel parasite-specific third site seems to exert a greater influence on cGMP-dependent kinase activation compared with sites 1 and 2 suggests a unique role for this site in the allosteric regulation of parasite PKG. These results can be potentially exploitable differences in the allosteric regulation of parasite PKG compared with PKGs of the vertebrate host [14].

MAPK

Mitogen-activated protein kinases (MAPKs) govern diverse cellular processes including proliferation, differentiation, survival and stress responses [53]. Regarding the role that MAPKs play in parasite life cycles, little is known, although such information is likely to be useful in understanding parasite pathobiology, and may be useful for drug discovery. *Toxoplasma gondii* is reported to possess two ERK MAPKs based on identification of immunoreactive protein kinases using anti-human ERK antibodies and supported by biochemical analyses [54].

Brumlik *et al.*, (2004) determined relative *MAPK-1* gene expression during tachyzoite to bradyzoite stage differentiation [55]. Stage-specific differences in *T. gondii* from MAPK-1 (TgMAPK-1) transcript accumulation are consistent with a role in stage differentiation as suggested by Wei *et al.*, (2002) [56]. However, further work is required to confirm any potential role for TgMAPK-1 in this regard.

The TgMAPK-1 amino acid sequence is also 35–45% homologous to human ERK MAPKs. Given the pivotal role of MAPKs in critical cellular processes including growth and differentiation, knowledge of *T. gondii* MAPK biology likely will lead to insights into the pathobiology of this medically important opportunist. On this basis, parasite MAPKs such as TgMAPK-1 may be important drug discovery targets. Unique structural features of parasite MAPKs compared to human MAPKs [57, 58] may be exploited to increase the possibility of developing relatively selective agents [55].

CDPK

TgCDPK1 is the major CDPK activity in the tachyzoite stage. Pharmacological evidence shows it is essential for tachyzoite motility and host cell invasion [59] and that it is involved in parasite egress from the infected host cell [60].

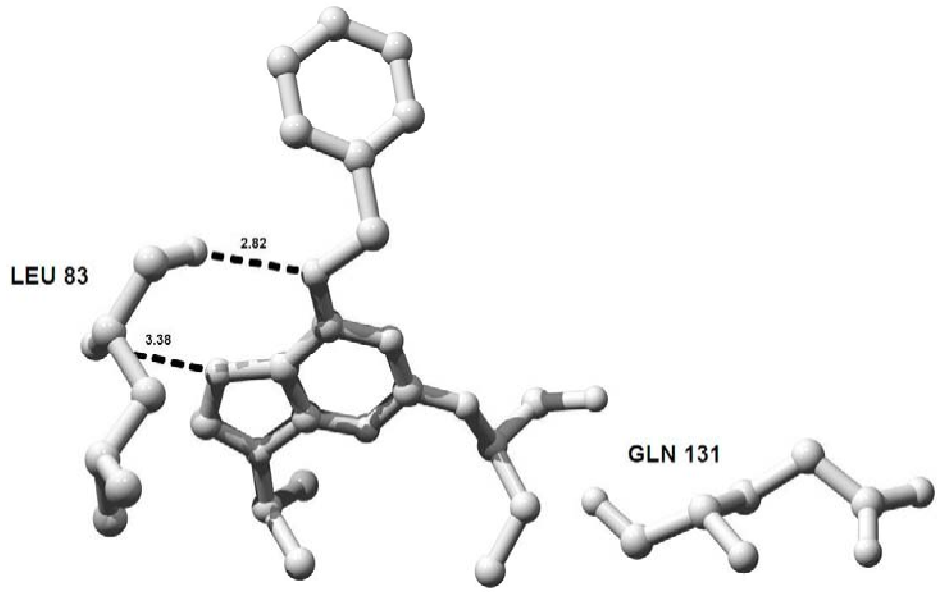


Fig. (3). Intermolecular hydrogen bonds at the ATP-binding pocket of the complexes of CDK2/R-roscotivine [117]. Distances are indicated in Å (10^{-10} m).

The essential role of CDPKs at different stages of the life cycles of *Toxoplasma* [61] and *Plasmodium*, combined with their absence from the parasites' vertebrate hosts, make members of the CDPK family attractive targets for pharmacological intervention [62].

Cryptosporidium species

The apicomplexan protist *Cryptosporidium parvum* is globally recognized as an intracellular parasite of humans and animals. This protozoan can cause self-limiting diarrhoea in immunocompetent individuals or life-threatening, prolonged opportunistic infections in immunocompromised patients. Transmission can occur from an infected person, animal or fecally contaminated environment [63]. To date, there is still no effective treatment available for *C. parvum* infections [64, 65]. In transplant patients, a course of antimicrobial therapy along with concurrent reduction in immunosuppression optimize immunologic status and may potentially lead to resolution of the infection [63].

This group of protists has at least three distinct life stage replication processes: sporogony, merogony and gamogony. Both sporogony and merogony are cell multiplication processes that produce more than two daughter cells in each cell cycle and differ from the host cell somatic duplication, which implies that a unique mechanism may be involved in DNA replication in apicomplexan parasites. To date, little work has been published on how *C. parvum* perform the processes associated with DNA metabolism. Replication protein A (RPA) participates in essential roles for DNA metabolism, which include replication, repair and recombination [66-68].

Previous studies in eukaryotic organisms have shown that the phosphorylation of RPA2 proteins might play an important role in regulating the biological functions of these organisms [69, 70]. In those studies, the phosphorylation of recombinant and native RPA2 proteins was achieved by DNA-PKs and/or cyclin-dependent protein kinases. The parasite's DNA replication proteins may serve as a novel target for drugs against cryptosporidiosis and possibly other apicomplexan-based diseases.

CONCLUDING REMARKS

Protein kinases have been extensively studied as targets for therapeutic intervention in many different disease indications [71-73].

Numerous efforts to identify small molecule inhibitors of cellular protein kinases have been aggressively pursued in several therapeutic areas. Unfortunately, there have been no successes in the area of infectious diseases, even insofar as the identification and validation of a protein kinase target. The cGMP-dependent protein kinases of apicomplexan parasites represent the first entry into this arena. Validation of this novel enzyme target for chemotherapeutic intervention provides impetus for future discovery efforts and establishes a precedent for the use of kinase inhibitors to treat infections of pathogenic microorganisms [52].

Largely due to the development of drug resistance to existing therapies, there is a particularly urgent need to develop new compounds to treat infection of *Plasmodium*. In the case of malaria, resistance to widely used 'first-line' therapies such as chloroquine or fansidar (a pyrimethamine/sulfa combination) has reached epidemic proportions [74, 75]. There is also a need for better control of opportunistic infections of *Toxoplasma gondii* and *Cryptosporidium parvum*, predominantly in immunocompromised patients. Primary treatment of *T. gondii* is achieved with a combination of pyrimethamine and sulfonamide drugs, a regimen that is often poorly tolerated [76, 77]. Currently there is no available therapy for the effective treatment of clinical cryptosporidiosis [78-81].

Recent drug discovery efforts are targeting CDKs to combat this infectious disease [82]. It is likely that *Plasmodium*-specific protein kinase inhibitors can be developed by screening a library of compounds [83]. The feasibility of production of a specific inhibitory compound against apicomplexan PKs [17].

The identification of novel drug targets has been greatly facilitated by the availability of the *P. falciparum* genomic database [41]. Two types of drug target can be identified from the genomic sequence: proteins that are entirely unique to the pathogen, and proteins that can be deduced from sequence conservation to play an essential role in pathogen survival [84].

Wiersma *et al.*, (2004) [81] have recently described a trisubstituted pyrrole, 4-[2-(4-fluorophenyl)-5-(1-methylpiperidine-4-yl)-1H-pyrrol-3-yl] pyridine (compound 1), which has broad spectrum activity against *T. gondii* in a murine toxoplasmosis model evaluated the efficacy of compound 1, a novel and selective inhibitor of protozoan cyclic GMP (cGMP)-dependent protein kinases, against a highly virulent form of *T. gondii* (RH strain) both *in vitro* and *in vivo*, using a murine model of acute toxoplasmosis [85]. Subsequent

studies in *T. gondii* validated PKG as a chemotherapeutic target, by showing that the gene is essential and by identifying catalytic site mutations that confer parasite resistance to compound 1 *in vitro* and in the mouse model [52,81].

The ability of the parasites to invade host cells is fundamental to the successful proliferation of obligate intracellular parasites such as the apicomplexa. Parasite protein kinases have been implicated in the control of invasion through previous studies with inhibitors. Staurosporine, a non-selective natural-product inhibitor of serine/threonine protein kinases, is a potent submicromolar inhibitor of *Plasmodium* red-blood cell invasion [86, 87] and *Toxoplasma* host cell invasion [88]. In *Toxoplasma* staurosporine inhibits the discharge of the microneme adhesin MIC2 [88] which mediates host cell attachment and invasion in *Toxoplasma* [89, 90]. *Toxoplasma* MIC2 is a homologue of the *Plasmodium* TRAP protein, which is essential for gliding motility and invasion by sporozoites [91, 92]. Inhibition of *Toxoplasma* MIC2 secretion, gliding motility, host cell attachment and invasion is observed with another compound, KT5926 [93]. In this case, a calcium-dependent protein kinase 1 (CDPK1) has been associated with the inhibitory effects of this compound [94]. These observations led to the investigation the potential role of PKG in parasite motility and invasion *T. gondii* tachyzoites, developmental stage that are readily cultured *in vitro* [81].

Furthermore, the spectrum of efficacy of compound 1, combined with the amino acid sequence conservation observed among apicomplexan parasite PKG enzymes compared with homologues from their vertebrate hosts [13], makes it likely that PKG catalytic site inhibitors will be useful against other apicomplexan parasites [52].

There is therefore a need for continued research aimed at the identification and development of new compounds for the treatment of toxoplasmosis either in monotherapy regimens or in combination with existing drugs. Widespread drug resistance exacerbates the problem and limits the options for effective malaria control. Current efforts to produce effective vaccines have not yet resulted in any significant success. Identification of drugs that interfere with parasite development could be a useful way to inhibit parasite growth in humans. Detailed knowledge of molecular mechanisms that control the life cycle of malaria parasite could provide crucial information needed to achieve this goal [24].

Several classes of CDK inhibitors have been developed including purines, paullones, and flavonoids that target the ATP binding site of CDKs [95-98]. Structural analysis of the binding of these inhibitors to the active site of CDKs aided the design of specific inhibitors [99]. Structural studies also identified key amino acids that are important for the specific interaction of the inhibitors to the CDK [100]. Comparison of plasmodial and mammalian CDK sequences reveals that there are structural differences within the ATP binding pocket that can be exploited to develop specific inhibitors. Specific inhibition is a very important consideration since there is significant homology within the protein kinase family [101]. Woodard *et al.*, (2003) have demonstrated that plasmodial CDKs have drug susceptibility differences from human CDKs suggesting that selective plasmodial CDK inhibitors can be developed [82,102, 103].

Chloroquine, mefloquine, and other frontline drugs for the treatment and prevention of malaria, are becoming increasingly ineffective [104]. Artemisinin analogues such as artesunate and arteether were later introduced and found to be quite effective, particularly against drug-resistant *P. falciparum,* but observations of drug-induced and dose-related neurotoxicity in animals have raised concern about the safety of these compounds for human use [104, 105]. Therefore, much effort is needed for discovery and development of new and less toxic antimalarial drugs.

Tryptanthrin (indolo[2,1-b]quinazoline-6,12-dione, **6**), an alkaloid isolated from the Taiwanese medicinal plant *Strobilanthes cusia*, and its substituted derivatives were revisited for a screening program conducted by the Walter Reed Army Institute of Research, Silver Spring, Maryland, U.S.A. [106-109]. The compounds displayed remarkable *in vitro* antimalarial activity against *P. falciparum*, both sensitive and multidrug-resistant strains [110]. The more potent analogs exhibit IC_{50} values in the range 0.43 to 10 ng/mL, about one-thousandth of the concentrations necessary to inhibit bacteria. Furthermore, the compounds are also found to be highly potent against strains of *P. falciparum* that are up to 5000-fold resistant to atovoquone, 50-fold resistant to chloroquine, and 20-fold resistant to mefloquine. Thus, this novel class of compounds has opened a new chapter for study in the chemotherapy of malaria [111].

Keenan *et al.*, (2005)[112] approach uses experimental and virtual screening methodologies to identify and refine chemical inhibitors and increase the success rate of discovering potent and selective inhibitors. The active pockets of the plasmodial CDKs are unique in terms of size, shape and amino acid composition compared with those of the mammalian orthologues. These differences exemplified through the use of screening assays, molecular modeling, and crystallography can be exploited for inhibitor design. To date, several classes of compounds including quinolines and oxindoles have been identified as selective inhibitors of the plasmodial CDK7 homologue, Pfmrk. From these initial studies and through the iterative rational drug design process, more potent, selective, and most importantly, chemically unique compound classes have been identified as effective inhibitors of the plasmodial CDKs and the malarial parasite [82, 112].

The lessons learned from several years of development of inhibitors against human CDKs may help in the discovery of a new generation of plasmodial CDKs. Several inhibitors of human CDKs have been reported, many have reached the nanomolar IC50 level, and present good selectivity [113-119].

Analysis of the crystallographic structures of binary complexes of enzymes and inhibitor provides a snapshot of the structure, which in certain point of view omits relevant features of the intermolecular interaction between the inhibitor and the enzyme, which should be considered in the development of more specific inhibitors. These features may be accessed using molecular dynamics simulations. Recent studies identified important features of the intermolecular interactions between R-roscovitine and related inhibitors and CDK2 (Fig. (**3**)) and CDK5 (Fig. (**4**)) [120]. The preference of the both CDKs for R-roscovitine over the S enantiomer, was also identified by the analysis of molecular dynamics simulations. Furthermore, the computer simulations showed that the cause of the stronger affinity for the R enantiomer is the presence of an intermolecular hydrogen bond between R-roscovitine and the CDKs not observed in S-roscovitine complexes. Two amino acid mutations in the ATP-binding pocket of CDK5/R-roscovitine that favor binding-enhanced electrostatic contributions, making the inhibitor more effective for CDK5 than for CDK2, were also identified in the molecular dynamics simulations. This suggests that the roscovitine-like inhibitors can be improved by enhancing their electrostatic interaction with the CDKs [120]. The use of the same molecular dynamics protocols may be able to identify the major structural features important for the inhibition of kinases.

Structural studies of human CDK2 complexed with inhibitors, using biocrystallography and structural bioinformatics [119-126], were able to explain the specificity and potency of the human CDK inhibitors, and suggest the use of structure-based design and combinatorial library design may improve the chance of discovering new inhibitors of CDK2 and also plasmodial CDKs.

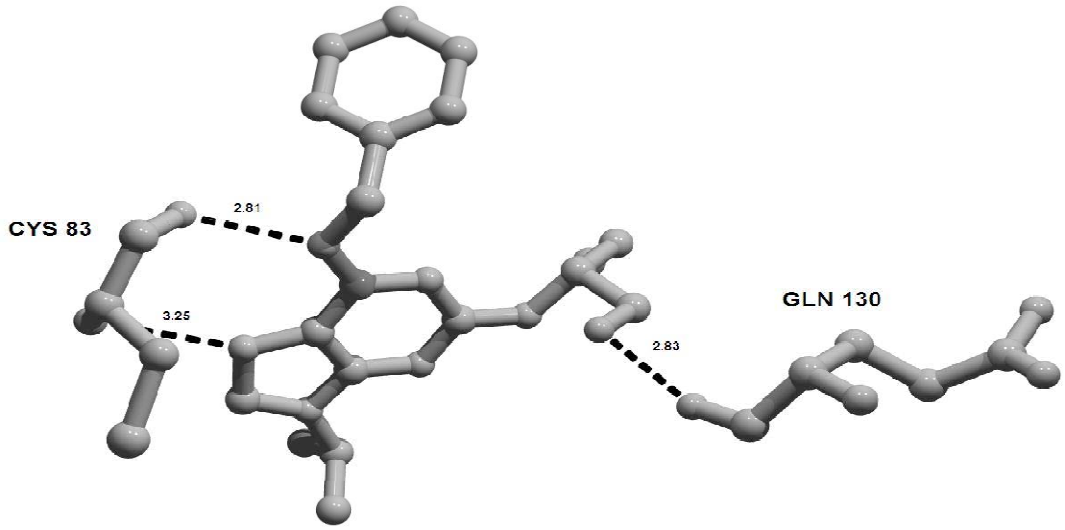


Fig. (4). Intermolecular hydrogen bonds observed in the complex of CDK5/R-roscotivine (PDB access code:1UNL). Distances are indicated in Å (10^{-10} m).

ACKNOWLEDGMENTS

This work was supported by grants from FAPESP (SMOLBNet 01/07532-0, 02/04383-7), CNPq, CAPES and Instituto do Milênio (CNPq-MCT). WFA (CNPq, 300851/98-7) is senior researcher of CNPq (Conselho Nacional de Pesquisas, Brazil).

ABBREVIATIONS

ATP	=	adenosine triphosphate
cAMP	=	cyclic-adenosine monophosphate
cdc	=	cell division cycle
CDK	=	cyclin-dependent kinase
CDPK	=	Ca^{2+}-dependent protein kinase
cGMP	=	cyclic guanosine monophosphate
crk	=	cdc2-related kinase
DDRT-PCR	=	differential display reverse transcriptase - polymerase chain reaction
DNA	=	deoxyribonucleic acid
ERK	=	extracellular signal-regulated kinase
IC50	=	inhibition concentration 50%
MAPK	=	mitogen-activated protein kinase
mrk	=	MO15-related kinase
mRNA	=	messenger ribonucleic acid
ORF	=	open reading frame
PDB	=	protein data bank
Pf	=	*Plasmodium falciparum*
PK	=	protein kinase
PKA	=	protein kinase A
PKB	=	protein kinase B
PKG	=	GMP-dependent protein kinase
RPA	=	replication protein A
Tg	=	*Toxoplasma gondii*
TPK	=	protein kinase from *Toxoplasma*

REFERENCES

[1] Trouiller, P., Torreele E., Olliaro, P., White, N., Foster, S., Wirth, D. and Pecoul, B. (2001) *Trop. Med. Int. Health*., **6**(11), 945-951.
[2] Sinski, E. and Behnke, J.M. (2004) *Pol. J. Microbiol*., **53**, 67-73.
[3] Belli, S.I., Walker, R.A., and Flowers, S.A. (2005) *Proteomics*, **5**(4), 918-924.
[4] Ross-Macdonald, P.B., Graeser, R., Kappes, B., Franklin, R. and Williamson, DH. (1994) *Eur. J. Biochem.,* **220**(3), 693-701.
[5] Doerig, C., Doerig, C., Horrocks, P., Coyle, J., Carlton, J., Sultan, A., Arnot, D. and Carter, R. (1995) *Mol. Biochem. Parasitol.,* **70**(1-2), 167–174.
[6] Li, J.L., Robson, K.J.H., Chen, J.L., Targett, G.A.T. and Baker, D.A. (1996) *Eur. J. Biochem.,* **241**(3), 805–813.
[7] Wastling, J.M. and Kinnaird, J.H. (1998) *Mol. Biochem. Parasitol.,* **94**(1), 143–148.
[8] Khan, F., Tang, J., Qin, C.-L. and Kim, K. (2002) *Mol. Microbiol.,* **45**(2), 321–332.
[9] Graeser, R., Wernli, B., Franklin, R.M. and Kappes, B. (1996) *Mol. Biochem. Parasitol.,* **82**(1), 37–49.
[10] Kinnaird, J.H., Bumstead, J.M., Mann, D.J., Ryan, R., Shirley, M.W., Shiels, B.R. and Tomley, F.M. (2004) *Int. J. Parasitol*., **34**(6), 683-92.
[11] Gould, K.L. (2000) *In Cyclin-dependent kinases in Protein Kinase Functions*. Jim Woodgett Ed.; Frontiers in molecular biology. Oxford University Press: GreadBritain., p. 276-302.
[12] Lohmann, S. M., Vaandrager, A. B., Smolenski, A., Walter, U. and De Jonge, H. R. (1997) *Trends Biochem. Sci.,* **22**(8), 307–312.
[13] Gurnett, A.M., Liberator, P.A., Dulski, P.M., Salowe, S.P., Donald, R.G.K., Anderson, J.W., Wiltsie, J., Diaz, C.A., Harris, G., Chang, B. *et al.* (2002) *J. Biol. Chem.,* **277**(18), 15913–15922.
[14] Donald, R.G.K. and Liberator, P.A. (2002) *Mol. Biochem. Parasitol.,* **120**(2), 165–175.
[15] Deng, W. and Baker, D. A. (2002) *Mol. Microbiol.,* **44**(5), 1141–1151.
[16] Diaz, C.A., Allocco, J., Powles, M.A., Yeung, L., Donald, R.G., Anderson, J.W. and Liberator, P.A. (2006) *Mol. Biochem. Parasitol*. **146**(1), 78-88.
[17] Deng, W., Parbhu-Patel, A., Meyer, D.J. and Baker, D.A. (2003) *Biochem. J.,* **374**, 559-65.
[18] Wolfe, L., Corbin, J. D. and Francis, S. H. (1989) *J. Biol. Chem.*, **264**(13), 7734–7741.
[19] Smith, J. A., Reed, R. B., Francis, S. H., Grimes, K. and Corbin, J. D. (2000) *J. Biol. Chem.,* **275**(1), 154–158.
[20] Taylor, M. K. and Uhler, M. D. (2000) *J. Biol. Chem*., **275**(36), 2805–28062.
[21] Schneider, A.G. and Mercereau-Puijalon, O. (2005) *BMC Genomics*, **6**(1), 30.
[22] Vanhaesebroeck, B. and Alessi, D. R. (2000) *Biochem. J.,* **346,** 561–576.
[23] Ridley, R. G. (2002) *Nature (London),* **415**(6872), 686–693.
[24] Kumar, A., Vaid, A., Syin, C. and Sharma, P. (2004) *J. Biol. Chem*., **279**(23), 24255-64.

[25] Carucci, D.J., Gardner, M.J., Tettelin, H., Cummings, L.M., Smith, H.O., Adams, M.D., Hoffman, S.L. and Venter, J.C. (1998) *Expert. Rev. Mol. Med*., **1998**, 1-9.
[26] Doerig, C., Endicott, J. and Chakrabarti, D. (2002) *Int. J. Parasitol*., **32**(13), 1575-1585.
[27] Harmse, L., van Zyl, R., Gray, N., Schultz, P., Leclerc, S., Meijer, L., Doerig, C. and Havlik, I. (2001) *Biochem. Pharmacol*., **62**(3)**,** 341-348.
[28] Holton, S., Merckx, A., Burgess, D., Doerig, C., Noble, M. and Endicott, J. (2003) *Structure (Camb).*, **11**(11), 1329-37.
[29] Keenan, S.M. and Welsh, W.J. (2004) *J. Mol. Graph. Model*., **22**(3), 241-7.
[30] Doerig, C., Endicott, J. and Chakrabarti, D. (2002) *Int. J. Parasit.,* **32**(13), 1575–1585.
[31] Knighton, D.R., Zheng, J.H., Ten-Eyck, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S. and Sowadski, J.M., (1991) *Science,* **253**(5018), 407–414.
[32] De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O. and Kim, S.H. (1993) *Nature,* **363**(6430), 595–602.
[33] MANHANI, K.K., ARCURI, H.A., SILVEIRA, N.J.F., UCHOA, H.B., DE AZEVEDO JR., W.F. AND CANDURI, F. (2005) *J. MOL. MODEL,* **12**(1), 42-8.
[34] Bracchi-Ricard, V., Barik, S., Delvecchio, C., Doerig, C., Chakrabarti, R. and Chakrabarti, D. (2000) *Biochem. J.,* **347**, 255-263.
[35] Li, J.L., Robson, K.J., Chen, J.L., Targett, G.A. and Baker, D.A. (1996) *Eur. J. Biochem.,* **241**(3), 805–813.
[36] Li, Z., Le Roch, K., Geyer, J.A., Woodard, C.L., Prigge, S.T., Koh, J., Doerig C. and Waters N.C. (2001) *Biochem. Biophys. Res. Commun.,* **288**(5), 1207–1211.
[37] Chen, Y., Jirage, D., Caridha, D., Kathcart, A.K., Cortes, E.A., Dennull, R.A., Geyer, J.A., Prigge, S.T. and Waters, N.C. (2006) *Mol. Biochem. Parasitol*. **149**(1), 48-57.
[38] Rangarajan, R., Bei, A., Henry, N., Madamet, M., Parzy, D., Nivez, M.P., Doerig, C. and Sultan, A. (2006) *Exp. Parasitol*., **112**(3), 202-7.
[39] Vinkenoog, R., Veldhuisen, B., Speranca, M.A., del Portillo, H.A., Janse, C. and Waters, A.P. (1995) *Mol. Biochem. Parasitol*., **71**(2), 233-241.
[40] Syin, C., Parzy, D., Traincard, F., Boccaccio, I., Joshi, M.B., Lin, D.T., Yang, X.M., Assemat, K., Doerig, C. and Langsley, G. (2001) *Eur. J. Biochem*., **268**(18), 4842-4849.
[41] Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M. and Barrell, B. (2002) *Nature*, **419**(6906), 498-511.
[42] Aravind, L., Iyer, L.M., Wellems, T.E. and Miller, L.H. (2003) *Cell*, **115**(7), 771-785.
[43] Chandran, V., Stollar, E.J., Lindorff-Larsen, K., Harper, J.F., Chazin, W.J., Dobson, C.M., Luisi, B.F. and Christodoulou, J. (2006) *J. Mol. Biol*., **357**(2), 400-10.
[44] Hotta, C.T., Gazarini, M.L., Beraldo, F.H., Varotti, F.P., Lopes, C., Markus, R.P., Pozzan, T. and Garcia, C.R. (2000) *Nat. Cell Biol.,* **2**(7), 466-468.
[45] Kawamoto, F., Alejo-Blanco, R., Fleck, S.L., Kawamoto, Y. and Sinden, R.E. (1990) *Mol. Biochem. Parasitol*., **42**(1), 101-108.
[46] Kawamoto, F., Fujioka, H., Murakami, R., Syafruddin, Hagiwara, M., Ishikawa, T. and Hidaka, H. (1993) *Eur. J. Cell Biol.,* **60**(1), 101-107.
[47] Li, J.L., Baker, D.A. and Cox, L.S. (2000) *Biochim. Biophys. Acta*., **1491**(1-3), 341-349.
[48] Farber, P.M., Graeser, R., Franklin, R.M. and Kappes, B. (1997) *Mol. Biochem. Parasitol*., **87**(2), 211-216.
[49] Le Roch, K., Sestier, C., Dorin, D., Waters, N., Kappes, B., Chakrabarti, D., Meijer, L. and Doerig, C. (2000) *J. Biol. Chem.,* **275**(12), 8952-8958.
[50] Garweg JG. (2005) *Parasite Immunol*., **27**(3), 61-68.
[51] Khan, F., Tang, J., Qin, C.L. and Kim, K. (2002) *Mol. Microbiol*., **45**(2), 321-332.
[52] Donald, R.G., Allocco, J., Singh, S.B., Nare, B., Salowe, S.P., Wiltsie, J. and Liberator, P.A. (2002) *Eukaryot Cell,* **1**(3), 317-328.
[53] Pokholok, D.K., Zeitlinger, J., Hannett, N.M., Reynolds, D.B. and Young, R.A. (2006) *Science*, **313**(5786), 533-6.
[54] Roisin, M.P., Robert-Gangneux, F., Creuzet, C. and Dupouy-Camet, J. (2000) *Parasitol. Res*., **86**(7), 588-598.
[55] Brumlik, M.J., Wei, S., Finstad, K., Nesbit, J., Hyman, L.E., Lacey, M., Burow, M.E. and Curiel, T.J. (2004) *Int. J. Parasitol*., **34**(11), 1245-54.
[56] Wei, S., Marches, F., Daniel, B., Sonda, S., Heidenreich, K. and Curiel, T. (2002) *Int. J. Parasitol*., **32**(8), 969-977.
[57] Dorin, D., Alano, P., Boccaccio, I., Ciceron, L., Doerig, C., Sulpice R., Parzy, D. and Doerig, C. (1999) *J. Biol. Chem.,* **274**(42), 29912–29920.
[58] Lin, D.T., Goldman, N.D. and Syin, C. (1996) *Mol. Biochem. Parasitol.,* **78**(1-2), 67–77.
[59] Kieschnick, H., Wakefield, T., Narducci, C. A. and Beckers, C. (2001) *J. Biol. Chem*., **276**(15), 12369-12377.
[60] Moudy, R., Manning, T.J. and Beckers, C. (2001) *J. Biol. Chem.*, **276**(44), 41492-41501.
[61] Nagamune, K. and Sibley, L.D. (2006) *Mol. Biol. Evol*. **23**(8), 1613-27.
[62] Billker, O., Dechamps, S., Tewari, R., Wenig, G., Franke-Fayard, B. and Brinkmann, V. (2004) *Cell,* **117**(4), 503-514.
[63] Tran, M.Q. (2005) *Clin. Nephrol*., **63**(4), 305-309.
[64] Okhuysen, P.C. and Chappell, C.L. (2002) *Int. J. Parasitol*., **32**(5), 517-525.
[65] Tzipori, S. and Griffiths, J.K. (1998) *Adv. Parasitol.,* **40**, 5-36.
[66] Millership, J.J. and Zhu, G. (2002) *Int. J. Parasitol*., **32**(12), 1477-1485.
[67] Zhu, G., Marchewka, M.J. and Keithly, J.S. (1999) *FEMS Microbiol. Lett*., **176**(2), 367-372.
[68] Millership, J.J., Cai, X. and Zhu, G. (2004) Microbiology, **150**(5), 1197-1205.
[69] Henricksen, L.A., Carter, T., Dutta, A. and Wold, M.S. (1996) *Nucleic Acids Res*., **24**(15), 3107-3112.
[70] Treuner, K., Findeisen, M., Strausfeld, U. and Knippers, R. (1999) *J. Biol. Chem.,* **274**(22), 15556-15561.
[71] Garcia-Echeverria, C., Traxler, P. and Evans, D. B. (2000) *Med. Res. Rev*., **20**(1)**,** 28-57.
[72] Johnson, L. N. (2001) *Ernst-Schering-Res-Found-Workshop,* **34,** 47-69.
[73] Sridhar, R., Hanson-Painton, O. and Cooper, D. R. (2000) *Pharmacol. Res.,* **17**(11)**,** 1345-1353.
[74] Sibley, C.H. and Macreadie, I. (2001) *IUBMB Life*., **52**(6), 285-289.
[75] Winstanley, P. (2002) *Int. J. Infect. Dis*., **6**(4), 246-252.
[76] Haverkos, H.W. (1987) *Am. J. Med.,* **82**(5), 907-914.
[77] Fung, H.B. and Kirschenbaum, H.L. (1996) *Clin. Ther*., **18**(6), 1037-1056.
[78] Chappell, C.L. and Okhuysen, P.C., (2002) *Curr. Opin. Infect. Dis.,* **15**(5), 523–527.
[79] Mead, J.R. (2002) *Drug Resist. Update,* **5**(1), 47–57.
[80] Leav, B.A., Mackay, M. and Ward, H.D. (2003) *Clin. Infect. Dis.,* **36**(7), 903–908.
[81] Wiersma, H.I., Galuska, S.E., Tomley, F.M., Sibley, L.D., Liberator, P.A. and Donald, R.G. (2004) *Int. J. Parasitol*., **34**(3), 369-380.
[82] Woodard, C.L., Li, Z., Kathcart, A.K., Terrell, J., Gerena, L., Lopez-Sanchez, M., Kyle, D.E., Bhattacharjee, A.K., Nichols, D.A., Ellis, W., Prigge, S.T., Geyer, J.A. and Waters, C. (2003) *J. Med. Chem*., **46**(18), 3877-3882.
[83] Doerig, C., Meijer, L. and Mottram, J. C. (2002) *Trends Parasitol*., **18**(8), 366–371
[84] Pattanaik, P., Raman, J. and Balaram, H. (2002) *Curr. Top. Med. Chem*., **2**(5), 483-505.
[85] Nare, B., Allocco, J.J., Liberator, P.A. and Donald, R.G. (2002) *Antimicrob. Agents Chemother*., **46**(2), 300-307.
[86] Ward, G.E., Fujioka, H., Aikawa, M. and Miller, L.H. (1994) *Exp. Parasitol.,* **79**(3), 480–487.
[87] Dluzewski, A.R. and Garcia, C.R. (1996) *Experientia,* **52**(6), 621–623.
[88] Carruthers, V.B., Giddings, O.K. and Sibley, L.D. (1999) *Cell Microbiol.,* **1**(8), 225–235.
[89] Brossier, F., Jewett, T.J., Lovett, J.L. and Sibley, L.D. (2003) *J. Biol. Chem.,* **278**(8), 6229–6234.
[90] Huynh, M.H., Rabenau, K.E., Harper, J.M., Beatty, W.L., Sibley, L.D. and Carruthers, V.B. (2003) *Eur. Mol. Biol. Org. J.,* **22**(9), 2082–2090.

[91] Sultan, A.A., Thathy, V., Frevert, U., Robson, K.J., Crisanti, A., Nussenzweig, V., Nussenzweig, R.S. and Menard, R. (1997) *Cell,* **90**(3), 511–522.
[92] Kappe, S., Bruderer, T., Gantt, S., Fujioka, H., Nussenzweig, V. and Menard, R. (1999) *J. Cell Biol.,* **147**(5), 937–944.
[93] Dobrowolski, J.M., Carruthers, V.B. and Sibley, L.D. (1997) *Mol. Microbiol.,* **26**(1), 163–173.
[94] Kieschnick, H., Wakefield, T., Narducci, C.A. and Beckers, C. (2001) *J. Biol. Chem.,* **276**(5), 12369–12377.
[95] Vesely, J., Havlicek, L., Strnad, M., Blow, J. J., Donella-Deana, A., Pinna, L., Letham, D. S., Kato, J., Detivaud, L. and Leclerc, S., *et al.* (1994) *Eur. J. Biochem.*, **224**(2), 771-786.
[96] Schultz, C., Link, A., Leost, M., Zaharevitz, D. W., Gussio, R., Sausville, E. A., Meijer, L. and Kunick, C. (1999) *J. Med. Chem.*, **42**(15), 2909-2919.
[97] Gray, N., Detivaud, L., Doerig, C. and Meijer, L. (1999) *Curr. Med. Chem.*, **6**(9)**,** 859-875.
[98] Schoepfer, J., Fretz, H., Chaudhuri, B., Muller, L., Seeber, E., Meijer, L., Lozach, O., Vangrevelinghe, E. and Furet, P. (2002) *J. Med. Chem.*, **45**(9)**,** 1741-1747.
[99] Legraverend, M., Tunnah, P., Noble, M., Ducrot, P., Ludwig, O., Grierson, D. S., Leost, M., Meijer, L. and Endicott, J. (2000) *J. Med. Chem.*, **43**(7)**,** 1282-1292.
[100] Noble, M. E. and Endicott, J. A. (1999) *Pharmacol. Ther.*, **82**(2-3), 269-278.
[101] Hanks, S. K., Quinn, A. M. and Hunter, T. (1988) *Science*, **241**(4861)**,** 42-52.
[102] Waters, N. C., Woodard, C. L. and Prigge, S. T. (2000) *Mol. Biochem. Parasitol.,* **107**(1)**,** 45-55.
[103] Xiao, Z., Waters, N. C., Woodard, C. L., Li, Z. and Li, P. K. (2001) *Bioorg. Med. Chem. Lett.*, **11,** 2875-2878.
[104] Vroman, J.A., Alvin-Gaston M. and Avery M.A. (1999) *Curr. Pharm. Design,* **5**(2), 101–138.
[105] Bhattacharjee A.K. and Karle. J.M. (1999) *Chem. Res. Toxicol.,* **12**(5), 422–428.
[106] Baker W.R. and Mitscher L.A. (1995) U.S. Patent 5 441, 955 .
[107] Eguchi, S., Takeuchi H. and Matsushita Y. (1992) *Heterocycles*, **33**(1), 153–156.
[108] Honda G. and Tabata. M. (1979) *Planta Medica,* **36**(1)**,** 85–90.
[109] Mitscher, L.A., Wong, W.C., DeMeulenaere, T., Sulko, J. and Drake, S. (1981) *Heterocycles,* **15**(2), 1017–1021.
[110] Pitzer, K.K., Kyle D.E. and Gerena L. (2001) Patent, 6 284 772 .
[111] Bhattacharjee, A.K., Hartell, M.G., Nichols, D.A., Hicks, R.P., Stanton, B., van Hamont, J.E. and Milhous, W.K. (2004) *Eur. J. Med. Chem*., **39**(1), 59-67.
[112] Keenan, S.M., Geyer, J.A., Welsh, W.J., Prigge, S.T. and Waters, N.C. (2005) *Comb. Chem. High Throughput Screen*., **8**(1), 27-38.
[113] De Azevedo, W.F.Jr.; Mueller-Dieckmann, H.J.; Schulze- Gahmen, U.; Worland, P.J.; Sausville, E. and Kim S.-H. (1996) *Proc. Natl. Acad. Sci. USA.*, **93**(7), 2735-2740.
[114] Kim, S.-H.; Schulze-Gahmen, U.; Brandsen, J. and De Azevedo, W. F. Jr. (1996) *Prog. in Cell Cycle Res.*, **2**, 137-145.
[115] Canduri, F.; Uchoa, H.B. and de Azevedo, W.F.Jr. (2004) *Biochem. Biophys. Res. Commun.*, **324**(2), 661-666.
[116] De Azevedo, W.F.Jr.; Canduri, F. and Silveira, N.J.F. (2002) *Biochem. Biophys. Res. Commun.*, **293**(1), 566-571.
[117] De Azevedo, W.F.Jr.; Leclerc, S.; Meijer, L.; Havlicek, L.; Strnad, M. and Kim, S.-H. (1997) *Eur. J. Biochem.*, **243**(1-2), 518-526.
[118] De Azevedo, W.F.Jr., Gaspar, R.T., Canduri, F.; Camera, J.C.Jr., and Silveira, N.J.F. (2002) *Biochem. Biophys. Res. Commun.*, **297**(5)**,** 1154-1158.
[119] De Azevedo, W. F. Jr., and Canduri, F. (2005) *Current Computer-Aided Drug Design,* **1**, 53-64.
[120] Zhang, B., Tan, V. B. C., Lim, K. M., Tay, T. E. and Zhuang, S. (2006) *J. Mol. Model (In press)*
[121] Basso, L. A., da Silva, L. H. P., Fett-Neto, A. G., de Azevedo, W. F. Jr., Moreira, I. S., Palma, M. S., Calixto, J. B., Astolfi Filho, S., dos Santos, R. R., Soares, M. B. and Santos, D. S. (2005) *Mem. Inst. Oswaldo Cruz,* ***100***(6), 475-506.
[122] De Azevedo, W. F., Canduri, F., Basso, L. A., Palma, M. S. and Santos, D. S. (2006) *Cell Biochem. Biophys.* **44**(3), 405-411.
[123] Oliveira, J. S., Pereira, J. H., Canduri, F.,Rodrigues, N. C., de Souza, O. N, de Azevedo Jr., W. F.,Basso, L. A. and Santos, D. S. (2006) *J. Mol. Biol.* **359**(3), 646–666.
[124] da Silveira, N. J., Uchoa, H. B., Canduri, F.,Pereira, J. H., Camera Jr., J. C., Basso, L. A., Palma, M. S., Santos, D. S. and de Azevedo Jr., W. F. (2004) *Biochem. Biophys. Res. Commun.*, **322**(1), 100–104.
[125] Uchoa, H. B., Jorge, G. E., Freitas da Silveira, N. J., Camera J. C., Canduri, F. and De Azevedo, WF. (2004) *Biochem. Biophys. Res. Commun.* **325**(4),1481–1486.
[126] da Silveira, N. J. F., Bonalumi, C. A., Uchoa, H. B., Pereira, J. H., Canduri, F.,Pereira, J. H. and de Azevedo Jr., W. F. (2006) *Cell Biochem. Biophys.* **44**, 366-374.

Received: December 12, 2005 Accepted: May 26, 2006 Updated: September 11, 2006

Shikimate Kinase: A Potential Target for Development of Novel Antitubercular Agents

José H. Pereira[†], Igor B. Vasconcelos[§], Jaim S. Oliveira[§], Rafael A. Caceres[¤], Walter F. de Azevedo Jr.[¤], Luis A. Basso[§*] and Diógenes S. Santos[§*]

[†]*Structural Biology Department, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA;* [§]*Centro de Pesquisas em Biologia Molecular e Funcional, Faculdade de Farmácia, Instituto de Pesquisas Biomédicas, Pontifícia Universidade Católica do Rio Grande do Sul. Avenida Ipiranga 6681, Tecnopuc, Partenon 90619-900, Porto Alegre, RS, Brazil and* [¤]*Faculdade de Biociências-PUCRS. Avenida Ipiranga 6681, 90619-900, Porto Alegre, RS, Brazil*

Abstract: Tuberculosis (TB) remains the leading cause of mortality due to a bacterial pathogen, *Mycobacterium tuberculosis.* However, no new classes of drugs for TB have been developed in the past 30 years. Therefore there is an urgent need to develop faster acting and effective new antitubercular agents, preferably belonging to new structural classes, to better combat TB, including MDR-TB, to shorten the duration of current treatment to improve patient compliance, and to provide effective treatment of latent tuberculosis infection. The enzymes in the shikimate pathway are potential targets for development of a new generation of antitubercular drugs. The shikimate path- way has been shown by disruption of *aroK* gene to be essential for the *Mycobacterium tuberculosis*. The shikimate kinase (SK) catalyses the phosphorylation of the 3-hydroxyl group of shikimic acid (shikimate) using ATP as a co-substrate. SK belongs to family of nucleo- side monophosphate (NMP) kinases. The enzyme is an / protein consisting of a central sheet of five parallel -strands flanked by - helices. The shikimate kinases are composed of three domains: Core domain, Lid domain and Shikimate-binding domain. The Lid and Shikimate-binding domains are responsible for large conformational changes during catalysis. More recently, the precise interactions be- tween SK and substrate have been elucidated, showing the binding of shikimate with three charged residues conserved among the SK se- quences. The elucidation of interactions between *Mt*SK and their substrates is crucial for the development of a new generation of drugs against tuberculosis through rational drug design.

Key Words: tuberculosis, malaria, shikimate kinase, shikimate pathway, drug design, drug target, *Mycobacterium tuberculosis*, *Plasmodium* sp

1. THE SHIKIMATE PATHWAY

The shikimate pathway was discovered as a biosynthetic route through the studies of Bernhard Davis and David Sprinson and their collaborators [1,2]. The shikimate pathway links metabolism of carbohydrates to biosynthesis of aromatic compounds through seven metabolic steps, where phosphoenolpyruvate and erythrose 4-phosphate are converted to chorismic acid [3,4]. Chorismic acid is a common precursor for the synthesis of aromatic compounds, such as aromatic amino acids, folate, ubiquinone and menaquinones. Amongst them, the only ones that can be synthesized by humans are tyrosine, which is synthesized from phenylalanine through a reaction catalyzed by phenylalanine hidroxylase enzyme, and ubiquinone from tyrosine through a cascade of eight aromatic precursors [5].

The molecular organization of the shikimate pathway enzymes varies between taxonomic groups [6]. Bacteria have seven individual polypeptides, which are encoded by separate genes. Plants have a molecular arrangement similar to bacteria [7], with the exception of dehydroquinase (DHQase, third enzyme) and shikimate dehydrogenase (fourth enzyme) which have been shown to be present as separate domains on a bifunctional polypeptide [8]. In fungi and apicomplexan parasites (*Toxoplasma gondii*) the shikimate pathway has been shown to include monofunctional 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase and chorismate synthase (CS) enzymes and a pentafunctional polypeptide termed AROM, which accounts for the remaining five shikimate pathway reactions [9].

The shikimate pathway enzymes are attractive targets for development of non-toxic antibacterial [10] and herbicides [11], because this pathway is essential for algae, higher plants, bacteria, fungi, whereas it is absent from mammals [12]. Thus, in the case of bacterial diseases, inhibition of any of shikimate pathway enzymes is unlikely to cause toxic side effects on the host. In addition, the importance of shikimate pathway can be indicated by the finding that deletion of the *aroA* gene, which codes EPSPS, causes *Streptomyces pneumoniae* and *Bordetella bronchiseptica* strains to be attenuated for virulence [13,14].

The shikimate pathway has also been discovered in apicomplexan parasites providing several targets for the development of new antiparasite drugs [15]. *In vitro* growth of apicomplexan parasites such as *Plasmodium falciparum* (malaria), *Toxoplasma gondii* (toxoplasmosis), and *Cryptosporidium parvum* (cryptosporidiosis) was inhibited by the herbicide glyphosate, a well-characterized inhibitor [16] of the shikimate pathway enzyme 5-enolpyruvyl shikimate 3-phosphate synthase (EPSPS), at concentration of 1-6 mM [15]. In *P. falciparum* and *T. gondii*, this inhibitory effect was reversed by co-addition o p-aminobenzoate (PABA) or folate suggesting that this pathway is essential for the biosynthesis of folate precursors. Moreover, several shikimate pathway enzymes has been detected in *T. gondii* and *P. falciparum* extracts [15,17] and a gene encoding a pentafunctional polypeptide AROM has been described in *T. goondii* indicating that a complete shikimate pathway is present in this parasite [18]. Whereas the shikimate pathway gives an array of compounds in bacteria, fungi and plants, only its folate biosynthesis function has been established is in apicomplexan parasites. All folate pathway enzymes, which convert GTP to deriva-

*Address correspondence to this author at the Centro de Pesquisas em Biologia Molecular e Funcional, Faculdade de Farmácia, Instituto de Pesquisas Biomédicas, Pontifícia Universidade Católica do Rio Grande do Sul. Avenida Ipiranga 6681, Tecnopuc, Partenon 90619-900, Porto Alegre, RS, Brazil; Tel: +55-51-33203629; Fax, +55-51-3320-3629; E-mail: diogenes@pucrs.br

1389-4501/07 $50.00+.00

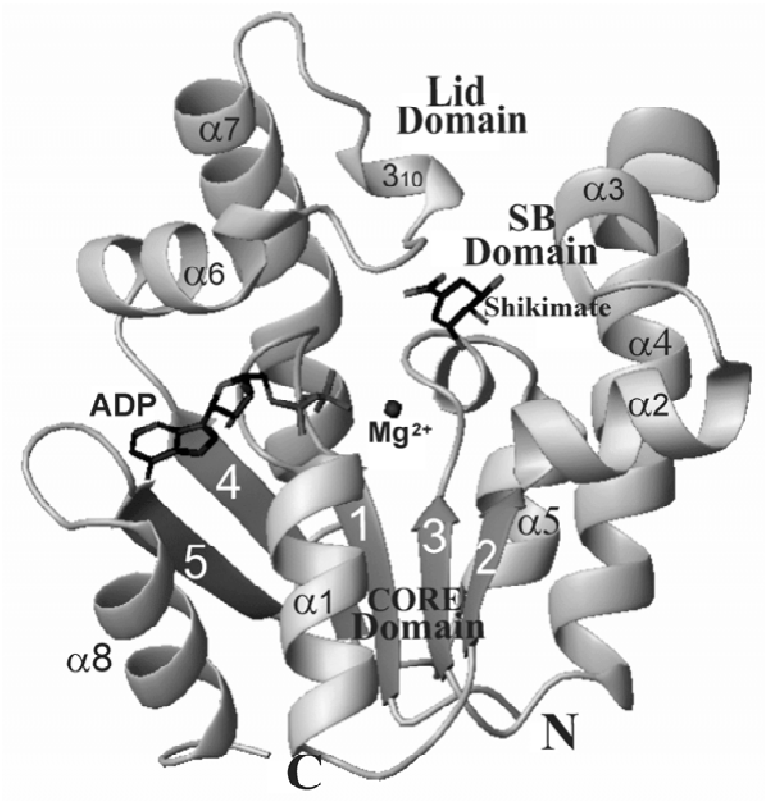


Fig. (1). Overall structure of Shikimate kinase from *M. tuberculosis* [61]. *Mt*SK displays an / -fold and the precise ordering of the strands 23145 in the parallel -sheet classifies *Mt*SK as belonging to the family as the NMP kinases. The SKs are composed of three domains: CORE, LID and Shikimate-binding (SB) domains. The positions of the ADP, Mg^{2+} and shikimate are shown in the *Mt*SK structure.

tives of tetrahydrofolate, were found in Apicomplexa when the complete genome was sequenced [19].

Two shikimic acid analogs, 6-*S*-fluorshikimate and 6-*R*-fluorshikimate, have been shown to inhibit *P. falciparum* growth and inhibition shown to be specific to the shikimate pathway [20]. Despite the completion of *P.falciparum* genome sequence [21], only a single gene encoding the chorismate synthase enzyme and a potential bifunctional SK/EPSP synthase protein has been identified in the genome annotation [22]. The coding DNA sequence of *P. falciparum* chorismate synthase has been cloned and the protein has been shown to be located on cytosol by immunological studies [23]. Moreover, chorismate synthase, which catalyzes the last step of shikimate pathway, has been shown to be required for *Plasmodium falciparum* growth, as disruption of expression by RNA interference decreased parasite growth [19]. It has been proposed that the missing shikimate pathway enzymes are either substituted by non-homologous enzymes that catalyze the same reaction or that the enzymes are homologous but too divergent to be identified [22].

In mycobacteria, the chorismic acid intermediate is a precursor for the synthesis of naphthoquinones, menaquinones, and mycobactins, besides aromatic amino acids [24]. The salycilate-derived mycobactins siderophores have been shown to be essential for *M. tuberculosis* growth in macrophages [25]. Particularly, in *Mycobacterium tuberculosis*, the shikimate pathway has been shown to be essential for the bacterial viability. The disruption of *aroK* gene, which codes for the shikimate kinase enzyme (SK), was only possible when the second functional copy of *aroK* was integrated into the chromosome. Moreover, excision of the second integrated copy of *aroK* by the L5 excisionase could not be achieved in a *M. tuberculosis* strain carrying the disrupted copy of *aroK* gene, but was possible in a strain carrying a wild-type copy [26].

2. SHIKIMATE KINASE

The shikimate kinase (SK; EC 2.7.1.71), the fifth enzyme of the pathway, catalyses the regiospecific phosphorylation of the 3-hydroxyl group of shikimic acid (shikimate) using ATP as a co-substrate. In *Escherichia coli*, the SK reaction is catalyzed by two isoforms: SK I encoded by the *aroK* gene [27] and SK II encoded by the *aroL* gene [28]. The major difference between the isoenzymes is their K_m for shikimate, 20 mM for the SK I and 0.2 mM for the SK II enzyme [29]. The SK II isoform appears to play a dominant role in the shikimate pathway, its expression is controlled by the *tyr*R regulator, and it is repressed by tyrosine and tryptophan [30,31]. The physiological role of SK I in *E.coli* is not clear. Since mutations in SK I are associated with sensitivity to the antibiotic mecillinam [32] it has been suggested that SK I may have an alternative biological role that is distinct and unrelated to its shikimate kinase activity [29]. As pointed out by Parish and Stoker [26], if *M. tuberculosis aroK*-encoded SK I possess a similar activity it is possible that disruption of this activity can account for the observed inability of *M. tuberculosis* to grow in the absence of a functional copy of *aroK* gene. However, the actual nature of second activity of *aroK* gene product remains to be established. Contrary to the presence of isoenzymes in *E. coli*, complete genome sequences of a number of bacteria, for example, *Haemophilus influenzae* and *Mycobacterium tuberculosis*, have revealed the presence of only one SK-coding gene. Most of these SKs appear to be encoded by *aroK* rather than *aroL* because their amino acid sequences have higher degree of identity with *E.coli* SK I. The kinetic parameters for *aroK*-encoded *M.tuberculosis* SK (*Mt*SK) are more similar to those of *aroL*-encoded *E.coli* SK II than to those of *aroK*-encoded *E.coli* SK I. Thus, the *Mt*SK K_m value (0.41mM) for shikimate suggests that not all *aroK*-encoded SKs have high K_m values for shikimate [33].

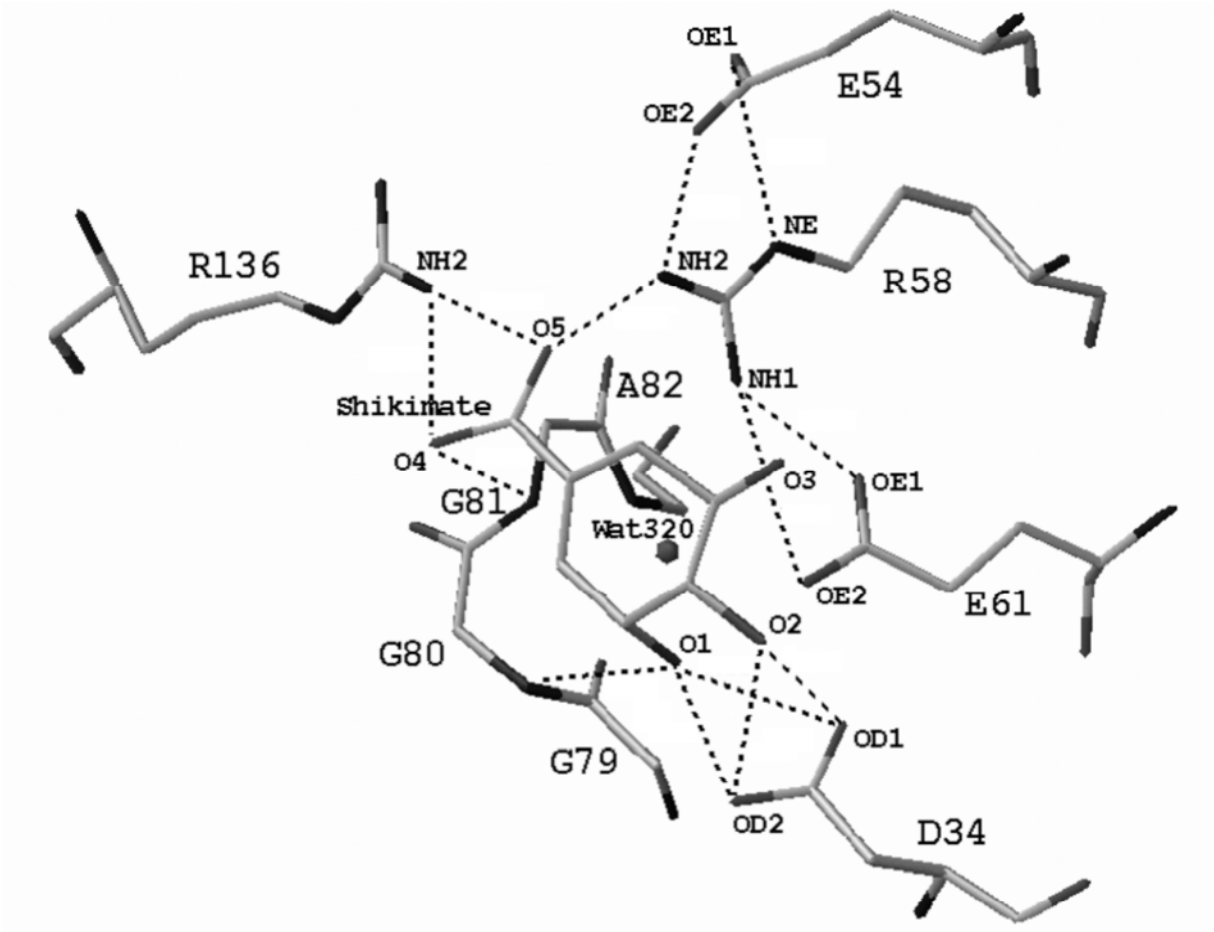


Fig. (2). Interactions involved in the binding of shikimate to the *Mt*SK active site (PDB access code: 1WE2) [61]. Hydrogen bonds are represented as broken lines. For clarity, the the water320-mediated hydrogens bonds between Gly79, Gly80, Gly81, Ala82, Arg58, and Glu61 protein residues and 3-hydroxyl group of shikimate are not shown.

2.1. The Enzyme Fold

Shikimate kinase displays an / -fold and consists of five central parallel -sheet with the strand order 23145, flanked by - helices [34] (Fig. (**1**)). The three crystal structures of SK from *Erwinia chrysanthemi* (*Erc*SK) [34,35] showed that SK belongs to the same structural family of nucleoside monophosphate (NMP) kinases for which structures are known for adenylate kinase (AK) [36,37], guanylate kinase [38], uridylate kinase [39] and thymidine kinase [40].

The NMP kinases are composed of three domains: CORE, LID and NMP-binding (NMPB) domains [41]. A characteristic feature of the NMP kinases is that they undergo large conformational changes during catalysis, for which AK is the most extensively studied [41]. There are two flexible regions of the structures that are responsible for movement: one is the NMP-binding site which is formed by a series helices between strands 2 and 3 of parallel - sheet and the other is the LID domain, a region of varied size and structure following the fourth -strand of the sheet [42,43]. In SK, the shikimate-binding (SB) domain corresponds to the NMPB domain of NMP kinases.

2.2. Functional Motifs

Three functional motifs of nucleotide-binding enzymes are recognizable in shikimate kinase, including a Walker A-motif (A-motif), a Walker B-motif (B-motif), and an adenine-binding loop. The Walker A-motif is located between the first -strand (1) and first -helix (1), containing the GXXXXGKT/S conserved sequence [44], where X represents any residue. This motif forms the phosphate-binding loop (P-loop), a giant anion hole which accommodates the -phosphate of the ADP by donating hydrogen bonds from several backbone amides [45]. The side chain of P-loop lysine have a catalytic role of stabilizing the pentavalent transition state of the -phosphoryl group as has been shown for adenylate kinase [46] and $p21^{ras}$ [47].

In addition to the Walker A-motif, it is observed a second conserved sequence ZZDXXG called Walker B-motif [44], where Z represents a hydrophobic residue. More recently, however, sequence and structural comparisons for all P-loop-fold proteins classified Shikimate Kinase in the DxD group of enzymes [48], which has a conserved DxD motif in strand 2. The Walker B motif consensus in shikimate kinases is ZZZTGGG and the second glycine has been implicated in hydrogen bonding to the phosphate of ATP. This motif is located on the C-terminal segment of the third strand (3) of the central -sheet The adenine-binding loop motif may be described as a sequence stretch of I/VDXXX(X)XP [33]. This motif forms a loop that wraps around the adenine moiety of ATP, connecting the 5-strand with the C-terminal -helix.

3. TUBERCULOSIS AND *MYCOBACTERIUM TUBERCULOSIS* SHIKIMATE KINASE

Tuberculosis (TB) remains the leading cause of mortality due to a bacterial pathogen, *Mycobacterium tuberculosis*. The interruption

of centuries of decline in case rates of TB occurred, in most cases, in the late 1980s and involved the USA and some European countries due to increased poverty in urban settings and the immigration from TB high-burden countries [49]. Thus, no sustainable control of TB epidemics can be reached in any country without properly addressing the global epidemic. It is estimated that 8.2 million new TB cases occurred worldwide in the year 2000, with approximately 1.8 million deaths in the same year, and more than 95 % of those were in developing countries [50]. Approximately 2 billion individuals are believed to harbor latent TB based on tuberculin skin test surveys [51], which represents a considerable reservoir of bacilli. According to a recent report compiled by the World Health Organization (WHO), the total number of new cases of tuberculosis worldwide in 2002 had risen to approximately 9 million [52]. A key river of the increase is the synergy with the HIV epidemic, which is having a devastating impact on some parts of the world mostly in the African Region, where 31% of new TB cases were attributable to HIV co-infection [50]. Another problem is the proliferation of multi-drug resistant (MDR) strains, defined as resistants to at least isoniazid and rifampicin, which are the most effective first-line drugs [53]. According to the 2004 Global TB Control Report of the World Health Organization, there are 300,000 new cases per year of MDR-TB worldwide, and 79 % of MDR-TB cases are now "super strains", resistant to at least three of the four main drugs used to treat TB [52]. The factors that most influence the emergence of drug-resistant strains include inappropriate treatment regimens, and patient noncompliance in completing the prescribed courses of therapy due to the lengthy standard "short-course" treatment or when the side effects become unbearable [54]. No new classes of drugs for TB have been developed in the past 30 years, reflecting the inherent difficulties in discovery and clinical testing of new agents and the lack of pharmaceutical industry in investing money and manpower for research in the area [55]. Hence, there is an urgent need to developing faster acting and effective new antitubercular agents, preferably belonging to new structural classes, to better combat TB, including MDR-TB, to shorten the duration of current treatment to improve patient compliance, and to provide effective treatment of latent tuberculosis infection [53].

In *M.tuberculosis,* the presence of the genes involved in the shikimate pathway began to be elucidated in the early 90s, with the cloning and characterization of *aro*A gene product, which codifies the EPSPS synthase [56]. Nevertheless, the complete genome sequence from *M.tuberculosis*, strain H37Rv reported by Cole *et al.* in 1998 [57] allowed the identification by sequence homology of all genes coding for shikimate pathway enzymes.

Four homologues to the shikimate pathway enzymes were located in a cluster containing the *aroD*-encoded type II DHQ dehydratase (Rv2537c), *aroB*-encoded DHQ synthase (Rv2538c), *aroK*-encoded type I shikimate kinase (Rv2539c), and *aroF*-encoded chorismate synthase (Rv2540c). The remaining homologues to shikimate pathway enzymes were annotated as follows: *aroG*-encoded class II phenylalanine-regulated DAHPS (Rv2178c), *aroE*-encoded shikimate dehydrogenase (Rv2552c), and *aroA*-encoded EPSP synthase (Rv3227).

The *aroK* structural gene is composed by 531 bp and codes a protein of 176 amino acids. The theoretical molecular mass of *Mt*SK enzyme subunit is 18.58 KDa. The first report of cloning and overexpression in soluble and functional form of *Mt*SK occurred in 2001 [58], where has been reported the PCR amplification *aroK* gene from genomic DNA of *M. tuberculosis* H37Rv strain, cloning in the plasmid pET-23a(+), and overexpression of *aroK*-encoded *Mt*SK protein in *Escherichia coli* BL21(DE3) host cells, without IPTG induction.

The crystal structure of a protein complexed with its substrates is of crucial importance for the rational design of inhibitors that target the enzyme. Although two crystal structures of SK from *M.tuberculosis* [33] have revealed the dynamic role of LID Domain in catalysis and the position of Mg^{2+} and ADP ligands in the structure, the precise positions and interactions between shikimate and *Mt*SK was not demonstrated because the shikimate-binding site was not occupied by the substrate or the electron density was not sufficient clear to position the shikimate molecule in the complex. In case of SK enzymes, a likely drawback of ATP-binding-site-based SK inhibitors would be their lack of specificity, owing to the common fold and similar ATP-binding site shared by many P-loop kinases [48]. Hence, trying to obtain a description of molecular interaction between *Mt*SK and shikimate, molecular-modeling and docking studies has been carried out and its results reported [59]. Despite these studies failed to predict all molecular interactions between *Mt*SK and shikimate [60], Asp34 and Arg136 residues and its hydrogen bonds implicated on shikimate binding are in agreement with the crystal structure of *Mt*SK-MgADP-Shikimate ternary complex reported afterwards [61]. The crystal structure of *Mt*SK-MgADP-Shikimate was the first crystallographic structure of SK with bound shikimate deposited in Protein Data Bank (PDB access code: 1WE2) [61]. Crystals were obtained by the hanging-drop vapour-diffusion method, and larger final concentration of ADP and shikimate in the drop (8.0mM) were used than those used by Gu *et al.* (2002) to obtain the crystals (4.0mM), in order to accurately determine the position of shikimate binding in the active site of *Mt*SK. The structure has been determined at 2.3 Å resolution, clearly reveling the amino-acids residues involved in shikimate binding. The molecular replacement method was used, using as a search model the structure of *Mt*SK-MgADP [33]. Almost at the same time of publishing of our article describing the structure of *Mt*SK-MgADP-Shikimate ternary complex, another article describing a similar structure of the ternary complex has been published (PDB access code: 1U8A) [60]. The crystal structures of *Mt*SK-MgADP-Shikimate [61] and *Mt*SK-ADP-Shikimate [60] ternary complexes have unequivocally revealed in detail the interactions of amino acid residues with bound shikimate and conformational changes upon substrate binding.

3.1. ADP/Mg^{2+} Interaction

Essentially all kinases require a Mg^{2+}-nucleotide complex as one of the enzyme substrates [62], an exception being the first partial reaction catalyzed by nucleoside diphosphate kinase, which can proceed independently of Mg^{2+} [63]. Nucleophilic attack on the -phosphate group of ATP will be most facilitated by meta-ion binding (Mg^{2+}) to the - and -phosphoryl groups, whereas departure of the leaving group will be most favoured in a structure with metal binding to the - and -phosphoryl groups [64]. Presumably Mg^{2+} also assists in orienting the -phosphoryl group of 'inline' with respect to the second substrate, creating the correct geometry to complete phosphoryl transfer [62]. The binding of this cation in many P-loop proteins such as myosin [65], elongation factor EF-Tu [66], p21-Ras [67] and the heterotrimeric G-proteins [68] involves hexa-coordination of the Mg^{2+} by two oxygen atoms (from the and -phosphates of the bound nucleotide), two water molecules, and two protein ligands. As in the *Mt*SK-MgADP structure (PDB code 1L4Y) [33], a typical six-coordination has been observed for Mg^{2+} in the *Mt*SK-MgADP-Shikimate structure, with some minor differences between the position of Mg^{2+} in the structures. In the *Mt*SK-MgADP-Shikimate structure Mg^{2+} interacts with a -phosphate oxygen of ADP, Ser16 OG of the Walker A motif and four water molecules [61]. Superimposition of shikimate-free *Mt*SK-MgADP structure on shikimate-complexed *Mt*SK-MgADP structure showed that in the ternary complex water1 and water4 are in equivalent positions but that shifts of 1.32 and 2.75 Å are observed for water2 and water3, respectively. In the *Mt*SK-MgADP structure, the interaction between Asp32 and Ser16 of the Walker A motif is *via* a bridging water molecule (water6), whereas in the ternary complex structure this interaction occurs directly *via* a hydrogen bond between the two residues [61], which accounts for the exclusion of water6 from the magnesium-binding site in the ternary

structure. In the ternary complex, the water1 molecule coordinated to the Mg^{2+} interacts directly with the chloride ion instead of interacting with Asp34 *via* a bridging water molecule (water5) as observed in the *Mt*SK-MgADP structure [33]. This chloride ion also interacts with the 3-hydroxyl group of shikimate and the backbone amide of Gly80 [61]. Thus, the different mode of interaction observed for residue Asp34 arises from the presence of shikimate, which leads to the exclusion of water5 from *Mt*SK active site [61]. The Mg^{2+} cation was not included in the final structure of *Mt*SK-ADP-Shikimate ternary complex reported by Dhaliwal et. al.(2004), since the best diffracting crystals grew in the absence of $MgCl_2$ [60]. Accordingly, the chloride ion observed in the active sites of both *Mt*SK binary complex [33] and *Mt*SK-MgADP-Shikimate ternary complex [61is absent in the Mg^{2+}-free structure of *Mt*SK-ADP-Shikimate ternary complex [60]. Thus, the absence of $MgCl_2$ in crystallization mixture probably have accounted for the small differences observed in conformation, position and molecular interactions of shikimate when the structure reported by Dhaliwal *et al*. is compared with *Mt*SK-MgADP-Shikimate-Cl structure [61]. In Mg^{2+}-free *Mt*SK structure the 3-hydroxyl group of shikimate is closer to -phosphoryl group of ADP than it is in the *Mt*SK-MgADP-Shikimate structure. Furthermore, two additional water-mediated interactions between protein residues and shikimate have been shown. The NZ atom of Lys15 residue forms a 2.6 Å hydrogen-bond with oxygen atom of another water molecule, which in turn interacts with the 3-hydroxyl group of shikimate, and the main-chain nitrogen of Leu119 residue forms a 2.9 Å hydrogen bond with the oxygen atom of a water molecule, which in turn interacts with 5-hydroxyl group of substrate [60]. In the *Mt*SK-MgADP-Shikimate structure, NZ atom of Lys15 cannot form a water-mediated interaction with 3-hydroxyl group of shikimate, since the chloride ion is bound to the active site cavity between them. The distance between NZ atom of Lys15 and the chloride ion is 3.94 Å, which in turn forms a 3.36-Å hydrogen bond with 3-hydroxyl group of shikimate [61].

In the *Mt*SK-MgADP structures [33], Lys15 forms a hydrogen bond with a -phosphate O1B atom and the chloride ion. The main-chain NH of Gly80 is hydrogen bonded to the chloride ion the binary (1L4Y) [33] and ternary complex structures [60,61]. These residues are located in vicinity of where the chemical reaction occur a may thus play a critical role in the transition state stabilization.

The molecular interactions that describe the ADP binding mode on enzyme are very similar in all available structures of *Mt*SK. The adenine moiety of ADP is sandwiched between Arg110 and Pro155 [33] and this interaction has also been observed in *Erc*SK [34], and in Adenilate kinase [69] and isoenzyme II [70]. Arg110 is located at he C-terminus of 6 where the LID domain starts. In *Mt*SK, Arg110 and Arg117 residues represent, respectively, the first and the last residue of a conserved motif of LID domain observed for P-loop kinases (typically RXX(X)R) [48]. In P-loop shikimate kinases the conserved motif of the LID domain has been proposed to be $R(X)_{6\text{-}9}R$ [61]. The Arg117 has been shown to interact with the and -phosphate groups of ADP [33,60,61], it generally interacts with the -phosphate of ATP bound to enzymes, and it may stabilize the transition state by neutralizing the developing negative charge on the - bridge O atom [71]. The Pro155 is the last residue of the adenine-binding loop motif (residues 148-155 in *Mt*SK), which was first recognized in Adenilate kinase and *Erc*SK [34] and has been described as an I/VD*XXX*(*X*)*X*P sequence stretch [33]. However, the second (aspartate) residue and the last (proline) residue are not conserved in the adenine-binding loops of *aroK*-encoded SKs from *Escherichia coli* [72] and *Campylobacter jejuni* (PDB acess code: 1VIA). However, they are recognizable from structural alignment analysis using *Mt*SK-MgADP as a reference structure [61]. In fact, it has been pointed out that the main-chain contacts with the adenine base and the presence of a structural motif independent of sequence may be the most important features for adenine binding in kinases [62]. The adenine-binding loop motif (residues 148-155 in *Mt*SK) forms a loop that wraps around the adenine moiety of ATP, connecting the 5-strand with the C-terminal 8-helix.

For catalysis, the three protein interacting residues Lys15, Arg117, and Arg136 have been proposed to be the most important [33]. Consistent with this proposal, the K15M mutant of *Erc*SK showed no detectable enzyme activity [35], although it was in fact a K15M/P115L double mutant. Contrary to the previously proposed, analysis of the *Mt*SK ternary complexes have revealed that Lys15 and Arg117 are the only positively charged residues located in the vicinity of where the reaction may occur and may therefore play critical roles in the stabilization of the transition state [33]. The Arg136 residue, instead, appears to interact with the carboxyl group of shikimate and probably, it is not involved in catalysis [60,61].

3.2. Interaction with Shikimate

The shikimate-binding domain, which follows strand 2, consists of helices 2 and 3 and the N-terminal region of helix 4 (Fig. (**1**)). In particular for *Mt*SK, the precise interactions between shikimate and SK have been elucidated, showing that the guanidinium groups of Arg58 and Arg136 and the NH backbone group of Gly81 interact with the carboxyl group of shikimate. The 3-hydroxyl group of shikimate forms hydrogen bonds with the carboxyl group of Asp34 and the main-chain NH group of Gly80 and a water molecule. The 2-hydroxyl group of shikimate hydrogen bonds to the side chain of Asp34 (Fig. (**2**)). The Glu61 is conserved in both *aro*K and *aro*L-encoded shikimate kinase enzymes. This residue is not directly involved in substrate binding, but it forms a hydrogen bond and a salt bridge with the conserved Arg58 and assists in positioning the guanidinium group of Arg58 for shikimate binding. In the ternary structure, Glu61 side chain also forms a water-mediated interaction with the 3-hydroxyl group of shikimate. Therefore, Glu61 plays an important role in the substrate-binding site [61]. The Glu54 carboxylate group also appears to anchor the guanidinium group of Arg58 for interaction with the shikimate carboxylate group; however, Glu54 is not conserved, except for *aro*K-encoded shikimate kinases [61]. In the *Mt*SK-ADP-shikimate structure [60], the main-chain nitrogen of Leu119 forms a 2.9 A H-bond with the water3 oxygen atom, which in turns interacts with the 5-hydroxyl group of shikimate, and the NZ atom of Lys15 forms a 2.6 A H-bond with the oxygen atom of water2, which also interacts with the 3-hydroxyl group of the substrate.

3.3. Conformational Changes Upon Substrate Binding

As pointed out above, NMP Kinases undergoes large conformational changes during catalysis, because their LID domain and the NMP binding-site are very flexible and can make large movements upon substrate binding. These structural changes act to position enzyme sidechains appropriately around the substrates and to sequester the substrates so as to prevent the hydrolysis of bound ATP or other phosphoryl-containing substrates prior to catalysis [41,62,64]. In addition, it has also been shown that these two domains are capable to make independently moves towards each other [41]. Previous studies have shown that hexokinases [73] and adenylate kinases [42,74], which are classified as NMP kinases, undergo a large conformational change during catalysis.

Several structural and spectroscopic studies have demonstrated that SKs undergo conformational changes on ligand binding. Circular dichroism spectra of unliganded and liganded *Erc*SK enzyme in the presence of 2mM shikimate or 2mM N-ATP (an non-hydrolisable ATP analogue) have shown that SK undergoes conformational changes upon ligand binding [34]. Moreover, fluorescence studies were performed using a single tryptophan residue (W54) as a report group, which is positioned close to the shikimate binding site. The addition of shikimate to protein solution caused a quenching in *Erc*SK protein fluorescence and a blue shift of 3 nm

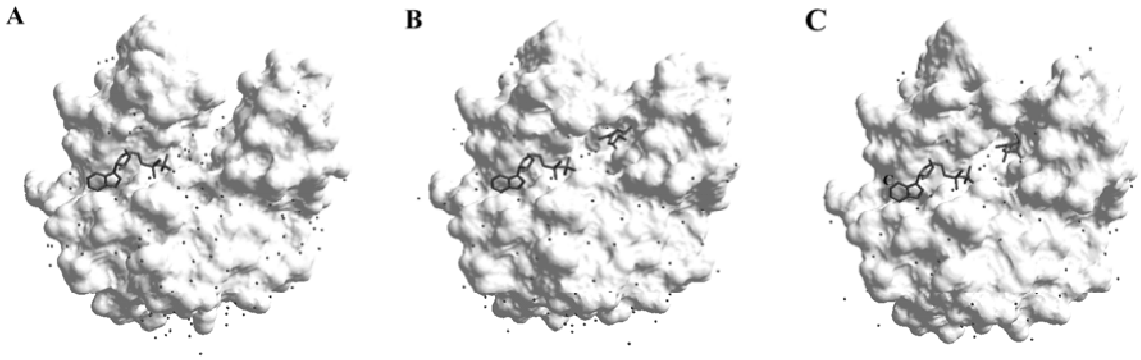


Fig. (3). Molecular surface of (A) *Mt*SK-MgADP (1L4U) [33], (B) *Mt*SK-MgADP-Shikimate (1WE2) [61], and (C) *Mt*SK-ADP-Shikimate (1U8A) [60] structures The molecular surface areas have been calculated using the program Swiss-PDBViewer v3.7 (www.expasy.org/spdbv), probe radius of 1.4 Å, and a fixed radius for all atoms. Calculated values are approximately 7246 $Å^2$ for *Mt*SK complexed with only MgADP (A), 6915 $Å^2$ for structure complexed with MgADP and Shikimate (B), and 7040 $Å^2$ for the Mg^{2+}-free structure complexed with ADP and Shikimate. Thus, approximately 330 $Å^2$ and 206 $Å^2$ of solvent-acessible surface are buried on shikimate binding to, respectively, *Mt*SK-MgADP-Shikimate and *Mt*SK-ADP-Shikimate ternary complexes. The shikimate binding leads to a counter movement of LID and SB domains and, consequently, to a partial closure of shikimate-binding pocket. Probably, the Mg^{2+} cation has a role in the closure of *Mt*SK active site. All atoms of shikimate and ADP, Mg^{+2} cation, and oxygen atoms of waters are colored grey.

in the emission maximum, consistent with the loop containing the tryptophan residue becoming more deeply buried within the protein following ligand-binding [75]. In addition, the averaged *B*-factor for all residues in crystal structure of *Erc*SK showed clear evidence of the flexibility of the molecule, where the temperature factors for both *Erc*SK molecules in the asymmetric unit (one with bound shikimate and other with unbound shikimate) indicate two regions of high mobility, corresponding to the shikimate binding-site and the LID domain and its flanking regions [34]. A comparison of the residue-averaged *B* factors between the crystal structures of *Erc*SK-MgADP, *Mt*SK-MgADP (1L4Y), and *Mt*SK-MgADP Pt-derivative (1L4U) binary complexes [33] shows that the enzyme from *M. tuberculosis* follows the same pattern of flexibility in the LID and SB domains as previously observed for *Erc*SK.

Another method used to evaluate conformational changes in proteins is the superposition of structures in order to compare different complexes of the same protein. To demonstrate conformational changes upon shikimate binding in *Mt*SK, alignment of C positions of the *Mt*SK-MgADP-Shikimate dead-end ternary complex and the *Mt*SK-MgADP binary complex structures were made [61], showing that the LID and SB domains undergoes noticeable concerted movements towards each other. In this alignment, which included all residues, were verified r.m.s. deviation values of 0.56 and 0.54 Å for 1L4U and 1L4Y, respectively. Residues 112-124, that comprises the LID domain, showed a r.m.s. deviation of 1.33 Å, and the SB doamin shift was somewhat smaller, with an r.m.s. deviation value of 0.74 Å for residues 33-61. Similar values were found when the same procedure was applied to the binary complex and the *Mt*SK-ADP-Shikimate structure [60], with a value of 0.7 Å for the overall structure (excluding residues 114-115), 1.5 Å for the LID domain and 0.4 Å for the SB domain. The residues directly involved in these movements are Val116, Pro118 and Leu119 (LID domain), Ile45, Ala46, Glu54, Phe57 and Arg58 (SB domain), where its side chains shifted upon shikimate binding [61]. A comparison of *Mt*SK-ADP-Shikimate ternary complex to *Mt*SK-MgADP binary complex showed, within of SB region, a shift of 0.9 Å of Arg58 residue towards the carboxylate group of shikimate. Phe49 residue moves approximately ~1.7 Å away from Phe57 and closer to the substrate, translating towards shikimate [60]. In addition, it has been proposed by Dhaliwal *et al.* (2004) that the observed changes in the orientation and position of Phe49 and Phe57 residues disrupt their strong ring stacking interactions and probably result from the Van Der Waals contacts made between shikimate with both phenylalanines [60]. However, in *Mt*SK-MgADP-Shikimate the strong ring stacking interaction between the phenylalanines is not disrupted in spite of the shift of Phe49 and Phe57 residues towards shikimate of, 0.49 Å and 0.79 Å, respectively. Moreover, there is a reduction in molecular surface area value the ternary complex compared to binary complex. The calculated value for *Mt*SK complexed with MgADP (1L4U) was approximately 7246 A^2 and for *Mt*SK complexed with MgADP and shikimate a value of 6915 A^2 was found. Based on these results, the difference between the complexes molecular surface areas is 330 A^2, which are buried on shikimate binding (Fig. (**3**)) [61]. A reduction in molecular surface area value has also been observed in the Mg^{2+}-free crystal structure of *Mt*SK-ADP-Shikimate ternary complex (Fig. (**3**)). Nevertheless, in this complex a molecular surface of only 206 A^2 is buried on shikimate binding, indicating that Mg^{2+} binding has a role in the closure of *Mt*SK active site. It is important to point out that both the *Mt*SK-MgADP bynary complexes [33] and *Mt*SK-MgADP-Shikimate [61] and *Mt*SK-ADP-Shikimate [60] ternary complexes have been crystallized in the same space group with similar unit-cell parameters and residues from LID and SB domains form no crystal contacts with symmetry-related *Mt*SK molecules. Thus, the conformational changes that have been described for *Mt*SK cannot merely be a reflection of the different crystal-packing arrangements.

The incomplete LID-domain closure observed in crystal structures of both *Mt*SK-MgADP-Shikimate [61] and *Mt*SK-ADP-Shikimate [60] ternary complexes suggests that -phosphate of ATP is necessary for the completion of the domain movement [34]. This find is not unexpected, since total active-site closure upon dead-end ternary complex formation would result in locking the enzyme active site in an inactive form in which shikimate substrate binding to *Mt*SK enzyme prior to MgADP dissociation from its active site would result in an inactive abortive complex. Several previously reported results are consistent with these structural data. Measurements of *Erc*SK intrinsic tryptophan fluorescence (Trp54) on shikimate binding to either *Erc*SK or the *Erc*SK-MgADP binary complex showed a modest synergism of binding between these substrates, since the dissociation constant value for shikimate (K_d = 0.72 mM) decreased to 0.3 mM in the presence of 1.5 mM ADP [75]. In addition, measurements of the quenching of protein fluorescence of the *aroL*-encoded *Erc*SK upon nucleotide binding demonstrated the dissociation-constant values for ADP and ATP to be 1.7 and 2.6 mM, respectively [35]. The *K*m for ATP (620 M) was found to be approximately four times lower than the dissociation constant in the absence of shikimate [35]. These results prompted the proposal that the conformational changes in the *Erc*SK enzyme associated with the binding of the first substrate led to an increas

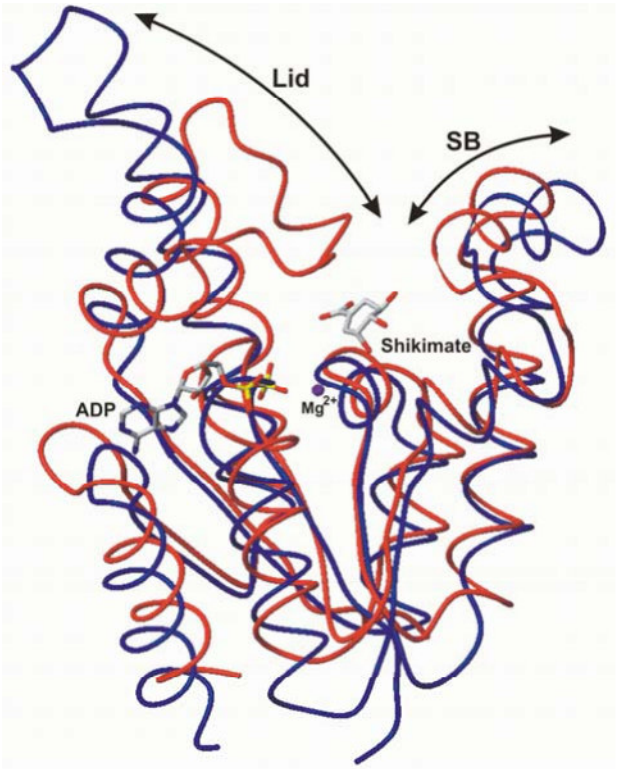


Fig. (4). Superposition of SK-MgADP-Shikimate ternary complex from *Mycobacterium tuberculosis* [61] (red line) and apo-SK (K15M/P115L) from *Erwinia chrysanthemi* [34] (blue line).

in the affinity for the second substrate. However, this synergism on substrate binding does not appear to hold for *Mt*SK since the apparent dissociation-constant values for ATP (89 M) and shikimate (440 M) are similar to their K_m values, 83 M for ATP and 410 M for shikimate considering either a rapid equilibrium random-order bi-bi enzyme mechanism or a steady-state-ordered bi-bi enzyme mechanism [33]. Moreover, no evidence for synergism between shikimate and ATP could be observed in substrate binding to *Erc*SK in a chloride buffer system [76].

The K15M *Erc*SK mutant has been crystallized in an open conformation that is proposed to presumably be equivalent to an apo-enzyme structure in which neither ADP (or ATP) nor shikimate would be bound [35]. This mutant was produced to evaluate the role of the conserved Lys15 of the Walker A motif. However, an unwanted mutation point mutation in the LID domain (Pro115Leu) was detected during the refinement of the model. Furthermore, the enzyme was crystallized with two independent molecules in the asymmetric unit and extensive contacts of neighboring LID domains lead to a stabilization of this part of the molecule that is not visible in the native crystal structure [34,35]. It therefore appears unwarranted to consider the double K15M/P115L *Erc*SK mutant a model for the apo enzyme. Since a better model for the apo-enzyme is not currently available, here we considered the double K15M/P115L *Erc*SK mutant as a model for the apo-SK. The superposition of *Mt*SK-MgADP-Shikimate and apo *Erc*SK (K15M+P115L) show the large conformational changes in the LID and SB domains associated with the binding of ADP and shikimate on *Mt*SK (Fig. (**4**)).

4. MOLECULAR DYNAMICS SIMULATIONS OF THE SHIKIMATE KINASE STRUCTURE

Molecular dynamics (MD) simulations are playing an important role in a wide area of research, which includes structural biology, condensed matter physics, material science, and bioinformatics. The use of molecular dynamics simulations allow monitoring the dynamics of individual atoms, thereby giving a unique insight into the molecule behavior that cannot be easily extracted from laboratory experiments. The application of molecular dynamics simulation allows prediction of structural flexibility features of complexes of enzyme and ligands. We performed molecular dynamics simulations of the complex MtSK-ADP-shikimate (system 1) and apoenzyme MtSK (system 2) in order to elucidate the influence of the shikimate on the overall structure of the MtSK. The root-mean square deviation (RMSD) of the positions for all backbone C atoms from their initial configuration as a function of simulation time for both systems are shown in Fig. (**5**). Analysis of the molecular dynamics simulations indicates the LID domain undergoes confor-

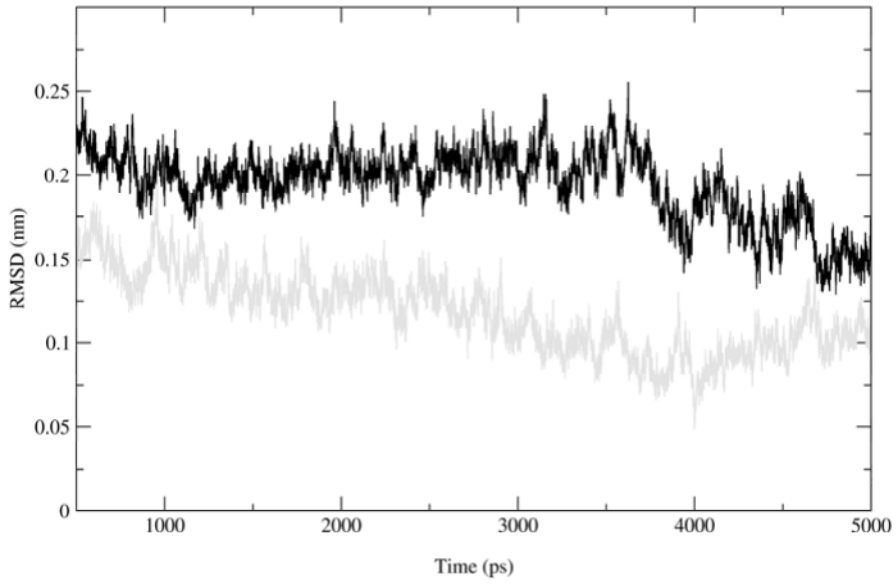


Fig. (5). Time-dependent C RMSD of both complexes with respect to the initial structures for MtSK-ADP-shikimate (system 1)(grey line) (a), MtSK, apoenzyme (system 2) (black line) (b). Molecular dynamics simulations were performed using the program GROMACS [97]. The starting geometries for the simulations of the systems 1 were generated from the X-ray structure obtained from the Protein Data Bank (PDB access code: 1WE2). For the simulations of the system 2 the atomic coordinates for the shikimate, ADP and ions were deleted from the PDB file of system 1.

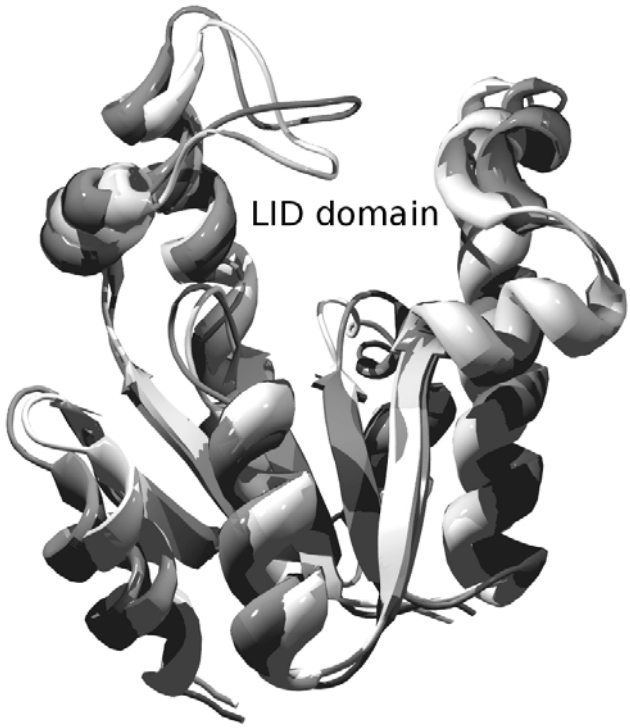


Fig. (6). Superposition of the atomic coordinates of the initial (light grey) and final (grey) structures for system 2. The atomic coordinates for the final structure were those at the end of 5ns molecular dynamics simulation.

mation change in the system 2 (without shikimate). The structure without the molecule of MtSK presents high flexibility for the residues involved in the LID domain in the system 2. Analysis of the residues participating in the LID domain in the system 1 indicates the presence of a hydrogen bond network involving residues 112-124 and the shikimate, which promotes the stabilization of the LID domain in the closed conformation. Molecular dynamics simulations of the system 2 for 5ns show the movement of the LID domain in the MtSK structure. Superposition of the atomic coordinates of the initial and final structures for system 2 after 5ns of molecular dynamics simulation is shown in Fig. (**6**). The LID domain is moved towards the solvent in the structure at the final of the 5ns molecular dynamics simulation, demonstrating that the molecular dynamics simulation results are in agreement with experimental X-ray crystallographic data (Fig. (**4**)), confirming that the binding of the shikimate to the SK structure is enough to promote the closure of the LID domain. On the other hand, the lacking of the shikimate in the MtSK promotes the opening of the LID domain. This structural feature should be considered in any structure-based design of SK inhibitors, which are competitive against shikimate. A detailed analysis of the molecular dynamics simulations for MtSK with different ligand is underway and will be published elsewhere.

5. CONCLUDING REMARKS

The disruption of gene *arok*, which codes for the shikimate kinase, has been shown to be essential for the viability of *M. tuberculosis.* The evidence that shikimate pathway is essential for *M. tuberculosis* even in the presence of exogenous supplements as p-aminobenzoate, p-hydroxibenzoate and aromatic amino acids reinforces its attractiveness as a drug target [26]. It is interesting to note that shikimic acid can be used to supply the aromatic amino acid and aromatic vitamin requirements of *E. coli* blocked in any of the first three steps of the shikimate pathway [77]. This ability has been related to the presence of the shikimic acid transport system encoded by the gene *shiA* [78]. Thus, there is the possibility that shikimic acid could be used as a supplement to allow the growth *in vitro* of mutant *M. tuberculosis* strains disrupted on *aroG*, *aroB, aroD* or *aroE* genes, which code for first four shikimate pathway enzymes in *M. tuberculosis*. These *aro*-disrupted *M. tuberculosis* strains could be used in experiments to determine if shikimate pathway is essential to *in vivo growth* of *M. tuberculosis.*

A comprehensive structural picture of the interactions between *M. tuberculosis* SK enzyme and Mg^{2+}, ADP and shikimate substrates have been obtained from crystallographic studies [33,60,61]. In addition, the LID and SB domains conformational changes upon ADP and shikimate substrates binding, which result in a partial closure of and solvent expulsion from *Mt*SK active site, has been described.

There has been considerable debate within the literature as whether enzyme-catalyzed phosphoryl transfer reactions operate primarily with an associative (S_N2-like) or dissociative (S_N1-like) transition state [62]. A criterion that has been used for distinguishing between these mechanisms is based on the geometry and reaction coordinate distances between the terminal phosphoryl group of ATP and the acceptor substrate in ground state enzyme-substrate or enzyme inhibitor complexes [79]. For the transition state to have some degree of associative character, these distances are expected to be less than the sum of a P-O van der Waals contacts and a P-O single bond (i.e in the order of approximately 4.9 Å) [62]. In dissociative mechanisms these distances would be expected to be longer than this to allow space for the intermediate monomeric metaphosphate to exist between the leaving group (ADP) and the entering group [79]. A crystal structure of *Mt*SK in complex with MgADP, Shikimate, and AlF_3 (a structural analog of -phosphoryl group transfered in the reaction) could permit a snapshot of the transition-state like structure of *Mt*SK. As it has been shown to other kinases [80-93], this structural study could distinguish between associative and dissociative transition states in the reaction catalyzed by *Mt*SK.

It has been recently described the structural alterations caused by binding of the shikimate, magnesium and chloride ions in the SK from *Mycobacterium tuberculosis* (MtSK). These new findings indicate that both ions and the shikimate influence the structural conformation of MtSK. The magnesium ion also appears to influence both the hydroxyl groups of shikimate molecule and some residues from active site. The chloride ion can influence the affinity of shikimate for MtSK, in the binding position of ADP in MtSK active site and also in the opening length of LID domain of the MtSK structure. The shikimate binding causes a closing of LID domain, which appears to influence drastically the crystallographic packing of the protein [94-95].

Structures of SKs deposited in the protein data bank (PDB) were superimposed and positions of residues involved in binding (Lys15, Asp34, Arg58, Glu61, Gly79, Gly80, Gly81, Arg136) between shikimate and SK are highly conserved [61]. The precise interactions between *Mt*SK and shikimate have been elucidated, showing the residues involved in the binding of substrate and the conformational changes upon substrate binding. Accordingly, the availability of the *M. tuberculosis* shikimate kinase structures complexed with shikimate provide crucial information for the design of non-promiscuous SK inhibitors that target both the shikimate- and ATP-binding pockets or uniquely, the shikimate-binding site. Moreover, the knowledge of functional factors that lead to active site closure could be used for designing inhibitors that force *Mt*SK to a closed conformation that would be unable to catalyze the phosphoryl transfer to shikimate. These inhibitors could block the biosynthesis of aromatic aminoacids [96] and other compounds (folate, micobactins, etc), which are essential for the growth and viability of the microorganism.

ACKNOWLEDGMENTS

Financial support for this work was provided by FAPESP (Proc. 02/05347-4, 01/07532-0, 04/00217-0) and Millennium Initiativ

Program MCT-CNPq, Ministry of Health - Secretary of Health Policy (Brazil) to DSS and LAB. DSS (304051/1975-06) WFA (CNPq, 300851/98-7), and LAB (520182/99-5) are research career awarded from the National Research Council of Brazil (CNPq).

ABBREVIATIONS

TB = Tuberculosis

Mt = *Mycobacterium tuberculosis*

SK = Shikimate kinase

DAHPS = 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase

CS = Chorismate synthase

EPSPS = 5-enolpyruvylshikimate-3-phosphate synthase

PABA = p-aminobenzoate

GTP = Guanosine triphosphate

ATP = Adenosine triphosphate

ADP = Adenosine diphosphate

Km = Michaellis-Menten constant

NMP = Nucleoside monophosphate

AK = Adenylate kinase

WHO = World Health Organization

HIV = Human immunodeficiency vírus

MDR = Multidrug-resistant

DHQ = Dehydroquinase

PCR = Polymerase chain reaction

IPTG = Isopropyl-beta-D-thiogalactoside

Mg = Magnesium

r.m.s. = Root mean square

Erc = *Erwinia chrysantemi*

AMP-PCP = Adenosine - , -methyleneadenosine 5'-triphosphate

AMP-PCP = , -imidoadenosine 5'-triphosphate

REFERENCES

[1] Davis, B.D. and Mingioli, E. S. (1953) *J. Bacteriol.* **66**(2), 129–136.

[2] Sprinson, D.B. (1960) *Adv. Carbohydrate Chem.* **15**, 235–270.

[3] Pittard, A.J. (1987) *In* F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol.**1**, pp. 368–394. American Society for Microbiology, Washington, D.C.

[4] Haslam, E. (1993) Shikimic acid: metabolism and metabolites. John Wiley & Sons, Chichester, United Kingdom.

[5] Folkers, K.(1996*). Biochem. Biophys. Res. Commun.,* **224**(2), 358-361.

[6] Coggins, J.R., Duncan, K., Anton, I.A., Boocock, M.R., Chaudhuri, S., Lambert, J.M., Lewendon, A., Millar, G., Mousdale, D.M., Smith, D.D., (1987) *Biochem. Soc. Trans.,* **15**(4), 754–759.

[7] Butler, J.R., Alworth, W.L., Nugent, M.J., (1974) *J. Am. Chem. Soc.,* 96, 1617–1618.

[8] Mousdale, D.M., Campbell, M.S., Coggins, J.R. (1987) *Phytochemistry,* **26**(10), 2665–2670.

[9] Duncan, K., Edwards, R.M., Coggins, J.R. (1987) *Biochem. J.,* **246**(2), 375–386.

[10] Davies, G.M., Barret-Bee, K.J., Jude, D.A., Lehan, M., Nichols, W.W., Pinder, P.E. *et al*. (1994). *Antimicrob. Agents Chemother*., **38**(2), 403-406.

[11] Coggins, J.R. (1989). The shikimate pathway as a target for herbicides. In Herbicides and Plant Metabolism (Dodge, A., ed.), Cambridge University Press, Cambridge, UK, pp. 97-112.

[12] Bentley, R. (1990). *Crit. Rev. Biochem. Mol. Biol.,* **25**(5), 307-384.

[13] McDevitt, D., Payne, D.J., Holmes, D.J., and Rosenberg, M. (2002) *J. Appl. Microbiol.,* **92**, 28S-34S.

[14] McArthur, J.D., West, N.P., Cole, J.N., Jungnitz, H., Guzman,C.A., Chin, J., Lehrbach, P.R., Djordjevic, S.P., and Walker, M.J. (2003) *FEMS Microbiol. Lett.,* **221**(1), 7-16.

[15] Roberts, F., Roberts, C.W., Johnson, J.J., Kyle, D.E., Krell, T., Coggins, J.R., Coombs, G.H., Milhous, W.K., Tzipori, S., Fergunson, D.J.P., Chakrebarti, D., and McLeod, R. (1998). *Nature*, **393**, 801-805.

[16] Kishore, G.M. & Shah, D.M. (1988) *Annu. Rev. Biochem.,* **57,** 627–663.

[17] Dieckmann, A., and Jung, A. (1986) *Mol. Biochem.Parasitol.*, **19**(2), 143-147.

[18] Campbell, S.A., Richards, T.A., Mui, E.J., Samuel, B.U., Coggins, J.R., McLeod, R., Roberts, C.W. (2004) *Int. J. Parasitol.*, **34**(1), 5-13.

[19] McRobert, L., McConkey, G.A. (2002) *Mol. Biochem. Parasitol.*, **119**(2), 273-278.

[20] McConkey G.A. (1999) *Antimicrob. Agents Chemother*., **43**(1), 175-177.

[21] Gardner, M. J., Shallom, S.J., Carlton, J.M., *et al*. (2002) *Nature*, **419**, 498-511.

[22] McConkey, G.A., Pinney, J.W., Westhead, D.R. (2004) *Trends Parasitol.*, **20**(2), 60-65.

[23] Fitzpatrick T, Ricken S, Lanzer M, Amrhein N, Macheroux P, Kappes B. (2001) *Mol. Microbiol.*, **40**(1), 65-75.

[24] Ratledge, C. (1982). The Biology of the Mycobacteria. *Academic Press, London*. **1**, 185-271.

[25] Voos, J.J., Rutter, K., Schroder, B.G., Su, H., Zhu, Y., and Barry, C.B. III. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 1252-1257.

[26] Parish, T., and Stoker, N. G. (2002). *Microbiology*, **148**, 3069-3077.

[27] Whipp, M.J. & Pittard, A.J. (1995) *J. Bacterial*. **177**(6), 1627-1629.

[28] Millar, G., Lewendon, A., Hunter, M.G. & Coggins, J.R. (1986). *Biochem. J*. **237**(2), 427-437.

[29] De Feyter, R.C. & Pittard, J. (1986) *J. Bacteriol*. **165**(1), 331-333.

[30] Ely, B. and Pittard, J. (1979) *J. Bacteriol.* **138**(3), 933–943.

[31] De Feyter, R.C., Davidson, B.E., and Pittard, J. (1986) *J. Bacteriol.* **165**(1)**,** 233–239.

[32] Vinella, D., Gagny, B., Joseleau-Petit, D., D'Ardi, R. & Cashel, M. (1996) *J. Bacteriol*. **178**(13), 3818-3828.

[33] Gu, Y., Reshetnikova, L., Li, Y., Wu, Y., Yan, H., Singh, S., Ji, X. (2002) *J. Mol. Biol*. **319**(3), 779-789.

[34] Krell, T., Coggins, J.R. & Lapthorn, A.J. (1998) *J. Mol. Biol.* **278**(5), 983–997.

[35] Krell, T., Maclean, J., Boam, D.J., Cooper, A., Resmini, M., Brocklehurst, K. *et al*. (2001) *Protein Sci*. **10**(6), 1137–1149.

[36] Dreusicke, D., Karplus, A. & Schulz, G.E. (1988) *J. Mol. Biol.* **199**(2), 359-371.

[37] Schlauderer, G.J. & Schulz, G.E. (1996) *Protein Sci*. **5**(3), 434-441.

[38] Stehle, T. & Schulz, G.E. (1990) *J. Mol. Biol*. **211**(1), 249-254.

[39] Müller-Dieckmann, H.-J. & Schulz, G.E. (1994) *J. Mol. Biol.* **236**(1), 361-367.

[40] Wild, K., Bohner, T., Aubry, A., Folkers, G. & Schulz, G.E. (1995) *FEBS Letters*, **368**(2), 289-292.

[41] Vonrhein, C., Schlauderer, G. J. & Schulz, G. E. (1995) *Structure*, **3**(5), 483–490.

[42] Müller, C.W., Schlauderer, G.J., Reinstein, J. & Schulz, G.E. (1996) *Structure*, **4**(2), 147-156.

[43] Gerstein, M., Schulz, G.E. & Chothia, C. (1993) *J. Mol. Biol.* **229**(2), 494-501.

[44] Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. (1982). *EMBO J.* **1**(8), 945–951.

[45] Smith, C.A. & Rayment, I. (1996) *Biophys. J.* **70**(4), 1590-1602.

[46] Reinstein, J., Schlichting, I., and Wittinghofer, A. (1990) *Biochemistry* **29**(32)**,** 7451–7459.

[47] Sigal, I.S., Gibbs, J.B., D'Alonzo, J.S., Temeles, G.L., Wolanski, B.S., Socher, S.H., and Scolnick, E.M. (1986) *Proc. Natl. Acad. Sci. USA* **83**(4)**,** 952–956.

[48] Leipe, D.D., Koonin, E.V., and Aravind, L. (2003) *J. Mol. Biol.* **333**(4), 781-815.

[49] Raviglione, M.C. (2003) *Tuberculosis*, **83**(1-3), 4-14.

[50] Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Willians, B.G., Raviglione, M.C., Dye, C. (2003). *Arch. Intern. Med*., **163**(9), 1009-1021.

[51] Dye C, Scheele S, Dolin P, Pathania V, Raviglione, M.C. (1999) *JAMA*, **282**(7), 677-686.

[52] World Health Organization: Global Tuberculosis Control. WHO Report 2004. ISBN 92 4 156264 1.
[53] Basso, L.A., and Santos, D.S. (2005) *Med. Chem. Reviews*.- Online. **2**, 393-413.
[54] Duncan, K. (2003) *Tuberculosis*, **83**(1-3), 201-207.
[55] O'Brien, R.J., Nunn, P.P. (2001) *Am. J. Resp. Crit. Care Med*., **163**(5)**,** 1055-1058.
[56] Garbe, T., Joens, C., Charles, I., Dougan, G., Young, D. (1990) *Journal of Bacteriology* **172**(12), 6774-6782.
[57] Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, K.; Osborne, J.; Quail, M.A.; Rajandream, M.-A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, R.; Squares, S.; Sulston, J.E.; Taylor, K.; Whitehead, S. and Barrell, B.G. (1998) *Nature*, **393**, 537-544.
[58] Oliveira, J.S., Pinto, C.A., Basso, L.A., Santos, D.S. (2001*) Protein Expr. and Purif.*, **22**(3)**,** 430-435.
[59] De Azevedo, W. F. Jr., de Oliveira, J. S., Basso, L. A., Palma, M.S., Pereira, J. H., Canduri, F., and Santos, D. S. (2002*). Biochem. Biophys. Res. Commun*., **295**(1), 142-148.
[60] Dhaliwal, B., Nichols, C.E., Ren, J., Lockyer, M., Charles, I., Hawkins, A.R., Stammers, D.K. *FEBS Letter* (2004), **574**(1-3), 49-54.
[61] Pereira, J.H., Oliveira, J.S, Canduri, F., Dias, M.V.B., Palma, M.S., Basso, L.A., Santos, D.S., and Azevedo Jr., W.F. (2004) *Acta Cryst. Section D* **60**, 2310-2319.
[62] Matte, A., Tari, L.W., and Delbaere, T.J. (1998) *Structure*, **6**(4), 413-419.
[63] Williams, R.L., Oren, D.A., Muñoz-Dorado, J., Inouye, S., Inouye, M., and Arnold, M. (1993) *J. Mol. Biol.*, **234**(4), 1230-1247.
[64] Jencks, W.P. (1975) *Adv. Enzymol*. **43**, 219-410.
[65] Smith, C.A., Reyment, I. (1995) *Biochemistry*, **34**(28)**,** 8973-8981.
[66] Berchthold, H., Reshetnikovca, L., Reiser, C.O.A., Schrimer, N.K., Sprinzl, M., Hilgenfeld, R. *Nature* (1993) **365**(6442)**,** 126-132.
[67] Pai, E.F., Krengel, U., Petski, G.A., Goody, R.S., Kabsch, W., Wittinghofer, A. . (1990) *EMBO J.,* **9**(8)**,** 2351-2359.
[68] Coleman, D.E., Berghuis A.M., Lee, E., Linder, M.E., Gilman, A.G., Sprang, S.R. (1994) *Science* **265**(5177)**,** 1405-1412.
[69] Abele, U. and Schulz, G. E. (1995) *Protein Sci.*, **4**(7), 1262-1271.
[70] Schlauderer, G.J. & Schulz, G.E. (1996) *Protein Sci*. **5**(3), 434-441.
[71] Hasemann, C. A., Istvan, E. S., Uyeda, K. & Deisenhofer, J. (1996) Structure, **4**(9), 1017-1029.
[72] Romanowski, M.J., Burley, S.K. (2002) *Proteins: Structure, Function, and Genetics*. **47**(4), 558-562.
[73] Bennett, W.S. and Steitz, T.A. (1980) *J. Mol.Biol.* **140**(2), 211-230.
[74] Schulz, G.E., Müller, C.W. and Diederichs, K. (1990) *J. Mol. Biol.* **213**(4), 627-630.
[75] Idziak, C., Price, N.C., Kelly, S.M., Krell, T., Boam, D.J., Lapthorn, A.J., and Coggins, J.R. (1997) *Biochem. Soc. Trans.***, 25**(4), S627.
[76] Cerasoli, E., Kelly, S.M., Coggins, J.R., Lapthorn, A.J., Clarke D.T. (2003) *Biochim. Biophys.Acta*, **1648**(1-2), 43-54.
[77] Pittard, A.J., and Wallace, B.J. (1966) *J. Bacteriol*., **92**(4), 1070-1075.
[78] Whipp, M.J., Camakaris, H., and Pittard, A.J. (1998) *Gene*, **209**(1-2), 185-192.
[79] Mildvan, A.S. (1979) *Adv. Enzymol. Relat. Areas Mol. Biol*., **49**, 103-126.
[80] Schlichting, I., and Reinstein, J. (1997) *Biochemistry*, **36**(31), 9290-9296.
[81] Xu, Y.-W., Moréra, S., Janin, J., and Cherfils, J. (1997) *Proc. Natl. Acad. Sci. USA*, **94**(8), 3579-3583.
[82] Scheffzek, K., and Wittinghofer, A. (1997) *Science*, **277**, 333-338.
[83] Basso, L. A., Pereira da Silva, L. H., Fett-Neto, A. G., de Azevedo Jr., W. F., Moreira, I. S., Palma, M. S., Calixto, J. B., Astolfi Filho, S., dos Santos, R. R., Soares, M. B. P., Santos, D. S. (2005) *Mem. Inst. Oswaldo Cruz,* **100**(6): 475-506.
[84] da Silveira, N. J. F., Uchoa, H. B., Canduri, F.,Pereira, J. H., Camera Jr., J. C., Basso, L. A., Palma, M. S., Santos, D. S., de Azevedo Jr., W. F. (2004) *Biochem Biophys. Res. Commun.*, **322**(1), 100–104.
[85] Uchoa, H. B., Jorge, G. E., da Silveira, N. J., Camera J. C., Canduri, F., De Azevedo, WF. (2004) *Biochem. Biophys. Res. Commun.* **325**(4),1481–1486.
[86] da Silveira, N. J. F., Bonalumi, C. A., Uchoa, H. B., Pereira, J. H., Canduri, F.,Pereira, J. H., de Azevedo Jr., W. F. (2006) *Cell Biochem. Biophys.* **44**(3), 366-374.
[87] De Azevedo, W.F.Jr.; Mueller-Dieckmann, H.J.; Schulze- Gahmen, U.; Worland, P.J.; Sausville, E.; Kim S.-H. (1996) *Proc. Natl. Acad. Sci. USA.*, **93**(7), 2735-2740.
[88] Kim, S.-H.; Schulze-Gahmen, U.; Brandsen, J.; De Azevedo, W. F. Jr. (1996) *Prog. in Cell Cycle Res*., **2**, 137-145.
[89] Canduri, F.; Uchoa, H.B.; de Azevedo, W.F.Jr. (2004) *Biochem. Biophys. Res. Commun.*, **324**(2), 661-666.
[90] De Azevedo, W.F.Jr.; Canduri, F.; Silveira, N.J.F. (2002) *Biochem Biophys. Res. Commun.*, **293**(1), 566-571.
[91] De Azevedo, W.F.Jr.; Leclerc, S.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S.-H. (1997) *Eur. J. Biochem.*, **243**, 518-526.
[92] De Azevedo, W.F.Jr., Gaspar, R.T., Canduri, F.; Camera, J.C.Jr., Silveira, N.J.F. (2002) *Biochem Biophys. Res. Commun.*, **297**(5)**,** 1154-1158.
[93] De Azevedo, W. F. Jr., Canduri, F. (2005) *Current Computer-Aided Drug Design,* **1**, 53-64.
[94] Gan, J., Gu, Y., Li, Y., Yan, H., Ji, X. (2006) *Biochemistry **45**,* 8539-8545.
[95] Dias, M. V. B., Faim, L. M., Vasconcelos, I. B., Oliveira, J. S., Basso, L. A., Santos, D. S., De Azevedo, W. F. *Acta Crystallographica F*. (In press).
[96] Ducati, R. D., Basso, L. A., Santos, D. S. (2007) *Current Drug Targets*, **8**(1) In press.
[97] van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. Berendsen, H. J. C. (2005) *J. Comp. Chem*. **26**(16),1701-1718.

Received: December 12, 2006 **Accepted: May 26, 2006** **Updated: September 11, 2006**

CDK9 a Potential Target for Drug Development

Fernanda Canduri[1], Patrícia Cardoso Peres[2], Rafael Andrade Caceres[2] and Walter Filgueira de Azevedo Jr.[2*]

[1]*Departamento de Química e Física Molecular, Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trabalhador Sãocarlense, 400. São Carlos-SP, Brazil;* [2]*Faculdade de Biociências-PUCRS-Av. Ipiranga,6681. Porto Alegre-RS Brazil*

Abstract: The family of Cyclin-Dependent Kinases (CDKs) can be subdivided into two major functional groups based on their roles in cell cycle and/or transcriptional control. CDK9 is the catalytic subunit of positive transcription elongation factor b (P-TEFb). CDK9 is the kinase of the TAK complex (Tat-associated kinase complex), and binds to Tat protein of HIV, suggesting a possible role for CDK9 in AIDS progression. CDK9 complexed with its regulatory partner cyclin T1, serves as a cellular mediator of the transactivation function of the HIV Tat protein. P-TEFb is responsible for the phosphorylation of the carboxyl-terminal domain of RNA Pol II, resulting in stimulation of transcription. Furthermore, the complexes containing CDK9 induce the differentiation in distinct tissue. The CDK9/cyclin T1 complex is expressed at higher level in more differentiated primary neuroectodermal and neuroblastoma tumors, showing a correlation between the kinase expression and tumor differentiation grade. This may have clinical and therapeutical implications for these tumor types. Among the CDK inhibitors two have shown to be effective against CDK9: Roscovitine and Flavopiridol. These two inhibitors prevented the replication of human immunodeficiency virus (HIV) type 1 by blocking Tat transactivation of the HIV type 1 promoter. These compounds inhibit CDKs by binding to the catalytic domain in place of ATP, preventing transfer of a phosphate group to the substrate. More sensitive therapeutic agents of CDK9 can be designed, and structural studies can add information in the understanding of this kinase. The major features related to CDK9 inhibition will be reviewed in this article.

Key Words: CDK9, drug-design, molecular modeling.

INTRODUCTION

Cell cycle progression is controlled by cyclin-dependent kinases (CDKs) [1-3]. CDKs (EC 2.7.11.22) are part of a family of serine/threonine protein kinases that play a key role in the regulation of progression through the cell division cycle. In each phase of the cell cycle CDKs phosphorylate distinct proteins and therefore are usually classified as either G1 (CDK4 and CDK6/cyclins D, CDK2/cyclin E), S (CDK2/cyclin A, CDK1/cyclin A), or G2/M (CDK1/cyclin B) phase specific [4,5]. Human CDKs are inactive as monomers and regulation of these kinases is tightly controlled at different levels, including interactions with negative and positive partners, activation by de/phosphorylation and subcellular localization.

Besides the role in cell cycle regulation, several CDKs participate in other physiological processes, such as neuronal function (CDK5, CDK11), apoptosis (CDK1, CDK5) and transcription (CDK2, CDK7, CDK8, CDK9, CDK11), which are also affected by pharmacological inhibitors of CDKs [6,7][9, 23]. Especially the latter one has recently attracted medicinal chemists, because roles of CDKs in many viral infections have been established. Small DNA viruses rely on host cell CDK2 for their replication, some require host expression apparatus (CDK2-activated) to synthesize replication proteins, while others encode their own cyclins that activate cellular CDKs, or contain proteins that directly recruits CDKs to the nascent viral transcripts to finish its synthesis [8].

Since deregulation of cyclins and/or alteration or absence of CDK inhibitors (CKIs) has been associated with many cancers, there is strong interest in chemical inhibitors of CDKs that could play an important role in the discovery of new family of antitumor agents [9]. Therefore, namely CDK2 and 4 became very often investigated as promising protein targets for development of novel anticancer agents [10-13]. In addition to the positive regulatory role of cyclins and CDK7, many negative regulatory proteins (CKIs) have been discovered [13]. CKIs fall into two broad categories based upon whether they bind solely to the catalytic subunit (CDK) or whether they bind to CDK complexes. The former category has been identified only in higher eukaryotic cells and is termed the INK family. The most studied INKs are the proteins p15, p16, p18, and p19. These inhibitory proteins bind to CDK4 and 6 and prevent binding of the cyclin partner [14]. Inhibitors of the second category bind to CDK complexes. In mammalian cells, this family includes protein p21, p27, and p57 [14]. Besides positive and negative regulation by phosphorylation and binding of cyclins and CKIs, CDKs interact with another class of regulatory proteins known as CskHs. This protein is also known as p9. It was identified two homologues in humans (CskHs1 and CskHs2) [13]. Despite many studies investigating its role in the cell cycle, the function of p9 is not yet understood. The structure of the complex CDK2-p9 has been determined and based on this structure it was proposed that p9 acts in targeting the

*Address correspondence to this author at the Faculdade de Biociências-PUCRS-Av. Ipiranga,6681. Porto Alegre-RS Brazil;
E-mail: walter.junior@pucrs.br

1573-4064/08 $55.00+.00

CDKs to other proteins or macromolecular assemblies [15]. It is clear that phosphorylation of serine, threonine and tyrosine residues is of major importance in all aspects of cell life and protein kinases are pharmacological targets [6].

ATP is the authentic cofactor of CDK it can be considered as a "lead compound" for discovery of CDK inhibitors. However, there are two major concerns: adenine containing compounds are common ligands for many enzymes in cells, thus, and adenine derivatives may inhibit other enzymes in the cells; second, any highly charged groups such as phosphates in ATP will prevent uptake by cells [16].

Over 80 CDK2 inhibitors have now been described, some of then presents IC50 at nanomolar concentrations [17]. During the past decade many potent and selective inhibitors of CDKs were developed in the SAR drug discovery programs, in spite of a relevant degree of active site similarity across the realm of protein kinases. The structures of binary complexes between CDK2 with several different inhibitors have been described using biocristallography and structural bioinformatics. Recently Canduri & Azevedo [18], proposed a major classification for CDK inhibitors. In this classification the chemical inhibitors of CDK2 are divided in two broad classes considering their specificities. Class I CDK inhibitors are those that are selective for CDK2 (and also CDK1, 5), such as olomoucine, roscovitine, purvalanol B, aminopurvalanol, hymenialdisine, indirubin-3'-monoxime, indirubin-5-sulfonate, SU9516 and alsterpaullone. Class II CDK inhibitors are those that are not selective for any specific CDK, such as deschloroflavopiridol and flavopiridol. Class II inhibitors are not specific for CDK2, although some of them present IC50 in nanomolar concentrations, such as flavopiridol. The effects of CDK inhibitors on the cell cycle progression and their potential value for the treatment of cancer have been extensively studied [6,19]. There are three main features that make CDK inhibitors potential anti-tumor agents: 1) They arrest cells in G1 or G2/M. 2) They trigger apoptosis, alone or in combination with other treatments. 3) In some case, inhibition of CDKs contributes to cell differentiation [20-23].

Most of the CDK inhibitors comprise of structurally distinct flat heterocyclic molecules, entering the active site and competing with ATP, a feature usually demonstrated by enzyme kinetics and CDK2-inhibitor co-crystals and CDK9-inhibitor homology models. Anti-CDK drugs possess inhibitory properties against cancer cells both *in vitro* and *in vivo* and some of them (e.g. trisubstituted purine roscovitine, chloroindolyl sulphonamide E7070 and 2-aminothiazole BMS-387032) are hence already evaluated in clinical trials as a new generation of anticancer chemotherapeutics [24]. Efficiency of several CDK inhibitors such as olomoucine, roscovitine and flavopiridol has been demonstrated on life cycles of several viruses and their activity attributed to inhibition of transcription [23,24,25,26]. However, also oncology offers a new application for these inhibitors. Despite both *in vitro* and *in vivo* efficiency of CDK2 inhibitors, a rationale for their anticancer application has been displaced by evidence that CDK2 knock-out mice are fully viable [27]. What is the proper reason for activity of these inhibitors is still not clear, but there are strong evidences that more CDKs must be targeted at the same time [17]. In particular, inhibition of CDKs mediating phosphorylation of the C-terminal domain (CTD) of RNA polymerase II is thought to markedly contribute to cytotoxic potency of such compounds, as shown with flavopiridol and roscovitine [28-31]. Pharmacological inhibition of transcription, resulting in rapid decline of the protein levels (especially those where both mRNA and protein have short half-lives), might be the explanation for induction of cell death. One such gene product is the anti-apoptotic factor Mcl-1crucial for the survival of a range of cell types, and its down-regulation either by roscovitine or siRNA is sufficient to induce apoptosis [32, 33]. More particularly, selective CDK9 inhibition could also serve as a potential therapeutic strategy against tumor invasion and metastasis, as has been recently demonstrated that inflammatory cytokine tumour necrosis factor-alpha promotes tumour progression through activation of matrix metalloproteinase [17, 34].

Recent progress on the development and analysis of CDK inhibitors already identified several structurally diverse compounds having the ability to block the CDKs' phosphorylating activity [17,35,36]. Newly identified, 4-phenylazo-3,5-diamino-pyrazole inhibitors were found to weakly inhibit the activity of CDK2/cyclin E [17]. Due to the relatively easy synthesis of its derivatives several new protein kinase inhibitors were developed based on this moiety. A total of thirty five 3,5-diaminopyrazole derivatives substituted with different arylazo side chains at position 4 were synthesized to illustrate their structure-activity relationships, and determine their *in vitro* antiproliferative properties on cancer cell lines [17]. This review is focused on the structural basis for specificity of human CDK9 by several different moieties such as flavone derivatives, purines and diaminopyrazole derivatives.

HUMAN CDKs

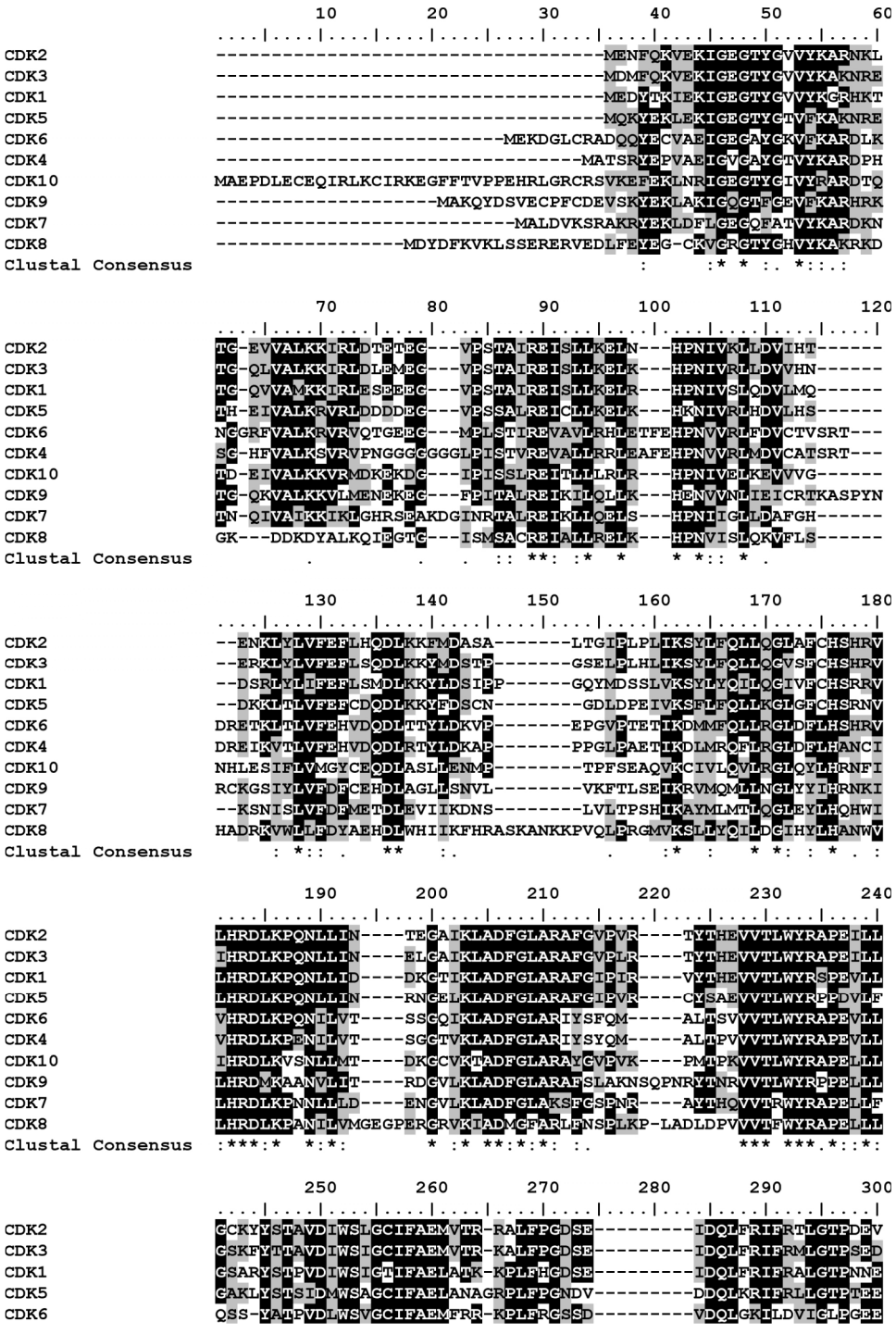
There are over a hundred close-related serine/threonine protein kinases identified in the human genome, however many of them are result of alternative splicing or present larger sequence, when compared with CDK1 and CDK2, which makes alignment clumsy. Fig. (**1**) shows the sequence alignment of 10 human CDKs. In this alignment we can easily visualize the proximity of CDKs. Analysis of the alignment human CDKs indicates the similarity of CDK9 with CDK8 and CDK7. The sequence dendrogram shown in Fig. (**2**), suggested that the CDK alignment joins the sequences in three sub-families, besides the CDK9 sub-family, there is one sub-family joining CDK1, 2, 3 and 5, and a third cluster joining CDK4, 6 and 10. This information has been successfully used to model CDK1 [37], CDK9 [38], and CDK5 [39], generating reasonable models. These models have been used to establish the structural basis for inhibition of these CDKs.

CLONING, EXPRESSION AND PURIFICATION OF CDK9

In order to perform structural studies, human CDK9 has been produced using pET vector system and *E. coli* BL21 (DE)pLys. The expression in different temperatures was used to yield soluble CDK9. The best condition was obtained by culturing the selected colonies with the plasmid

pET23a(+)::cdk9 in LB medium at 30 °C and 0.1mM IPTG. The expression at 37 °C resulted in high amounts of precipitated proteins; at lower temperature (28 and 20 °C), the expression resulted in low amounts of protein. In respect to IPTG, we found that cultures with lower IPTG concentrations produced appropriate amounts of soluble protein. Almost 40mg.mL^{-1} of total protein was synthesized and the soluble proteins were recovered after centrifugation [40]. The CDK9 obtained following this procedure was used for CD studies, which corroborates the molecular model previously obtained for CDK9 [17].

Circular dichroism is an important technique to estimate secondary structure content, to validate protein homology modeling, and for monitoring conformational changes of proteins due to drug binding. This technique has been employed to assess the CDK9 structure in solution. Analysis of the CD spectrum showed that it contains 30–34% helix, 33–36% beta, and 21–29% of coil. Such CD data permitted to obtain information about the CDK9 structures of the regions not modeled [40]. The results suggest that human recombinant CDK9 maintains secondary structures and is viable to functional and structural studies.



(Fig. 1. Contd….)

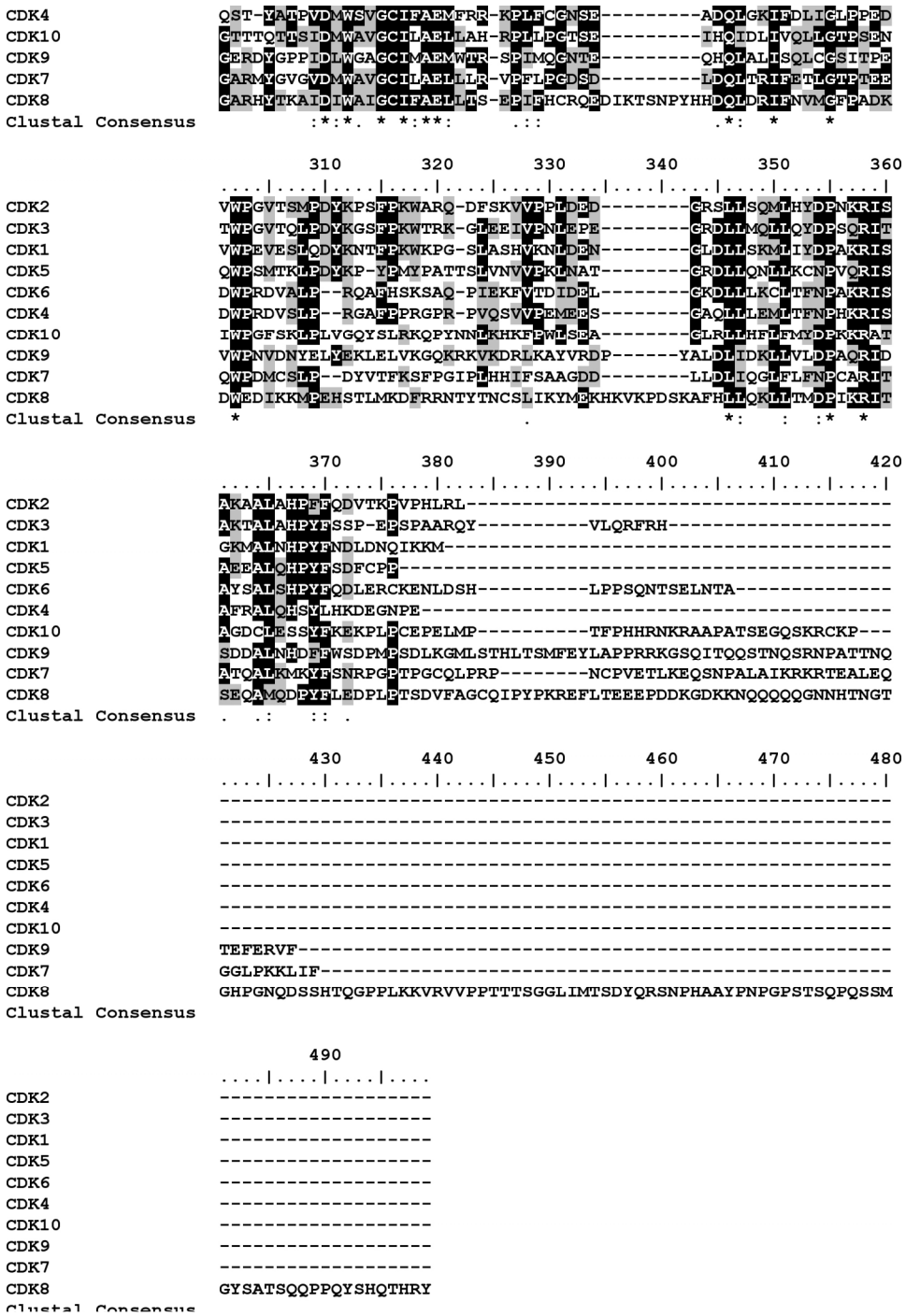


Fig. (1). Sequence alignment of 10 human CDKs generated using BioEdit [58]. The alignment was performed using ClustalW [54].

CDK9 BIOLOGICAL FUNCTIONS

Global Transcriptional Elongation Factor

Transcription by RNA polymerase II (pol II) is a multistep process including preinitiation, initiation, promoter clearance, elongation, and termination [41].

P-TEFb, a positive transcription elongation factor stimulates transcription elongation by phosphorylating Ser2 of the 52 repeats of heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser of the carboxyl terminal domain (CTD) of the RNA pol II [42], preventing its arrest. P-TEFb is a heterodimeric complex comprised of cyclin-dependent kinase 9 (CDK9) and a regulatory cyclin subunit of the T (T1, T2a, T2b or K1).

Tak and CDK9

Human immunodeficiency virus type 1 (HIV-1) encodes for a protein, Tat, which dramatically activates viral transcription. This simulation occurs through the interaction of

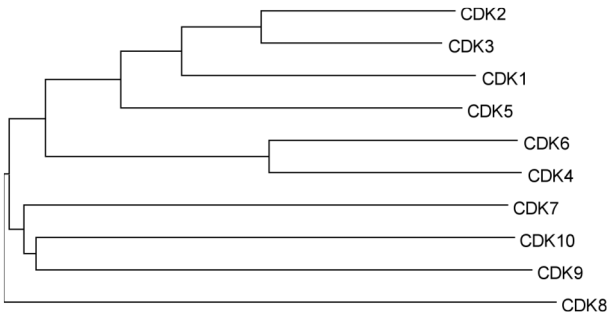


Fig. (2). Sequence dendrogram of the alignment obtained for 10 human CDKs. The alignment was performed using ClustalW [54].

the protein with the transactivator response (TAR) RNA element located at the 5´ end of nascent retroviral transcripts. A Tat-associated kinase (TAK) was isolated, with the ability to phosphorylate CTD. It was identified that CDK9 is the kinase of the TAK complex, showing that TAK is analogous to the P-TEFb [43]. In absence of Tat, phosphorylation of CTD occurs on Ser 2 ad Ser 5; Cdk9 phosphorylates Ser 2, while Cdk7 phosphorylates Ser 5. In presence of Tat, on the other hand, the substrate specificity of Cdk9 is altered, and this kinase phosphorylates both the Ser residues. This suggests that the ability of Tat to increase transcription elongation may be due to its ability to modify the substrate specificity of the Cdk9 complex [44].

CDK9 model

The sequence of Cell division protein kinase 9 (EC 2.7.11.22) (EC 2.7.11.23) (CDK9) consists 372 amino acids with a molecular weight of 42,778 Da and a theoretical pI of 8.97. There is no crystallographic structure available for human CDK9, however the similarity between CDK9 and CDK2 sequence makes CDK2 structure a reasonable template for modelling of CDK9. Azevedo *et al.* using the atomic coordinates of CDK2 generated a homology model for human CDK9, that has also been used to study the structural basis for interaction between CDK9 and inhibitors [38]. To model CDK9 structure it was used Parmodel, a free webserver dedicated to molecular modeling [45]. The alignment of CDK2 (template) and CDK9 (target) is shown in Fig. (**1**). Fifteen residues from the N-terminal and 45 residues from the C-terminus were removed from the CDK9 model, since there is no good template for these fragments. Analysis of the CDK9 structure shows that it belongs to the class of alpha+beta structures and investigation of the fold indicates two alpha+beta domains with a larger C-terminal mostly alpha helical, which is characteristic of the protein-kinase fold. Fig. (**3**) shows the structure of human CDK9 in complex with ATP. As expected the structural model of CDK9 shows a bilobal structure, with smaller N-terminal domain consisting of a sheet of five antiparallel -strands and a single large -helix. The larger C-terminal domain consists primarily of -helices. It contains a pseudo-4-helical bundle, a small -ribbon and two additional -helices. The ATP-binding pocket is found in the cleft between the two lobes. The adenine base is positioned in a hydrophobic pocket between the -sheet of the small domain. In the binary complex of CDK9-ATP the ATP phosphates are held in position by ionic and hydrogen-bonding interactions with several residues, including Lys33, Asp145, and the backbone amides of the glycine-rich loop (residues 10-17). It was observed that ATP binding to CDK9 appears to induce a slight closure of the cleft by a 2^{o} hinge movement around an axis parallel to the longitudinal axis of the ATP molecule, as observed for the crystallographic structure of CDK2 [38]. All inhibitors and ATP bind in the deep cleft between the two domains. Sequence alignment of 10 close related human CDKs (Fig. (**1**)) indicates that these CDKs share several conserved structures, such as: Glycine loop, T loop, and P loop. Fig. (**4**) shows the intermolecular contact between CDK9 and ATP.

Fig. (3). Structure of human CDK9 in complex with ATP solved by molecular modeling [55]. Figure generated using Pymol [56], the protein is presented as ribbon diagram (light gray) and ligand ATP is presented as stick model (dark gray).

Inhibition of CDK9

Based on the structural models obtained for the complexes between CDK9 and different ligands we may highlight few structural features identified on these models that are important for inhibition of CDK9. There is structural information available for CDK9 in complex with flavopiridol [38] and with CAN508 [17], both inhibitors present low molecular weight, below 600 Da, and they are all flat and hydrophobic heterocycles. Fig. (**5**) shows the structural formula for flavopiridol and CAN508. Furthermore, all CDK inhibitors studied so far act by competing with ATP for binding in the CDK ATP-binding pocket [18]. Fig. (**6**) shows the interaction of these two inhibitors with ATP-binding pocket. Most of the contacts are hydrophobic and the complexes present few intermolecular interactions. The inhibitors are enclosed in a hydrophobic pocket formed by Ile10, Ala31, Val64, and Leu134. This hydrophobic environment is conserved in all human CDKs, with one exception at position 31, which is replaced for Asp in the CDK8 sequence. We used the numbering scheme shown for CDK2. Analysis of the contact area between inhibitor and CDK9 indicates that inhibitors with low IC50 values (higher affinity for CDK9) present higher contact areas and higher number intermolecular hydrogen bonds. One of the most studied CDK inhibitor is flavopiridol, which is also one of the most potent inhibitors of HIV-1 transcription and virus replication identified to date.

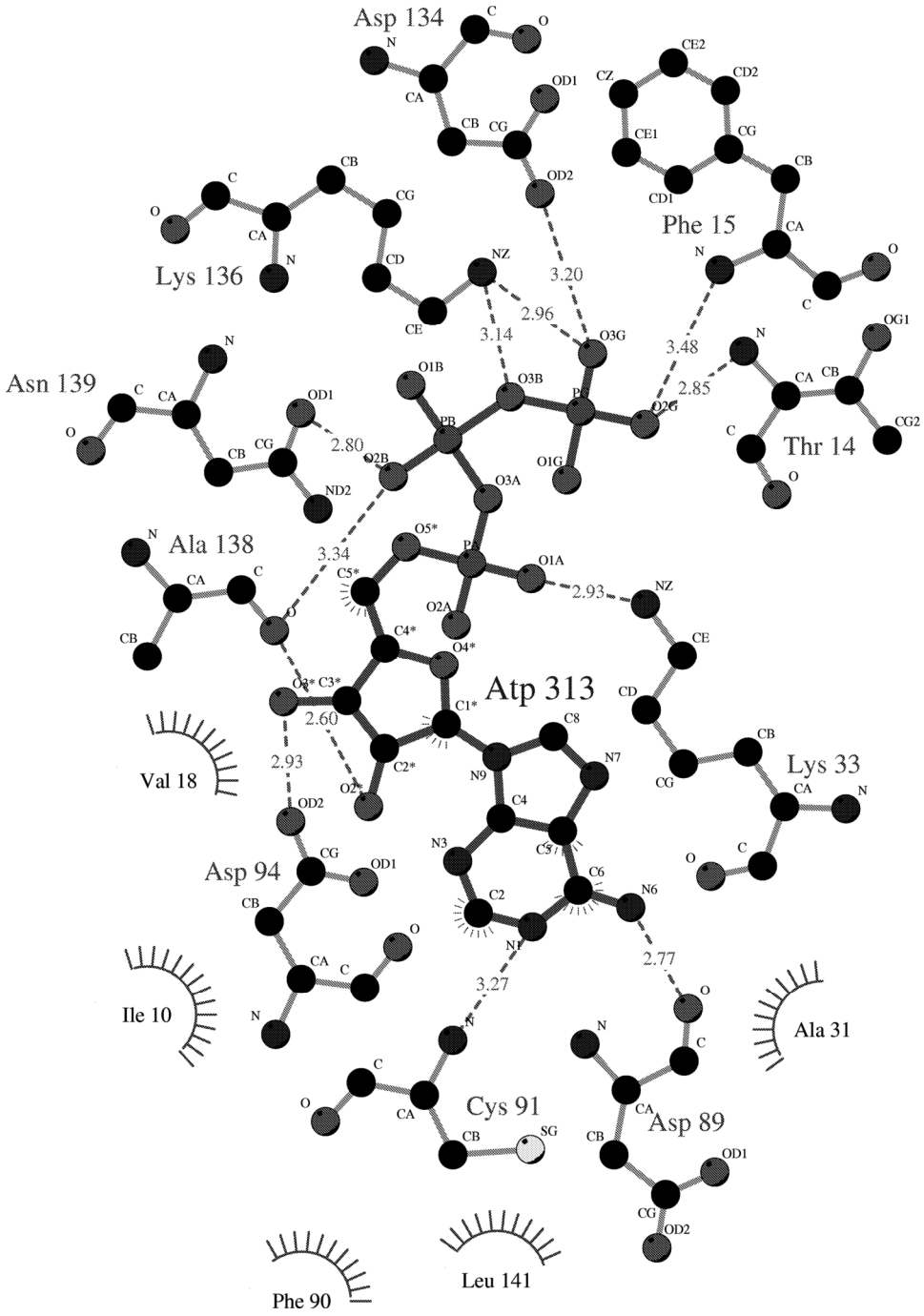


Fig. (4). Intermolecular contacts between CDK9 and ATP. Figure generated with Ligplot [57], the dashed lines shows the hydrogen bonds among CDK9 residues and ATP ligand and its length.

A primary target of flavopiridol is the positive transcription elongation factors b (P-TEFb) kinase. A close inspection of the CDK9 model did not reveal the presence of any additional binding pockets. Furthermore, attempt to dock CDK inhibitors with CDK9 identified the same ATP pocket as a binding site [38]. CDK9-flavopiridol model clearly reveals a very tight binding of the inhibitor to the enzyme, which may essentially inactivate the enzyme, in such scenery a competitive kinetic mechanism is possible.

Analysis of the binary complexes of CDK2 and CDK9 with flavopiridol indicates that the contact areas for the complexes of CDK2 and CDK9 with flavopiridol are 320 $Å^2$ d 332 $Å^2$ respectively, which may partially explain the lower IC50 value observed for CDK9. In addition, the CDK9–flavopiridol complex shows a higher number of intermolecular hydrogen bonds, which also indicates that flavopiridol has higher affinity for CDK9. The structure of the binary complex indicates that the flavopiridol is tightly bound to the ATP-binding pocket and since no further binding sites were

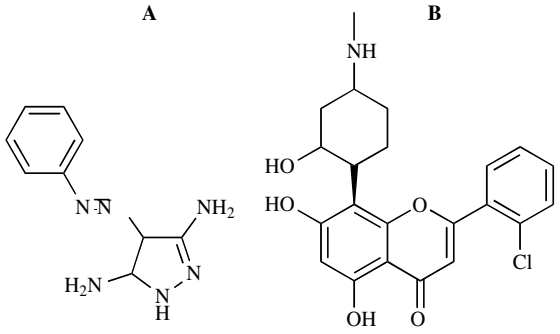


Fig. (5). Structural formula for CAN508 (4-Arylazo-3,5-diamino-pyrazole) (A) and flavopiridol (B).

identified in the CDK9 structure we have strong structural evidence that flavopiridol is a competitive inhibitor with ATP, as observed for the CDK2. Superposition of the complexes of flavopiridol with CDK2 and CDK9 is shown on Fig. (**7**). The benzopyran ring of flavopiridol occupies approximately the same region in both complexes. This region is the same occupied by the purine ring of ATP in the CDK2-ATP complex [16]. Superposition of the CDK2-ATP onto CDK9-flavopiridol structure indicates that the two ring systems overlap approximately in the same plane, but the benzopyran is rotated about 52.5° relative to the adenine in ATP, measured as the angle between the carbon–carbon bonds joining the two cycles in benzopyran and adenine rings, respectively. Two strong salt bridges are observed between Glu 92–Lys 20 and Glu 92–Lys 29 in the CDK9–flavopiridol structure, these salt bridges involve residues from two different lobes of the CDK9 structure and they are not conserved on the CDK2 structure, Lys 20 and 29 in the N-terminal lobe and Glu 92 in the C-terminal lobe. This strong electrostatic interaction brings the two lobes of CDK9 closer, which increases the contact area between enzyme and inhibitor. Fig. (**8**) shows the position of the salt bridges on the CDK9 structure. Furthermore the side chains of Glu 92, Lys 20, and Lys 29 brings the phenyl of Phe 90 to a position deeper in the binding pocket, which also contributed to increase the contact area between flavopiridol and CDK9.

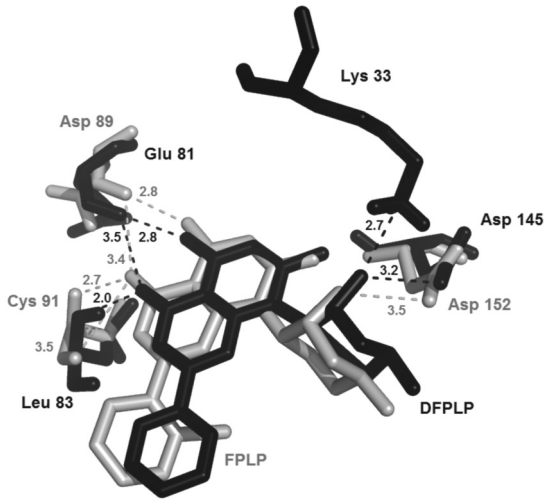


Fig. (7). Details of superposition of the complexes of deschloroflavopiridol (DFPLP) with CDK2 and flavopiridol (FPLP) with CDK9 showing intermolecular hydrogen bonds involving ligand and residues, indicated with dashed lines. The figure was generated with program Pymol [55]. CDK2 and DFPLP are presented as black stick and CDK9 and FPLP as light gray stick.

Most recently, a structural model for the binary complex of CDK9-CAN508 has been proposed using the same protocol described for the modeling of CDK9-flavopiridol [17]. This inhibitor, CAN508 (4-Arylazo-3,5-diaminopyrazole), proved to be a competitive inhibitor of CDK2**-**cyclin E with respect to ATP [17]. Cellular activity of CAN508 blocks proliferation of several tested cancer cell lines. Roscovitine a well-known CDK inhibitor, which has similar affinity for CDK9 [24] causes accumulation of p53 in cells treated at lower concentrations [17]. Analysis of the CDK selectivity

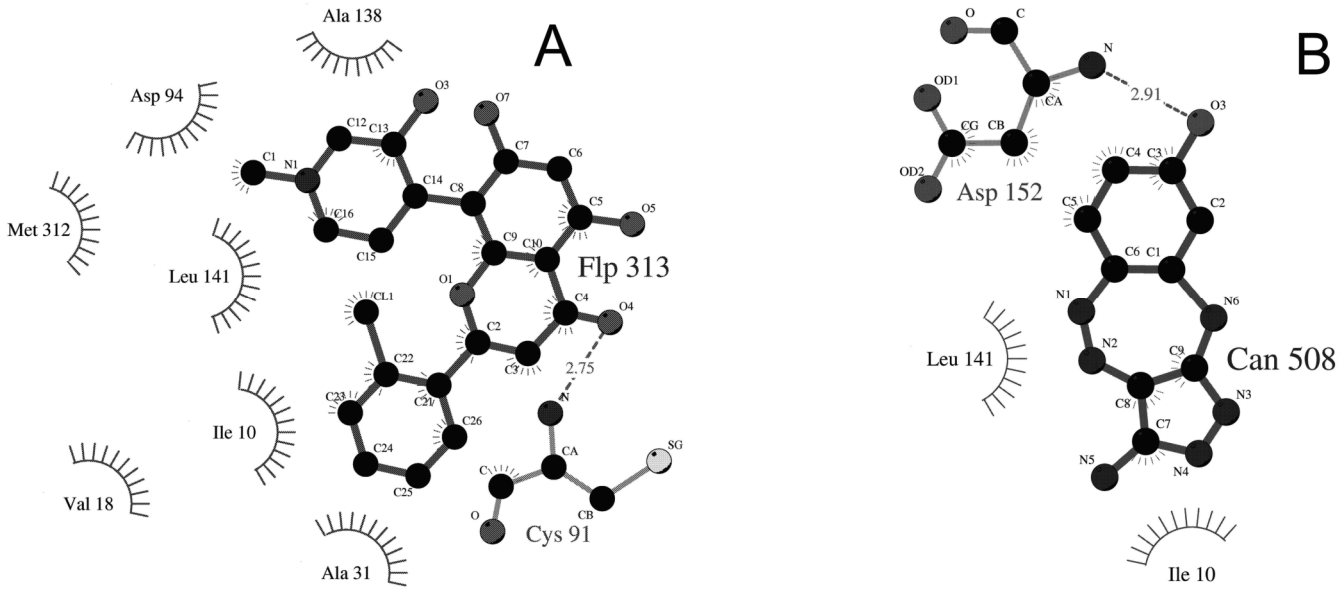


Fig. (6). Intermolecular hydrogen bonds for flavopiridol (A) and CAN508 (B) with ATP-binding pocket. Figure generated with Ligplot [57].

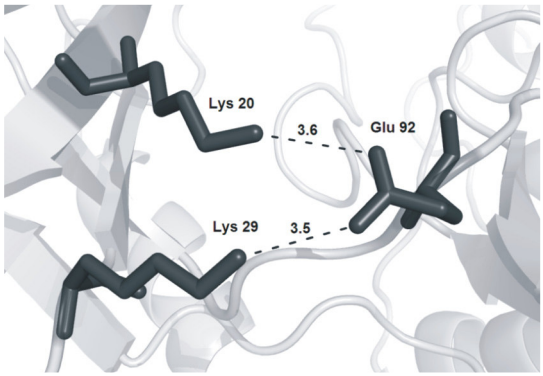


Fig. (8). Salt bridges, observed between Glu 92–Lys 20 and Glu 92–Lys 29 in the CDK9–flavopiridol. Figure generated with program Pymol [51], the CDK9 is presented as ribbon (light gray) and residues Lys 20, Lys 29 and Glu 92 are presented as stick (dark gray).

profiles indicates that inhibition of other CDKs (probably *via* combined effects on CDK2, CDK7, and CDK9) may be responsible for the antiproliferative and pro-apoptotic potency of CDK inhibitors [17]. However, the cellular activity of several 4-phenylazo-3,5-diamino-1*H*-pyrazole derivatives significantly exceeded that of olomoucine, thus, they represent good lead compounds for further structural optimization of specific CDK9 inhibitors [17], with applications in development of anticancer and anti-viral drugs.

The selectivity of CAN508 towards CDK9-cyclin T1 found during kinase assays prompted an investigation into the structural basis for this selectivity. Analysis of the modeled interactions between CAN508 and CDK9 (Fig. (**6**)) indicates the participation of Asp152 (which corresponds to Asp145 in CDK2) in an intermolecular hydrogen bond with the phenol- OH of the inhibitor. Intermolecular hydrogen-bond interactions corresponding to those with Glu81 and Leu83 observed in the CDK2-CAN508 [17] complex appear to be absent in the CDK9-CAN508 model. This was the first CDK inhibitor to show no participation of the molecular fork of CDK in intermolecular hydrogen bonds, previously described to be present in intermolecular hydrogen bonds in the structures of all CDK-inhibitor complexes studied so far [38]. In the CDK9 complex structure, CDK9 interactions with CAN508 are characterized by predominantly hydrophobic and van der Waals interactions between the protein and the 4-phenylazo moiety. Most of the intermolecular contacts between CDK9 and CAN508 involve a hydrophobic pocket formed by the residues Ile25, Val33, Phe103, Cys116, Leu156, Ala166, and Asp167. A structural comparison of both complexes (CDK2-CAN508 and CDK9-CAN508) indicates that the specificity of CAN508 for CDK9 does not stem from the establishment of intermolecular hydrogen bonds, because the CDK9-CAN508 complex presents a smaller number of hydrogen bonds when compared with CDK2-CAN508 complex. The most striking difference between the two complexes concerns the intermolecular contact area, which is significantly higher in the CDK9 complex (190 $Å^2$) than in the CDK2 complex (174 $Å^2$). Structural inspection of the interaction between CAN508 and CDK9 suggests that the inhibitor's specificity against CDK9 is probably due to the 4-phenylazo moiety, which affords a higher intermolecular contact area with CDK9 than that observed for the CDK2-CAN508 complex. Especially interesting is the presence of two salt bridges (Glu107-Lys35 and Glu107-Lys44) in the CDK9-CAN508 structure; these were also observed in the CDK9-flavopiridol complex and involve residues at interface between both lobes of the CDK9 structure, the overall result is an increase in the contact area between the CDK9 and CAN508. These salt bridges are not observed in the CDK2-CAN508 complex. A tighter ATP-binding pocket observed in the CDK9-CAN508 complex makes possible a higher number of intermolecular van der Waals contacts between CDK9 and CAN508.

The analysis of the charge distribution of the binding pockets indicates the presence of some charge complementarity between inhibitor and enzyme, nevertheless most of the binding pocket is hydrophobic in both structures. The overall analysis of the intermolecular interaction between CDK9 and inhibitors strongly indicates that we would expect higher intermolecular contact area and higher number of intermolecular hydrogen bonds for inhibitors with low IC50. This qualitative observation may help the guidance on exploring large chemical libraries, selecting potential new CDK inhibitor based on the intermolecular hydrogen bond pattern and contact area.

MOLECULAR FORK

It has been observed in several structures of CDK2 and CDK9 complexed with inhibitors the participation of a molecular fork, composed by a C=O group of Glu81 (Asp in CDK9) and the N-H and C=O group of Leu83 (Cys in CDK9), in intermolecular hydrogen bonds between CDK2 and CDK9 and the inhibitors. There are eighty structures of CDK2 complexed with inhibitors (search performed on the PDB [46] on January 2007). Analysis of the intermolecular hydrogen bond distances between CDK2 and all inhibitors that have atomic coordinates deposited in the PDB and also for the complexes of CDK2 with IPA, olomoucine, roscovitine, and flavopiridol strongly indicates the conservation of at least one member of the molecular fork in hydrogen bonding. This molecular fork, composed of two hydrogen bond acceptors (C=O) and one hydrogen bond donor (N-H), allows a wide range of different molecules to dock on to the ATP binding pocket, such as: olomoucine, isopentenyladenine, and roscovitine [47,48,39], staurosporine [49], purvalanols [50], indirubins [51], hymenialdisine [49], UCN-01[52], NU2058 [18], and diaminopyrazole derivatives [17]. All these inhibitors have pairs of hydrogen bond partners that show complementarity to the molecular fork on CDK2 and CDK9, most of them involving at least two hydrogen bonds with the molecular fork. The relative orientation of the inhibitor in the ATP-binding pockets of CDK2 and CDK9 locate one hydrogen bond donor close to C=O in Glu81 (Asp in CDK9) and/or Leu83 (Cys in CDK9), and an acceptor close to N-H in Leu83 (Cys in CDK9). Differences in the sequence of the molecular fork, observed between CDK2 and CDK9, is not expected to cause major differences in the intermolecular hydrogen bonds of the molecular fork, since it involves only main-chain atoms in the hydrogen bonds. Such simple paradigm is conserved in all CDK2-inhibitor complex structures solved so far. In all binary models obtained by

homology modeling for CDK9, this molecular fork is also conserved [40].

COMPLEX CDK9-CYCLIN

Human CDKs are inactive as monomers, and their activation needs binding to cyclins, a diverse family of proteins whose levels oscillate during the cell cycle, in the case of CDK9 the partner is (cyclin T or cyclin K). There is no structural information for cyclin T1, however recently the cyclin K structure has been solved by X-ray crystallography [53]. The structure of cyclin K opened the possibility of modeling of cyclin T1, since the sequence identity between them is over 42 %. This modeling allowed studies of the interaction of CDK9 and cyclin T1, using molecular modeling approaches. Two isoforms of CDK9 have been identified in the human genome cells, CDK9 (42kDa) and CDK9 (55kDa), which differ in their N-terminal domains. The association of these isoforms with cyclins K, T1, T2a, and T2b, results in the formation of eight structurally and functionally distinct P-TEFb. This binary complex shows that their CDK9-interacting surfaces present great structural differences, which could potentially be exploited for the design of cyclin-targeted inhibitors of the CDK9–cyclin K and CDK9–cyclin T1 complexes.

REFERENCES

[1] Hunt, T. Curr. Opin. Cell Biol.,**1989**, *1*, 268.
[2] Fang, F. & Newport, J. W. Cell, **1991**, *66*, 731.
[3] Norbury, C. & Nurse, P. Annu. Rev. Biochem., **1992**, *61*, 441.
[4] Sherr, C. J. *Science*, **1996**, *274*, 1672.
[5] Morgan, D. O. *Annu. Rev. Cell Dev. Biol.*, **1997**,*13*, 261.
[6] Knockaert, M.; Greengard, P.; Meijer, L. *Trends Pharmacol. Sci.,* **2002**, *23*, 417.
[7] Loyer, P.; Trembley, J. H.; Katona, R.; Kidd, V. J.; Lahti, J. M. *Cell. Signalling,* **2005**, *17*, 1033.
[8] Schang, L. M. *Curr. Drug Targets: Infect. Disord.,* **2005**, *5*, 29.
[9] De Azevedo, W.F.Jr.; Mueller-Dieckmann, H.J.; Schulze-Gahmen, U.; Worland, P.J.; Sausville, E.; Kim S.-H. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 2735.
[10] Dai, Y.; Grant, S. (2003) *Curr. Opin. Pharmacol.,* **2003***, 3*, 362.
[11] Sielecki, T. M.; Boylan, J. F.; Benfield, P. A.; Trainor, G. L. *J. Med. Chem.*, **2000**, *43*, 1.
[12] Fischer, P. M.; Lane, D. P. *Curr. Med.* Chem., **2000**, *7*, 1213.
[13] Richardson, H.E.; Stueland, C.S.; Thomas, J.; Russel, P.; Reed, S. I. *Genes Dev.*, **1990**, *4*, 1332.
[14] Peter, M. *Prog. Cell Cycle Res.*, **1997**, *3*, 99.
[15] Bourne, Y.; Watson, M. H.; Hickey, M. J.; Holmes, W.; Rocque, W.; Reed, S. I.; Tainer, J. A. *Cell*, **1996**, *84*, 863.
[16] Kim, S.-H.; Schulze-Gahmen, U.; Brandsen, J.; De Azevedo, W. F. Jr. *Prog. in Cell Cycle Res.*, **1996**, *2*, 137.
[17] Krystof, V.; Cankar, P.; Frysova, I.; Slouka, J.; Kontopidis, G.; Dzubak, P.; Hajduch, M.; Srovnal, J.; de Azevedo, W.F.Jr.; Orsag, M.; Paprskarova, M.; Rolcik, J.; Latr, A.; Fischer, P.M.; Strnad, M. *J. Med. Chem.*, **2006**, *49*, 6500.
[18] Canduri, F. and De Azevedo, W. F. *Current Computer-Aided Drug Design.,* **2005***, 1,* 53.
[19] Malumbres, M.; Barbacid, M. *Natl. Rev. Can.*, **2001**, *1*, 222.
[20] Soni, R.; O'Reilly, T.; Furet, P.; Muller, L.; Stephan, C.; Zumstein-Mecker, S.; Fretz, H.; Fabbro, D.; Chaudhuri, B. J. *Natl. Can. Inst.,* **2001**, *93*, 436.
[21] Damiens, E.; Baratte, B.; Marie, D.; Eisenbrand, G.; Meijer, L. *Oncogene.*, **2001**, *20*, 3786.
[22] Edamatsu, H.; Gau, C.L.; Nemoto, T.; Guo, L.; Tamanoi, F. *Oncogene*, **2000**, *19*, 3059.
[23] Matushansky, I.; Radparvar, F.; Skoultchi, A.I. *Proc. Natl. Acad. Sci. USA,* **2000**, *97*, 14317.
[24] Benson, C.; Kaye, S.; Workman, P.; Garrett, M.; Walton, M.; de Bono, J. *Br. J. Cancer,* **2005**, *92*, 7.
[25] Chao, S.-H.; Fujinaga, K.; Marion, J.E.; Taube, R.; Sausville, E.A.; Senderowicz, A.M.; Peterlin, B.M.; Price, D.H. J. *Biol. Chem.,* **2005***, 275*, 28345.
[26] Agbottah, E.; de La Fuente, C.; Nekhai, S.; Barnett, A.; Gianella-Borradori, A.; Pumfery, A.; Kashanchi, F. *J. Biol. Chem.,* **2005***, 280*, 3029.
[27] Ortega, S.; Prieto, I.; Odajima, J.; Martin, A.; Dubus, P.; Sotillo, R.; Barbero, J.L.; Malumbres, M.; Barbacid, M. *Nat. Genet.,* **2003***, 35*, 25.
[28] Ljungman, M.; Paulsen, M.T. *Mol. Pharmacol.,* **2001***, 60*, 785.
[29] Chao, S.-H.; Price, D.H. *J. Biol. Chem.,* **2001**, *276*, 31793.
[30] Demidenko, Z.N.; Blagosklonny, M.V. *Cancer Res.*, **2004**, *64*, 3653.
[31] Whittaker, S.R.; Walton, M.I.; Garrett, M.D.; Workman, P. *Cancer Res.,* **2004**, *64*, 262.
[32] MacCallum, D.E.; Melville, J.; Frame, S.; Watt, K.; Anderson, S.; Gianella-Borradori, A.; Lane, D.P.; Green, S.R. *Cancer Res.,* **2005**, *65*, 5399.
[33] Lacrima, K.; Valentini, A.; Lambertini, C.; Taborelli, M.; Rinaldi, A.; Zucca, E.; Catapano, C.; Cavalli, F.; Gianella-Borradori, A.; MacCallum, D. E.; Bertoni, F. *Ann. Oncol.,* **2005**, *16*, 1169.
[34] Shan, B.; Zhuo, Y.; Chin, D.; Morris, C.A.; Morris, G.F.; Lasky, J.A. Cells. *J. Biol. Chem.,* **2005***, 280*, 1103.
[35] Havlicek, L.; Fuksova, K.; Krystof, V.; Orsag, M.; Vojtesek, B.; Strnad, M. *Bioorg. Med. Chem.,* **2005**, *13*, 5399.
[36] Moravcova, D.; Krystof, V.; Havlicek, L.; Moravec, J.; Lenobel, R.; Strnad, M. *Bioorg. Med. Chem. Lett.,* **2003***, 13*, 2989.
[37] Canduri, F.; Uchoa, H.B.; de Azevedo, W.F.Jr. *Biochem. Biophys. Res. Commun.,* **2004***, 324*, 661.
[38] De Azevedo, W.F.Jr.; Canduri, F.; Da Silveira, N.J.F. *Biochem. Biophys. Res. Commun,.* **2002***, 293*, 566.
[39] De Azevedo, W.F.Jr.; Gaspar, R.T.; Canduri, F.; Camera, J.C.Jr.; Silveira, N.J.F. *Biochem Biophys. Res. Commun.*, **2002**, *297*, 1154.
[40] Leopoldino A.M.; Canduri F.; Cabral H.; Junqueira M.; de Marqui A.B.T.; Apponi L.H.; da Fonseca I.O.; Domont G.B.; Santos D.S.; Valentini S.; Bonilla-Rodriguez G.O.; Fossey M.A.; de Azevedo Jr. W.F.; Tajara E.H.; *Protein Expression Purification,* **2006**, *47*, 614.
[41] Sims, R.J.; Belotserkovskaya, III.R.; Reinberg, D. *Genes Dev.*, **2004**, *18***,** 2437.
[42] Mortillaro, M.J.; Blencowe, B.J.; Wei, X.; Nakayasu, H,; Du, L.; Warren, S.L.; Sharp, P.A.; Berezney, R. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 8253.
[43] De Falco, G.; Giordano, A. *Cancer Biol Ther*., **2002**, *1*, 342.
[44] Fu J, Yoon HG, Qin J, Wong J. *Mol. Cell Biol*., **2007**, *27*, 4641.
[45] Uchoa, H.B.; Jorge, G.E.; Freitas Da Silveira, N.J.; Camera, J.C.,Jr.; Canduri, F.; De Azevedo, W.F.,Jr. *Biochem. Biophys. Res. Commun.,* **2004***, 325*, 1481.
[46] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic. Acids Res.*, **2002**, *28*, 235.
[47] Noble, M.E.; Endicott, J.A.; Johnson, L.N. *Science*, **2004**, *303*, 1800.
[48] Schultze-Gahmen, U.; Brandsen, J.; Jones, H.D.; Morgan, D.O.; Meijer, L.; Vesely, J. *Proteins*, **1995**, *22*, 378.
[49] Meijer, L.; Raymond, E. *Acc. Chem. Res.*, **2003**, *36*, 417.
[50] Losiewicz, M.D.; Carlson, B.A.; Kaur, G.; Sausville, E.A.; Worland, P.J. *Biochem. Biophys. Res. Commun.*, **1994**, *201*, 589.
[51] Kaur, G.; Stetler-Stevenson, M.; Sebers, S.; Worland, P.; Sedlacek, H.; Myers, C.; Czech, J.; Naik, T.; Sausville, E. *J. Natl. Can. Inst.*, **1992**, *84*, 1736
[52] Senderowicz, A.M.; Sausville, E.A. *J. Natl. Canc. Inst.*, **2000**, *92*, 376.
[53] Baek, K.B.; Brown, R.S.; Birrane, G.; Ladias, J.A. *J. Mol. Biol.,* **2007**, *366*, 563.
[54] Thompson, J.D.; Higgins, D.G.; Gibson T.J. *Nucleic Acids Res.*, **1994**, *22*, 4673.
[55] De Azevedo, W.F.Jr.; Canduri, F.; Silveira, N.J.F. *Biochem. Biophys. Res. Commun.,* **2002**, *293*, 566.
[56] DeLano, W.l. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA, **2002**. [http://www.pymol.org]
[57] Wallace, A.C.; Laskowski, R.A.; Thornton, J.M. *Prot. Eng.*, **1995**, *8*, 127.
[58] Tippmann H. F. *Brief Bioinform.*, **2004**, *5*, 82.

JrnlID 894 ArtID 291 Proof# 1 - 04/03/2008

J Mol Model
DOI 10.1007/s00894-008-0291-2

ORIGINAL PAPER

Molecular modeling and dynamics studies of cytidylate kinase from Mycobacterium tuberculosis H37Rv

Rafael Andrade Caceres · Luís Fernando Saraiva Macedo Timmers · Ana Luiza Vivan · Cristopher Zandoná Schneider · Luiz Augusto Basso · Walter Filgueira De Azevedo Jr. · Diogenes Santiago Santos

Received: 10 December 2007 / Accepted: 12 February 2008

Abstract Bacterial cytidylate kinase or cytidine monophosphate kinase (CMP kinase) catalyses the phosphoryl transfer from ATP to CMP and dCMP, resulting in the formation nucleoside diphosphates. In eukaryotes, CMP/UMP kinase catalyses the conversion of UMP and CMP to, respectively, UDP and CDP with high efficiency. This work describes for the first time a model of bacterial cytidylate kinase or cytidine monophosphate kinase (CMP kinase) from mycobacterium tuberculosis (MtCMPK). We modeled MtPCMPK in apo form and in complex with cytidine 5′-monophosphate (CMP) to try to determine the structural basis for specificity. Comparative analysis of the model of MtCMPK allowed identification of structural features responsible for ligand affinities. Analysis of the molecular dynamics simulations of these two systems indicates the structural features responsible for the stability of the structure, and may help in the identification of new inhibitors for this enzyme.

Keywords Cytidine 5′-monophosphate · Enzyme substrate specificity · Molecular dynamics · Molecular modeling · Mycobatcrium tuberculosis

Abbreviations

TB	tuberculosis
CMPK	cytidine monophosphate kinase
MDR-TB	multidrug resistant TB
MD	molecular dynamic

Introduction

Tuberculosis (TB) remains one of the deadliest diseases worldwide. World Health Organization estimates that one third of the population is infected with mycobacterium tuberculosis, the causative agent of TB [1]. It has been estimated that there are approximately 8 million new cases of TB every year and 2 million deaths occur each year [2]. The reemergence of TB is due to the high incidence of AIDS, the proliferation of drug-resistant strains and the decline in health care structures and national surveillance. It has been pointed out that the emergence and spread of multi-drug resistant strains of M. tuberculosis (MDR-TB), defined as resistant to at least isoniazid and rifampicin, could threaten global TB control [3]. More recently, a survey of the frequency and distribution of extensively drug-resistant (XDR) TB cases showed that during 2000–2004, of 17,690 TB isolates, 20% were MDR and 2% were XDR [4]. XDR-TB is defined by the World Health

RA. Caceres
Programa de Pós Graduação em Biologia Celular e Molecular,
Pontifícia Universidade Católica do Rio Grande do Sul,
Porto Alegre, RS, Brazil

R. A. Caceres · L. F. S. Macedo Timmers · A. L. Vivan · W. F. De Azevedo Jr. (✉)
Faculdade de Biociências, Laboratório de Bioquímica Estrutural,
Pontifícia Universidade Católica do Rio Grande do Sul,
Porto Alegre, RS, Brazil
e-mail: walter.junior@pucrs.br

C. Z. Schneider · L. A. Basso · D. S. Santos
Centro de Pesquisas em Biologia Molecular e Funcional,
Instituto de Pesquisas Biomédicas, Pontifícia Universidade
Católica do Rio Grande do Sul,
Porto Alegre, RS, Brazil

D. S. Santos
e-mail: diogenes@pucrs.br

JmlD 864 ArtID 291 Proof# 1 - 04/03/2008

Organization as resistance to at least rifampin and isoniazid plus resistance to the fluoroquinolones and to at least one of the injectable drugs capreomycin, kanamycin and amikacin [5]. XDR-TB has a wide geographic distribution [6] and high fatality rate [7]. The emergence of XDR-TB worldwide raises the bleak prospect of virtually untreatable TB. There is thus an urgent need for the development of new chemotherapeutic agents to treat TB.

CMP kinases belong to a family called nucleoside monophosphate kinases (NMK), which are key enzymes in the metabolism of nucleotides [8]. Substrate specificity studies on recombinant human UMP/CMP kinase (pyrimidine nucleoside monophosphate kinase) has shown that UMP and CMP are far better substrates than dCMP [9]. Bacterial cytidylate kinase or cytidine monophosphate kinase (CMP kinase) catalyses the phosphoryl transfer from ATP to CMP and dCMP, resulting in the formation of nucleoside diphosphates. CMP kinase has been shown to be essential for bacterial growth in Streptococcus pneumoniae [10]. The Staphylococcus aureus CMP kinase crystal structure has recently been resolved [11]. The crystal structure of Escherichia coli CMP kinase resembles those of other NMP kinases sharing common features such as a central five-stranded β-sheet connected by α-helices, a fingerprint sequence of Glu-x-x-Gly-x-Gly-Lys (P-loop), and an anion hole in the central cavity for substrate binding [12]. However, E. coli CMP kinase contains an insertion of 40 residues and a short LID domain, differing from other NMP kinases [12]. These differences could allow the rational design of drugs specific for CMP kinase of pathogenic bacteria, such as M. tuberculosis. However, no report has yet described the three-dimensional structure of M. tuberculosis CMP kinase. In the present work we modeled the structure of CMPK from M. tuberculosis (MtCMPK). It was analyzed the structure of MtCMPK in complex with cytidine-5′-monophosphate (CMP) to identify the structural basis for the substrate interactions in the binding site and to predict the enzyme substrate specificity. The structural features and the structural stability were assessed by molecular dynamic (MD) simulation.

Materials and methods

Molecular modeling

Homology modeling is usually the method of choice when there is a clear relationship of homology between the sequences of a target protein and at least one experimentally determined three-dimensional structure. This computational technique is based on the assumption that tertiary structures of two proteins will be similar if their sequences were related, and it is the approach most likely to give accurate results [13].

For modeling of the MtCMPK we used restrained-based modeling implemented in the program MODELLER 9v1 [14]. This program is an automated approach to comparative modeling by satisfaction of spatial restrains [15]. The modeling procedure begins with alignment of the sequence to be modeled (target) with related known three-dimensional structures (templates). This alignment is usually the input to the program and the output is a three-dimensional model for the target sequence containing all main-chain and side-chain non-hydrogen atoms [16].

The high degree of primary sequence identity between MtCMPK (target) and of E.coli indicates that these crystallography structures are good models to be used as templates for MtCMPK enzyme (target). The alignment of the MtCMPK (target) and E. coli CMPK is shown in Fig. 1 [17].

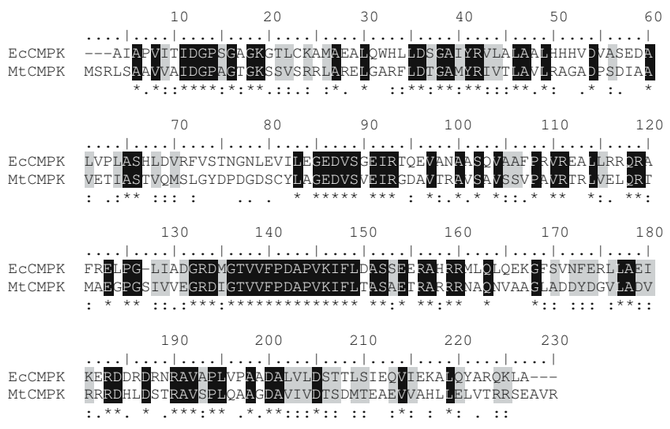


Fig. 1 Sequence alignment for EcCMPK and MtCMPK The sequence identity between EcCMPK and MtCMPK is □40%. The alignment was performed using ClustalW and edited with BioEdit [17]

A total of 1000 models were generated for each binary complex and the final models were selected based on the MODELLER [14] objective function.

Evaluation of binding affinity

Analysis of the interaction between a ligand and a protein target is still a scientific endeavor. The affinity and specificity between a ligand and its protein target depend on directional hydrogen bonds and ionic interactions, as well as on shape complementarity of the contact surfaces of both partners [18]. We used the programs X-SCORE [19], SCORE [20] and PEARLS [21] to evaluate the binding affinity of the ligand against E. coli CMPK and MtCMPK.

Analysis of the models

The overall stereochemical quality of the final models for each enzyme of the MtCMPK was assessed by the program PROCHECK [22, 23] and the objective function supplied of the MODELLER [14].

Molecular dynamic simulations

MD simulations were performed with the GROMACS [24] package using the Gromos 96.1 (43A2) force field. The CMP topology was generated with the PRODRG program [25]. Accurate force fields are essential for reproducing the conformational and dynamic behavior of condensed-phase systems, the Gromos 96.1 force fields are well parameterized for proteins but the parameters for small molecules are still limited for simulations of more complicated biological systems. Accordingly, GAMESS was used for the atomic charges in the CMP molecule [26], which were submitted to single-point ab initio calculations at RHF 6-31G* level in order to obtain Löwdin derived charges. Manipulation of structures was performed with Swiss-PDBViewer v3.7 program [27]. The first system was composed by apoenzyme MtCMPK (system 1) and the second by MtCMPK enzyme, three sulphate ions and CMP ligand (system 2). The simulations of two systems were performed by a time period of 3 ns. Na^+ counter ions were added to both systems (six Na^+ ions on system 1 and 14 Na^+ ions on system 2) using Genion Program of the GROMACS simulation suite to neutralize the negative charge density of the systems.

Each structure was placed in the center of a truncated cubic box filled with simple point charge (SPC/E) water molecules [28], containing 14,993 water molecules for system 1 and 16,741 water molecules for system 2. The initial simulation cell dimensions were 44.78×47.23×49.38 Å for system 1 and 51.37×48.76×55.23 Å for system 2, and had the protein solvated by a layer of water molecules of at least 10 Å length in all directions in both systems. During the simulations, bonds lengths within the proteins were constrained by using LINCS algorithm [29]. The SETTLE algorithm was used to constrain the geometry of water molecules [30]. In the MD protocol, all hydrogen atoms, ions, and water molecules were first subjected to 1000 steps of energy minimization by steepest descent followed by 500 steps of conjugate gradient to remove close van der Waals contacts. The systems were then submitted to a short molecular dynamic with position restrains for a period of 20 ps and afterwards performed a full molecular dynamics without restrains. The temperature of the system was then increased from 50 to 300 K in 5 steps (50 to 100 K, 100 to 150 K, 150 to 200 K, 200 to 250 K, 250 to 300 K), and the velocities at each step were reassigned according to the Maxwell-Boltzmann distribution at that temperature and equilibrated for 10 ps except the last part of termalization phase that were for 40 ps. Energy minimization and MD were carried out under periodic boundary conditions. The simulation was computed in the NPT ensemble at 300 K with the Berendsen temperature coupling and constant pressure of 1 atm with isotropic molecule-based scaling [31]. The LINCS algorithm, with a 10^{-5} Å tolerance, was applied to fix all bonds containing a hydrogen atom, allowing the use of a time step of 2.0 fs in the integration of the equations of motion. No extra restraints were applied after the equilibration phase. The electrostatic interactions between nonligand atoms were evaluated by the particle-mesh Ewald method [32] with a charge grid spacing of □ 1.0 Å and the charge grid was interpolated on a cubic grid with the direct sum tolerance set to 1.0×10^{-5}. The Lennard-Jones interactions were evaluated using a 9.0 Å atom-based cutoff [33].

All analyses were performed on the ensemble of system configurations extracted at 0.5-ps time intervals from the simulation and MD trajectory collection was initiated after 1 ns of dynamics to guarantee a completely equilibrated evolution. The MD simulation and results analysis were performed on a personal computer Intel Core 2 Duo E6300 - 1,86 GHz and 4 Gb RAM.

The convergences of the different simulations were analyzed in terms of the secondary structure, root mean-square deviation (RMSD) from the initial models structures, and root mean-square fluctuation (RMSF).

For the RMSFs were calculated relative to the last 2 ns averaged backbone structures, and all coordinate frames from the trajectories were first superimposed on the initial conformation to remove any effect of overall translation and rotation.

Results and discussion

Quality of the models

There is no crystallographic structure available for MtCMPK, however the sequence identity (38.70%) be-

Fig. 2 Tertiary structure of the apoenzyme MtCMPK. The structure is presented as a ribbon diagram. The structure contains ten alpha-helices surround by seven-stranded β-sheet. The image was generated using Pymol [22]

tween MtCMPK and E. coli CMPK (PDB access code: 1KDO) sequences makes E. coli CMPK structure a good template for modeling of MtCMPK. The atomic coordinates of crystallography structures of template were hence used as basic models for MtCMPK modeling. The atomic coordinates of all water molecules were removed from the templates.

The analysis of the Ramachandran diagram φ - ψ plots for the template (E. coli CMPK) was used to compare the overall stereochemical quality of the MtCMPK structure against those of the template solved by biocrystallography. The homology model has over 92.1% of the residues in the most favorable regions

Overall description

The structural model of the MtCMPK contains a seven beta-ribbon motif. The beta-ribbons are composed of residues 7–12, 33–37, 69–75, 78–84, 127–132, 145–150 and 202–206. Ten alpha - helices surround the beta - ribbon structure. The helical regions are composed of residues 19–30, 38–51, 58–66, 89–91, 95–105, 109–120, 134–138, 153–167, 173–189 and 211–225. Figure 2 shows a schematic drawing of the MtCMPK structure (monomer). As described in the M. tuberculosis H37Rv genome annotation [34], the model of MtCMPK consists of 230 amino acids with a predicted molecular mass of 24,177.29 Da and a theoretical pI value of 5.05. The E. coli CMPK consists of 223 amino acids with a predicted molecular mass of 24,746.34 Da and a theoretical pI value of 5.56. Analysis of both structures indicates the conservation of the residues that make intermolecular hydrogen bonds with cytidine-5′-monophosphate.

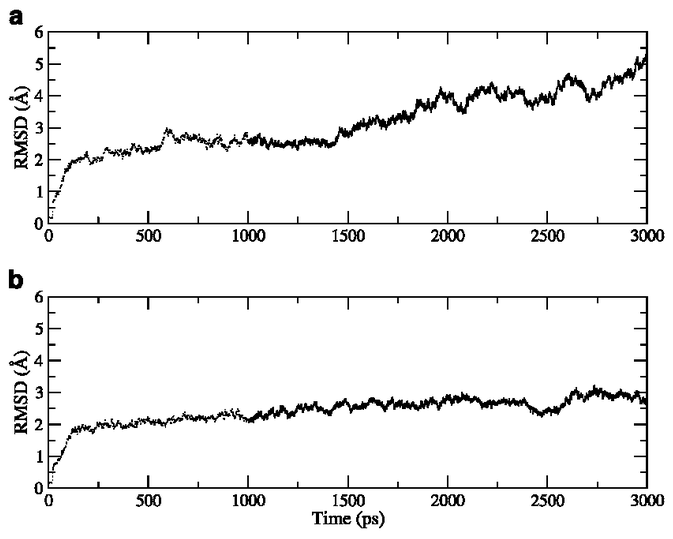


Fig. 3 Graphical representation of root-mean-square deviation (RMSD) of all Cα from starting structure of models as a function of time. The graphic a shows the RMSD of apoenzyme MtCMPK, b graphic shows the RMSD of MtCMPK-CMP complex. The dashed line gives the equilibration phase, the solid line shows the last 2 ns of calculation

Fig. 4 Superposition of the average during the last of 2000 snapshots with the initial minimized structure of MtCMPK-CMP. The structures are presented as ribbon diagram. The average structure is colored light gray; the initial structure is colored dark gray

MD simulations

We performed molecular dynamics simulations of the MtCMPK structure in the apo form (system 1) and the complex MtCMPK:CMP (system 2) to elucidate the influence of CMP on the overall structure of the MtCMPK. The root-mean square deviation (RMSD) of the positions for all backbone C-alpha atoms from their initial configuration as a function of simulation time for all systems were calculated and are shown in Fig. 3. Analysis of this figure indicates that the structure without ligand presents high RMSD when compared with the structure of the binary complex.

As shown in Fig. 3a system 1 presents an increase in RMSD values during overall MD simulation showing a relative stability between 250 and 1500 ps. On the other hand, system 2 after a rapid increase during the first 250 ps reaches a plateau of 2.6±0.2 Å value for RMSD of all C-alpha atoms. Figure 3b, suggests that that 2980 ps unrestrained simulation was sufficient for stabilizing a fully relaxed MtCMPK-CMP model. Accordingly, the MtCMPK-CMP binary complex structure appears to be more stable than the apo form of the enzyme.

Superposition of the average structure of system 2 with the initial model (Fig. 4) does not show major conformational changes from the initial model, which is consistent with the relatively low RMSD value. System 1 presents high RMSD values, however its secondary structure was kept (Fig. 5), and these high values are due to the flexibility of the apo form of MtCMPK.

The flexibilities of the proteins were assessed by the RMSF values from MD of the trajectory which reflects the flexibility of each atom residue in a molecule (Fig. 6). The major backbone fluctuation occurs in the loop region and in the region surrounding the beta-alpha-beta fold, whereas regions with the low RMSF correspond exclusively to the rigid beta-alpha-beta fold. These results indicate the stability of our model structures.

LID domain

The MtCMPK structure model as E. coli CMP kinase contains a short LID domain [12]. Shikimate Kinase from M. tuberculosis (MtSK), a member of NMP kinase family,

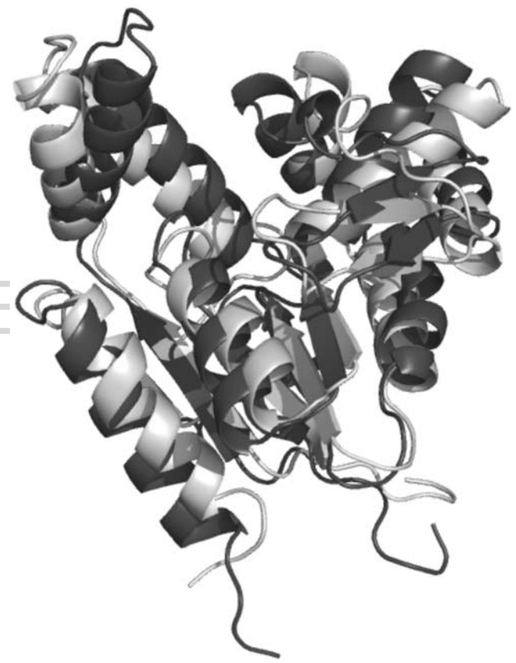


Fig. 5 Superposition of the average during the last of 2000 snapshots with the initial minimized structure of aporenzyme MtCMPK. The structures are presented as ribbon diagram. The average structure is colored light gray; the initial structure is colored dark gray

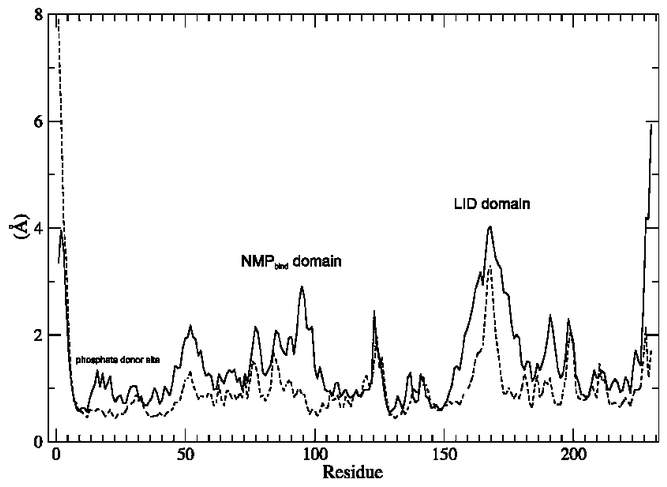


Fig. 6 Graphical representation of root-mean-square fluctuations (RMSF) of all Cα from starting structure of models as a function of time. The graphic shows the RMSF of apoenzyme MtCMPK and of MtCMPK-CMP complex. The average of the last 2 ns of calculation is given in the dashed line apoenzyme MtCMPK and the solid line shows MtCMPK-CMP complex

has three domains: the CORE, LID and NMP-binding domains. A comparison between the crystal structure of MtSK structure in complex with MgADP and Erwinia chrysanthemi SK suggested a concerted conformational change of the LID and shikimate binding (SB) domains upon nucleotide binding [35, 36]. More recently, a comparison between the structures of MtSK-MgADP-shikimate dead-end ternary complex and the MtSK-MgADP binary complex showed that the LID and SB domains undergo concerted movements toward each other [37], resulting in an additional closure of the enzyme active site. The RMSF values calculated for both systems described here suggest that a short LID domain in MtCMPK may undergo conformational changes as observed for the structure of MtSK (Fig. 6).

The LID of MtCMPK corresponds to the Arg160-Asp174 domain. The large mobility of LID domain is consistent with the need for this region to undergo conformational changes upon substrate binding to shield the enzyme active site from water in order to avoid ATP hydrolysis [38]. In the LID domain the behavior differences between system 1 and system 2 is evident and the LID domain appears in a more open conformation in apoenzyme than in MtCMPK-CMP (Fig. 7), demonstrating that the CMP induces a major stability of this region, as the substrate binding loop and is involved in substrate entrance and exit.

The NMP$_{bind}$ domain

The NMP$_{bind}$ domain has also been described as undergoing motions when it binds the phosphoryl acceptor substrate. The NMP$_{bind}$ domain of MtCMPK corresponds to the segment Gly38-Glu123. As GMP kinases from yeast, the MtCMPK has a two-stranded β sheet in the NMP$_{bind}$ domain [12], contrary to EcCMPK that contains a three-

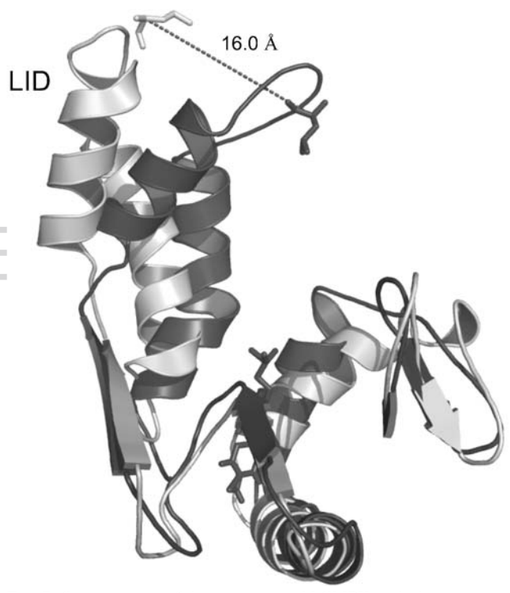


Fig. 7 Superposition of the apoenzyme MtCMPK with the structure MtCMPK-CMP showing the LID domain, the distance of Leu169 (C-alpha) between both structures is shown. The structure of the apoenzyme MtCMPK is colored light gray and MtCMPK-CMP structure is colored dark gray

stranded antiparallel β sheet [11]. The RMSD for the last 2 ns between MtCMPK and MtCMPK-CMP is 9.02 Å demonstrating a large motion in this region. These data are is in agreement with RMSF, which thus confirms the differences between ligand-free enzyme and enzyme in complex with CMP.

The sulphate ion bound in the phosphate donor site

MtCMPK contains the classical mononucleotide-binding motif: a β strand (β1 for MtCMPK) followed by a helix (α1) connected by a glycine-rich loop with a strongly conserved fingerprint sequence Gly-X-X-Gly-X-Gly-Lys (P loop), where X stands for any amino acid [8]. This motif forms a giant anion hole also known as P-loop. In MtCMPK, this motif corresponds to residues Gly13-X-X-Gly16-X-Gly18-Lys. The RMSF values indicate the high flexibility of helix α1 (Lys19-Leu30) and the phosphate donor site (Gly13-Gly18) in apo form. The presence of CMP causes a major stability and approximation of this structure when compared with its ligand-free model. It has been suggested that negatively charged ions with tetrahedral geometry (such as sulphate and phosphate ions) can inhibit kinases from binding nucleotides bearing a β-phosphate [39].

Interaction with cytidine-5′-monophosphate

The specificity and affinity between enzyme and its inhibitor depend on directional hydrogen bonds and ionic interactions, as well as on shape complementarity of the contact surfaces of both partners [40]. Analysis of the hydrogen bonds between cytidine-5′-monophosphate and MtCMPK reveals ten intermolecular hydrogen bonds and 14 with EcCMPK, the residues involved in the interaction with ligand are showed in Table 1. Analysis of the affinity constants between complex protein-ligand calculated by the program X-SCORE, SCORE and PERALS (Table 2) reveals that the pKd value of 4.93 calculated by SCORE program is more accurate as compared to the experimental value of 4.45 pKd units [41]. EcCMPK has higher affinity for CMP than MtCMPK, which is consistent with the larger number of hydrogen bonds for the latter as compared with the former.

Conclusions

The molecular models of MtCMPK show that the interaction with CMP is favorable as observed for EcCMPK. The results obtained are in agreement with experimental data [12] demonstrating that the LID domain, NMP_{bind} domain and phosphate donor site have low mobility when complexed with CMP and large motion in ligand-free form. Moreover, the data presented here suggest that the mode of action for MtCMPK may be similar to MtSK, with the open/closed conformational changes of the LID domain and the substrate binding loop involved in substrate entrance and exit.

Table 1 Intermolecular contacts of EcCMPK and MtCMPK with CMP

Residues/ligands	EcCMPK	Residues/ligands	MtCMPK
Met133	SD-N19 -> 3.67 Å	Arg133	NH2-O3 -> 3.01 Å
Arg41	NH2-O4 -> 3.50 Å		NE-N14 -> 3.13 Å
Arg181	NH1-O10 -> 2.96 Å		N-N19 -> 3.06 Å
Asp185	OD2-O10 -> 2.81 Å	Arg183	NE-O10 -> 2.99 Å
	OD1-O12 -> 2.63 Å	Asp187	OD2-O10 -> 2.85 Å
Arg110	NH1-O16 -> 2.73 Å		OD1-O12 -> 2.84 Å
	NH2-O16 -> 3.30 Å	Arg111	H2-O16 -> 3.09 Å
	NH2-N17 -> 3.14 Å	Arg190	NH1-O16 -> 2.96 Å
Arg188	NH1-O16 -> 2.87 Å		NH2-O16 -> 2.89 Å
	NH2-N17 -> 2.75 Å	Asp134	OD1-N19 -> 2.85 Å
Ser36	OG-N19 -> 2.82 Å		–
Arg131	N-N19 -> 3.05 Å		–
	NH2-O3 -> 3.48 Å		–
Asp132	OD1-N19 -> 3.16 Å		–
Total Number of Hydrogen Bonds	14		10

Table 2 pKds for EcCMPK and MtCMPK

	XSCORE	SCORE	PEARLS
EcCMPK	6.31 (4.45)	4.93 (4.45)	6.03 (4.45)
MtCMPK	5.96	4.69	5.14

In parentheses show the experimental value for EcCMPK complexed with CMP [35]

The results were generated using XSCORE [16], SCORE [17] and PEARLS [18]

JmlID 894_ArtID 291_Proc# 1 - 04/03/2008

Acknowledgements This work was supported by grants from Millennium Institutes (CNPq-MCT) to DSS, LAB and WFA. WFA, LAB and DSS are research career awardees from the National Research Council of Brazil (CNPq).

References

1. World Health Organization (2005) World health organization report: global tuberculosis control - surveillance, planning, financing. Geneva, Switzerland
2. Dye C (2006) Lancet 367:938–940
3. Pablos-Mendez A, Gowda DK, Frieden T (2002) Bulletin of the World Health Organization 80(6):489–500
4. CDC (Centers for Disease Control and Prevention) (2006) Morb Mortal Wkly Rep 55:301–305
5. WHO: Report WHO/HTM/TB/2006.375, (2006) Geneva
6. Dorman SE, Chaisson RE (2007) Nature 113:295–298
7. Singh JA, Upshur R, Padayatchi N (2007) PLoS Medicine 4(1): 50
8. Leipe DD, Koonin EV, Aravind L (2003) J Mol Biol 333:781–815
9. Liou JY, Dutschman GE, Lam W, Jiang Z, Cheng YC (2002) Cancer research 62:1624–1631
10. Yu L, Mack J, Hajduck PJ, Kakavas SJ, Saiki AYC, Lerner CG, Olejniczak ET (2003) Protein Sci 12:2606–2621
11. Dhaliwal B, Ren J, Lockyer M, Charles I, Hawkins AR, Stammers DK (2006) Acta Crystallogr F 62:710–715
12. Briozzo P, Golinelli-Pimpaneau B, Gilles AM, Gaucher JF, Burlacu-Miron S, Sakamoto H, Janin J, Bârzu O (1998) Structure 6:1517–1527
13. Kroemer RT, Doughty SW, Robinson AJ, Richards WG (1996) Protein Eng 9(6):493–498
14. Sali A, Blundell TL (1993) J Mol Bio 234(3):779–815
15. Sali A, Overington JP (1994) Protein Sci 3(9):1582–1596
16. Canduri F, Uchoa HB, de Azevedo WF Jr, (2004) Biochem Biophys Res Commun 324(2):661–666
17. Hall TA (1999) Nucleic Acids Symp Ser 41:95–98
18. de Azevedo WF, Mueller-Dieckmann JH, Schulze-Gahmen U, Worland PJ, Sausville E, Kim SH (1996) Proc Natl Acad Sci USA 93:2735–2740
19. Wang R, Lai L, Wang S (2002) J Comput-Aided Molecular Des 16:11–26
20. Wang R, Liu L, Lai L, Tang Y (1998) J Mol Model 4:379–394
21. Han LY, Lin ZR, Zheng CJ, Cao ZW, Xie B, Chen YZ (2006) J Chem Inf Model 46:445–450
22. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291
23. Uchoa HB, Jorge GE, Da Silveira NJF, Camera JC, Canduri F, De Azevedo WF (2004) Biochem Biophys Res Comum 325:1481–1486 DOI 10.1016/j.bbrc.2004.10.192
24. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) J Comp Chem 26 p. 1701–1718 http://www.gromacs.org
25. (a) van Aalten DMF, Bywater B, Findlay JBC, Hendlich M, Hooft RWW, Vriend GJ (1996) Comput Aided Mol Des 10, 255–262 (b) http://davapc1.bioch.dundee.ac.uk/programs/prodrg/prodrg.html
26. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA (1993) J Comput Chem 14: 1347–1363
27. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
28. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, The Netherlands, pp 331
29. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) J Comput Chem 18:1463
30. Miyamoto S, Kollman PA (1992) J Comput Chem 13:952
31. Chowdhuri S, Tan ML, Ichiye T (2006) J Chem Phys 125:144513
32. Darden T, York D, Pedersen LA (1993) J Chem Phys 98: 10089–10092
33. Norberto de Souza O, Ornstein RL (1999) J Biomol Struct Dyn 16:1205–1218
34. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG (1998) Nature 393(1998):537–544
35. Gu Y, Reshetnikova L, Li Y, Wu Y, Yan H, Singh S, Ji S (2002) J Molec Biol 319:779–789
36. Stehle T, Schultz GE (1992) J Mol Biol 224:1127–1141
37. Pereira JH, Oliveira JS, Canduri F, Dias MVB, Palma MS, Basso LA, Santos DS, Azevedo Jr WF (2004) Acta Cryst D 60:2310–2319
38. Jencks WP (1975) Adv Enzymol Relat Areas Mol Biol 43: 219–410
39. Gan J, Gu Y, Li Y, Yan H, Ji X (2006) Biochemistry 45:8539–8545
40. de Azevedo WF, Canduri F, Fadel V, Teodoro LG, Hial V, Gomes RA (2001) Biochem Biophys Res Commun 287(1):277–281
41. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) Nucleic Acids Res Jan 1:32, Database issue: D431–3

Mensagem Nº:	142 de 167
De:	PSFBeditor@jhu.edu
Para:	wfdaj@uol.com.br, walter.junior@pucrs.br
CC:	analuvivan@yahoo.com.br, rafael.caceres@pucrs.br,ramon@df.ibilce.unesp.br, borgesjc@iqsc.usp.br, joao@ibilce.unesp.br,cramos@lnls.br, carlos.ramos@pesquisador.cnpq.br, wfdaj@uol.com.br,walter.junior@pucrs.br, luiz.basso@pucrs.br, diogenes@pucrs.br
Data:	Mon, 11 Feb 2008 10:01:47 -0500 (EST)
Assunto:	**PROTEINS: Manuscript Prot-00276-2007.R2 Accepted**

PROTEINS: Structure, Function, and Bioinformatics
11-Feb-2008

Dear Prof. Walter de Azevedo Jr.:

Your manuscript entitled "Structural studies of prephenate dehydratase from Mycobacterium
tuberculosis H37Rv by SAXS, ultracentrifugation and computational analysis" has passed all
required peer review and has been recommended to me by the Editorial Board. I am pleased
to accept the paper for publication in the next available issue of PROTEINS.

Along with this e-mail you will receive another with instructions for production of your
article. I look forward to seeing it in press.

Congratulations on submitting such an excellent study.

Sincerely,

Eaton E. Lattman
Editor-in-Chief

PROTEINS: Structure, Function, and Bioinformatics
The Johns Hopkins University
Department of Biophysics
Baltimore, MD 21218 U.S.A.

EDITORIAL MANAGER'S COMMENTS TO AUTHOR:
Editorial Manager: 1
Comments to the Author:
(There are no comments)

REVIEWERS' COMMENTS TO AUTHOR:
Reviewer: 1
Comments to the Author
The manuscript is now readable.

Reviewer: 2
Comments to the Author
(There are no comments)

----- Original Message -----
From: Dunn,Ben M
To: Walter Filgueira de Azevedo Junior
Sent: Monday, November 19, 2007 5:22 PM
Subject: RE: Revised version of PPL

Dear Dr. Walter Filgueira de Azevedo Junior:

Your revised manuscript has been accepted and sent to the publisher for inclusion in an upcoming issue of Protein and Peptide Letters. You will receive a .PDF file as the page proofs. Please make any necessary correction and send back to the publisher without copying to me.

Best Wishes, Ben Dunn, E-i-C, PPL and CPPS

Ben Dunn

Distinguished Professor of Biochemistry & Molecular Biology

New email to be used in the future: bdunn@ufl.edu

Old email address with college.med will no longer work

Structural bioinformatics study of PNP from Listeria monocytogenes

Timmers LFSM[+], Caceres RA[+], Basso LA[+,‡], Santos DS[‡], and De Azevedo Jr WF[+].

[+] Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande Sul, Porto Alegre – RS, Brazil.

[‡] Centro de Pesquisas em Biologia Molecular e Funcional, Instituto de Pesquisas Biomédicas, Pontifica Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil.

Corresponding author:
Av. Ipiranga, 6681, CEP 90619-900, Porto Alegre, Rio Grande do Sul, Brasil,

walter.junior@pucrs.br

Telephone:

+55 51 33203500

Abstract

This work describes for the first time a model of Purine Nucleoside Phosphorylase from Listeria monocytogenes (LmPNP). We modeled the complexes of LmPNP with ligands in order to determine the structural basis for specificity. Comparative analysis of the model of LmPNP allowed identification of structural features responsible for ligand affinities.

Key-words: Purine Nucleoside Phosphorylase, molecular modeling, bioinformatics, drug design, Listeria monocytogenes**.**

Abbreviations

PNP – Purine Nucleoside Phosphorylase

HsPNP – Human Purine Nucleoside Phosphorylase

LmPNP – Listeria monocytogenes Purine Nucleoside Phosphorylase

RMSD – Root Mean Square Deviation

pI – isoeletric point

Da – Dalton

kDa – Kilodalton

Introduction

The genus Listeria comprises seven species, of which Listeria monocytogenes is an important human pathogen. L. monocytogenes is widespread in the environment and has been recovered from soil, water, vegetation, and animal products. L. monocytogenes is a gram-positive bacterium, and facultative intracellular parasite able to survive in macrophages and to invade a variety of normally non-fagocytic cells, such as epithelial cells, hepatocytes, and endothelial cells [1]. Listeria is easily engulfed by macrophages and phagocytes; however, this species is not only able to survive inside the macrophages – protected from other host's immune defenses – but it is also able to multiply intracellularly. The most important complication of infection is listeriosis that is now recognized as uncommon but serious infection primarily of neonates, pregnant women, and immunocompromised hosts. Immunocompromised individuals are particularly vulnerable to this intracellular pathogen [1]. Infection may also occur in healthy individuals.

Purine Nucleoside Phosphorylase (PNP) has been proposed as target for development of antibacterial drugs. PNP catalyzes the cleavage of N-ribosidic bonds of the purine ribonucleosides and 2-deoxyribonucleosides in the presence of inorganic orthophosphate as a second substrate. This reaction generates the purine base and ribose(deoxyribose)-1-phosphate [2]. PNP is specific for purine nucleosides in the β-configuration and cleaves the glycosidic bond with inversion of configuration to produce α – ribose – 1 – phosphate [3]. PNP is a ubiquitous enzyme of purine metabolism that functions in the salvage pathway, including those of Apicomplexan parasites

[4]. PNP is classified as belonging to the class I of Nucleoside Phosphorylase (NP-I) [5].

Drugs that inhibit human PNP activity have the potential of being utilized as modulators of the immunological system to treat leukemia, autoimmune diseases, and rejection in organ transplantation [6]. PNP has been submitted to intensive studies focused on the identification of new inhibitors, most of them related to human PNP.

In the present work we modeled the structure of PNP from Listeria monocytogenes (LmPNP). It was analyzed three binding sites present in the structures of human PNP and LmPNP. The analysis was carried out with different ligands in order to identify the structural basis for the specificity of different ligands against PNPs. Furthermore, the understanding of the rules governing the specificity of different ligands against PNP could be used to help in the design of more specific inhibitors. Moreover, an understanding of these rules should assist in the rational-based design of agents to treat Listeria monocytogenes infection.

Materials and Methods

Molecular modeling

Homology modeling is usually the method of choice when there is a clear relationship of homology between the sequence of a target protein and at least one experimentally determined three-dimensional structure. This computational technique is based on the assumption that tertiary structures of two proteins will be similar if their sequences were related, and it is the approach most likely to give accurate results [7].

For modeling of the LmPNP complexed with acyclovir, guanine, 7-methyl-6-thio-guanosine, 3-deoxyguanosine, guanosine and inosine the following crystallographic structures were used as templates, 1PWY [8], 1V2H [9], 1YRY [10], 1V45, 1RFG [11] and 1RCT [12], respectively. The web server PARMODEL was used to model the binary complexes [13]. PARMODEL is a parallelized version of the MODELLER [14]. The modeling procedure begins with alignment of the sequence to be modeled (target) with related known three-dimensional structures (templates). This alignment is usually the input to the program and the output is a three-dimensional model for the target sequence containing all main-chain and side-chain non-hydrogen atoms [15].

The high degree of primary sequence identity between LmPNP (target) and of Human PNP(HsPNP) indicates that these crystallography structures are good models to be used as templates for LmPNP enzyme (target). The alignment of the LmPNP (target) and human PNP is shown in **FIGURE (1)** [16].

```
10        20        30        40        50        60
....|....|....|....|....|....|....|....|....|....|....|....|
HsPNP  ENGYTYEDYKNTAEWLLSHTKHRPDVAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRSTV 60
LmPNP  ---MSLEKVNEAVAKIRESYNGTPKIGLILGSGLGVLADEVNNPTKLSYSEIPHFPVSTV 57
: *. :::.  : .  :  *::.:* ***** *:*::.:.  :.*****:** ***

70        80        90       100       110       120
....|....|....|....|....|....|....|....|....|....|....|....|
HsPNP  PGHAGRLVFGFLNGRACVMMQGRFHMYEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGLN 120
LmPNP  EGHAGQFVFGFLENKEVVAMQGRFHFYEGYSMQDVTFPVRVMKELGVEVLVVTNAAGGVN 117
****::*** *:.:  * ******:****.: .*******:: ***:.*********:*

130       140       150       160       170       180
....|....|....|....|....|....|....|....|....|....|....|....|
HsPNP  PKFEVGDLMLIRDHINLPGFSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQM 180
LmPNP  ELYSAGDLMLISDHIN---FTGTNPLIGPNDEHFGPRFPDMSEAYNLALRVDARLIAQEL 174
:..**:*** ****   *:* *** *****:** *** **:**: ::*  *    :::

190       200       210       220       230       240
....|....|....|....|....|....|....|....|....|....|....|....|
HsPNP  GEQRELQEGIYVMVAGPSFETVAECRVLQKLGADAVGMSTVPEVIVARHCGLRVFGFSLI 240
LmPNP  N--LTIREGVYAGFSGPTYETPAEIQMMRTLGADAVGMSTVPEVIIANHAGLRVLGISCI 232
.    ::**.*. .:***::** ** :::::.***************:*.*.****:*:* *

250       260       270       280
....|....|....|....|....|....|....|....|....|....
HsPNP  TNKVIMDYESLEK-ANHEEVLAAGKQAAQKLEQFVSILMASIPLPDKAS 288
LmPNP  TN---MAAGILDQPLSHEEVIETTDQVRSTFLQVVKAIVAKIS------ 272
**   *    *::  .* **: : .*. ..: *:*. ::*.*.
```

Figure 1: Sequence alignment for Human Purine Nucleoside Phosphorylase and PNP of Listeria monocytogenes. The multiple alignment was performed using ClustalW and edited with BioEdit [16].

A total of 1000 models were generated for each binary complex and the final models were selected based on stereochemical quality. All optimization process was performed on a Beowulf cluster with 16 nodes (BioComp, AMD Athlon XP 2100+, BioComp, Brazil).

Evaluation of binding affinity

Analysis of the interaction between a ligand and a protein target is still a scientific endeavor. The affinity and specificity between a ligand and its protein target depend on directional hydrogen bonds and ionic interactions, as well as on shape complementarity of the contact surfaces of both partners [17]. The program SCORE [18] was used to evaluate the binding affinity of the ligands against HsPNP and LmPNP. According to this method, the binding affinity of the ligand can be separated into individual atom contributions. Each ligand atom obtains a score, called the atomic binding score, indicating its role in the binding process. The program reads the structure, assigns atom types and parameters, performs the calculation, and gives the dissociation constant of the given protein–ligand complex. The computational results are fed into a text file in which the detailed information of each ligand atom, including the atomic binding score, is tabulated. These data were used to evaluate the correlation coefficient between the affinities against both PNPs to verify possible resemblance in the structural basis for specificity against these enzymes.

Analysis of the models

The overall stereochemical quality of the final models for each enzyme of the LmPNP was assessed by the program PROCHECK [19]**.** Atomic models were superposed using the program LSQKAB from CCP4 [20] and the intermolecular hydrogen bonds were assessed by the program LIGPLOT [21].

Results and Discussion

Quality of the models

Although there is no crystallographic structure available for LmPNP, the similarity between LmPNP and HsPNP sequence makes HsPNP structure a reasonable template for modelling of LmPNP. Furthermore, there are several binary complexes between human PNP and different ligands, which provide templates to model binary complexes of LmPNP against these ligands. The atomic coordinates of crystallography structures of templates were used as basic models for modeling of the LmPNP. The atomic coordinates of all waters were removed from the templates.

The analysis of the Ramachandran diagram $\phi - \psi$ plots for the templates (HsPNP) was used to compare the overall stereochemical quality of the LmPNP structures against those of templates solved by biocrystallography. The homology models present over 88.7% of the residues in the most favorable regions. The rmsd values for superposition of the substrate binding sites of human PNP and LmPNP are presented in **Table 1**.

Table 1: The amino acid identity, and rmsd value after superposition of HsPNP (template) and LmPNP (target).

Ezyme	PDB access code	Resolution (Å)	References	rmsd (Å) Site / Protein
PNP	1PWY	2,8	[8]	0.09 / 0.39
	1RCT	2,8	[12]	0.11 / 0.46
	1RFG	2,9	[11]	0.10 / 0.48
	1V45	2,86	---	0.12 / 0.43
	1V2H	2,7	[9]	0.10 / 0.39
	1YRY	2,8	[10]	0.10 / 0.40

Overall description

The structural model of LmPNP indicates that it belongs to the class α/β, consisting of a mixed β–sheet and helices. It presents a distorted β-barrel surrounded by α-helices. The structure contains an eight-stranded mixed β—sheet and a five-stranded mixed β-sheet, which join to form a distorted β-barrel. The residues making up the eight-stranded sheet are 23 – 27, 42 – 44, 63 – 69, 72 – 78, 105 – 116, 124 – 132, 179 – 186, and 225 – 233. The five-stranded sheet consists of residues 112 – 116, 129 – 132, 183 – 186, 206 – 209, and 231 – 233. Seven α-helices surround the beta-sheet structure. The α-helices are composed of residues 3 – 16, 32 – 36, 89 – 102, 161 – 174, 194 – 203, 213 – 222, and 251 – 270. **Figure (2)** [22] show schematic drawings of the LmPNP structure (monomer). **Figure (3)** shows a representation of the active-site of LmPNP. The binary complexes were modeled with six ligands (guanine, guanosine, 3-deoxyguanosine, 7-methyl-6-thio-guanosine, inosine and acyclovir). The high identity between human HsPNP and LmPNP (~51%) classifies LmPNP as a member of the Nucleoside Phosphorylase-I group, more specifically as Low-molecular-mass (low-mm) homotrimers, with M_r~ 80 – 100 kDa, specific for catalysis of 6 – oxopurines and their nucleosides.

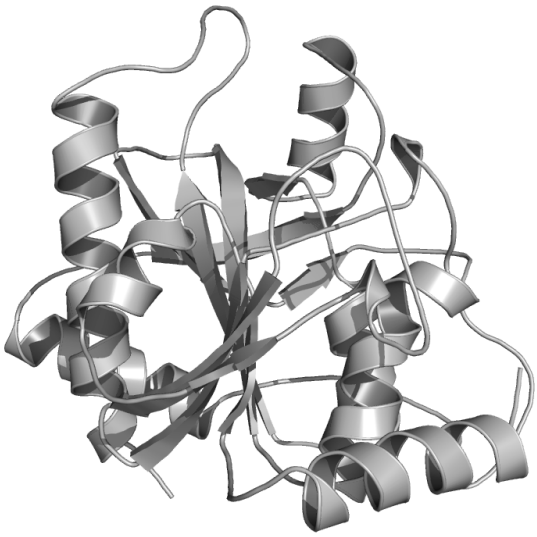


Figure 2: Tertiary structure of the LmPNP. The structure is presented as ribbon diagram. The structure contains an eight-stranded mixed β—sheet a five-stranded mixed β-sheet, which join to form a distorted β-barrel. The image was generated using Pymol [22].

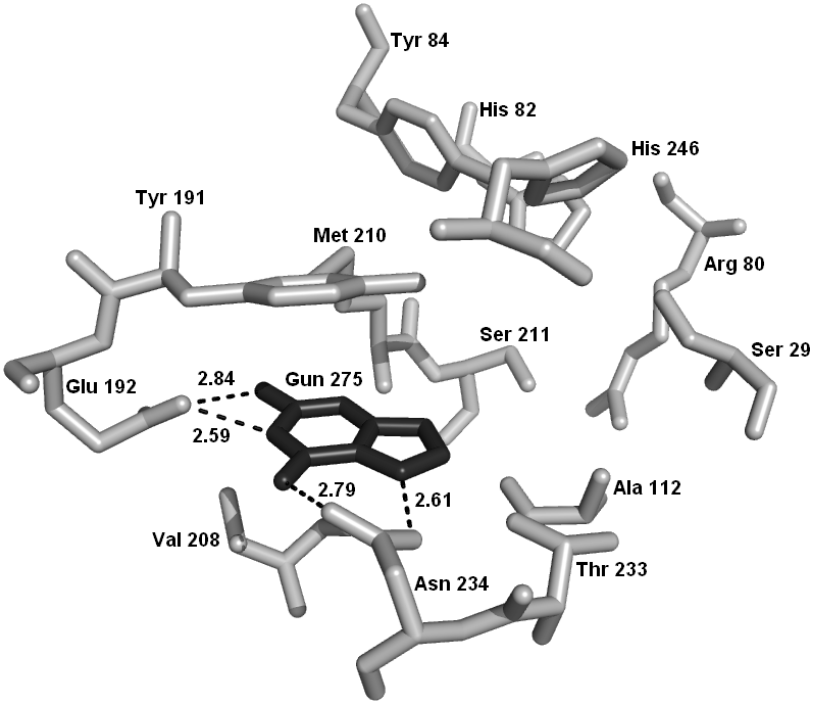


Figure 3: Active site of the LmPNP with 3-deoxyguanosine. The residues are presented in light gray and the inhibitor (3-deoxyguanosine) is presented in dark gray. The image was generate using Pymol [22].

The LmPNP (EC 2.4.2.1) consists of 272 amino acids with a molecular weight of 29485.6 Da and a theoretical pI of 4.86. The HsPNP consists of 288 amino acids with a molecular weight of 32016.7 Da and a theoretical pI of 6.50. Analysis of the structure of both PNPs indicates that, despite conservation of the ribose-binding site, three mutations were observed in other sites, one in the purine-binding site and two in the second phosphate-binding site. Phe200, Arg148 and Gln144 in the human PNP [23] were replaced with Tyr191, Ile141 and Thr137 in LmPNP, respectively, these differences are shown in **Table 2** (HsPNP) and **Table 3** (LmPNP). Furthermore, superposition of the LmPNP onto HsPNP indicates that a small region involving residues 249-252 presents a small helical region not observed in the HsPNP structure (residues 260-263) **Figure (4)** shows the superposition of LmPNP against HsPNP. The sequence alignment indicates that this region is not conserved in both PNPs, as shown in **Figure (1)**. Analysis of the propensities to form helix using Chow and Fasman scale [24] was not conclusive, it indicates only a slightly higher propensity to form turn in the HsPNP sequence. Molecular dynamics simulations may contribute to gain further insights on the stabilization of this region.

Table 2: Intermolecular contacts of HsPNP with ligands.

Residues/Ligands	Guanine	Guanosine	3-deoxyguanosine	7-methyl-6-thio-guanosine	Acyclovir	Inosine
Glu201	OE2-N1 -> 2.46 Å	OE2-N1 -> 2.49 Å	OE2-N1 -> 2.78 Å	OE2-N1 -> 2.91 Å	OE2-N1 -> 3.03 Å	OE2-O6 -> 3.33 Å
	OE2-N2 -> 2.80 A	OE1-N2 -> 2.63 A	OE1-N2 -> 2.75 A	OE1-N2 -> 2.70 A	OE1-N2 -> 2.55 A	
Asn243	ND2-O6 -> 2.81 A	ND2-N7 -> 3.16 A	ND2-O6 -> 3.34 A	---	ND2-N7 -> 2.91 A	ND2-O6 -> 2.87 A
		ND2-O6 -> 2.89 Å	OD1-N7 -> 2.78 Å		ND2-O6 -> 2.65 Å	ND2-N7 -> 3.30 Å
His257	---	---	ND1-O5` -> 2.80 A	---	---	ND1-O5* -> 2.99 A
Tyr88	---	OH-O3* -> 3.28 A	---	OH-O3* -> 2.76 A	---	OH-O3* -> 3.03 A
Ala116	---	O-O2* -> 3.19 A	---	---	---	O-O2* -> 3.42 A
Met219	---	N-O2* -> 3.08 A	N-O2` -> 3.07 A	N-O2* -> 3.05 A	---	N-O2* -> 3.05 A

Table 3: Intermolecular contacts of LmPNP with ligands.

Residues/Ligands	Guanine	Guanosine	3-deoxyguanosine	7-methyl-6-thio-guanosine	Acyclovir	Inosine
Asn234	OD1-N7 -> 2.61 Å	ND2-O6 -> 2.89 Å	OD1-N7 -> 2.80 Å		ND2-N7 -> 2.89 Å	ND2-O6 -> 2.91 Å
	ND2-O6 -> 2.79 Å	ND2-N7 -> 3.17 Å	ND2-O6 -> 3.31 Å	---	ND2-O6 -> 2.75 Å	ND2-N7 -> 3.27 Å
Glu192	OE2-N1 -> 2.59 A	OE2-N1 -> 2.60 A	OE2-N1 -> 2.79 A	OE2-N1 -> 2.96 A	OE2-N1 -> 2.99 A	OE2-O6 -> 3.28 A
	OE2-N2 -> 2.84 Å	OE1-N2 -> 2.91 Å	OE1-N2 -> 2.95 Å	OE1-N2 -> 2.81 Å	OE1-N2 -> 2.67 Å	
Tyr191	---	OH-O5* -> 2.58 Å	OH-O5' -> 2.44 Å	OH-O5* -> 2.62 Å	---	---
Met210	---	N-O2* -> 3.12 Å	N-O2' -> 3.04 Å	N-O2* -> 3.09 Å	---	N-O2* -> 3.08 Å
Tyr84	---	OH-O3* -> 3.21 Å	---	OH-O3* -> 2.89 Å	---	OH-O3* -> 3.06 Å
Ala112	---	O-O2* -> 3.18 Å	---	---	---	O-O2* -> 3.39 Å
His246	---	---	ND1-O5` -> 2.79	---	---	ND1-O5* -> 2.99 Å

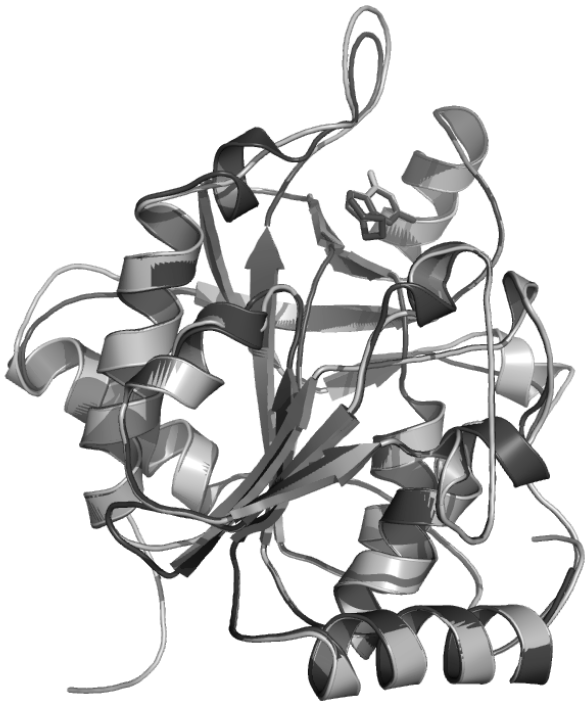


Figure 4: Superposition of complex HsPNP-acyclovir (light gray) against the complex LmPNP-acyclovir (dark gray). The image was generated using Pymol [22].

Differences between unligand PNPs

The superposition of apo-HsPNP on apo-LmPNP structure showed small differences, which are observed in the region presented in **Figure (5)**. Analysis of the propensities to form alpha-helix by Chou and Fasman method [24] in both PNPs is not conclusive for this conformation change present in the LmPNP apoenzyme, as observed for the region (133-136), previously discussed.

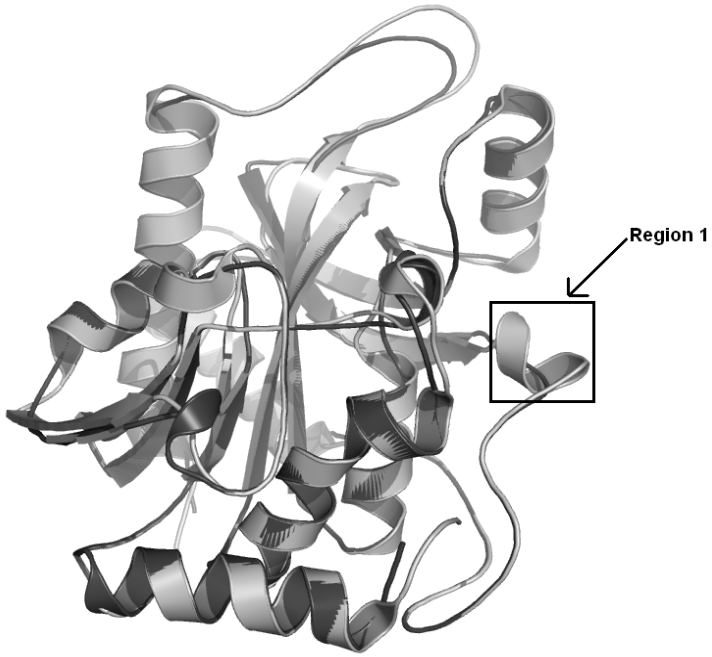


Figure 5: Superposition of HsPNP apoenzyme (light gray) against the LmPNP apoenzyme (dark gray). Region 1 indicates the conformation change. The image was generated using Pymol [22].

Differences in the apoenzyme and the binary complexes

In order to evaluate the possible differences in the LmPNP structure due to ligand binding we superposed the structure of the apoenzyme, modeled using HsPNP as template (PDB access code: 1M73) [23], against the six binary complexes. The rmsd for C-alpha superposition range from 0.73 Å to 0.96 Å. In all binary complexes the highest differences are observed in two regions, region 1 from 57 to 59 and region 2 from 249 to 252. The region 1 presents large conformation differences in all six binary complexes, and region 2 presents the highest differences only for complex LmPNP-acyclovir. This all together indicates the flexibility of these two regions which allows different

ligands to bind to the PNP active site. The highest rmsd values are observed for the PNP-3-deoxyguanosine complex 0.96 Å. This region undergoes a conformational change upon ligand binding. **Figure (6)** illustrates the superposition of LmPNP apoenzyme against the complex LmPNP-3-deoxyguanosine, which clearly shows see both regions.

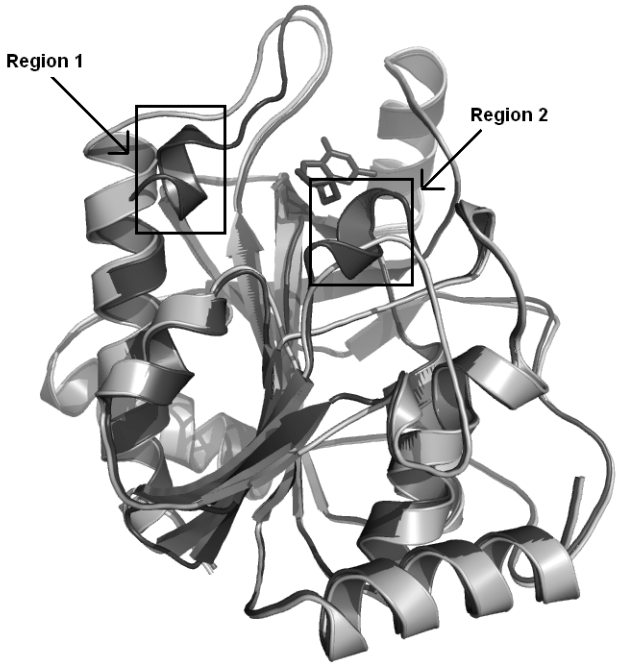


Figure 6: Superposition of LmPNP apoenzyme (light gray) against the LmPNP-3-deoxyguanosine complex (dark gray). Regions 1 and 2 indicate the conformational changes due to ligand binding. The image was generated using Pymol [22].

Interactions with ligands

The binding affinity for a protein-ligand complex for human PNP and LmPNP were calculated with the program SCORE and the results are shown in **Table 4**. Correlation coefficient between the affinity constants for HsPNP and LmPNP is 0.64. The affinity of the six ligands against HsPNP and LmPNP strongly indicates that there is high correlation between the affinities against

both PNPs. **Figures (7A-D)** show the intermolecular contacts between enzyme and ligands for complexes with highest affinity constants. The ligand inosine presents the highest affinity against both PNPs. **Table 2** and **Table 3** shows the intermolecular hydrogen bonds for all complexes studied in this article. We can clearly see that the intermolecular hydrogen bonds involving Glu201/Glu191 and Asn243/Asn233 are conserved in all complexes.

Table 4: pKds for HsPNP and SpnPNP. The results were generated using SCORE [18].

PNPs / Ligands	Acyclovir	Guanine	Guanosine	3-deoxyguanosine	7-methyl-6-thio-guanosine	Inosine
HsPNP	4.04	5.22	4.92	5.76	3.44	4.39
LmPNP	3.57	4.52	5.31	5.11	4.38	4.11

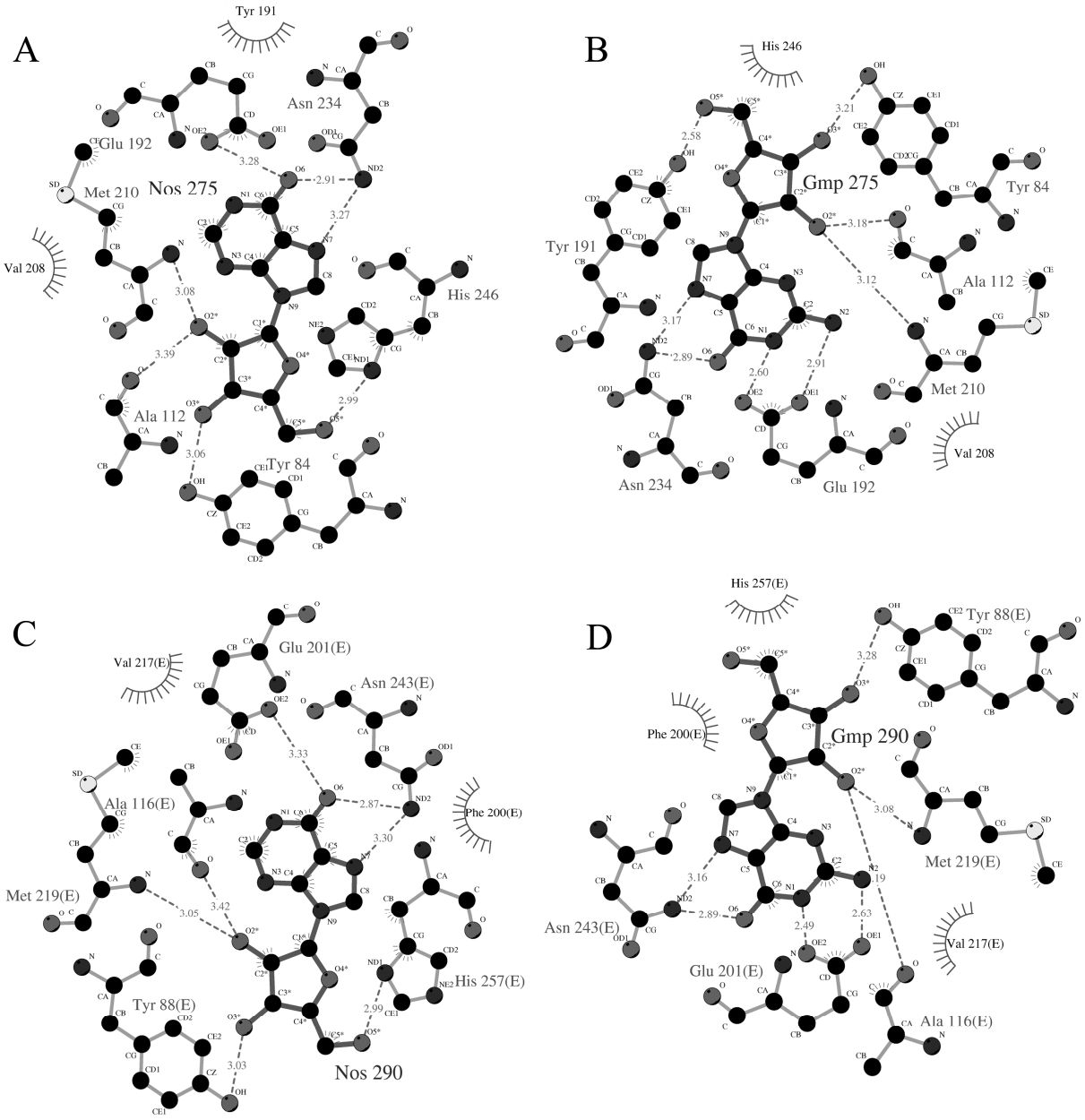


Figure 7: Intermolecular contacts of binary complexes. (A) LmPNP with inosine, (B) LmPNP with guanosine, (C) HsPNP with inosine, (D) HsPNP with guanosine. The images were generate using Ligplot [21].

Conclusions

Structural analyses of human PNP and LmPNP have shown three mutations in the binding-pocket, one in the purine-binding site and two in the second phosphate-binding site. These mutations appear to have small influences in the affinity constants of both PNPs. Furthermore, computationally determined affinity constant of the inhibitor 3-deoxyguanosine against LmPNP and HsPNP strongly indicates that this inhibitor presents high affinity against LmPNP, suggesting a possible new lead compound against Listeria monocytogenes. Future kinetics experiment may confirm this prediction.

Acknowledgments

Financial support for this work was provided by Millennium Initiative Program MCT-CNPq, Ministry of Health-Department of Science and Technology (Brazil) to D.S.S., W.F.A. and L.A.B. D.S.S. (CNPq, 304051/1975-06), L.A.B. (CNPq, 520182/99-5), W.F.A. (CNPq, 300851/98-7) are research career awardees from the National Council for Scientific and Technological Development of Brazil. RC thanks Hewlett-Packard of Brazil for a scholarship.

References:

[1] O'Neil HS, Marquis H. (2006) Infect Immun., 74(12), 6675-81.

[2] Kalckar HM (1947). J. Biol. Chem., 167 (2), 429-443.

[3] Canduri F, Fadel V, Basso LA, Palma MS, Santos DS, de Azevedo Jr. WF (2005). Biochem. Biophys. Res. Commun., 327, 646-649

[4] Bzowska A, Kulikowska E, Shugar D (2000). Pharmacol. Ther., 88, 349-425.

[5] Pugmire MJ, Ealick SE (2002). Biochem. J., 361(Pt 1), 1-25.

[6] Silva RG, Nunes JES, Canduri F, Borges JC, Gava LM, Moreno FB, Basso LA, Santos DS (2007). Current Drugs Target, 8, 413-422.

[7] Kroemer RT, Doughty SW, Robinson AJ, Richards WG (1996). Protein Eng., 9 (6), 493-498.

[8] dos Santos DM, Canduri F, Pereira JH, Dias MVB, Silva RG, Mendes MA, Palma MS, Basso LA, de Azevedo WF, Santos DS (2003). Biochem. Biophys. Res. Commun., 308(3), 553-559.

[9] de Azevedo WF, Canduri F, dos Santos MD, Pereira JH, Dias MVB, Silva RG, Mendes MA, Basso LA, Palma MS, Santos DS (2003). Biochem. Biophys. Res. Commun., 312(3), 767-772.

[10] Silva RG, Pereira JH, Canduri F, de Azevedo Jr. WF, Basso LA, Santos DS (2005). Arch. Biochem. Biophys., 442, 49-58.

[11] Canduri F, Silva RG, Dos Santos DM, Palma MS, Basso LA, Santos DS, de Azevedo WF (2005). Acta Crystallogr. D Biol. Crystallogr., 6, 856-862.

[12] Canduri F, dos Santos DM, Silva RG, Mendes MA, Basso LA, Palma MS, de Azevedo Jr. WF. D.S. Santos (2004). Biochem. Biophys. Res. Comm., 313, 907-914.

[13] Uchoa HB, Jorge GE, Da Silveira NJF, Camera JC, Canduri F, De Azevedo WF (2004). Biochem. Biophys. Res. comum., 325, 1481-1486.

[14] (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Bio., 234(3), 779-815.

[15] Canduri F, Uchoa HB, de Azevedo Jr. WF (2004). Biochem Biophys Res Commun., 324(2), 661-6.

[16] Hall TA (1999). Nucleic Acids Symp. Ser., 41, 95-98.

[17] De Azevedo WF, Mueller-Dieckmann JH, Schulze-Gahmen U, Worland PJ, Sausville E, Kim SH (1996). Proc. Natl. Acad. Sci. USA., 93, 2735-2740.

[18] Wang R, Liu L, Lai L, Tang Y (1998). J. Mol. Mod., 4, 379-394.

[19] Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993). J. Appl. Crystallogr., 26, 283-291.

[20] Bailey S (1994). Acta Crystallogr. D Biol. Crystallogr. 50: 760-763.

[21] Wallace AC, Laskowski RA, Thornton JM (1995). Protein Eng, 8(2), 127-34.

[22] Delano WL, Lam JW (2005). Abstr. Pap. Am. Chem. Soc., 230: 1371-1372.

[23] Azevedo WF, Canduri F, Dos Santos DM, Silva RG, Oliveira JS, Carvalho LPS, Basso LA, Mendes MA, Palma MS, Santos DS (2003). Biochem. Biophys. Res. Commun., 308, 545-552.

[24] Chou PY, Fasman GD (1978). Adv Enzymol Relat

45-148.

8. REFERÊNCIAS

[1] Hanahan D. and Weinberg R. A. **The Hallmarks of Cancer**. *Cell*, 2000(100):57-70.

[2] Seth A. and Watson D. K. **ETS transcription factors and their emerging roles in human cancer**. *Eur J cancer*, 2005(41):2462-2478.

[3] Sementchenko V. I. And Watson D. K. **Ets target genes: past, present and future**. *Oncogene*, 2000(19):6533-6548.

[4] Sharrocks, A. D. **The ETS-domain transcription factor family**. *Nature Reviews – Molecular Cell Biology*. 2001(2):827-837.

[5] Garvie, C.W., Hagman, J., Wolberger, C. **Structural studies of Ets-1/Pax5 complex formation on DNA.** *Mol.Cell* . 2001(8):*1267-1276.*

[6] Malik, S & Roeder, R. G. **Dynamic regulation of pol II transcription by the mammalian mediator complex**. *TRENDS in Biocem Sci*. 2005(5)256-263.

[7] Sims, R. J., Mandal, S. S. & Reinberg, D. **Recent highlights of RNA-polimerase-II_mediated trasncription**. *Current Opinion in Cell Biology*. 2004(16)263-271.

[8] Oikawa, T. & Yamada, T. **Molecular biology of the Ets family of transcription factors**. *Gene*. 2003(303):11-34.

[9] Sato, Y. **Role of Ets Family Transcription Factors in vascular Development and Angiogenesis**. *Cell Structure and Function*. 2001(26)19-24.

[10] Verger, A. & Duterque-Coquillaud, M. **When ETS trasncription factors meet their partners**. *BioEssays*. 2002(24):362-370.

[11] Wasylyk, B; Hagman, J & Gutierrez-hartmann, A. **ETS transcription factors: nuclear effectors of the Ras-MAP-kinase signalin pathway**. *TIBS*. 1998(23):213-216.

[12] Hsu, T.; Trojanowska, M & Watson, D. K. **ETS proteins in biological control and cancer**. *Journal of Celular Biochemistry*. 2004(91):896-903.

[13] Oikawa, T. **ETS transcription factors: Possible targets for cancer therapy**. *Cancer Sciences*. 2004(8):626-633.

[14] Wang, Y.; Feng, L.; Said, M.; Balderman, S.; Fayazi, Z.; Liu, Y.; Ghosh, D. & Gulick, A. **Analysis of the 2.0 Á crystal structure of the protein-DNA complex of the human PDEF ETS domain bound to the prostate specific ntigen regulatory site**. *Biochemistry*. 2005(19):7095-7106.

[15] Gromihs, M. M.; Sibers, J. G.; Selvaraj, S.; Kono, H. & Sarai, A. **Intermolecular and intramolecular readout mechanisms in protein-DNA recognition.** *Journal of Molecular Biology*. 2004(337):285-294.

[16] Szymczyna, B. R. & Arrowsmith, C. H. **DNA binding specififity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition.** *The Journal of Biological Chemistry*. 2000(37):28363-28370.

[17] Ogden, T. H. & Rosenberg, M. S. **Alignment and Topological Accuracy of the Direct Optimization approach via POY and Traditional Phylogenetics via ClustalW + PAUP*.** *Syst. Biol.* 2007(2):182-193.

[18] Hall, T. A.**BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symposium Series.* 1999(41):95-98.

[19] Caceres, R. A.; Nunes, C. P.; de Azevedo, W. F.; Basso, L. A. & Santos, D. S. **Molecular dynamics and structural studies of the ETS domain–DNA complexes.** *Current Bioinformatics.* 2007 (*submitted*)

[20] M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. **The Protein Data Bank.** *Nucleic Acids Research,* 2000:235-242.

[21] Mo, Y., Vaessen, B., Johnston, K., Marmorstein, R. **Structure of the elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA.** *Nat.Struct.Biol. 2000(7):*292-297.

[22] Uchôa, H. B.; Jorge G. E.; da Silveira, N. J.; Câmera Jr. J. C.; Canduri, F.; de Azevedo, W. F.; **Parmodel: a web server for automated modeling of proteins**. *Biochem. and Biophys. Research Communications.* 2004(325):1481-1486.

[23] Sali, A. & Blundell T. L. **Comparative protein modeling by saisfaction of spatial restrains.** *J. Mol. Biol.* 1993(234):779-815.

[24] Fiser, A.; Do, R. K.; Sali, A . **Modeling o loops in protein structures.** *Protein Sci*. 2000(9):1753-1773.

[25] Bourne, P. E. & Weissig, H. **Structural Bioinformaics**. Wiley-Liss, 2003.

[26] Garvie, C.W., Pufall, M.A., Graves, B.J., Wolberger, C. **Structural Analysis of the Autoinhibition of ETS-1 and its Role in Protein Partnerships.** *J.Biol.Chem.* 2002(277):*45529-45536.*

[27] Bowie, J. U.; Luthy, R.; Eisenberg, D. **A method to identify protein sequences that fold into a known three-dimensional structure**. *Scienci*. 1991(25):164-170.

[28] Luthy, R.; Bowie, J. U.; Eisnberg, D. **Assessment of protein moels with three-dimensional profiles.** *Nature*. 1992(356):383-385.

[29] Marti-Renom, M. A.; Madhusudhan, M. S.; Sali, A. **Alignment protein sequences by their profiles.** *Protein Sci.* 2004(13):1071-1087.

[30] Berendsen, H. J. C.; Van der Spoel, D.; Van Drunen, R. **GROMACS: a message-pssing parallel molecular dynamics implementation**. *Comp. Phys. Commun.* 1995(91):43-56.

[31] Lindahl, E.; Hess, B.; Van der Spoel, D. **GROMACS 3.0: a package for molecular simulation and trajectory** analysis. *J. Mol. Mod.* 2001(7):306-317.

[32] Chen, R.; Li, L.; Weng, Z. **ZDOCK: an initial-stage protein-docking algorithm.** *Poteins.* 2003(52):80-87.

[33] Fanelli, F. & Ferrari, S. **Prediction of MEF2A-DNA interface by rigid body docking: A tool for fast estimation of protein mutational effects on DNA binding.** *J. Structural Biology.* 2006(153):278-283.

[34] Wallace A. C., Laskowski R. A. & Thornton J. M. **LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions.** *Prot. Eng.* 1995(8):127-134.

[35] Guex, N & Peitsch, M. C. **SWISS_MODEL and the Swiss-Pdb Viewer: na enviroment for comparative protein modeling.** *Electrophoresis.* 1997(18)2714-2723.

[36] deLano, W. L. **The PyMOL Molecular Graphics System.** *DeLano Scientific,* San Carlos, CA, USA, 2002. [http://www.pymol.org]

[37] Humphrey, W.; Dalke, A. & Schulten, K. **VMD – Visual Molecular Dynamics.** *J. Mol. Graphics.* 1996(14):33-38.