

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E EVOLUÇÃO**

**CLONAGEM E CARACTERIZAÇÃO MOLECULAR
DA FOSFORRIBOSIL PIROFOSFATO SINTASE
(PRPP sintase) DE CANA-DE-AÇÚCAR**

Susana Andréa Sculaccio

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Genética e Evolução do Centro de Ciências Biológicas e da Saúde da Universidade Federal de São Carlos como parte dos requisitos para obtenção do título de Mestre em Genética e Evolução, área de concentração: Genética e Evolução.

SÃO CARLOS – SP

2002

**CLONAGEM E CARACTERIZAÇÃO MOLECULAR
DA FOSFORRIBOSIL PIROFOSFATO SINTASE
(PRPP sintase) DE CANA-DE-AÇÚCAR**

Susana Andréa Sculaccio

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

S473cc

Sculaccio, Susana Andréa.

Clonagem e caracterização molecular da fosforribosil pirofosfato sintase (PRPP sintase) de cana-de-açúcar / Susana Andréa Sculaccio. -- São Carlos : UFSCar, 2003. 128 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2002.

1. Biologia molecular. 2. Clonagem. 3. Biosíntese. 4. Purinas. 5. Fosforribosil pirofosfato. I. Título.

CDD: 574.88 (20ª)

ORIENTADOR

Prof. Dr. Otavio Henrique Thiemann

**"... O ACASO VAI ME PROTEGER
ENQUANTO EU ANDAR
DISTRÁÍDO...".**

Titãs

A minha mãe

Ana Maria Moraes Sculaccio

Com todo carinho e amor

*Ao meu pai Fernando, aos meus irmãos
Edson, Joselaine, Naiage e a caçulinha
Tatiane. Dedico também aos meus avós
Limpio e vó Landa que já partiram e aos
meus avós Salvador e Vó da Zia.*

AGRADECIMENTOS

À Universidade Federal de São Carlos e ao Departamento de Genética e Evolução.

À Universidade de São Paulo e ao Instituto de Física da USP de São Carlos.

Ao meu orientador e amigo Prof. Dr. Otavio Henrique Thiemann por todo o apoio, paciência, orientação e principalmente por ter me aceitado como sua aluna.

Ao grupo de Biofísica da USP e a todos os professores e funcionários que dele fazem parte.

Ao grupo de Cristalografia, ao Prof. Dr. Richard, Prof. Dr. Glaucius.

Aos meus colegas e amigos que contribuíram para meu desenvolvimento profissional e principalmente com meu crescimento como ser humano.

A Prof^a e amiga Leila por ter me oferecido todas as chances de crescer e principalmente por ter aberto as portas para meu desenvolvimento profissional, principalmente por acreditar em mim e confiar sem nem ao menos me conhecer.

À minha grande Bel que me adotou com todo seu carinho e que tem um lugar especial na minha vida. Obrigada por ter me ensinado meus primeiros passos.

À minha amiga Andressa por todos os anos de companheirismo, pelas várias noites de filosofia e gandaia. Com certeza sempre estará presente, sendo praticamente uma irmã.

À minha amiga Adelise por todos os “poucos” anos de amizade que dispensam qualquer comentário e é lógico não posso esquecer sua irmã e minha amiga Andréa, afinal de contas elas nasceram juntinhas e até hoje comigo estão.

Aos meus cunhados Cleyton e Andréa por toda a ajuda, carinho e preocupação comigo e com minha família. A todos os meus tios (as), primos (as) e agregados pelo amor que dedicaram a mim. Ao Seu Chico e a Marlene por toda ajuda e carinho oferecidos.

Às minhas amigas de república: Zildinha, Juliana, Fernanda, Mariana, Anelise e Daiane, cada qual com seu significado especial que me ensinaram a viver e a sentir como a vida pode ser boa e que se tornaram minha família. Obrigada amiguinhas.

A Ritinha da Edusp, onde passei grande parte da graduação recebendo apoio, puxões de orelha e várias outras coisas que só nós sabemos, ao Jairão que fez a felicidade da Ritinha e por tabela a minha. Agradeço também a minha amiga Denise companheira das companheiras.

Ao Zé e a Carol que para mim sempre serão o casal 20. Ao Ricardinho e a Fabi.

Ao meu amigo Cícero com carinho e com aquele abraço apertado.

À Mel por todos os Caasos da vida, que dispensam qualquer comentário. Valeu!

Ao meu grande amigo Junior que com certeza seja na Alemanha ou no Brasil sempre terá meu carinho e admiração.

Às secretárias e amigas Jaciara e Sueli; a Claudia, Maria Helena, Ester e ao Fernando. Às secretarias da Genética, em especial a Regiane.

Aos técnicos e amigos Beto, Augusto, Eli e Mineiro. Ao Gustavo por ter me agüentado e me ajudado com o computador.

À Prof^a Ana Paula, Prof. Otaciro e ao Prof. Flávio.

À Bianca, Dulce, Sandra Cruz, Maria, Raquel, Teresa e Betinha por toda ajudam, companheirismo e alegrias. Vocês são muito especiais. Não poderia me esquecer da Wanda que também teve grande participação.

Às minhas amigas Cer e Lir e obrigada por permitirem que eu fosse a Sur.

Aos sujeitinhos sem qualidade Marcão, Wanius e Ney.

Ao grande Antônio e ao João Renato por todo carinho, pelos momentos de alegria, bandejões e ao João pela ajuda com minhas figuras.

Ao meu amigo Diet, meu exemplo de pessoa de paz.

Ao meu amigo PH por toda ajuda moral nas noites que passei escrevendo.

Ao meu amigo Alécio por todas as conversas e desabafos.

Ao meu amigão Leandro por tudo, principalmente pela aventuras de carro no caminho da federal.

Ao meu amigo Artur pelos incentivos e ao Peninha (Valeu!)

Ao Single (Humberto), Castilho, André Ambrósio, Stefano, Pavão, Lucas, Monique, Calgaro, Andréinha, Jabah, Cristininha, João Possato, Aparecido, Rosa, Hamilton, Márcio, ao meu “cunhado” Luciano, Carlos, Rafael, Sandra M., Alessandra, Priscila, Heitor, Assoeiro, Cris Nordi, Patricinha (por todos os DCEs), Marli, Jack, Queite e Roberto, Neuza, Fabiana, Julio, Débora, Roni, Ane, Edson, Eugênio, Milton, a Cecília e ao André, a Célia e a Rosana, Alessandra Da Roz e ao Fábio, César (“veio”), Euclides, Alexandre (Zulu), Luis, Henrique, Daniela, Camilo, Fabrício, Ana, Rodrigo, Patrícia, Wilson, Dani, Claudia e Claudete, Rômulo, Alexandre (B...da) e Sauli, Ceará, Ivan, Fred, Fábio, Mario Jancso, Gê, Tomich, Adriana, Ricardo, Débora, Marcel, Sheila, Claudia e a todos aqueles que de uma forma ou de outra contribuíram na realização deste trabalho e no meu desenvolvimento.

SUMÁRIO

LISTA DE ABREVIATURAS	i
LISTA DE FIGURAS	iii
LISTA DE TABELAS	vi
RESUMO	vii
ABSTRACT	viii
1 - INTRODUÇÃO	2
1.1 - PROJETOS GENOMA	3
1.2 - PORQUE A ESCOLHA DA CANA-DE-AÇÚCAR	5
1.2.1 - Histórico.....	5
1.2.2 - Principais características da cana-de-açúcar	6
1.3 - PROJETO GENOMA DE ESTs DE CANA-DE-AÇÚCAR (SUCEST)	7
1.3.1 - Análise dos dados obtidos através do seqüenciamento – “DATA MINING”	9
1.4 - BIOSÍNTESE DE PURINONUCLEOTÍDEOS	11
1.4.1 - Síntese <i>de novo</i> dos nucleotídeos purínicos	12
1.4.2 - Mecanismo da síntese <i>de novo</i>	14
1.4.3 - Mecanismo da via de síntese <i>de novo</i>	21
1.4.4 - Via de recuperação	22
1.4.5 - Retroinibição da biossíntese de nucleotídeos purínicos.....	23
1.5 - A ENZIMA FOSFORRIBOSIL PIROFOSFATO SINTASE (PRPPase)	24
2 - OBJETIVOS	28
3 - MATERIAIS E MÉTODOS	30
3.1 - ESTUDOS REALIZADOS COM A SEQÜÊNCIA DO GENE <i>prs</i> DA CANA-DE-AÇÚCAR	31
3.1.1 - Obtenção dos clones	31
3.1.2 - Extração do DNA plasmidial dos clones obtidos.....	32
3.2 – SEQUENCIAMENTO	33
3.2.1 - Protocolo de precipitação da reação de sequenciamento	36
3.3 - ANÁLISE DAS SEQÜÊNCIAS OBTIDAS	37

3.4 - ALINHAMENTO DA SEQÜÊNCIA DO GENE <i>prs</i> DA CANA-DE-AÇÚCAR COM OUTROS ORGANISMOS.....	38
3.5 - CONSTRUÇÃO DA ÁRVORE FILOGENÉTICA.....	39
3.6 - CLONAGEM DO GENE <i>PRS</i> NO VETOR DE EXPRESSÃO pCRT7 NT/CT TOPO.....	40
3.6.1 - Construção dos oligonucleotídeos e amplificação do gene	41
3.6.2 - Purificação do DNA amplificado	43
3.6.3 - Clonagem do gene <i>prs</i> no vetor pCRT7 CT/NT TOPO.....	44
3.6.4 - Transformação do ligante ao vetor pCRT7 TOPO NT/CT	47
3.7 - EXPRESSÃO RECOMBINANTE DA PROTEÍNA CODIFICADA PELO GENE <i>PRS</i> DA CANA-DE-AÇÚCAR.....	49
3.8 – PURIFICAÇÃO.....	51
3.8.1 - Precipitação com sulfato de amônio	52
3.8.2 - Cromatografia líquida de afinidade	53
3.8.3 - Cromatografia por exclusão molecular em Superdex 200 HR 10/30	55
3.9 - CARACTERIZAÇÃO DA ENZIMA PRPP SINTASE.....	56
3.9.1 - Determinação da massa molecular.....	56
3.9.2 - Espalhamento dinâmico de luz (DLS).....	57
3.9.3 - Determinação do ponto isoelétrico	58
3.9.4 - Dicroísmo circular	59
3.9.5 - Clivagem com a enzima EnteroquinaseMAX (EKMAX)	62
3.9.6 - Seqüenciamento do N-terminal da proteína	63
3.10 – CRISTALIZAÇÃO.....	64
4 - RESULTADOS	68
4.1 - CLONES SELECIONADOS E EXTRAÇÃO DO DNA PLASMIDIAL.....	68
4.2 - RESULTADOS DA ANÁLISE DAS SEQÜÊNCIAS DOS CLONES.....	70
4.3 - SEQÜENCIAMENTO DO GENE DA PRPP SINTASE (CLONE 3).....	72
4.4 - ALINHAMENTO DA SEQÜÊNCIA DE AMINOÁCIDOS DA PRPP SINTASE COM SUAS HOMÓLOGAS.....	75
4.5 - CLONAGEM E EXPRESSÃO DA PRPP SINTASE.....	81
4.5.1 - Testes de indução com a proteína PRPP sintase.....	85

4.6 - PURIFICAÇÃO DA PROTEÍNA PRPP SINTASE	86
4.6.1 - Precipitação com sulfato de amônio	86
4.6.2 - Cromatografia líquida de afinidade	87
4.6.3 - Purificação por exclusão molecular em coluna Superdex 200 HR 10/30	88
4.7 - CARACTERIZAÇÃO DA PROTEÍNA PRPP SINTASE.....	90
4.7.1 - Análise da determinação da massa molecular.....	90
4.7.2 - Experimento de espalhamento dinâmico de luz.....	92
4.7.3 - Determinação do ponto isoelétrico	93
4.7.4 - Experimento de dicroísmo circular	93
4.7.5 - Clivagem com a enzima EnteroquinaseMAx (EK).....	95
4.7.6 - Determinação N-terminal	96
4.8 - ENSAIOS DE CRISTALIZAÇÃO.....	97
5 - DISCUSSÃO E CONCLUSÕES.....	101
6 - PERSPECTIVAS FUTURAS.....	112
7 - REFERÊNCIAS BIBLIOGRÁFICAS.....	114
ANEXO I.....	123
ANEXO II.....	125

LISTA DE ABREVIATURAS

- ADP:** Adenosina difosfato
- AICAR:** 5-aminoimidazol-4-carboxiamida ribonucleotídeo
- AIR:** 5-Aminoimidazol ribonucleotídeo
- AMP:** Adenosina monofosfato
- amp:** ampicilina sódica
- APRT:** Adenina-fosforribosil-transferase
- ATP:** Adenosina trifosfato.
- BLAST:** “Basic Local Alignment Search Tool” - Algoritmo de procura de identidades seqüenciais depositadas no banco de dados do NCBI
- BSA:** Albumina serina bovina
- CAPS:** “(3-[ciclohexilamino]-1-propanesulfonic acid)” - (3-[ciclohexilamino]-1- ácido propano sulfônico).
- CD:** Dicroísmo Circular
- cDNA:** DNA complementar
- Cluster:** Seqüências similares agrupadas em uma única seqüência.
- Corante Ponceau:** “3-hidroxi-4-[2-sulfo-4-(sulfophenyllaso) phenyllazo]-2; 7-naphthalenedisulfonic acid” - 3-hidroxi-4-[2-sulfo-4-(sulfofenillaso) fenillazo]-2; 7-ácido naftaleno dissulfônico.
- dATP:** Adenina desoxirribosídeo trifosfato
- ddNTP:** Desoxirribonucleicos que não apresentam a hidroxila no carbono 3.
- DLS:** “Dynamic Light Scattering” – Espalhamento Dinâmico de Luz.
- DNA:** Ácido desoxirribonucleico
- dNTP:** Desoxinucleotídeo trifosfato
- EDTA:** Ácido etileno diaminotetraacético
- EKMax:** Enzima EnterokinaseMax
- EST:** “Expressed Sequence Tag” – Etiqueta de seqüências transcritas
- FAICAR:** N-Formilaminoimidazol-4-carboxiamida ribonucleotídeo
- FGAM:** Formilglicinamida ribonucleotídeo
- FGAR:** Formilglicinamida ribonucleotídeo
- GAR:** Glicinamida ribonucleotídeo
- Gblocks:** Programa que faz seleção de blocos conservados a partir de alinhamentos múltiplos para usar em análise filogenética.
- GMP:** Guanosina Monofosfato
- GTP:** Guanosina trifosfato
- HEPES:** N-2-Hidroxietilpiperazina-N’-ácido etano sulfônico
- HGP:** “Human Genome Project” – Projeto Genoma Humano

HGPRT: Hipoxantina-guanina-fosforribosil-transferase

IEF: “Isoelectric Focusing” – Eletroforese de focalização isoeétrica

IMP: Inosina Monofosfato ou Inosinato

IPTG: Isopropil β - D-tiogalactopiranosídeo

LB: Meio de cultura Luria-Bertani

Mes: “(2-[N-morpholino]ethanesulfonic acid)” - (2-[N-morfolino] ácido etano sulfônico)

ML: “Maximum-Likelihood Method”

NCBI: “National Center for Biotechnology Information”

Ni-NTA: Níquel - Ácido Nitriloacético

ONSA: “Organization Nucleotide Sequencing Analysis” - Organização de Análise e Seqüenciamento de Nucleotídeos

ORF: “Open Read Frame” – Fase aberta de leitura

PCR: “Polymerase Chain Reaction” – Reação em Cadeia da Polimerase

PEG: Polietileno glicol

pI: Ponto isoeétrico

PP_i: Pirofosfato

PRPP: “Phosphoribosilpyrophosphate” – Fosforribosil pirofosfato

PRS: Fosforribosil sintetases

PRSS3: “Probability of Randon Shuffle” – Algoritmo de embaralhamento de seqüências de Monte Carlo.

PRTases: Fosforribosil transferases.

PVDF: “Polivinylidene Fluoride” – Fluoreto de polivinilideno

RMS: “Root Mean Square” - Desvio Médio Quadrático

RNA: Ácido ribonucléico

SAICAR: *N*-Succinil-5-aminoimidazol-4-carboxiamida ribonucleotídeo

SDS: Detergente “Sodium Dodecilsulfate” - Dodecil sulfato de sódio

SDS-PAGE: “Sodium Dodecilsulfate Polyacrilamide Gel Electrophoresis” - eletroforese desnaturante em gel de poliacrilamida contendo SDS

SELCON2: “Self-Consistent Method” - Programa de análise de espectros de CD que estima o conteúdo de estrutura secundária da proteína.

SUCEST: “Sugarcane Expressed Sequence Tag”

TAE: Tampão de eletroforese composto de tris, ácido acético, EDTA

TBE: Tampão utilizado no seqüenciamento contendo tris, ácido bórico, EDTA

Temed: Tetrametiletilediamina

Tris: Hidróximetil-aminometano

Triton X 100: t-octilfenoxietoxietanol

UTR: “Untranslated Region” - Região não traduzida

XPRT: Xantina-fosforribosil-transferase

LISTA DE FIGURAS

FIG1: DNA até a formação do cromossomo.....	3
FIG2: Cana-de-açúcar.....	6
FIG3: Representação esquemática da estratégia do “Data Mining”.....	10
FIG4: Origem dos átomos do anel purínico.....	12
FIG5: Síntese do PRPP.....	13
FIG6: Passo 1.....	14
FIG7: Passo 2.....	15
FIG8: Passo 3.....	15
FIG9: Passo 4.....	16
FIG10: Passo 5.....	16
FIG11: Passo 6 e 7.....	17
FIG12: Passo 8 e 9.....	18
FIG13: Passo 10.....	19
FIG14: Passo 11.....	19
FIG15: Biossíntese do AMP e GMP a partir do IMP.....	20
FIG16: Esquema representativo do mecanismo da via da síntese <i>de novo</i>	21
FIG17: Via de recuperação e interconversão, em células de mamíferos.....	22
FIG18: As diversas reações de transferência catalisadas pelas PRtases estão mostradas.....	23
FIG19: Mecanismo de retroinibição da síntese de purinonucleotídeos.....	23
FIG20: Representação esquemática da estratégia para caracterização das proteínas DA via da síntese <i>de novo</i>	30
FIG21: Estratégia para reação de seqüenciamento automatizado de DNA.....	34
FIG22: Estratégia de seqüenciamento.....	36
FIG23: Esquema representativo da estratégia de amplificação.....	42
FIG24: Vetor pCRT7 CT TOPO.....	44
FIG25: Vetor pCRT7 NT TOPO.....	45
FIG26: Esquema utilizado na clonagem do gene <i>prs</i> da cana-de-açúcar.....	46
FIG27: Esquema da interação dos resíduos vizinhos no “tag” 6XHis e a matriz Ni-NTA.....	53

FIG28: Estrutura química do imidazol e da histidina.....	54
FIG29: Esquema representativo do arranjo de um cristal.....	65
FIG30: Esquema do método de cristalização usando a técnica gota suspensa.....	65
FIG31: Alinhamento de uma seqüência conhecida padrão contra seqüências EST do banco de dados do SUCEST usando tBLASTn.....	68
FIG32: Extração do DNA plasmidial.....	70
FIG33: Representação esquemática do alinhamento de confirmação dos clones seqüenciados.....	71
FIG34: Alinhamento representativo das fases de leituras (“Frames”).....	71
FIG35: Alinhamento múltiplo realizado através do programa SeqMan.....	73
FIG36: Seqüência de nucleotídeos e aminoácidos da PRPP sintase da cana-de-açúcar.....	74
FIG37: Alinhamento múltiplo de seqüências representativas de PRS de plantas.....	76
FIG38: Reconstrução filogenética de proteínas PRS representativas.....	80
FIG39: Resultado da amplificação do DNA.....	81
FIG40: Representação esquemática da estratégia de amplificação.....	82
FIG41: Experimento de amplificação do DNA.....	82
FIG42: Experimento de amplificação do gene da PRPP sintase.....	83
FIG43: Resultado do teste de expressão com o clone 3b da PRPP sintase.....	84
FIG44: Resultados dos testes de expressão da proteína PRPP sintase.....	85
FIG45: Resultado do fracionamento com sulfato de amônio da PRPP sintase (3bNT) em 15% SDS-PAGE.....	86
FIG46: Perfil cromatográfico da eluição da proteína em coluna Ni-NTA.....	7
FIG47: Resultado da purificação através de cromatografia de afinidade Ni-NTA.....	88
FIG48: Perfil cromatográfico da eluição da PRPP sintase na coluna Superdex 200 HR 10/30.....	89
FIG49: Resultado da purificação em coluna de exclusão molecular Superdex 200 HR 10/30.....	89
FIG50: Cromatografia de calibração com a coluna Superdex 200 HR 10/30.....	90
FIG51: Curva de calibração da coluna Superdex 200 HR 10/30.....	91

FIG52: Histograma representativo do R_H da proteína PRPP sintase.....	92
FIG53: Determinação do ponto isoelétrico da proteína PRPP sintase.....	93
FIG54: Espectro de CD da proteína PRPP sintase.....	94
FIG55: Experimento de clivagem da proteína PRPP sintase com a enzima EKMAx.....	95
FIG56: Produto de digestão da proteína com a enzima EKMAx submetido a seqüenciamento N-terminal.....	96

LISTA DE TABELAS

TABELA 1: Frequência dos clones EST inteiros.....	11
TABELA 2: Clones a serem analisados.....	32
TABELA 3: Mistura da reação de sequenciamento.....	35
TABELA 4: Condições do ciclo da reação de sequenciamento.....	35
TABELA 5: Oligonucleotídeos sintetizados para reação de sequenciamento.....	36
TABELA 6: Oligonucleotídeos utilizados na amplificação.....	41
TABELA 7: Condições da reação de amplificação dos genes.....	42
TABELA 8: Condições do ciclo da reação de PCR para amplificação do gene..	43
TABELA 9: Reação utilizada para adição de deoxiadenosina (A).....	46
TABELA 10: Mistura de reação para inserção do produto de PCR ao vetor pCRT7 NT/CT TOPO.....	47
TABELA 11: Padrão de massa molecular utilizado na coluna Superdex 200 HR 10/30.....	56
TABELA 12: Padrões de pI utilizados constituintes do kit geral de calibração..	59
TABELA 13: Descrição dos tampões utilizados no estudo da variação de pH por Dicroísmo Circular (CD).....	61
TABELA 14: Clones selecionados para análise.....	69
TABELA 15: Resultado obtido a partir do alinhamento para obtenção da fase de leitura significativa.....	72
TABELA 16: Composição dos aminoácidos.....	75
TABELA 17: Alinhamento da PRPP sintase com PRS de organismos homólogos.....	78
TABELA 18: Seleção das colônias resultantes da ligação do inserto ao vetor, verificados por PCR de amplificação.....	84
TABELA 19: Padrões de massa molecular utilizado na coluna Superdex 200 HR 10/30 e seus respectivos volumes de eluição.....	91
TABELA 20: Desconvolução dos espectros de CD através do programa SELCON2.....	94
TABELA 21: Ensaios de cristalização com Fatorial I – Hampton Research.....	98
TABELA 22: Ensaios de cristalização com Fatorial II – Hampton Research.....	98

RESUMO

Fosforribosilpirofosfato sintetase (PRS – EC: 2.7.6.1) é uma enzima de central importância em muitas vias metabólicas em todas as células. A PRS de plantas não tem sido investigada em grandes detalhes. Entretanto, as evidências acumuladas indicam que estas enzimas formam uma família complexa de isoenzimas com localização subcelular (mitocôndria, citoplasma e núcleo) e diferentes características de dependência a fosfato. O gene *prs* foi clonado e a sequência determinada a partir de uma biblioteca de cDNA da cana-de-açúcar identificadas por um genoma de planta (SUCEST). O gene *prs* da cana-de-açúcar contém uma fase aberta de leitura de 984 pb que codifica uma proteína de 328 aminoácidos com massa molecular de 36,6 kDa. A sequência de aminoácidos deduzida tem 77 e 78% de identidade com a PRS4 de *Arabidopsis thaliana* e *Spinacia oleracea*. A reconstrução filogenética de PRS homólogas selecionadas indica que esta enzima pode ser uma PRS independente de fosfato. Uma proteína foi expressa em *Escherichia coli*, purificada e encontrado os elementos de estrutura secundária e arranjos quaternários, consistentes com PRS homólogas conhecidas. A disponibilidade da enzima PRS a partir de uma outra planta e a possibilidade de expressar a proteína em larga escala pode fornecer a base para análises funcionais e estruturais desta importante enzima.

ABSTRACT

Phosphoribosylpyrophosphate synthetase (PRS - EC: 2.7.6.1) is an enzyme of central importance in several metabolic pathways in all cells. So far the plant PRS enzymes have not been investigated in great detail. However, accumulated evidences indicate that those enzymes form a complex family of isoenzymes with subcellular localization (mitochondria, cytoplasm and nucleus) and different phosphate dependence characteristics. The *prs* gene was cloned and had the sequence determined from a sugarcane cDNA library clone identified by the plant genome effort (SUCEST). The sugarcane *prs* gene contains a 984 bp open reading frame that encodes 328 amino acids protein with a calculated molecular weight of 36.6 kDa. The predicted amino acid sequence has 77 and 78% amino acid identity to the PRS4 of *Arabidopsis thaliana* and *Spinacia oleracea* respectively. The phylogenetic reconstruction of selected PRS homologues indicates that this enzyme may be a phosphate-independent PRS isoenzyme. The PRS protein was expressed in *Escherichia coli*, purified to homogeneity and found to retain secondary structure elements and quaternary arrangement consistent with the known PRS homologues. The availability of the PRS enzyme from another plant and the possibility of expressing the protein in large quantities should provide the basis for a functional and structural analysis of this important enzyme.

INTRODUÇÃO



1 - INTRODUÇÃO

A biologia, com a descoberta da estrutura molecular do DNA com Watson & Crick na década de 50, ganhou impulsos para várias transformações. Houve principalmente um avanço da Biologia Molecular, que passou a se preocupar não como funciona um gene, ou uma proteína isoladamente, mas como um sistema de genes e proteínas atua nos organismos vivos. Oferecendo um conjunto de ferramentas e tecnologias aos cientistas, proporcionou a possibilidade de estudar, fora da célula, o mecanismo bioquímico complexo que rege as interações entre as moléculas, dos ácidos nucléicos a proteínas, podendo não só analisá-los e interpretá-los como também modificá-los. Estes avanços incluem desenvolvimentos de métodos de clonagem, amplificação e seqüenciamento de DNA.

O desenvolvimento de duas técnicas de seqüenciamento, por Frederick Sanger e Maxam e Gilbert em 1977 [1] permitiu pela primeira vez a obtenção da seqüência completa dos genes. As técnicas envolvem mecanismos diferentes de seqüenciamento de DNA, sendo uma por terminação da cadeia enzimática e a outra por degradação química, respectivamente.

Em 1995, devido a esses avanços tecnológicos Hamilton Smith e Craig Venter seqüenciaram o primeiro genoma completo da história, o de *Haemophilus influenza*, contendo exatamente 1.830.137 pares de bases (pb).

Pesquisadores das mais diversas áreas começaram aperfeiçoar as técnicas de seqüenciamento, permitindo assim, o seqüenciamento completo do DNA de organismos complexos, como por exemplo, o do homem, dando início então, ao

que se conhece por “Projetos Genoma”. A partir destes projetos, foram sendo criadas ferramentas computacionais que possibilitaram não só seqüenciar os genes, mas também analisar e interpretar os dados obtidos, dando espaço a uma nova área que mescla biologia com computação, conhecida como Bioinformática.

1.1 - PROJETOS GENOMA

Genoma é a seqüência completa do DNA de um organismo, incluindo os genes e as seqüências intergênicas. Os genes são seqüências específicas de bases que contém instruções para codificação das proteínas necessárias para todos os organismos. O DNA é

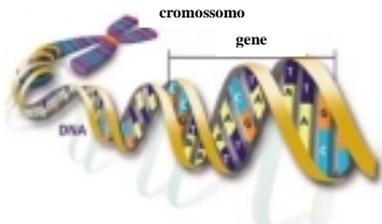


FIG1: DNA até a formação do cromossomo composto de quatro diferentes bases que se repetem milhões ou bilhões de vezes ao longo do genoma. A ordem em que essas bases se posicionam é de extrema importância, pois é esta seqüência que faz os organismos serem distintos uns dos outros. É então nos Projetos Genoma que se irá descobrir como ocorre o processamento de toda esta maquinaria. Algumas áreas de aplicação dos Projetos Genoma incluem a área de medicina molecular para compreensão e tratamento de doenças, arqueologia, antropologia, evolução e migração humana, agricultura, etc.

Estes projetos são constituídos por grupos nacionais e internacionais. Os quais formam o que se conhece por consórcios de laboratórios, financiados por órgãos públicos ou privados, compostos por laboratórios responsáveis pelo seqüenciamento dos genes e outros pela análise computacional dos mesmos.

Entre os grupos internacionais tem-se a “HGP - U. S. Human Genome Project” (consórcio público), e o grupo “Celera Genomics Corporation” (grupo privado), responsáveis pelo seqüenciamento e publicação do Genoma Humano com cerca de 3,5 bilhões de pares de bases seqüenciados [2].

Os primeiros genomas estudados incluem: a bactéria *Escherichia coli*, com um genoma de 4 milhões de pares de bases; a mosca de frutas *Drosophila melanogaster* com 100 milhões de pb; o nematódeo *Caenorhabditis elegans* com 100 milhões de pb; o camundongo *Mus musculus* com 3 bilhões de pb; e o genoma da levedura *Saccharomyces cerevisiae* com 12 milhões de pb. Há também o genoma do vegetal *Arabidopsis thaliana* que é uma pequena erva daninha da família da mostarda que se tornou um organismo experimental modelo para o estudo genético molecular de plantas. Este genoma contém 100 milhões de pb.

Os Projetos Genoma de âmbito nacionais são financiados pela Fundação de Apoio e Pesquisa do Estado de São Paulo (FAPESP), e fazem parte do programa ONSA (Organização de Análise e Seqüenciamento de Nucleotídeos). Através desse programa já foi obtida a seqüência completa do genoma da bactéria *Xylella fastidiosa* com 2.679.305 de pb [3], que foi a primeira bactéria fitopatogênica seqüenciada no Brasil. Alguns outros projetos estão em andamento, como o Projeto Genoma do Câncer Humano, Genomas Clínicos, *Xanthomonas citri*, *Schistosoma mansoni*, Genomas agrônômicos e ambientais. Já finalizado e de extrema importância é o Projeto Genoma da Cana-de-Açúcar (SUCEST) [4].

1.2 - PORQUE A ESCOLHA DA CANA-DE-AÇÚCAR

1.2.1 - HISTÓRICO

A cana-de-açúcar é originária da Ásia e tem nome científico *Saccharum sp.* No Brasil passou a ser cultivada em larga escala em meados do século XVI, tornando-se um produto importante e atraente para o mercado mundial, principalmente por ser o elemento básico para a obtenção de açúcar e de vários tipos de álcool, entre eles o etílico, desempenhando um importante papel sob o aspecto infra-estrutural e socioeconômico do país.

Os núcleos iniciais de produção canavieira localizavam-se na região Nordeste, ao redor da Bahia e em Pernambuco, expandindo-se para região Sudeste, onde sua produção passou a se concentrar no Rio de Janeiro e em São Paulo.

A condição climática tropical brasileira favoreceu e permitiu a formação de grandes latifúndios monocultores, fundamentando a economia até os dias atuais.

Em 1995, $1,2 \times 10^9$ toneladas de cana foram produzidas numa área estimada em 18 milhões de hectares, e foram usadas principalmente na produção de açúcar e como fonte de energia.

A cana-de-açúcar é uma das mais importantes culturas de planta, sendo cultivada em áreas tropicais e subtropicais em mais de 80 países em torno do mundo. O Brasil é o responsável por 25% desta produção, metade concentrada no Estado de São Paulo (<http://sucest.lad.ic.unicamp.br>).

1.2.2 - PRINCIPAIS CARACTERÍSTICAS DA CANA-DE-AÇÚCAR

É uma planta pertencente à família das gramíneas, sendo um grupo amplamente distribuído de plantas vasculares com uma grande facilidade de adaptação. É classificada como sendo pertencente à classe das Monocotiledôneas; ordem Ciperales; família Poaceae. Seu gênero é o *Saccharum* e somam cerca de 30 espécies.

A cana-de-açúcar (fig2) é uma planta ereta, perene, rizomatosa, formando touceiras. De colmo cilíndrico, externamente glarbo, de coloração variável e internamente com feixes vasculares totalmente primários e amplamente diversos. Possui entrenós retos ou em ziguezague e de comprimento, espessura e formas variadas, podendo crescer de 3 a 6 metros. Podem estar ou não revestidas com camada cerosa. Os nós apresentam-se protuberantes ou constrictos. As folhas são simples, alternadas, estreito-lanceonadas. O fruto é seco do tipo cariopse e com semente de endosperma abundante (<http://encarta.msn.com>)



FIG2: Cana-de-açúcar

O fato de o Brasil contribuir com 25% da produção de cana-de-açúcar faz com que esta cultura represente uma importante atividade agro-industrial, principalmente para o Estado de São Paulo. Sua importância econômica se torna significativa, pois a cana-de-açúcar é produzida tanto para consumo interno como para exportação, oferecendo como bioprodutos o açúcar, melado, rum, álcool, combustível e forragem animal, e a partir dos resíduos dos talos o papel.

As espécies de cana-de-açúcar cultivadas comercialmente são o resultado de diversos cruzamentos interespecíficos, envolvendo as espécies *Saccharum*

officinarum, *S. barberi*, *S. sinense*, e as espécies selvagens *S. spontaneum* e *S. robustum* e da manipulação humana por técnicas de melhoramento genético. Este melhoramento busca plantas mais resistentes a períodos de secas prolongados, e menos susceptíveis ao ataque de patógenos.

O fato de a cana-de-açúcar ter se originado do cruzamento de múltiplas espécies resultou numa planta com um genoma muito complexo, contendo um número de cromossomos geralmente na ordem de $2n = 70-120$. Esta complexibilidade torna o mapeamento dos caracteres desejáveis para o melhoramento genético molecular uma tarefa difícil.

Os resultados originados pelo seqüenciamento da cana-de-açúcar podem ser explorados pela comunidade científica em estudos que relacionam o uso dos genes como fonte de marcadores para características significantes da agricultura, e também oferecem bases de estudos de crescimento e desenvolvimento da planta que podem mais adiante ser usados para resolver questões fisiológicas, bioquímicas, patológicas e por último formação da planta (<http://sucest.lad.ic.unicamp.br>). Isto justificou a necessidade da criação do Projeto Genoma da Cana-de-açúcar.

1.3 - PROJETO GENOMA DE ESTs DE CANA-DE-AÇÚCAR (SUCEST)

O SUCEST (**SU**garCane **E**xpressed **S**equences **T**ag) faz parte do consórcio brasileiro ONSA ([Simpson e Perez 1998](#)), o qual engloba 74 laboratórios de pesquisa no Estado de São Paulo, e foi fundado pela Fapesp com cerca de 200 pesquisadores com o objetivo de implementar o Programa Genoma (<http://>

sucest.lad.ic.unicamp.br). Nestes laboratórios incluem-se os de seqüenciamento e os de análise e armazenamento dos dados seqüenciados.

O principal objetivo deste projeto foi empreender um programa de EST (**Expressed Sequence Tag**) em larga escala através do seqüenciamento randômico de clones originados de bibliotecas de DNA complementar (cDNA). Essas bibliotecas foram preparadas a partir de tecidos diferenciados (calo, raiz, caule, folhas etioladas, flores, sementes, etc) originados de 30 diferentes cultivos de cana-de-açúcar derivadas da hibridização das espécies *S. officinarum* (2n=80) e *S. spontaneum* (2n=40-128) [4]. Este projeto não somente objetivou a realização do seqüenciamento para concretizar as metas do consórcio, mas também o estabelecimento da tecnologia de seqüenciamento em larga escala.

O SUCEST construiu um banco de dados contendo 238.000 ESTs originadas de 26 bibliotecas de cDNA. Essas bibliotecas foram construídas a partir da extração do mRNA dos vários tecidos. Os cDNA foram clonados no vetor pSPORT, que é um vetor multifuncional próprio para construção de bibliotecas [4]. Para um total de 291.689 clones de cDNA a extremidade 5' de cada gene foi seqüenciada e em alguns a extremidade 3'. Essas seqüências quando depositadas no banco de dados do SUCEST após receberem um tratamento chamado de limpeza das seqüências (“trimming”), gerou 238.000 ESTs de cana-de-açúcar. O fato de existirem clones que foram seqüenciados mais de uma vez faz com que existam muitas seqüências redundantes depositadas no banco de dados. Com o intuito de diminuir esta redundância, e facilitar o manuseio das seqüências, optou-se por agrupar seqüências similares em uma consenso (“cluster”) resultando num banco de dados com 43.141 “clusters” depositados [4, 5].

1.3.1 - ANÁLISE DOS DADOS OBTIDOS ATRAVÉS DO SEQÜENCIAMENTO – “DATA MINING”

O “Data Mining” é composto por uma coordenação de 50 grupos responsáveis pela análise e catalogação dos resultados obtidos após o seqüenciamento completo do Genoma da Cana-de-açúcar. Este trabalho é direcionado de acordo com o tipo de pesquisa que cada grupo desenvolve.

No nosso laboratório, por exemplo, a pesquisa desenvolvida envolve a via metabólica de purinonucleotídeos, que é de central importância para todas as células. Em muitos organismos, os purinonucleotídeos são sintetizados *de novo* a partir de precursores não nucleotídeos tais como aminoácidos, amônia e dióxido de carbono. As purinas também podem ser recicladas através da via de recuperação.

A proposta desse “Data Mining” no esforço de identificação e catalogação dos ESTs de cana-de-açúcar gerados no banco de dados do SUCEST foi de identificar diversos clones pertencentes ao metabolismo de purinonucleotídeos, enfocando dois temas:

- 1 – Identificação de genes envolvidos na síntese *de novo* de purinas.
- 2 – Identificação de diversos genes envolvidos na via de recuperação de purinas.

A identificação de diversos genes envolvidos nos onze passos da via metabólica da síntese *de novo* (fig6, p. 22) permitirá um estudo mais detalhado, com o objetivo de entender o mecanismo das enzimas da cana-de-açúcar

envolvidas na síntese de purinas [6]. Isto tornaria possível usar estas enzimas como alvo para desenvolvimento de drogas as quais possibilitem combater agentes fitopatogênicos, como tem sido aplicado a muitos parasitas e ao câncer.

A estratégia utilizada para realização deste trabalho segue o seguinte fluxograma da figura 3 [6].

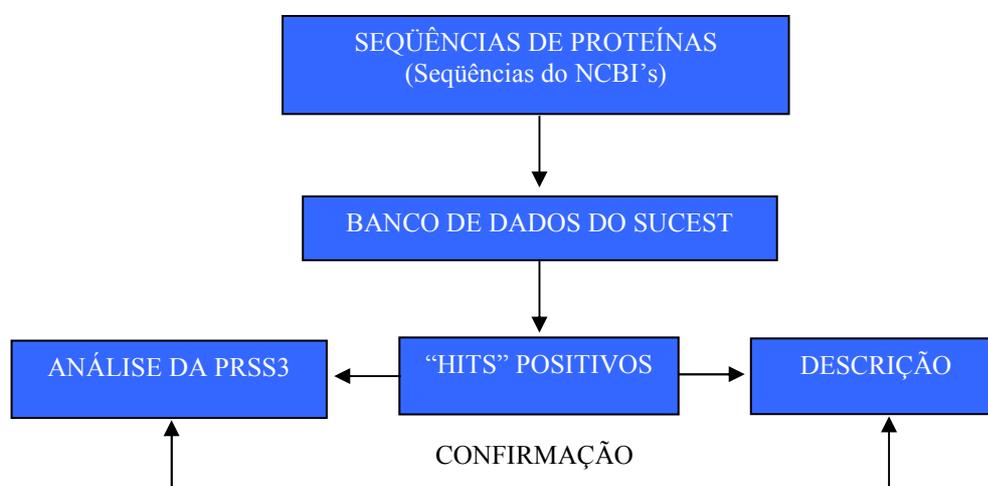


FIG3: Representação esquemática da estratégia do “Data mining”.

Sequências representativas de cada uma das onze enzimas participantes da via da síntese *de novo* de purinas foram selecionadas do banco de dados do NCBI (“National Center for Biotechnology Information”). Estas sequências de peptídeos foram usadas para separar as ESTs de cana-de-açúcar do banco de dados do SUCEST utilizando o programa tBLASTn (<http://www.expasy.ch>). Os resultados positivos foram confirmados pela implementação PRSS3 (“Probability of Random Shuffle”) do algoritmo “shuffling” Monte Carlo, resultando em uma tabela de valores estatisticamente significantes [6]. Este estudo permitiu encontrar clones EST, possibilitando uma análise rápida da presença de sequências inteiras ou parciais,

podendo-se realizar estudos funcionais da proteína expressa. A tabela 1 fornece os dados sobre a quantidade de seqüências EST inteiras e seqüências EST parciais.

TABELA 1: Frequência dos clones EST inteiros

ENZIMA	SEQ. TOTAL	SEQ. PARCIAL
PRPP sintase	7	9
Glutamina-PRPP-amidotransformilase	0	2
GAR sintetase	2	6
GAR transformilase	1	7
FGAM sintetase	0	3
AIR sintetase	1	2
AIR carboxilase	0	3
SAICAR sintetase	1	1
Adenilosuccinato liase	0	5
AICAR transformilase-IMP ciclohidrolase	1	2
TOTAL	13	40

Os resultados permitiram identificar um caso potencial de polimorfismo em cana-de-açúcar, o da proteína PRPP sintase, enzima iniciadora da via da síntese *de novo* de purinas.

1.4 - BIODISSÍNTESE DE PURINONUCLEOTÍDEOS

Os nucleotídeos têm importantes papéis em todas as células. Eles são precursores do DNA e RNA. São importantes na biossíntese de intermediários como o ATP e GTP. Eles são componentes dos cofatores NAD, FAD, e coenzima A, e também exercem papéis de reguladores metabólicos [7].

Há dois tipos de vias, como já mencionado, a via da síntese *de novo* e a via de recuperação. A síntese *de novo* de nucleotídeos purínicos começa com seus precursores metabólicos: aminoácidos, ribose-5-fosfato, CO₂ e NH₃. A via de

recuperação recicla as bases livres (G, A, T, C, U) e nucleosídeos liberados da degradação hidrolítica dos ácidos nucleicos [8].

1.4.1 - SÍNTESE *DE NOVO* DOS NUCLEOTÍDEOS PURÍNICOS

O anel purínico é construído a partir de vários precursores (fig4).

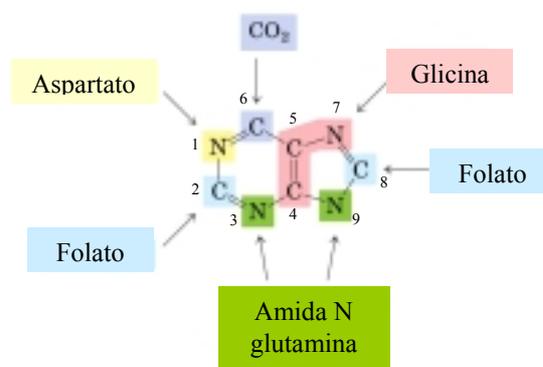


FIG4: Origem dos átomos no anel purínico. Glicina fornece C-4, C-5 e N-7. Aspartato o N-1. A glutamina os átomos de N-3 e N-9 que vem de da cadeia lateral da amida. O C-2 e C-8 vêm dos derivados ativos do tetra-hidrofolato. E o C-6 é doado pelo CO₂ [7, 8].

John Buchanan e G. Robert Greenberg foram os primeiros a trabalhar com a via da biossíntese de purinas. Esta via é bastante conservada dentre os vários organismos até hoje estudados.

Trata-se de uma via central constituída por onze passos enzimáticos, divididos em dois estágios:

- ✚ Formação do 5-aminoimidazol ribonucleotídeo a partir do PRPP.
- ✚ Formação do Inosina monofosfato (IMP) a partir do 5-aminoimidazol ribonucleotídeo (AIR).

O ponto inicial desta via é a síntese do substrato 5-fosforribosil-1-pirofosfato (PRPP), culminando com a formação de Inosina monofosfato (IMP). O IMP formado é convertido aos purinonucleotídeos Guanosina monofosfato (GMP) e Adenosina monofosfato (AMP) na via de interconversão.

Sabe-se que a porção ribose fosfato dos nucleotídeos purínicos e pirimidínicos é fornecida pela PRPP.

A PRPP é um intermediário notável em muitas vias de síntese de nucleotídeos. É sintetizado a partir da ribose-5-fosfato derivado da via pentose-fosfato numa reação catalisada pela ribose fosfato pirofosfoquinase (PRPP sintetase). Também esta presente na síntese de aminoácidos como Histidina e Triptofano.

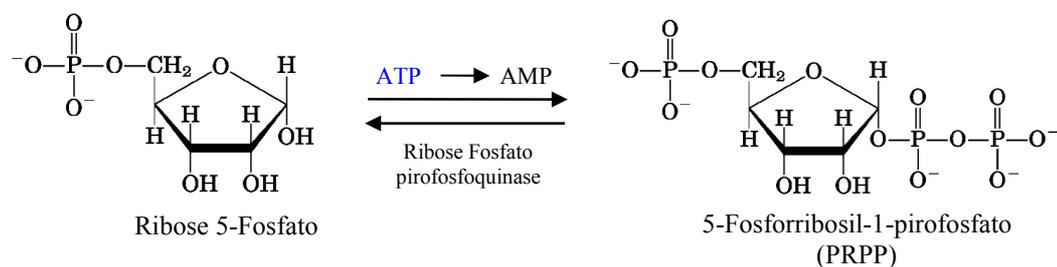


FIG5: Síntese do PRPP. Formação do substrato PRPP a partir da ribose-5-fosfato através da enzima PRPP sintase.

A síntese enzimática da PRPP foi descoberta por Arthur Kornberg e seus colaboradores, que mostraram que a reação se realiza pela transferência direta do pirofosfato a partir do ATP para a ribose-5-fosfato, como descrita no esquema acima [9].

1.4.2 - MECANISMO DA SÍNTESE *DE NOVO*

A via da síntese *de novo* constitui-se de onze passos enzimáticos subdivididos em dois estágios sendo o primeiro determinado por cinco passos enzimáticos descritos a seguir.

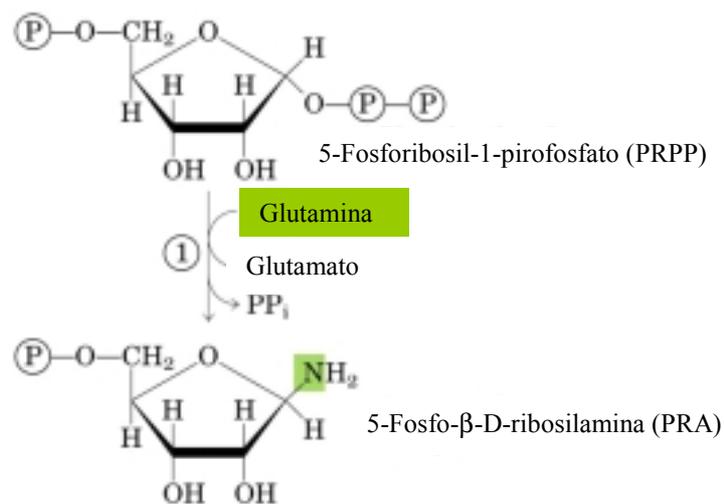


FIG6: Passo1 - Ocorre uma substituição do PPi por uma amina.

Esta é considerada a etapa reguladora onde ocorre a formação da 5-fosfosilamina a partir de PRPP e Glutamina (fig6). O grupamento amida da cadeia lateral da glutamina desloca o pirofosfato ligado ao C-1. Essa reação é causada pela hidrólise do pirofosfato dando início a reação de formação do anel purínico.

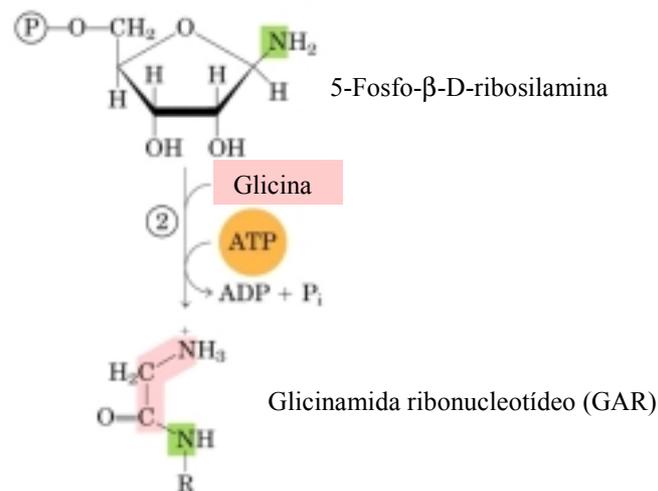


FIG7: Passo 2 – Adição de glicina

A glicina liga-se a fosforribosilamina produzindo *glicinamida ribonucleotídeo* (fig7). Um ATP é hidrolisado na formação de uma ligação amida entre a carboxila da glicina e a amina da fosforribosilamina. Ocorre a formação de um intermediário acil-fosfato.

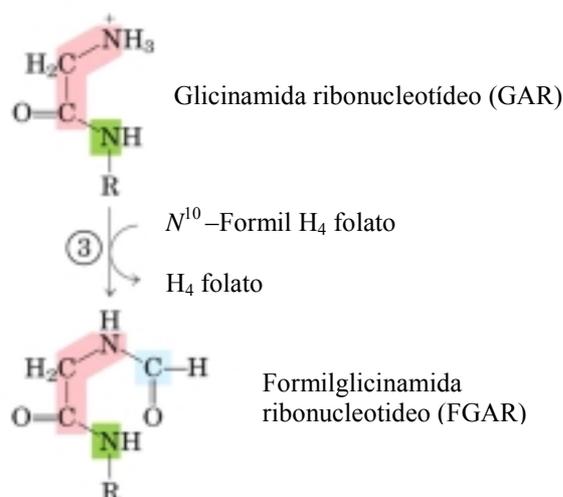


FIG8: Passo 3 - Formilação pelo N^{10} -formil tetra-hidrofolato

Os derivados de folato servem como intermediários ativos em várias etapas da formação de nucleotídeos purínicos e pirimidínicos. Nesta etapa então, ocorre a formilação do amino α terminal da glicina pelo N^{10} -formil tetra-hidrofolato formando α - N -formil glicinamida ribonucleotídeo (fig8).

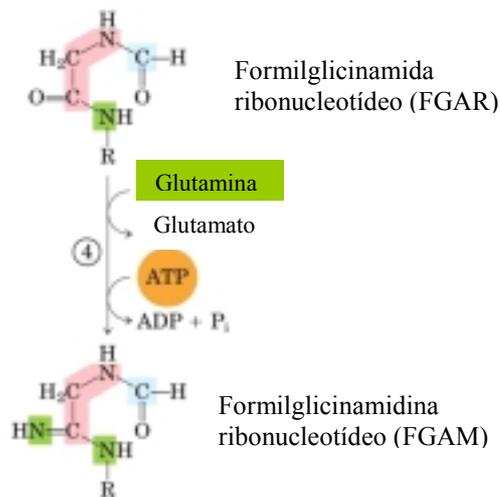


FIG9: Passo 4 - A glutamina contribui com um nitrogênio.

Nessa reação ocorre a hidrólise de um ATP e a cadeia lateral da glutamina doa um átomo de nitrogênio (fig9).

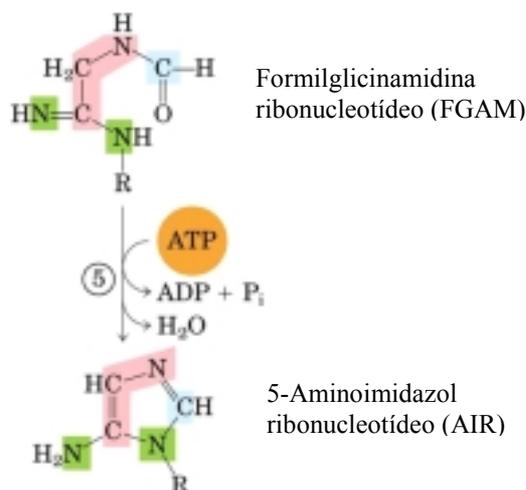


FIG10: Passo 5 - Desidratação e fechamento do anel.

Antes da desidratação e fechamento do anel purínico há a formação do anel de imidazol do núcleo purinas, como o *5-aminoimidazol ribonucleotídeo* (fig10).

Neste ponto três dos seis átomos necessários para o fechamento do anel já se encontram posicionados iniciando-se então o segundo estágio constituído dos seis últimos passos enzimáticos para a formação do Inosinato a partir do 5-aminoimidazol ribonucleotídeo.

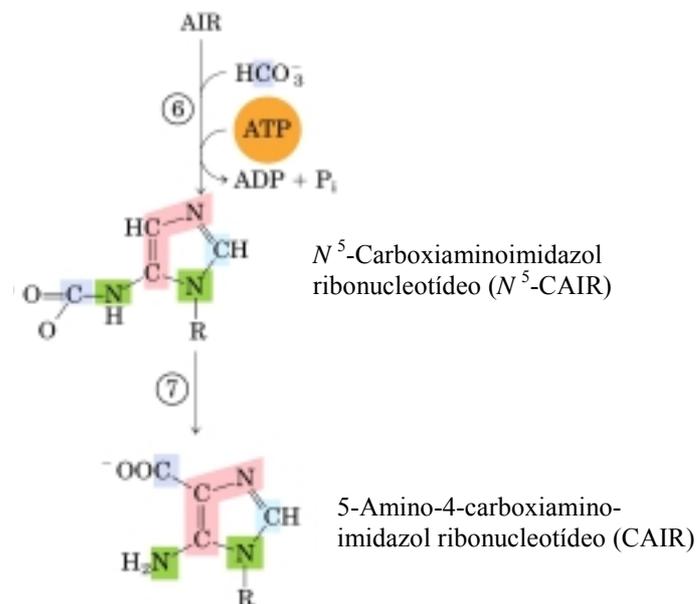


FIG11: Passo 6 e 7 - Um grupo carboxil é adicionado e em seguida ocorre a carboxilação.

Através da carboxilação (adição de um grupo carboxil) do aminoimidazol ribonucleotídeo é introduzido o átomo de carbono seguinte formando o *5-aminoimidazol-4-carboxilato ribonucleotídeo* (fig11). Esta carboxilação utiliza-se do bicarbonato, geralmente presente em soluções aquosas. Ocorre um rearranjo que transfere o carboxilato do grupo amino exocíclico para a posição 4 do anel de imidazol.

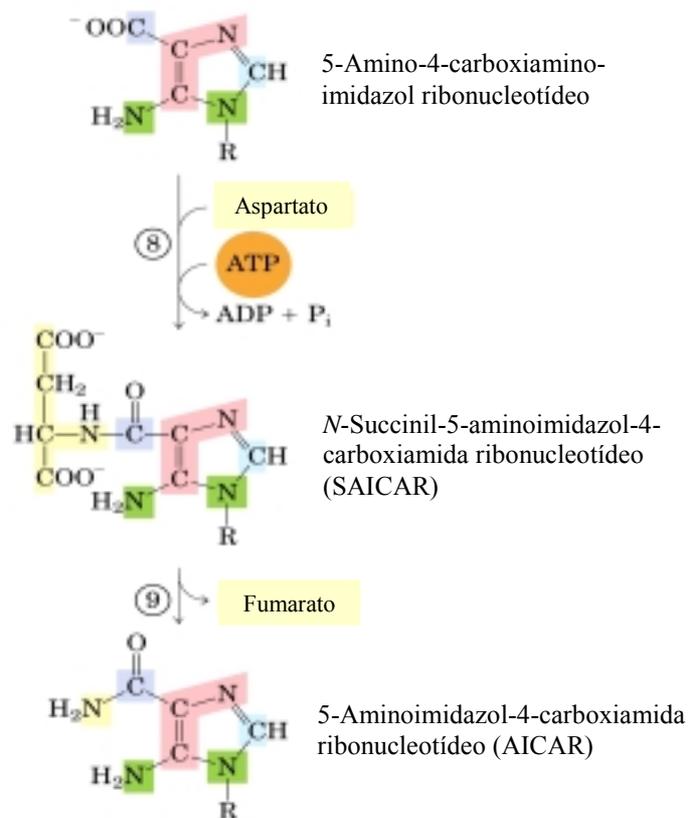


FIG12: Passo 8 e 9 - Caracteriza-se pela doação de um grupo amina do aspartato.

O aspartato contribui com seu átomo de nitrogênio para o anel purínico. Dessas reações tem-se a conversão do carboxilato em uma amida. Esta reação ocorre devido ao fato da amina do aspartato reagir com a carboxila 5-aminoimidazol-4-carboxilato ribonucleotídeo. Na formação dessa ligação amida uma molécula de ATP é hidrolisada (fig12). Em seguida o esqueleto de carbono do aspartato é liberado como fumarato e o 5-aminoimidazol-4-carboxiamida ribonucleotídeo é produzido.

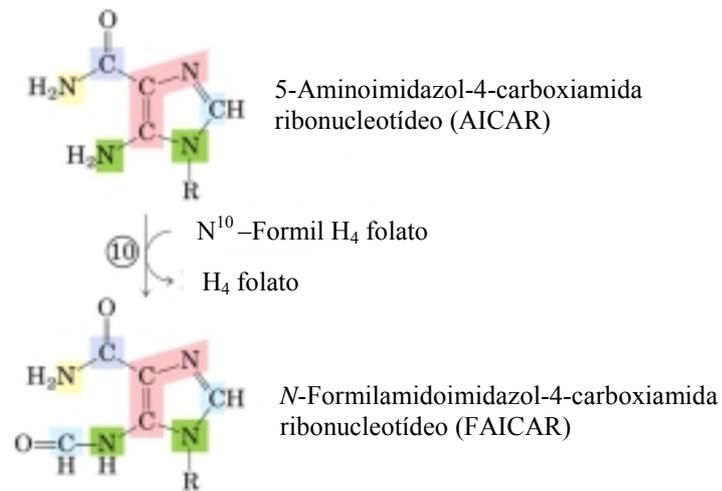


FIG13: Passo 10 - Formilação pelo N^{10} -formil tetra-hidrofolato.

O N^{10} -formil tetra-hidrofolato fornece o último átomo do anel purínico tendo como resultado o *5-formamidoimidazol-4-carboxamida ribonucleotídeo* (fig13).

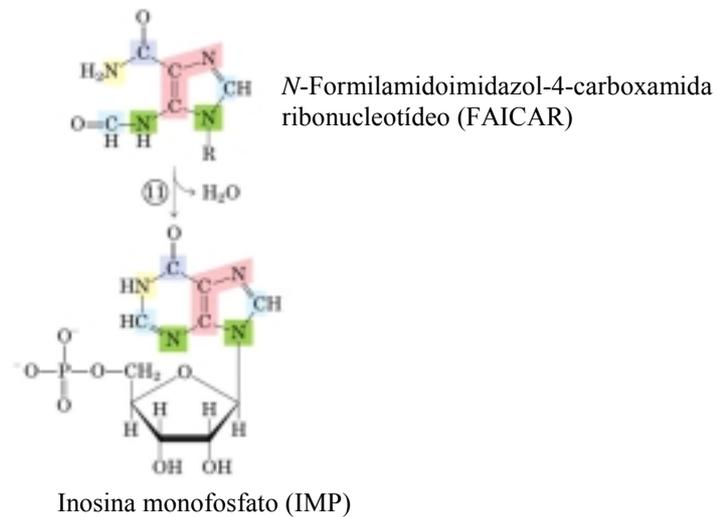


FIG14: Passo 11: Último passo da cadeia enzimática dos purinonucleotídeos. Ocorre a desidratação e fechamento do anel.

A 5-formamidoimidazol-4-carboxamida ribonucleotídeo sofre desidratação e ocorre o fechamento do segundo anel formando *Inosina monofosfato* (IMP), que contém o anel purínico completo e tem como base purínica a Hipoxantina (fig14).

O Inosinato é o responsável pela formação dos ribonucleotídeos AMP e GMP. Este processo de conversão pode ser visualizado na figura 15:

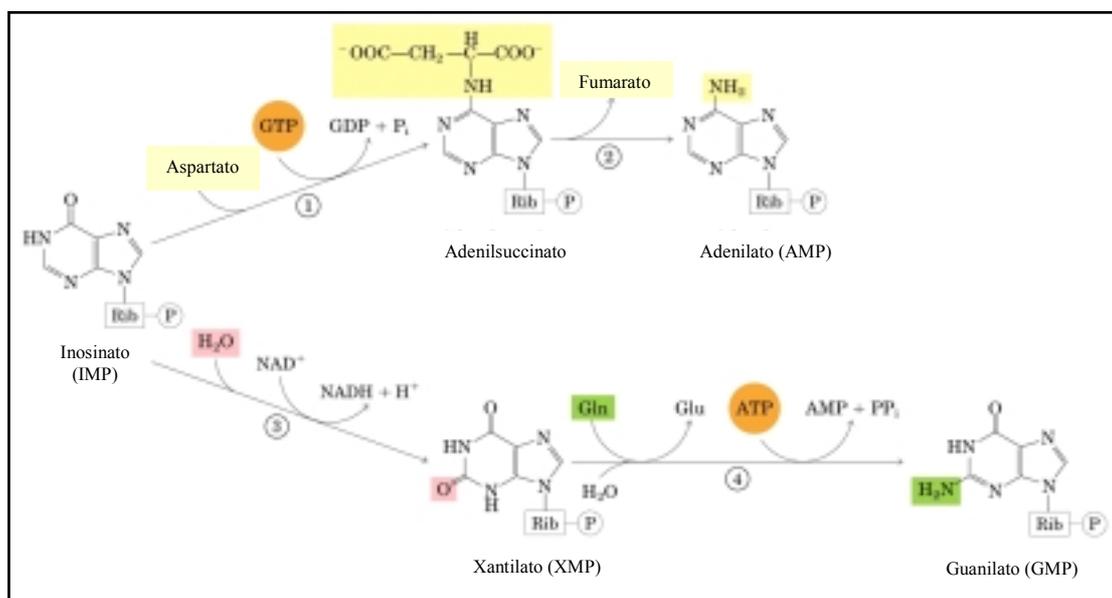


FIG15: Biossíntese do AMP e GMP a partir do IMP.

Para converter o IMP em AMP o aspartato insere um grupo amina no lugar de um Oxigênio carbonílico no C-6.

1.4.3 - MECANISMO DA VIA DE SÍNTESE *DE NOVO*

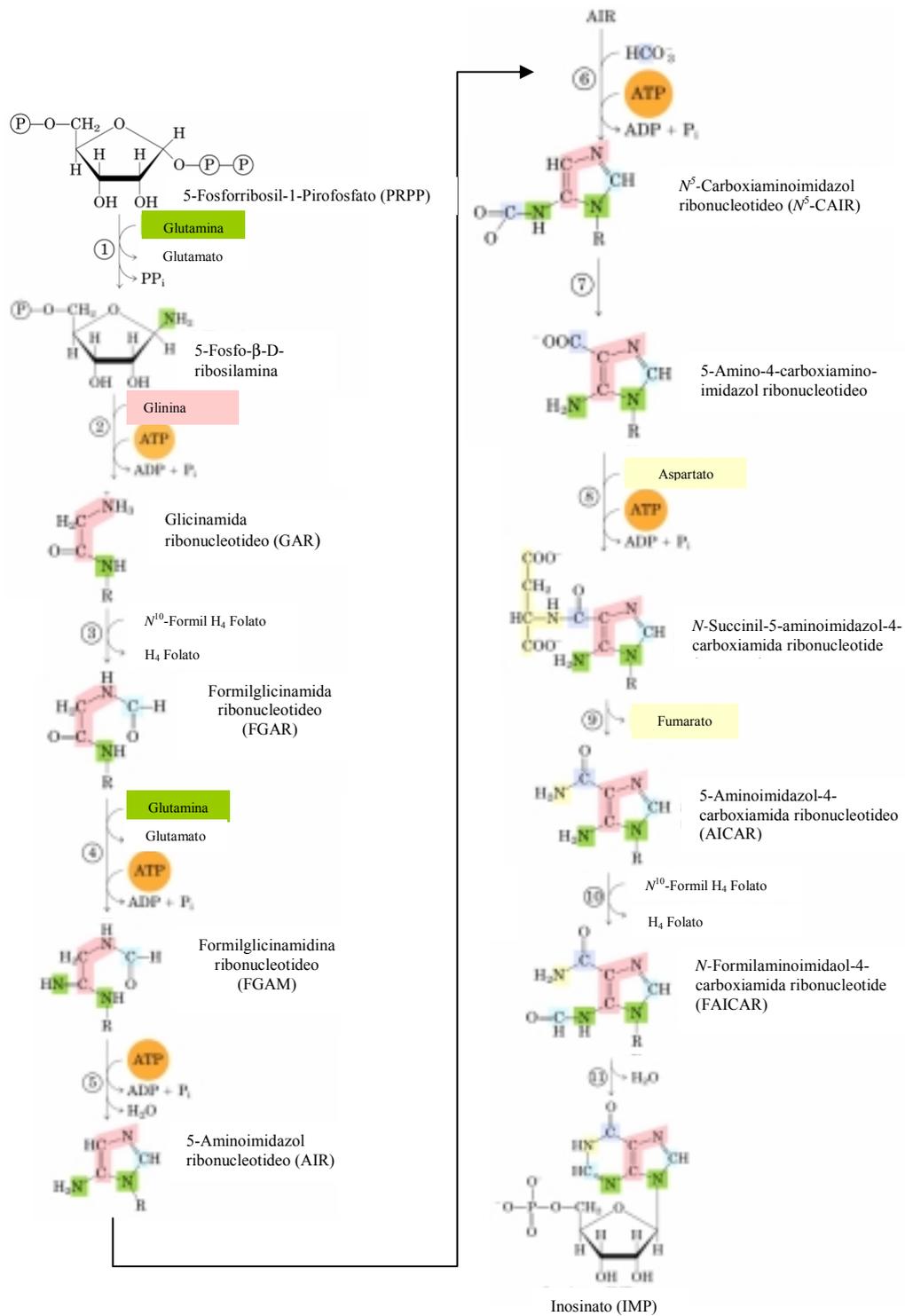


FIG16: Esquema representando o mecanismo da via da síntese *de novo*. Têm-se os onze passos enzimáticos iniciando com a PRPP e culminando para a formação de IMP

1.4.4 - VIA DE RECUPERAÇÃO

A via de recuperação é uma segunda via metabólica (fig17) que utiliza nucleotídeos pré-formados (adenina, guanina e/ou hipoxantina) derivados da degradação de ácidos nucléicos ou nucleotídeos livres. Na via de recuperação [10] as purinas livres reagem diretamente com o PRPP sendo convertidas nos nucleotídeos-5'-monofosfatos, AMP e IMP.

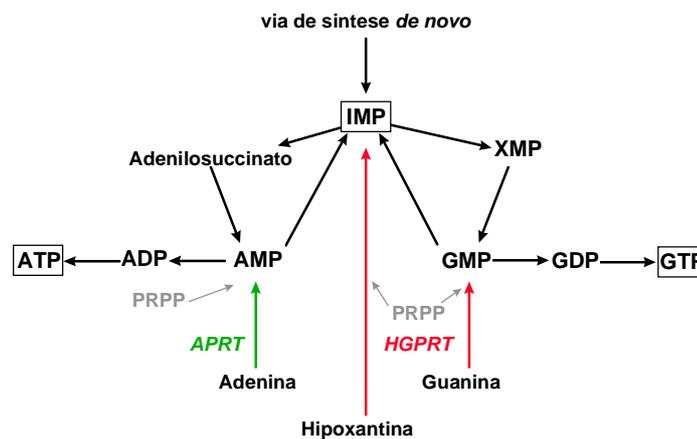


FIG17: Via de recuperação e interconversão, em células de mamíferos. APRT-adenina-fosforribosil-transferase; HGPRT-hipoxantina-guanina-fosforribosil-transferase; PRPP-5'-fosforribosil-1-pirofosfato.

Estas reações são catalisadas por enzimas distintas: adenina-fosforribosil-transferase (APRT), hipoxantina-guanina-fosforribosil-transferase (HGPRT) e em alguns organismos a xantina-fosforribosil-transferase (XPRT) nas reações reversíveis mostradas na figura 17 [10, 11].

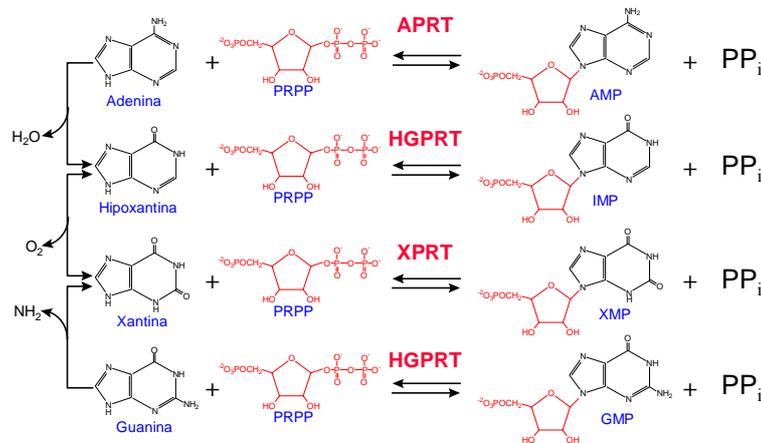


FIG18: As diversas reações de transferência catalisadas pelas PRTases estão mostradas. PRPP 5'-fosforribosil-1-pirofosfato, APRT adenina-fosforribosil-transferase; HGPRT hipoxantina-guanina-fosforribosil-transferase; XPRT xantina-guanina-fosforribosil-transferase; PP_i pirofosfato.

1.4.5 - RETROINIBIÇÃO DA BIOSÍNTESE DE NUCLEOTÍDEOS PURÍNICOS

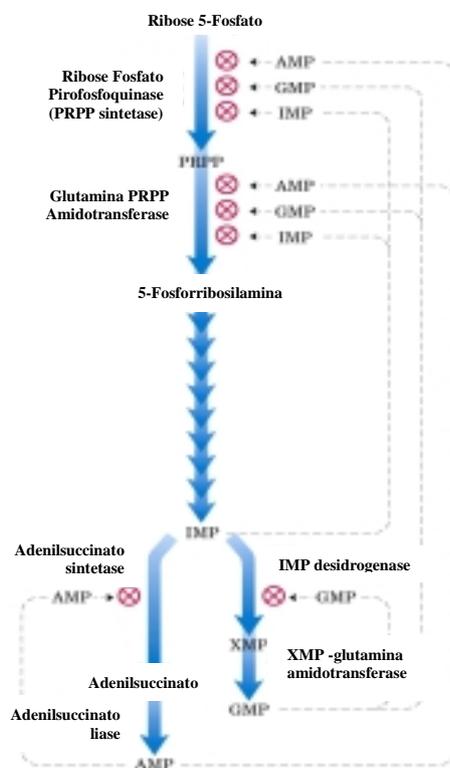


FIG19: Mecanismo de retroinibição da síntese de purinonucleotídeos

A retroinibição é um mecanismo de regulação da biossíntese dos purinonucleotídeos em vários pontos (fig19). Cada produto formado pode ter papel inibitório em passos diferenciados da síntese. Verifica-se que os produtos AMP, GMP e IMP inibem a enzima PRPP sintetase que deixa de produzir o substrato PRPP, doador da unidade ribose fosfato aos nucleotídeos. A seguir se tem a etapa regulatória da síntese, que é a formação

da 5-fosforribosilamina pela conversão da PRPP, através da transferência da amina da cadeia lateral da glutamina. Este ponto também pode ser inibido pelos produtos AMP, GMP e IMP. É importante salientar que o AMP e GMP atuam sinergisticamente na inibição da enzima glutamina-PRPP-amidotransferase. O último local de retroinibição é na síntese do IMP que é um ponto de ramificação na biossíntese de AMP e GMP. O produto AMP inibe a adenilsuccinato sintetase não ocorrendo à conversão do IMP no seu precursor imediato. O mesmo acontece como o GMP, o qual não converte o IMP em seu precursor xantilato. Para converter o IMP em AMP é necessário o substrato GTP, já para converter IMP em GMP é necessário o substrato ATP. Relacionando reciprocamente estes dois substratos verifica-se um equilíbrio na síntese destes dois ribonucleotídeos.

1.5 - A ENZIMA FOSFORRIBOSIL PIROFOSFATO SINTASE (PRPP sintase)

Foi descoberta em pombos, e mais adiante foi sendo identificada em galinhas, ratos e em *E. coli*. Sua ocorrência também tem sido demonstrada células tumorais [12, 13] espécies microbianas [14, 15] e em tecidos humanos [16].

Pouco se conhece a respeito de sua localização subcelular em células eucarióticas. Tem sido mostrado [17] que 80% da atividade está localizada nas frações mitocondriais e 15% na fração nuclear. Um resultado muito interessante é a provável localização citoplasmática da PRPPase [18]. Sugere-se também que a PRPPase possa ocorrer nos espaços periplasmáticos de bactérias [19]. Em todos os casos a maior parte da atividade enzimática da PRPPase parece estar localizada no citoplasma.

Acredita-se que a PRPP sintetase é essencial a todos os organismos de vida livre, e conseqüentemente estes organismos possuem no mínimo um gene *PRS* específico que codifica para a esta enzima. Apenas dois microorganismos parasitas intracelulares *Chlamydia trachomatis* e *Rickettsia prowazekii* não possuem PRPP sintetase [20, 21].

Em estudos realizados com organismos eucariontes verificam-se a presença de mais de um gene *PRS*. Os mamíferos *Homo sapiens* e *Rattus norvegicus* possuem três e duas isoformas respectivamente [23, 22]. A levedura *Saccharomyces cerevisiae* possui cinco isoformas [24], produtos os quais são homólogos as PRPP sintetases de mamíferos e bactérias [25, 26]. O seqüenciamento completo dos genomas de *Schizosaccharomyces pombe* e do nematódeo *Caenorhabditis elegans* revelaram a existência de duas isoformas [27].

Há também trabalhos que visam à caracterização detalhada da PRPP sintetase de bactérias como *Escherichia coli* e *Salmonella typhimurium* [28, 29]. Em organismos mutantes de *E. coli*, que crescem na ausência da atividade desta enzima quando suplementado com compostos da PRPP requeridos na via como nucleosídeos purínicos e pirimidínicos, NAD, His e Trp [30].

Em geral, as PRPP sintetases usam ATP como doador difosforil tendo como substrato MgATP. As enzimas de *E. coli*, *S. typhimurium*, e mamíferos também requerem Mg^{2+} livre. Todas as enzimas PRPP necessitam de P_i para sua atividade catalítica e para sua estabilidade [29, 31, 32]. Mas essa dependência ao P_i apresenta um comportamento diferenciado em algumas isoenzimas de plantas.

Esta enzima apresenta-se em muitas formas ativas e tem diferentes efetores: Mg^{2+} e fosfato inorgânico (P_i) como ativadores; e ADP, 2,3-difosfoglicerato

ou GDP como inibidores competitivos ou não competitivos. Sabe-se que em bactérias a enzima é altamente inibida pelo ADP que se liga ao sítio ativo da enzima. Em *Bacillus subtilis* e mamíferos essa inibição só ocorre pelo GDP e também pelo ADP [33].

A enzima se apresenta em oligômeros com subunidades que muitas vezes variam de 33 a 40 kDa. Estudos estruturais da PRPP sintetase de *Bacillus subtilis*, em um complexo com análogos do ativador fosfato e inibidor alostérico, demonstraram que a forma funcional da proteína é um hexâmero. Suas subunidades individuais enovelam-se em dois domínios. O sítio ativo está localizado entre esses domínios e inclui resíduos das duas subunidades. Em adição identificaram-se também resíduos importantes para a ligação de substratos e efetores, sugerindo que a estrutura possui um novo modelo de regulação alostérica [33]. Outros organismos apresentam a PRPP sintetase sob forma de tetrâmero [34] ou octâmero [35].

Quando os estudos envolvem a PRPP sintetase de plantas tem-se que algumas das propriedades diferem da “clássica” PRPP sintetase de outros organismos, principalmente bactérias e mamíferos, e que esta engloba uma nova classe de enzima.

Pesquisas desenvolvidas com plantas como *Arabidopsis thaliana* [27] tem revelado um número de cinco novos genes desta família. Estes cinco genes *PRS* variam em comprimento de 1190 a 1382 pb. As isoenzimas 1, 2 e 5 da PRPP sintetase apresentam 96% de similaridade quando comparadas com as demais isoenzimas desta espécie. Por sua vez as isoenzimas 3 e 4 são 71% similares. Pode-se dizer que a *Arabidopsis thaliana* possui duas classes de isoenzimas *PRS* que compreendem as enzimas 1, 2, 5 em uma classe e 3 e 4 em outra, atualmente descritas. A principal

característica destas classes é que as isoenzimas 3 e 4 aparecem independentes do fosfato (P_i) para a sua atividade.

Realizaram-se também trabalhos com *Spinacia oleracea*, verificando que esta apresenta quatro novas isoenzimas PRS com características similares a *Arabidopsis thaliana*. Da mesma forma as isoenzimas 1 e 2 requerem P_i para atividade catalítica enquanto as demais apresentam-se independentes do P_i . A PRS2 localiza-se nos cloroplastos, a PRS3 pode ser transportada para mitocôndria, e a PRS4 pode ser localizada no citossol. Análises da seqüência dos genes das isoenzimas 1 e 2 e as isoenzimas 3 e 4 apresentam 88% e 75% de similaridade respectivamente [36]. Através da literatura verifica-se que tanto as isoenzimas 3 e 4 de *Arabidopsis thaliana* quanto *Spinacia oleracea* fazem parte da nova classe de enzimas PRPP específicas de plantas que independem de P_i para sua atividade catalítica. E esta nova classe recebeu o nome de PRPP sintase.

A PRPP sintase é uma das mais importantes enzimas da via da síntese *de novo*, sendo responsável pela síntese do substrato PRPP utilizado em todas as reações da PRTases, tornando-se de central importância no metabolismo celular, principalmente porque participa de oito diferentes vias metabólicas [37]. Atualmente tem sido pouco descritas em plantas, o que fez com que esta enzima tornar-se o principal alvo de estudo deste trabalho. O conhecimento desta enzima da cana-de-açúcar envolvida na síntese de purinas abre possibilidades de uso de tal enzima no descobrimento de drogas que combatam agentes fitopatogênicos.

OBJETIVOS

2 - OBJETIVOS

Este trabalho tem como objetivo estudar uma das enzimas envolvidas na via de purinonucleotídeos da cana-de-açúcar como modelo na caracterização físico-química e possivelmente estrutural da Fosforribosil Pirofosfato Sintase (PRPP sintase). Para isso alguns objetivos específicos foram traçados:

- ✚ Clonagem do gene *prs* que codifica para a proteína PRPP sintase.
- ✚ Seqüenciamento completo do gene.
- ✚ Análise comparativa a partir de alinhamentos da seqüência completa do gene. *PRS* da cana-de-açúcar com genes *prs* de plantas e outros organismos.
- ✚ Análise filogenética.
- ✚ Expressão do gene *prs* em *Escherichia coli*.
- ✚ Purificação da proteína PRPP sintase.
- ✚ Caracterização da proteína PRPP sintase.

MATERIAL E MÉTODOS

3 - MATERIAIS E MÉTODOS

O trabalho foi desenvolvido de acordo a metodologia descrita na figura 20:

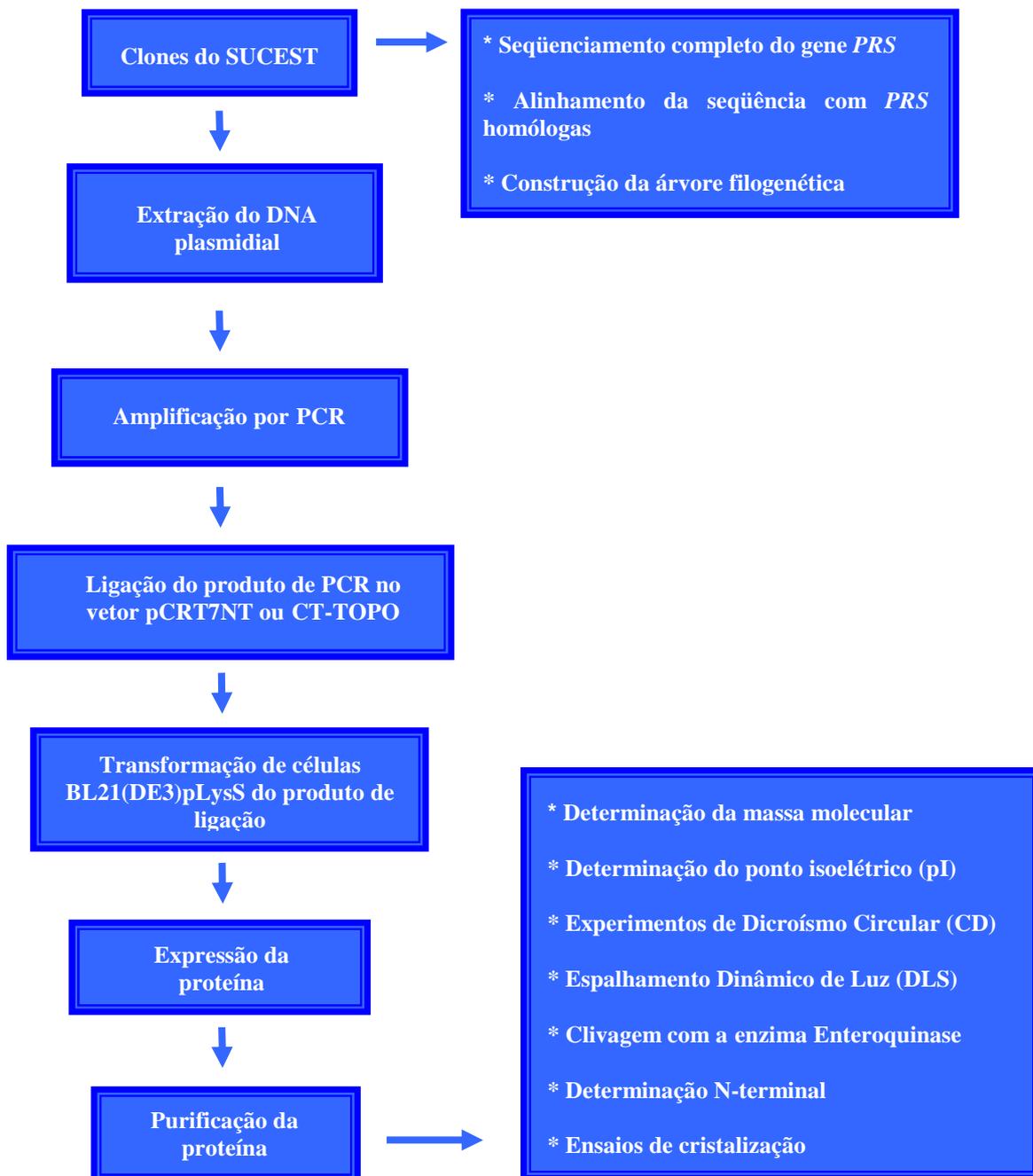


FIG20: Representação esquemática da estratégia para caracterização das proteínas da via da síntese *de novo*.

3.1 - ESTUDOS REALIZADOS COM A SEQUÊNCIA DO GENE *prs* DA CANA-DE-AÇÚCAR

3.1.1 - OBTENÇÃO DOS CLONES

A estratégia utilizada pelo “Data Mining” [6] foi aplicada com o intuito de identificar “clusters” da cana-de-açúcar para cada passo da via da síntese *de novo* de purinas. Dessa forma identificaram-se diversos clones da cana-de-açúcar pertencentes ao metabolismo de purinonucleotídeos, obtidos de seqüências expressas (EST) depositadas no banco de dados do SUCEST. Estes clones referem-se as seguintes enzimas da via de síntese *de novo* de purinas:

- 🌈 5-fosforribosil-1-pirofosfato sintetase (PRPPase);
- 🌈 fosforribosilglicinamida transformilase (GARase);
- 🌈 fosforribosilamidoimidazol-succinocarboxiamida sintetase (SAICARase);
- 🌈 fosforribosil-aminoimidazol sintetase (AIRase).

Os critérios utilizados para escolha destes clones foram:

- 🌈 Identidade com proteínas de *Arabidopsis thaliana* e outras plantas;
- 🌈 Grau de identidade;
- 🌈 N-terminal contido no EST da cana-de-açúcar mesmo não ocorrendo o alinhamento.

Os clones consistentes com estes critérios foram requisitados a Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP/SUCEST) de acordo

com as normas estabelecidas para transferência de material biológico e estão apresentados na tabela 2.

TABELA2: Clones a serem analisados.

NOME	CLONE	PROTEÍNA	FONTE	FASE LEITURA
A	SCCCCL3120E04	SAICAR	ARABIDOPSIS	2
B	SCCCCL3140E09	AIR	ARABIDOPSIS	-
C	SCCCRZ1003F04	AICAR	GALINHA	2
1	SCJLRZ1024F09	PRPP	ESPINAFRE	1
2	SCJLAM2093F02	PRPP	ESPINAFRE	2
3	SCACRZ3035C11	PRPP	ARABIDOPSIS	1
4	SCCCST3C04C09	GAR	ARABIDOPSIS	-
5	SCQSAM2100E05	SAICAR	ARABIDOPSIS	2
6	SCEQRT2095D07	AIR	ARABIDOPSIS	1
7	SCEQRT2025F06	SAICAR	ARABIDOPSIS	-
8	SCBGLR1047H09	SAICAR	ARABIDOPSIS	3
9	SCVPRT2083H04	PRPP	ARABIDOPSIS	1
10	SCVPRT2083H10	PRPP	ARABIDOPSIS	-

SC – SugarCane; **CC, JL, AC, QS, EQ, BG** e **VP** – código dos laboratórios envolvidos no projeto genoma; **CL3, RZ1, AM2, ST3, RT2** e **LR1** – código da biblioteca de cDNA; **120** – código da placa; **E04** – posição da placa.

3.1.2 - EXTRAÇÃO DO DNA PLASMIDIAL DOS CLONES OBTIDOS.

Uma vez selecionado os clones realizou-se a extração do DNA plasmidial seguida de duas etapas: a) provocou-se a lise das membranas celulares permitindo a liberação do DNA; b) realizou-se um ou mais tratamentos químicos e enzimáticos para eliminar contaminantes como DNA genômico, RNA, proteínas, macromoléculas, etc.

Fez-se um inóculo das colônias crescidas em meio LB (“Luria Bertane Medium”) sólido referente a cada clone adquirido em 5mL de meio de cultura LB

líquido [38] contendo 100 µg/mL (concentração final) de ampicilina {EMS Indústria Farmacêutica} deixando crescer durante a noite sob agitação a 300 rpm a 37°C no aparelho “C24 Incubator Shaker” {New Brunswick Scientific}. Após o crescimento das células, 500 µL de cada cultura foi estocada a -80°C em solução estéril glicerol 65% contendo Tris/HCl 25 mM pH 8,0 e sulfato de magnésio 100 mM na proporção 1:1.

Com o restante da cultura realizou-se a extração do DNA plasmidial de cada clone utilizando o protocolo “Wizard Plus SV Miniprep DNA Purification Systems” {Promega} de acordo com as instruções do fabricante. A extração foi analisada por eletroforese em gel de agarose {BioRad} 0,8% TAE [1X] em cuba horizontal da “Life Technologies Gibco BRL Horizontal Gel Electrophoresis Apparatus” a 80V. Utilizou-se 1kb “Plus DNA Ladder” {Gibco BRL} como padrão de massa molecular. Em seguida cada DNA extraído foi seqüenciado com oligonucleotídeo T7 para posterior análise.

3.2 - SEQÜENCIAMENTO

A estratégia de seqüenciamento utiliza-se do mecanismo de síntese do DNA pela DNA polimerase. Este método também é chamado de método da terminação da cadeia [39], pois usa um procedimento enzimático para sintetizar a cadeia de DNA de comprimento variado, interrompendo a síntese do DNA com uma das quatro bases (A, G, C, T) marcadas originando o tamanho do fragmento randomicamente. Cada reação de seqüenciamento necessita de um DNA molde; uma pequena seqüência de nucleotídeos (~20 pares de bases) chamada de

oligonucleotídeo (“primer”); da DNA polimerase que inicia a síntese de uma nova fita de DNA, no qual o oligonucleotídeo é hibridizado ao molde; uma solução chamada “Big Dye Terminator” {Perkin Elmer} que contém dNTPs (desoxirribonucleotídeos) e ddNTPs (dideoxirribonucleotídeos). A hidroxila 3’ do oligonucleotídeo reage como deoxinucleotídeo trifosfato (dNTP) formando uma nova ligação fosfodiéster. Os ddNTPs não apresentam a hidroxila no carbono 3, dessa forma interrompem o crescimento da cadeia, gerando fragmentos de diferentes tamanhos com o último nucleotídeo marcado com uma fluorescência de comprimento de onda específica (fig21). Os fragmentos de DNA marcados são separados por eletroforese em gel de poliacrilamida 5% e lidos pelo Sequenciador Automático “ABI 377 DNA Sequencer” {Perkin Elmer}.

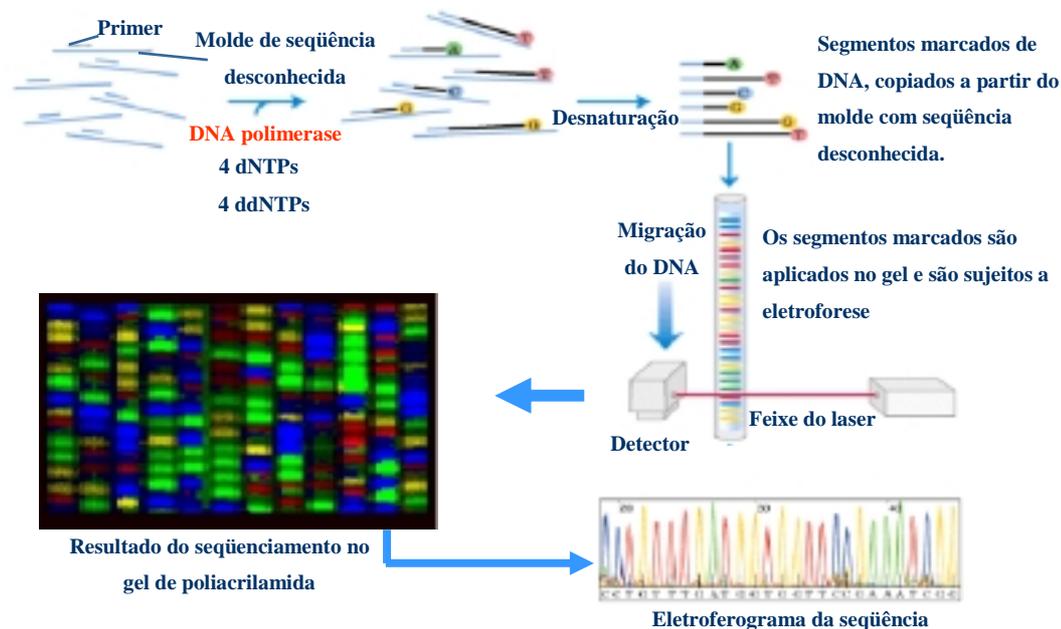


FIG21: Estratégia para reação de sequenciamento automatizado de DNA [40]. Esta estratégia utiliza o método de Sanger que faz uso do mecanismo da síntese do DNA pela DNA polimerase.

A partir da extração do DNA de cada clone fez-se a reação de seqüenciamento (tabela 3) por PCR (Reação em Cadeia da Polimerase) [41], com o objetivo de confirmar os clones adquiridos.

TABELA3: Mistura da reação de seqüenciamento.

Mistura de reação de PCR	Quantidade
DNA (estimado através de gel de agarose 1%)	200- 500ng
Oligonucleotídeo (5pmol/ μ l)	2 μ l
“Big Dye TM Terminator” (dNTP+ ddNTP* + Taq polimerase)	4 μ l
H ₂ O q. s. p.	10 μ l

* N= bases A, G, C, T.

A mistura de reação foi executada em um termociclador “PTC-100TM Programmable Thermal Controller” {MJ Research Inc.} sendo submetida ao programa mostrado na (tabela 4).

TABELA4: Condições do ciclo da reação de seqüenciamento.

Condições do ciclo da reação						
CICLO	1x	35x			1x	
T ^o C	96	96 ¹	50 ²	60 ²	72 ³	4
TEMPO	2 min	45 s	30 s	4 min	10 min	infinito

1-Desnaturação

2- Anelamento

3- Extensão

Para realização deste procedimento torna-se necessário sintetizar oligonucleotídeos específicos que possibilitem seqüenciar as extremidades 5' e 3' e também segmentos internos dos genes. Os oligonucleotídeos sintetizados {Gibco BRL} estão listados na tabela 5.

TABELA5: Oligonucleotídeos sintetizados para a reação de seqüenciamento.

OLIGONUCLEOTÍDEO	SEQÜÊNCIA
T7	5'-TAATACGACTCACTATAGGG-3'
SP6	5'-ATTTAGGTGACACTATAG-3'
OligoA	5'-AGCTCCCTGACGCAGAAAATATC-3'
OligoB	5'-ACTGGTAGATGGGCTTGAATC-3'

A estratégia utilizada no seqüenciamento dos clones empregando os oligonucleotídeos da tabela 3 pode ser visualizada no esquema da figura 22:

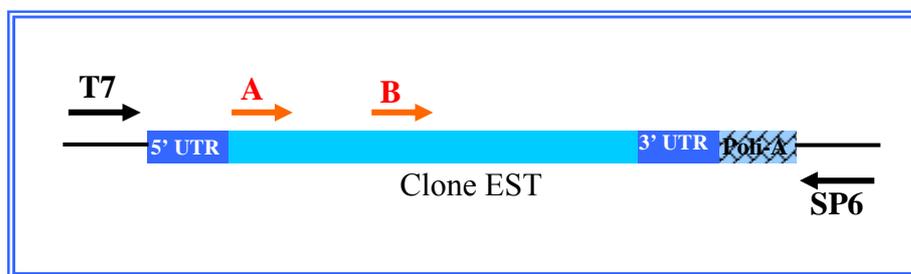


FIG22: Estratégia de seqüenciamento. T7, A, B e SP6 são os oligonucleotídeos utilizados para seqüenciar as extremidades do gene e sua região interna. 5' e 3' UTR representam as regiões não transcritas do gene. Poli A – região de poliadenilação do gene.

3.2.1 - PROTOCOLO DE PRECIPITAÇÃO DA REAÇÃO DE SEQÜENCIAMENTO

A precipitação do DNA marcado foi realizada com a finalidade de eliminar os reagentes da reação, permitindo assim a obtenção do DNA isento de impurezas. Utilizou-se isopropanol {Mallinckrodt} para realizar a precipitação uma vez que este apresenta uma ação precipitante bastante eficaz. Alguns sais, entretanto são pouco solúveis em isopropanol e acabam precipitando junto com o DNA, sendo necessário realizar lavagens com etanol {Mallinckrodt} para eliminar esses tipos sais.

Após a precipitação, as amostras de DNA foram ressuspensas em 3 µL de tampão contendo 50 mg de “Blue Dextran 2000” {Sigma}, 1 mL Formamida {Merck} e EDTA {Sigma} 0,5 M pH 8,0 e desnaturadas a 96 °C por 2 min. Aplicou-se 1 µL das amostras no gel de seqüenciamento composto por Uréia {Sigma} 6 M, “Long Ranger 5% Gel Solution” {BMA}, TBE [1X] {USB}, Persulfato de amônia 0,05% {Amresco}, Temed 0,07% {Amresco}. A eletroforese foi realizada por sete horas no seqüenciador automático “ABI Prism 377 DNA Sequencer” {PE Applied Biosystems}. O resultado do gel de seqüenciamento foi fornecido através de programa “ABI Prism 377-96 Collection” {PE Applied Biosystems} responsável pela coleta dos dados seqüenciados. Em seguida as seqüências foram analisadas através do programa “Sequencing Analysis V. 3.4” {PE Applied Biosystems}, que captura os dados coletados de cada amostra corridas no gel e automaticamente extrai as bases de cada seqüência, que são fornecidas em arquivo texto ou em forma de eletroferogramas (fig21).

3.3 - ANÁLISE DAS SEQÜÊNCIAS OBTIDAS

De posse dos vários eletroferogramas obtidos do seqüenciamento de cada gene foi possível alinhar e corrigir as seqüências eliminando as bases não identificadas através do programa “SeqMan” pertencente ao software DNASTAR (Lasergene Sequence Analysis Software – DNASTar) obtendo-se uma única seqüência consenso.

Com as seqüências consenso de nucleotídeos, alinhou-se cada uma delas com as seqüência de nucleotídeos do banco de dados do SUCEST através do programa BLAST (site: www.ncbi.nlm.nih.gov/gorf/wblast2.cgi), utilizando o “BLAST Two Sequences” (BLASTn) com o objetivo de verificar se os clones adquiridos correspondiam aos clones solicitados a FAPESP/SUCEST.

Utilizando o programa “Translate” (<http://www.expasy.ch/>) cada seqüência consenso foi traduzida, obtendo-se as fases de leitura (“Frames”) de cada gene. Cada fase de leitura foi alinhada com a proteína de *A. thaliana* e *S. oleracea* do banco de dados do SUCEST, através do programa BLASTp, verificando a fase de leitura de cada proteína que continha códon iniciador (ATG-metionina).

3.4 - ALINHAMENTO DA SEQÜÊNCIA DO GENE *prs* DA CANA-DE-AÇÚCAR COM OUTROS ORGANISMOS

Utilizando uma das ferramentas do programa BLAST (“Blast two sequences”) realizou-se um alinhamento dois a dois da seqüência da enzima PRPP sintase da cana-de-açúcar com plantas homólogas como a *Arabidopsis thaliana*, *Spinacia oleracea*, *Oryza sativa*, e com outros organismos como *Homo sapiens*, *Saccharomyces cerevisiae*, *Escherichia coli*, construindo uma tabela de “scores” e de valores de identidade e similaridade (tabela 17), realizando análises comparativas.

A seqüência de aminoácidos da enzima PRPP sintase da cana-de-açúcar (número de acesso AF536565) e PRSSs homólogas a partir das isoformas 1 a 5 de *Arabidopsis thaliana* (CAA58717; CAA 63552; CAB43552; CAB43553 e

AC004521 respectively [27]) , isoformas 1 a 3 de *Homo sapiens* (BAA00733 [42]; P11908 [43] e AAB59463 [44]), isoformas 1 a 5 de *Saccharomyces cerevisiae* (CAA49674; CAA52436; CAA52437; CAA84888 e NP_014580 [25]), *Rattus norvegicus* (NP_036766 [45]), *Escherichia coli* (BAA36065 [28]), *Oryza sativa* (BAB89498) [46], isoformas 1 to 4 de *Spinacia oleracea* (CAB43599; CAB43600; CAB43601 e CAB43602 [36, 47]), *Mycoplasma pneumoniae* (P75044 [48]), *Synechocystis sp* (NP_442534), *Schizosacharomyces pombe* (NP593826), e *Candida albicans* (P46585) foram alinhadas através do programa CLUSTALX [49]. Mais adiante refinamentos manuais foram realizados para otimizar o alinhamento, considerando a estrutura secundária predita a partir de polipeptídios individuais. O programa “Predict Protein” (EBI com <http://www.ebi.ac.uk/searches/searches.html>) e o programa “PHDsec” foram usados para deduzir estas predições da estrutura secundária.

O programa Gblocks [50] foi usado para eliminar posições de alinhamento ruim e regiões divergentes.

3.5 - CONSTRUÇÃO DA ÁRVORE FILOGENÉTICA

A construção da árvore filogenética foi realizada a partir uso dos dados do alinhamento das seqüências obtidas através do programa “CLUSTAL X” juntamente com a utilização de vários outros programas utilizados para estudos filogenéticos.

O conjunto de dados resultante do alinhamento foi analisado pelos métodos da máxima verossimilhança - ML (“Maximum-Likelihood Method”)

através do algoritmo do quarteto “puzzling” implementado pelo programa TREE-PUZZLE [51], pelo método da Parcimônia usando o programa PROTPARS (PHYLIP) [52], e pelo método da Distância utilizando os programas PROTDIST e NEIGHBOR (PHYLIP). A estabilidade dos nós da árvore foi testada pelo “Bootstrapping” (1000 ciclos). Para obtenção de uma árvore consenso usou-se o programa CONSENSE (PHYLIP). E para visualização da árvore consenso obtida utilizou-se do programa TreeView [53].

3.6 - CLONAGEM DO GENE *prs* NO VETOR DE EXPRESSÃO pCRT7 NT/CT TOPO

A clonagem de DNA é um procedimento de amplificação de fragmentos idênticos do DNA alvo. As bactérias contêm um DNA circular conhecido como plasmídeo, que se replica de forma autônoma. Assim, genes de interesse podem ser introduzidos nestes plasmídeos e clonados em bactérias. Esses plasmídeos são usados como vetores de clonagem e possuem o que chamamos de fatores de resistência a antibióticos. Esses antibióticos são responsáveis pela seleção das bactérias que contêm os plasmídeos.

Para a produção de um DNA recombinante tanto o plasmídeo quanto os genes de interesse são digeridos por enzimas de restrição que reconhecem seqüências específicas do DNA (seqüências de reconhecimento) gerando extremidades compatíveis. Estas extremidades sendo do tipo “adesivas” podem ligar-se as duas fitas do DNA (plasmídeo e inserto) uma vez que é complementar. Dessa forma pode então ocorrer à ligação de ambas extremidades já que estas

possuem seqüências complementares. Os vetores juntamente com seus insertos são introduzidos em células bacterianas as quais tornam-se transformadas, possibilitando obter milhões de cópias de DNA de interesse incorporado ao plasmídeo.

3.6.1 - CONSTRUÇÃO DOS OLIGONUCLEOTÍDEOS E AMPLIFICAÇÃO DO GENE

A construção de oligonucleotídeos específicos para a extremidade 5' dos genes tornou-se necessária para otimizar as condições de amplificação por PCR [41] da região genômica de interesse e a purificação deste DNA amplificado. Os oligonucleotídeos {GIBCO BRL} sintetizados estão listados na tabela 6.

TABELA 6: Oligonucleotídeos utilizados na amplificação

PROTEÍNA	CLONE	OLIGOS	SEQUÊNCIA
PRPPase	RZ1024F09	TOPO1	ATGAGGAACGACACGAGGCTGCGC
	RZ3035C11	TOPO3(A)	ATGGAGGTCGTCGGCGCCGCGAAG
	RT2083H04	TOPO9	ATGGAGGTCGTCGGCGCCGCGAAG
AIRase	RT2095D07	TOPO6	ATGACCACCGACCGTCTCTCTCCA
SAICARase	AM2100E05	TOPO2	ATGTCCTCTTCTTCTCTCCGC

Estes oligonucleotídeos foram utilizados na reação de amplificação por PCR dos genes em estudo. Este procedimento resume-se em utilizar dois pequenos fragmentos de DNA complementar as extremidades 3' e 5' do gene de interesse que irão hibridizar com esta região servindo de ponto de ancoragem

para a enzima Taq DNA-polimerase, responsável pela replicação do DNA (fig23).

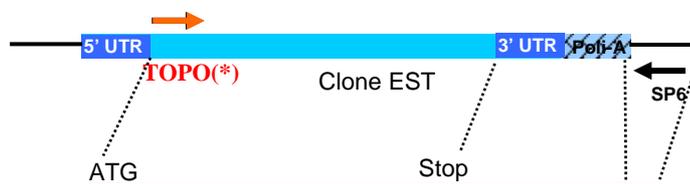


FIG23: Esquema representativo da estratégia da amplificação. Foi utilizado o DNA extraído. De cada um dos clones em estudo e o (*) representa cada um dos diferentes oligonucleotídeos construídos (tabela 4). 5' e 3' UTR representam as regiões não transcritas do gene. Poli A – região de poliadenilação de gene.

A reação de amplificação foi feita com os genes 1, 3 e 9 (PRPPase), 5 (SAICARase), 6 (AIRase) (tabela 2), seguindo as condições descritas na tabela 7.

TABELA7: Condições da reação de amplificação dos genes.

Mistura da reação de amplificação	Volume (μL)
Tampão sem Mg [10 x]	5
dNTP 10mM (N= A, G, C, T)	1
MgCl ₂ 50mM	1,5
Oligonucleotídeo SP6 100 pmol/ μL (3' do gene)	1
Oligonucleotídeo TOPO(*) 100 pmol/ μL (5' do gene)	1
DNA molde	1
Taq DNA Polimerase (5U/ μL)	0,2
H ₂ O q.s.p.	50

(*) = 1, 3(A), 9 oligonucleotídeos construídos para amplificar a PRPPase. Os oligonucleotídeos 5 e 6 para amplificar a SAICARase e a AIRase respectivamente.

A reação foi submetida a um programa de PCR (tabela 8), executado no termociclador PTC-100TM Programable Thermal Controller.

TABELA8: Condições do ciclo da reação de PCR para amplificação dos genes.

Condições do ciclo da reação						
CICLO	1x	25x			1x	
T ^o C	95	95	55	72	72	4
TEMPO	3 min	30 s	30 s	2 min	10 min	infinito

O resultado da amplificação foi analisado a partir da aplicação de 30 μ L da reação de amplificação em gel de agarose preparativo {BioRad} 0,8% TAE [1X].

3.6.2 - PURIFICAÇÃO DO DNA AMPLIFICADO

O gel de agarose apresenta uma matriz porosa que faz com que os fragmentos de DNA de diferentes massas moleculares migrem com velocidade, inversamente proporcionais ao seu tamanho. As moléculas de DNA migram para o pólo positivo, já que sua carga é negativa devido à ionização dos grupos fosfatos. Moléculas de menor massa molecular migram com maior facilidade, sendo possível dessa forma separar e estimar o tamanho do fragmento de DNA.

A purificação do DNA amplificado dos clones 1, 3 e 9 (PRPPase), 5 (SAICARase), 6 (AIRase) (tabela2), foi realizada em gel preparativo utilizando-se o “Prep-A-Gene DNA Purification Systems Master Kit” {BioRad} de acordo com as instruções do fabricante, a técnica consiste em extrair os fragmentos de DNA selecionados do gel de agarose, eliminando todas as impurezas e contaminantes. O resultado foi analisado em gel de agarose 0,8% TAE [1X].

3.6.3 - CLONAGEM DO GENE *prs* NO VETOR pCRT7 CT/NT TOPO

O “pCR T7 TOPO TA Cloning Kits” {Invitrogen} permite a inserção direta do produto de PCR amplificado pela *Taq* DNA polimerase no vetor plasmidial de expressão. O DNA clonado neste vetor será expresso produzindo uma proteína recombinante em fusão N-terminal ou C-terminal com uma seqüência de 6 histidinas (fig24 e 25).

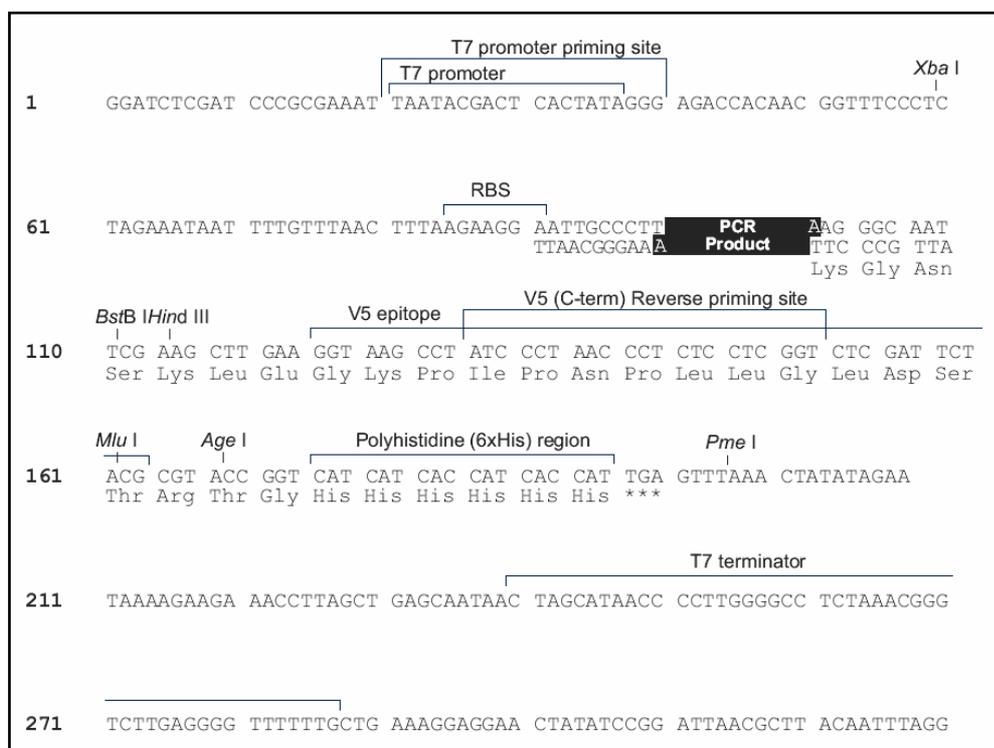


FIG24: Vetor pCRT7/CT TOPO {Invitrogen}. Este vetor é utilizado como vetor de clonagem e expressão contendo o tag de 6XHis na região C-terminal.

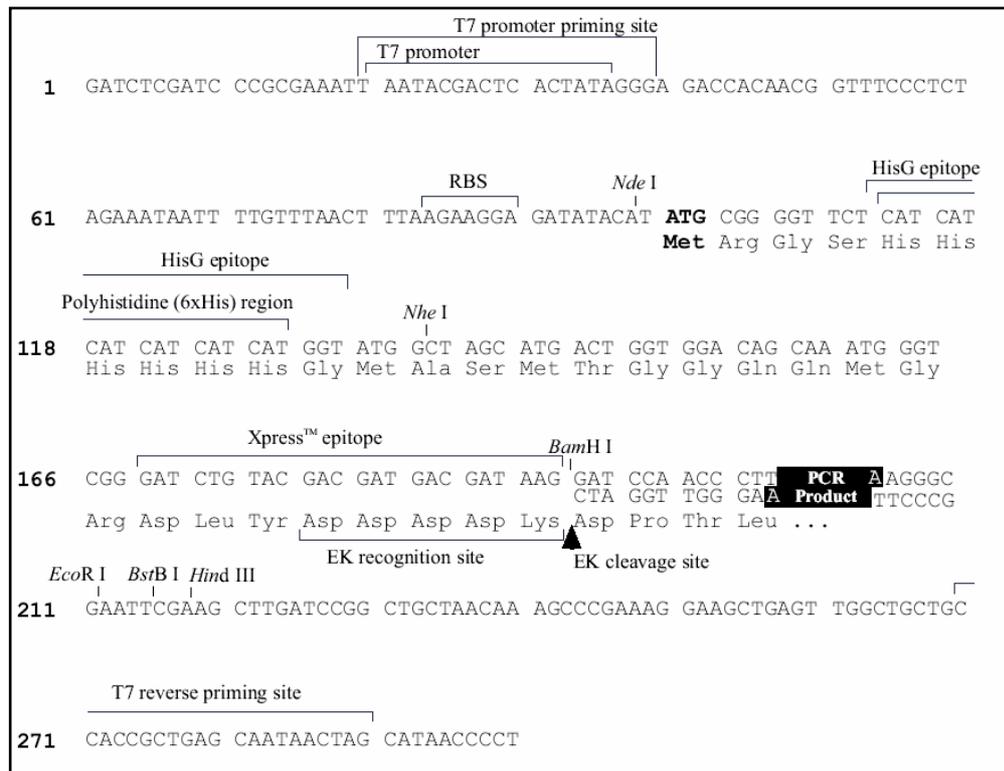


FIG25: Vetor pCRT7/NT TOPO {Invitrogen}. Este vetor é utilizado como vetor de clonagem e expressão contendo o tag de 6XHis na região N-terminal.

O vetor pCRT7 TOPO é um vetor linearizado que contém uma deoxitimidina (T) na extremidade 3' e uma topoisomerase covalentemente ligada a ela. A topoisomerase I liga-se ao DNA fita dupla com sítios específicos e cliva o esqueleto fosfodiéster depois do 5'-CCCTT em uma fita. A energia a partir da quebra do esqueleto é conservada pela formação de uma ligação covalente entre o fosfato 3' da fita clivada e um resíduo tirosil (Tyr-274) da topoisomerase. O fosfo-tirosil liga-se entre o DNA e a enzima podendo subsequenteemente ser atacado pelo 5'OH da fita original clivada, invertendo a reação e liberando a topoisomerase. A *Taq* polimerase adiciona uma desoxiadenosina (A) a

extremidade 3' do produto de PCR independente do molde de DNA permitindo que este produto se ligue eficientemente ao vetor, como apresentado na figura 26.

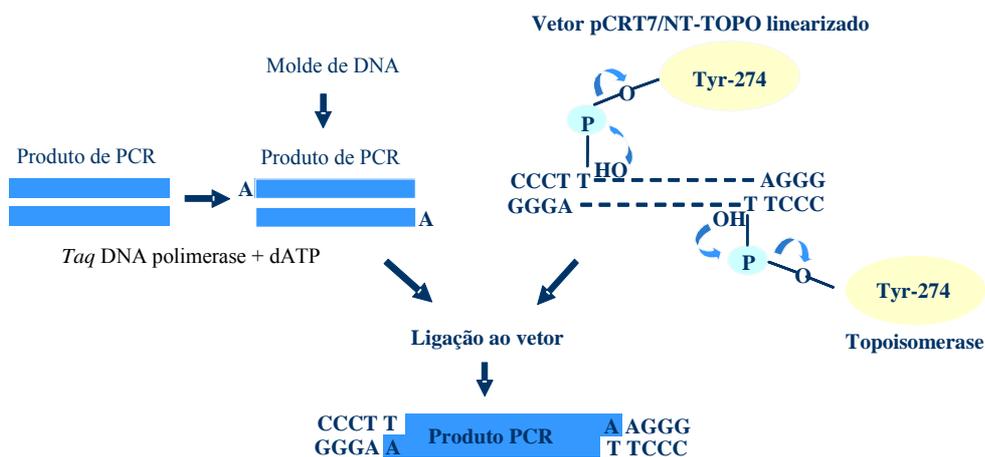


FIG26: Esquema utilizado na clonagem do gene *prs* da cana-de-açúcar.

Antes de realizar a reação de ligação do inserto ao vetor é necessário então adicionar desoxiadenosina (A) (tabela 9) às extremidades do inserto purificado para que o mesmo possa ligar-se covalentemente ao vetor de clonagem.

TABELA 9: Reação utilizada para adição de desoxiadenosina (A).

Mistura da reação de adição	Volume (μL)
DNA purificado	7
Taq polimerase 1U	1
dATP 10 mM	1
Tampão PCR [10x]	1
Total	10

A reação descrita na tabela 9 foi incubada em um termociclador por 15 min a 72°C.

Uma vez adicionada, a desoxiadenosina torná-se possível realizar a reação de ligação do inserto ao vetor pCRT7 CT/NT TOPO que segue descrita na tabela 10.

TABELA 10: Mistura de reação para inserção do produto de PCR ao vetor pCRT7 NT/CT TOPO

REAÇÃO DE LIGAÇÃO	VOLUME (μL)
Produto de PCR	4
Solução salina	1
Vetor TOPO	1
Volume total	6

soluções do kit TOPO TA Cloning {Invitrogen}

A reação descrita na tabela 9 foi incubada em um termociclador por 30 min a 22°C.

Para este experimento utilizou-se os clones 1, 3 e 9 (PRPPase), 5 (SAICARase), 6 (AIRase) (tabela2).

3.6.4 - TRANSFORMAÇÃO DO LIGANTE AO VETOR pCRT7 TOPO NT/CT

Na transformação as bactérias são submetidas ao choque térmico que facilita a entrada do DNA exógeno. Para aumentar a eficiência de transformação, tais bactérias passam por um tratamento com cloreto de cálcio [54] que aumenta a permeabilidade da membrana celular. As bactérias submetidas a tal tratamento passam a ser denominada célula competente.

Todo o volume da mistura de ligação (6 μ L) foi utilizado na transformação de células competentes *E. coli* TOP10F' ("Kit One Shot Chemical Transformation" – Invitrogen) de acordo com as instruções do fabricante. Estas células competentes não contêm o gene codificante para a T7 RNA Polimerase, promovendo um hospedeiro para propagação estável e manutenção do plasmídeo recombinante. Sabe-se que a presença da T7 RNA Polimerase pode conduzir a expressão do gene desejado na falta do indutor e se este gene for tóxico ao hospedeiro *E. coli*, poderá resultar a instabilidade do plasmídeo, crescimento reduzido ou a morte das células. Por isso é recomendado o uso prévio destas células competentes para caracterização, propagação e manutenção.

Uma vez transformada as células selecionou-se 4 colônias representativas de cada gene que foram inoculadas em 5 mL de meio LB contendo ampicilina (100 μ g/mL) e após uma noite de crescimento a 37°C a 300 rpm, 250 μ L da cultura foram estocados a -80°C. Do restante da cultura fez-se à extração do DNA.

Antes dos experimentos de expressão, a orientação de cada construção recombinante foi checada por amplificação de PCR com a combinação de oligonucleotídeos específicos **SP6**, **TOPO** e **T7**. As condições da reação foram às mesmas descritas nas tabelas 5 e 6. Os resultados de análise foram verificados a partir de gel de agarose 0,8% TAE [1X].

Esta análise de orientação de ligação do inserto ao vetor também foi realizada pelo seqüenciamento da extremidade 5' do gene através do oligonucleotídeo T7 como descrito na seção 3.2.

3.7 - EXPRESSÃO RECOMBINANTE DA PROTEÍNA CODIFICADA PELO GENE *prs* DA CANA-DE-AÇÚCAR.

Selecionou-se o produto recombinante pCRT7 TOPO CT e NT juntamente com um controle negativo (plasmídeo religado) e cada um dos mesmos foi submetido à transformação da linhagem de células competentes BL21(DE3) pLysS {Invitrogen}. A designação DE3 significa que a linhagem contém o lisógeno lambda DE3 que carrega o gene para a T7 RNA Polimerase sob o controle do promotor *lacUV5*. O IPTG (Isopropil β -D-tiogalactopiranosídeo – Invitrogen) torna-se necessário para induzir a expressão da T7 RNA Polimerase. O plasmídeo pLysS carregado por esta linhagem produz T7 lisozima que reduz o nível basal de expressão do gene de interesse. O gene pLysS confere resistência ao antibiótico cloranfenicol {AristonTM}. Esta linhagem é usada somente para expressão.

Os candidatos 1, 3 e 9 (PRPPase), 5 (SAICARase), 6 (AIRase) (tabela2) da proteína codificada pelo gene *prs* foram submetidas a testes de expressão.

Para realizar os testes de expressão fez-se um inóculo da colônia transformante em 10 mL LB contendo ampicilina (100 μ g/mL) e cloranfenicol (34 μ g/mL) incubado durante a noite (16 horas) sob agitação de 300 rpm a 37°C. Após crescimento das células inoculou-se 10 mL de meio LB líquido com 500 μ L do inóculo que ficou sob agitação a 37°C até atingir a D.O._{600nm} de aproximadamente 0,5. Antes da adição de IPTG a concentração final de 1mM retirou-se uma alíquota de 500 μ L que foi congelada a -20°C. Esta alíquota representa o controle não induzido. Decorrido o período de 5 horas de indução

retirou-se uma alíquota de 1 mL. As alíquotas retiradas foram centrifugadas por 2 min a 10.000 g, os sobrenadantes foram descartados. O precipitado do controle não induzido foi ressuspenso em 100 µL em tampão de amostra TA [1X] contendo β-mercaptoetanol {Pierce}. Já os precipitados resultantes das induções foram ressuspenso em 500 µL de Tampão de Lise (fosfato de potássio 500 mM pH 7, 8, 400 mM NaCl, 100mM KCl, 10% glicerol, 0,5% triton X-100 e 10 mM Imidazol) e submetidos a sonicação por 6 ciclos, sendo uma sonicação de 30 segundos com intervalos de descanso de 1 minuto usando o aparelho “550 Sonic Dismembrator” {Fisher Scientific}. As células, após a lise, foram centrifugadas separando a fração solúvel (sobrenadante) e insolúvel (precipitado) por uma centrifugação de 5 min a 15.000 g a temperatura ambiente. As alíquotas foram submetidas à eletroforese em gel de poliacrilamida {Pharmacia Biotech} 15% em condições desnaturantes pelo método de Laemmli [38], utilizando um sistema vertical Mini V 8.10 da Gibco-BRL. Os marcadores de massa molecular {Sigma} utilizados foram: albumina sérica bovina (66 kDa), ovalbumina (45 kDa), anidrase carbônica (30 kDa), inibidor de tripsina de soja (20,1 kDa) e citocromo c (12,4 kDa). O gel foi revelado com 0.010% de “Comassie brilliant blue” G-250 e descorado com ácido acético 10%.

Com base nos resultados dos testes de expressão e com o objetivo de otimizar a expressão da proteína foram testadas duas condições de temperatura de indução a 37°C e a 20°C, variando também a concentração de IPTG em 0; 0,1; 0,5 e 1mM. Uma vez estabelecidas às condições de expressão mais favoráveis, começou-se a expressar a proteína recombinante em larga escala. Esta expressão em larga escala foi feita a partir 3 litros de cultura (meio LB + Amp 100 µg/mL +

cloranfenicol 34 $\mu\text{g/mL}$) que depois de atingida a DO_{600} 0,5 foi induzida com IPTG 1mM por um período de indução de 5 horas a temperatura de 37°C a 250 rpm. As células foram centrifugadas por 15 min a 8000 g a 4°C . O sobrenadante foi descartado e o precipitado celular ressuspendido em 20 mL por litro de cultura em Tampão de Lise (fosfato de potássio 500 mM pH 7, 8, 400 mM NaCl, 100 mM KCl, 10% glicerol, 0,5% triton X-100 e 10 mM Imidazol) e submetido à lise celular como descrito anteriormente. A amostra lisada foi centrifugada a 4°C por 20 min a 20.000 g. O sobrenadante resultante foi então submetido à purificação e o precipitado descartado.

3.8 - PURIFICAÇÃO

Métodos cromatográficos têm sido utilizados para identificar e isolar proteínas de interesse. As proteínas são purificadas através de sucessivos fracionamentos considerando propriedades como solubilidade, carga iônica, massa molecular, ponto isoelétrico, propriedades de adsorção e especificidade de ligação a outras moléculas biológicas.

As técnicas cromatográficas empregam colunas verticais nas quais as amostras são colocadas de modo que as moléculas protéicas se distribuem em duas fases, uma fixa ou estacionária e outra móvel. A fase móvel corresponde ao solvente cujo fluxo arrasta a maioria das moléculas. A fase fixa por sua vez inclui um elemento imóvel (papel de filtro, resinas carregadas, camadas porosas, etc.) acondicionado de modo a interagir com a mistura de proteínas ou outras moléculas [55]. Os métodos cromatográficos mais utilizados são os da exclusão

por tamanho, troca iônica e afinidade [56]. Neste trabalho, para purificar a proteína de interesse utilizaram-se os métodos de fracionamento com sulfato de amônio, cromatografia de afinidade em resina de níquel e cromatografia por exclusão molecular em coluna Superdex 200 HR 10/30.

3.8.1 - PRECIPITAÇÃO COM SULFATO DE AMÔNIO

Realizou-se fracionamento com sulfato de amônio como primeiro passo na eliminação de contaminantes. A precipitação com o sulfato de amônio, também chamada “salting-out”, refere-se a solubilidade das proteínas que é diminuída em altas concentrações de sais. A relação de solubilidade com a concentração de sal difere de uma proteína para outra, a adição de um sal em altas concentrações pode seletivamente precipitar algumas proteínas enquanto outras permanecem em solução. O sulfato de amônio é frequentemente usado para esta proposta porque possui alta solubilidade em água e não interfere no enovelamento da maioria das proteínas. A precipitação com sulfato de amônio ocorre devido à neutralização das cargas da superfície da amostra pelo sal e pela diminuição da concentração efetiva da água [57].

Ao sobrenadante obtido após a clarificação do lisado adicionou-se inicialmente 30% de sulfato de amônio ficando sob agitação por 20 min a 4°C. Após dissolução total do sal centrifugou-se a amostra por 15 min a 20.000 g. O sobrenadante foi separado do precipitado que por sua vez foi ressuspenso em 10 mL de tampão de lise. Ao sobrenadante resultante foi adicionada a próxima concentração de sulfato de amônio aumentada de 5% repetindo-se todo processo

descrito acima por mais quatro vezes até que a última concentração de sulfato de amônio adicionada fosse de 55%. Durante todo o experimento alíquotas do sobrenadante e do precipitado foram retiradas e dialisadas contra o tampão A (Fosfato de potássio 50 mM pH 7,8, NaCl 400 mM e Imidazol 10 mM) e submetidas à eletroforese SDS-PAGE 15% para posterior análise.

3.8.2 - CROMATOGRAFIA LÍQUIDA DE AFINIDADE

Como o gene que expressa a proteína PRPP sintase foi clonado no vetor pCRT7 TOPO NT que possui uma cauda de 6XHis N-terminal que tem capacidade de ligação a metal divalente Ni^{2+} a purificação da proteína foi realizada em coluna de cromatografia de afinidade “Ni-NTA Superflow” {Quiagen}. Na cromatografia de afinidade as proteínas são separadas por suas especificidades de ligação. A cauda de 6XHis permite a ligação ao Ni-NTA (Níquel-Ácido Nitriloacético). A cauda não interfere com a estrutura ou função da maioria das proteínas. A interação da cauda 6XHis com a matriz Ni-NTA pode ser visualizada na figura 27.

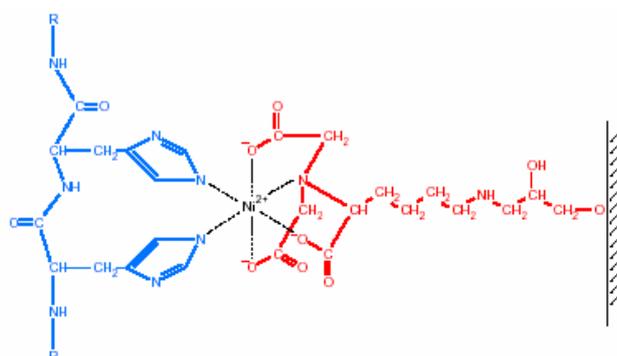


FIG27: Esquema da interação dos resíduos vizinhos no tag 6XHis e a matriz Ni-NTA.

Neste tipo de purificação utiliza-se o composto imidazol {Sigma} para realizar a eluição da proteína, já que o anel imidazólico é parte integrante da estrutura da histidina e este anel liga-se aos íons níquel imobilizado pelos grupos NTA na matriz (fig28).

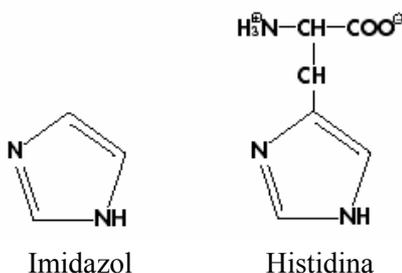


FIG28: Estrutura química do imidazol e da histidina.

Com base nestes critérios a proteína precipitada pelo sulfato de amônio ressuspendida em 20 mL em Tampão de Lise foi dialisada a 4°C em “Spectra/Por Molecularporous Dialysis Membrane” {Spectrum} com limite de exclusão molecular de 12 a 16 kDa, contra tampão **A** (Fosfato de potássio 50 mM pH 7, 8, NaCl 400 mM e Imidazol 10 mM) para retirada do sulfato de amônio. A amostra foi centrifugada por 15 min a 3000 g a 4°C antes de ser aplicada à coluna de afinidade imobilizada Ni-NTA (5 mL de resina). A coluna foi equilibrada com 4 volumes de coluna (cv) do tampão **A** e a amostra foi aplicada. Em seguida fez-se uma lavagem com 2 cv de tampão **A** para retirada das proteínas que não se ligaram a resina. A proteína recombinante foi eluída da coluna através da aplicação de um gradiente linear de 10 a 500 mM de imidazol {Sigma} em tampão **A** sob 100 min com um fluxo de 0,5 mL/min monitorada a 280 nm em um detector de absorvância LKB U.V. As frações eluídas foram analisadas em gel 15% SDS-PAGE.

3.8.3 - CROMATOGRAFIA POR EXCLUSÃO MOLECULAR EM SUPERDEX 200 HR 10/30

A Superdex é uma coluna para realização de cromatografia por exclusão molecular de proteínas e outras biomoléculas. A matriz da coluna é um polímero com poros de tamanho selecionado. Proteínas com maior massa molecular migram mais rápido do que proteínas com menor massa molecular, pois aquelas não entram nos poros passando através da coluna pelo volume de exclusão. As proteínas menores, por sua vez, entram nos poros da resina e saem vagarosamente através da coluna. A resolução da cromatografia por exclusão molecular é determinada por três características: distribuição do tamanho dos poros, volume do poro e tamanho da partícula. As resinas são quimicamente estáveis sobre pH de 3 a 12 durante o uso normal. Pode ser usada com qualquer sistema cromatográfico seja manual ou automatizado.

Nesta etapa de purificação utilizou-se a coluna Superdex 200 HR 10/30 {Amersham Pharmacia Biotech} de 24 mL de volume de coluna acoplada ao sistema automatizado de cromatografia AKTA Explorer10 {Amersham Pharmacia Biotech}. Esta coluna é usada em separações de proteínas na faixa de 10 a 600 kDa de massa molecular. As frações da amostra, após eluída da coluna de níquel e contendo a proteína, foram agrupadas e dialisada como anteriormente descrito em tampão **B** (fosfato de potássio 50 mM pH 7,8 contendo 150 mM NaCl). Após diálise, a amostra foi concentrada para 500 µl por ultrafiltração em centriprep e centricon 10 {Amicon}, centrifugada por 10 min a 15.000 g a temperatura ambiente e aplicada na coluna previamente equilibrada com tampão **B**. O fluxo de

corrida utilizado foi de 0,5 mL/min e o andamento da purificação foi monitorado a 280 nm. As frações eluídas foram analisadas em gel 15% SDS-PAGE.

3.9 - CARACTERIZAÇÃO DA ENZIMA PRPP SINTASE

3.9.1 - DETERMINAÇÃO DA MASSA MOLECULAR

Para se determinar o valor da massa molecular da proteína e, possivelmente, confirmar seu valor teórico, foi necessário realizar uma cromatografia de exclusão molecular em coluna Superdex 200 HR 10/30 com aplicação de uma mistura de proteínas com diferentes massas moleculares, nas mesmas condições da cromatografia realizada com a proteína descrita na seção 3.8.3.

A tabela 11 apresenta todas as proteínas padrão utilizadas na cromatografia de exclusão molecular Superdex 200 HR 10/30.

TABELA 11: Padrão de massa molecular utilizado na coluna Superdex 200 HR 10/30

Padrões de MW	MW (Da)
Ferritina	440.000
β -Amilase	152.000
Albumina sérica bovina	66.000
Ovalbumina	45.000
Gliceraldeído-3-fosfato desidrogenase	36.000
Anidrase Carbônica	29.000
Ribonuclease A	13.000

Uma vez realizada a separação destas proteínas na coluna determinou-se o volume de eluição (V_e) de cada uma delas e com este volume de eluição e o valor da massa molecular pode-se construir um gráfico utilizando o programa Origin 6.1 (“Data Analysis and Technical Graphics”). A partir da linearização da curva obtida através do programa obteve-se a equação da reta como sendo $(Y) = A + B * (X)$ sendo Y o log de MW e X o volume de eluição, possibilitando determinar o valor da massa molecular aparente da proteína em estudo.

3.9.2 - ESPALHAMENTO DINÂMICO DE LUZ (DLS)

O fenômeno do espalhamento de luz é interessante no processo de investigação da massa molecular e da homogeneidade das macromoléculas [58].

Considerando esta técnica como uma ferramenta importante na caracterização da proteína realizou-se um experimento envolvendo o espalhamento dinâmico de luz com o objetivo de estimar a massa molecular da proteína e verificar como a amostra se apresenta. Utilizou-se o equipamento “ProteinSolutions DynaPro”. A amostra depois de purificada pela cromatografia por exclusão molecular em Superdex 200 HR 10/30 foi submetida à concentração em centricon 10 até 3 mg/mL. Utilizou-se 30 μ L da amostra em tampão **B** que foi centrifugada por 10 min a 10.000 g. O equipamento realiza várias medidas e faz uma seleção das melhores medidas, que são representadas na forma de um histograma. Este histograma fornece o raio hidrodinâmico (R_H) da partícula estimando-se a massa molecular da proteína e seu grau de homogeneidade. Este

experimento foi realizado pela Dr^a. Raquel Kelly Bortoleto Bugs do Instituto de Física de São Carlos da Universidade de São Paulo.

3.9.3 - DETERMINAÇÃO DO PONTO ISOELÉTRICO

As proteínas são moléculas anfóteras, carregam cargas positivas ou negativas dependendo do pH. A rede de cargas na proteína é uma soma algébrica de todas as cargas das cadeias laterais dos aminoácidos que as compõe. Há um pH específico para todas as proteínas na qual esta somatória de cargas é zero ou nula e a este valor de pH dá-se o nome de ponto isoelétrico (pI) [59]. Em valores de pH acima do pI da proteína estas são negativamente carregadas e em valores de pH abaixo do seu pI elas são positivamente carregadas. (“Protein Electrophoresis Technical Manual” – Amersham Pharmacia Biotech)

A determinação do pI é uma técnica de alta resolução que separa proteínas de acordo com seus pontos isoelétricos. Um gradiente de pH estável é estabelecido em um gel de poliacrilamida.

Quando uma mistura de proteínas é aplicada num meio com variação de pH e sujeita a um campo elétrico, ela pode inicialmente mover-se através do gel em direção ao eletrodo com a carga oposta. A proteína irá mover-se até o ponto onde o valor de pH for igual ao seu pI. Nesse ponto, sendo a somatória de cargas nula a molécula não irá migrar mais.

Para a realização deste experimento utilizou-se o sistema “Phast System” {Pharmacia LKB Biotechnology} que utiliza um “Phast Gel IEF 3-9” {Pharmacia LKB Biotechnology} que permite a determinação do pI situado na faixa de pH

que abrange de 3 a 9. Este é um método rápido (30 min) e reproduzível. Aplicou-se 4 μ L da proteína concentrada (3 mg/mL) em tampão **B** ao gel de poliacrilamida IEF. A determinação do pI da proteína foi realizada sob condições padrão do equipamento e a coloração foi realizada com “Coomassie Blue” G-50 0,1% em solução aquosa de metanol 25% e ácido acético 5% . Os padrões de pI utilizados que constituem o kit geral de calibração {Amersham Pharmacia Biotech} estão descritos na tabela 12.

TABELA 12: Padrões de pI constituintes do kit geral de calibração

Padrões de pI	pI
Tripsinogênio	9,30
Banda básica de lectina	8,65
Banda média de lectina	8,15
Banda ácida de lectina	8,15
Mioglobina básica	7,35
Mioglobina ácida	6,85
Anidrase carbônica humana	6,55
Anidrase carbônica bovina	5,85
β Lactoglobulina A	5,20
Inibidor de tripsina de soja	4,55

3.9.4 - DICROÍSMO CIRCULAR

O Dicroísmo Circular (CD) é uma técnica apropriada para estimar a percentagem dos componentes da estrutura secundária das proteínas (hélice- α , folhas β , voltas β e seqüências aleatórias).

A espectrometria de CD baseia-se no desvio da luz circularmente polarizada incidente em compostos assimétricos (quirais). Essa luz circularmente

polarizada consiste de dois feixes de onda plana, linearmente polarizados, ortogonais entre si e fora de fase por 90° . A onda linearmente polarizada pode ser decomposta em duas componentes individuais circularmente polarizadas: uma à esquerda e a outra à direita. Quando a luz incide em um composto assimétrico esta é absorvida e cada um dos componentes interage de forma distinta em relação a absorvidade específica dos diferentes compostos. Dessa maneira pode-se então definir o CD como a propriedade que consiste na interação distinta para a luz circularmente polarizada a direita e para a luz circularmente polarizada a esquerda emitida com igual intensidade, e o sinal de CD como sendo a diferença entre a absorção da luz circularmente polarizada à esquerda e a direita.

O espectro de CD é dividido em duas regiões: na região do UV próximo e na região do UV distante. Na região de UV próximo que compreende o intervalo para o comprimento de onda entre 250 a 300 nm, verifica-se a contribuição dos aminoácidos aromáticos e também das pontes dissulfeto. Já na região de UV distante que compreende o intervalo de 190 a 250 nm é dominada pela transição das cadeias peptídicas da proteína e esta região que nos permite estimar a composição da estrutura secundária e acompanhar as possíveis mudanças na estrutura que possam ocorrer durante o experimento. O espectro de CD pode ser considerado como sendo do tipo hélice α quando apresenta um espectro contendo duas bandas negativas em 208 e 222 nm e uma banda positiva em 192 nm. Já o espectro característico de folhas β apresenta uma banda negativa em 216 nm e uma banda positiva em 195 nm [62].

Com base nesses conceitos realizou-se um experimento com uma amostra de 0,2 mg/mL da proteína recombinante purificada com o objetivo de determinar

seus componentes de estrutura secundária e analisar a estabilidade da proteína em variações de pH.

As medidas foram realizadas no espectropolarímetro Jasco J-720 {“Jasco Corporation, Japão”} variando-se o comprimento de onda de 195 a 250 nm com uma média de 16 medidas consecutivas usando uma cubeta cilíndrica de quartzo de 1 mm de caminho óptico em tampão fosfato de potássio 50 mM pH 7,8 contendo 50 mM NaCl. Para o estudo da influência do pH na amostra, alíquotas da proteína foram dialisadas contra diferentes tampões com diferentes pHs, mantendo a concentração da proteína de 0,2 mg/mL. A concentração dos tampões foi de 20 mM e os mesmos estão descritos na tabela 13.

TABELA 13: Descrição dos tampões utilizados no estudo da variação de pH por Dicroísmo Circular (CD)

TAMPÃO	pH
Glicina	2,5
Glicina	3,5
Acetato de sódio	4,0
Acetato de sódio	5,0
Bis Tris	6,0
Bis Tris	7,0
Fosfato de potássio	7,8
Glicina	9,0

A análise dos espectros de CD foi realizada através do programa “Self Consistent” SELCON-2 [62] com um banco de dados de 33 proteínas diferentes usadas para desconvolução da contribuição espectral de cada estrutura secundária. Na desconvolução o espectro de CD é incluído em uma matriz com dados

espectrais de CD e é selecionada inicialmente uma estrutura de uma proteína referencial ao acaso. A solução inicial é substituída por outra e o processo é repetido até que a autoconsistência seja alcançada [62].

3.9.5 - CLIVAGEM COM A ENZIMA ENTEROQUINASEMax (EKMax)

Como a proteína contém uma cauda de seis histidinas foi necessário realizar uma clivagem por digestão proteolítica com a enzima EnteroquinaseMax - EKMax {Invitrogen}. A enteroquinase é uma serino protease altamente específica que pode ser usada para digerir produtos de fusão separando a cauda de histidina da proteína de interesse. A enzima reconhece a sequência (Asp)₄ Lys e cliva após o resíduo de lisina. Esta sequência de clivagem está presente em muitos vetores de expressão. Genes clonados dentro do sítio de múltipla clonagem destes vetores expressam proteínas de fusão recombinante N-Terminal. A proteína nativa pode ser liberada a partir da proteína de fusão N-Terminal pela digestão com a EKMax.

Realizou-se uma série de reações piloto para determinar empiricamente a melhor condição de digestão da cauda de histidina com a enzima EKMax. As reações piloto foram testadas com a proteína em tampão fosfato de potássio 50 mM, pH 7,8 em duas temperaturas: a 18 e a 4°C. Para este experimento utilizou-se 10 µg da proteína, 6 µL do tampão EKMax [10X]{Invitrogen} e quatro diferentes quantidades da enzima EKMax expressa em unidades (U) - 0U ; 0,1U; 1U e 4U. Foi feito um volume de reação de 60 µL e 3 alíquotas de 20 µL foram retiradas de 12 em 12 horas para posterior análise em 15% SDS-PAGE. A reação

de digestão foi interrompida adicionando-se 10 μ L de tampão de amostra TA contendo β -mercaptoetanol e fervidas por 5 min.

3.9.6 - SEQÜENCIAMENTO DO N-TERMINAL DA PROTEÍNA

O seqüenciamento N-terminal de uma proteína é determinado usando repetidos ciclos da reação de degradação de Edman [63] que remove seqüencialmente um aminoácido por vez da extremidade amínica de um peptídeo.

Em cada ciclo o N-terminal da proteína é modificado com fenil isotiocianato que reage com o radical amino terminal não carregado do peptídeo, formando um derivado feniltiocarbamilado (PTC-N-terminal). O resíduo PTC-N-terminal em condições levemente ácidas é clivado formando um aminoácido anilintiazolino (ATZ) derivado do resíduo N-terminal original e uma proteína cujo penúltimo resíduo agora leva o N-terminal. O aminoácido ATZ por ser instável é convertido no aminoácido feniltiohidantoina (PTH). O resíduo N-terminal (n-1) do polipeptídeo é agora disponível para um outro ciclo que repete todos estes três passos [64].

Em instrumentos automatizados, a proteína é usualmente adsorvida ou unida por uma membrana ou suporte de fibra de vidro. O aminoácido ATZ depois de convertido em aminoácido PTH é então separado em uma coluna de fase reversa em um HPLC (Cromatografia Líquida de Alta Pressão) e o aminoácido é detectado por absorbância UV e identificado de acordo com o tempo de eluição através de métodos cromatográficos [64].

De posse dessas informações e com o objetivo de identificar o produto resultante da clivagem proteolítica com EKMax e com interesse de se determinar o N-Terminal da proteína fez-se uma clivagem da cauda de 6xHis usando 1 mg da proteína para 100 U da enzima EKMax por 32 horas de incubação a 18°C. Cerca de 50 µL da amostra, após a clivagem, foi submetida a 15% SDS-PAGE e eletrotransferida para uma membrana de PVDF (“Polyvinylidene Fluoride” - Bio-Rad) utilizando um tampão específico de transferência (Tris-base 48 mM, CAPS 39 mM e SDS 0,035% (m/v), 800 mL H₂O e 200 mL metanol) por um período de 5 h a 150V [65]. A membrana foi corada com corante Ponceau a 0,5% em ácido acético 0,1% por 5 min, e descorada com ácido acético 1%. As bandas imobilizadas da proteína foram recortadas da membrana e enviadas à análise da seqüência de aminoácidos NH₂-Terminal pelo Laboratório do Prof. Dr. Lewis Joel Greene do Centro de Química de Proteínas da Escola de Medicina de Ribeirão Preto. O equipamento utilizado nesta análise foi o “Procise Model 491” {PE Applied Biosystems} utilizando o método da Degradação de Edman [63]

3.10 - CRISTALIZAÇÃO

A cristalografia de proteínas tem oferecido uma poderosa ferramenta para elucidação detalhada da estrutura das proteínas e vem assumindo um papel importante no estudo de vários problemas biológicos. A cristalografia permite o estudo da estrutura tridimensional de moléculas através da realização de experimentos de difração de raios X. A difração de raios X é a técnica mais utilizada para se estudar a estrutura atômica de uma proteína.

A determinação da estrutura tridimensional de uma proteína por difração de raios X constitui-se de um processo trabalhoso, que tem como etapa crítica à obtenção de um mono cristal de proteína. Um cristal de proteína consiste de moléculas idênticas que se repetem em um arranjo tridimensional ordenado e periódico (fig29).

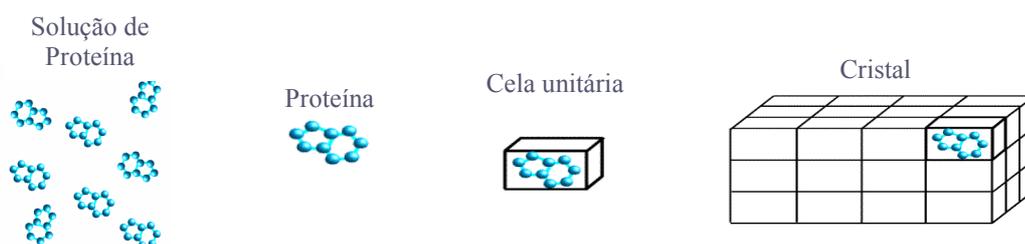


FIG29: Esquema representativo do arranjo de um cristal.

Para obtenção de um cristal de proteína existem vários métodos, entre eles o método de Difusão de Vapor, no qual duas soluções tendem ao equilíbrio pela difusão do vapor de água. Neste método a técnica mais utilizada é a da Gota Suspensa [66] (fig30) no qual a gota da solução de proteína mais agente precipitante é exposta gradualmente à saturação pela perda de água da gota para o largo reservatório que contém a solução precipitante, dessa maneira possibilitando o arranjo das moléculas de forma ordenada, originando o cristal [67].

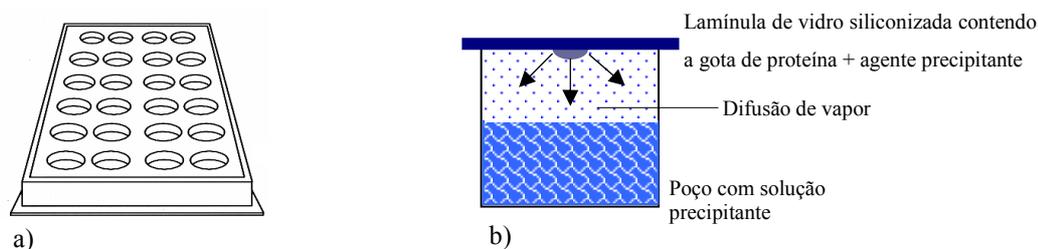


FIG30: Esquema do método de cristalização usando a técnica da gota suspensa. a) Modelo da caixa de cristalização. Geralmente são plásticas, apresentam 24 poços marcados de diâmetro de 16 mm e cada poço comporta cerca de 2 mL de solução. **b)** Modelo de um poço de cristalização.

Os ensaios de cristalização foram iniciados após a obtenção da proteína pura com cauda de histidina. Foi adicionada 500 μL de cada solução precipitante em cada poço e em uma lamínula siliconizada fez-se uma gota contendo 3 μL de solução precipitante (solução do poço) mais 3 μL da solução de proteína a uma concentração de 8,6 a 9,0 mg/mL em tampão fosfato de potássio 50 mM contendo 150 mM de cloreto de sódio. A concentração foi determinada pelo método de Bradford [68] e os ensaios de cristalização foram realizados a temperaturas de 4 e 18°C.

Para a realização deste experimento utilizou como agentes precipitantes os Fatoriais “Crystal Screen Reagent Formulation I” e “Crystal Screen Reagent Formulation II” {HAMPTON RESEARCH}.

RESULTADOS

4 - RESULTADOS

4.1 - CLONES SELECIONADOS E EXTRAÇÃO DO DNA PLASMIDIAL.

Seqüências das proteínas de genes de plantas participantes da via da síntese de purinas foram usadas para encontrar ESTs no banco de dados do SUCEST [6]. Essa busca foi realizada levando em consideração alguns critérios entre estes a fidelidade do alinhamento N-terminal da proteína de interesse com proteínas conhecidas. Segue na figura 31 um exemplo de como essa busca foi realizada com cada clone.

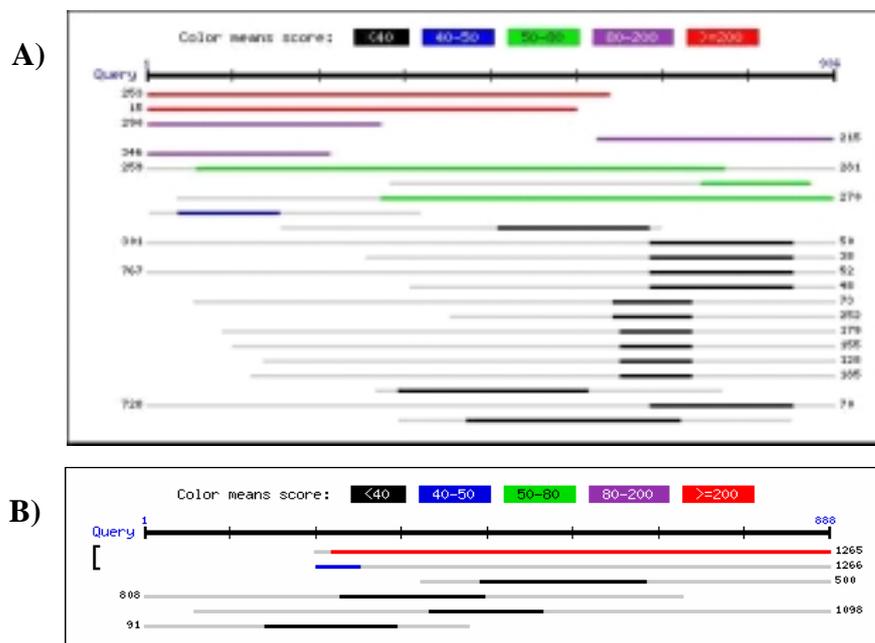


FIG31: Alinhamento de uma seqüência conhecida padrão contra seqüências EST do banco de dados do SUCEST usando tBLASTn. Alinhamento da região 5' (N-terminal da proteína) dos clones do SUCEST com seqüências conhecidas que foram consideradas como clones com seqüências completas em potencial. **(A)** é um exemplo de um alinhamento de uma seqüência inteira da PRPP sintetase de *Arabidopsis thaliana* contra o banco de dados do SUCEST. **(B)** é um exemplo de um “cluster” EST parcial de FGAM sintetase de *Glicina Max* contra o banco de dados do SUCEST.

Com esta busca vários clones envolvidos na via metabólica de purinonucleotídeos foram selecionados, especialmente clones homólogos a genes de *Arabidopsis thaliana* e a *Spinacia oleracea* criando um conjunto de informações que podem ser verificados na tabela 1 descrita na Introdução.

A partir deste resultado selecionaram-se treze clones EST com seqüências inteiras da cana-de-açúcar pertencentes a esta via metabólica (tabela 14). Estes clones são representantes de três das onze enzimas envolvidas na via da síntese *de novo*: a 5-Fosforribosil-1-Pirofosfato sintetase (PRPPase), Fosforribosil-aminoimidazol sintetase (AIRase) e Fosforribosilaminoimidazol-Succinocarboxamida sintetase (SAICARase).

TABELA14: Clones a selecionados para análise.

NOME	CLONE	PROTEÍNA	FONTE
A	SCCCCL3120E04	SAICAR	ARABIDOPSIS
B	SCCCCL3140E09	AIR	ARABIDOPSIS
C	SCCCRZ1003F04	AICAR	GALINHA
1	SCJLRZ1024F09	PRPP	ESPINAFRE
2	SCJLAM2093F02	PRPP	ESPINAFRE
3	SCACRZ3035C11	PRPP	ARABIDOPSIS
4	SCCCST3C04C09	GAR	ARABIDOPSIS
5	SCQSAM2100E05	SAICAR	ARABIDOPSIS
6	SCEQRT2095D07	AIR	ARABIDOPSIS
7	SCEQRT2025F06	SAICAR	ARABIDOPSIS
8	SCBGLR1047H09	SAICAR	ARABIDOPSIS
9	SCVPRT2083H04	PRPP	ARABIDOPSIS
10	SCVPRT2083H10	PRPP	ARABIDOPSIS

O DNA plasmidial de todos os clones foi extraído e o resultado pode ser verificado na figura 32.

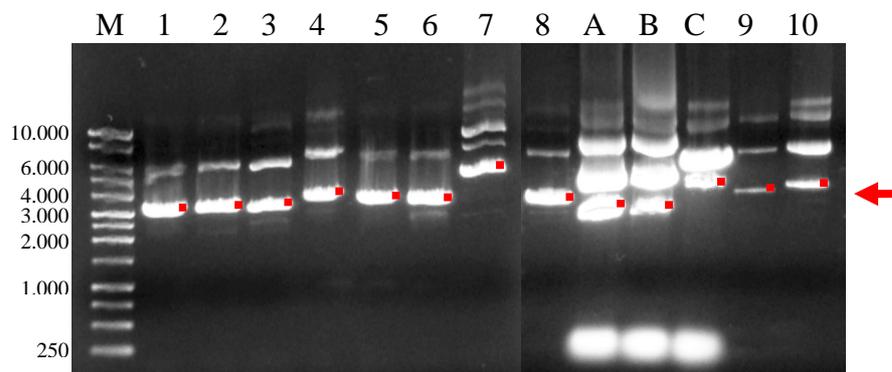


FIG32: Extração do DNA plasmidial. Analisada em gel de agarose 0,8% TAE [1X]. A numeração do gel refere-se ao DNA plasmidial dos clones descritos na tabela 1. M representa o marcador de massa molecular de 1 kb Ladder. A seta e o símbolo (■) representam os DNAs extraídos.

As bandas mostradas no gel representam os fragmentos de DNA extraídos com massa molecular entre 3.000 a 4.000 pb, com exceção do clone 7 com massa molecular entre 5.000 a 6.000 pb.

4.2 - RESULTADOS DA ANÁLISE DAS SEQUÊNCIAS DOS CLONES

O sequenciamento da extremidade 5' dos 13 clones foram devidamente corrigidas (dados não mostrados) e alinhadas cada qual com a respectiva sequência de *Arabidopsis thaliana* e *Spinacia oleracea* selecionadas. Isto permitiu verificar a fidelidade dos clones recebidos como exemplificado no esquema de alinhamento da PRPP sintase do clone 3 adquirida com a sequência da PRPP sintetase da cana-de-açúcar depositado no SUCEST (fig33), verificando fidelidade dos clones adquiridos. Apenas com os clone 10 (VPRT2083H10) e clone 7 (SCEQRT2025F06) não foi encontrada similaridade, sendo estes descartados. Os demais clones representavam os clones esperados

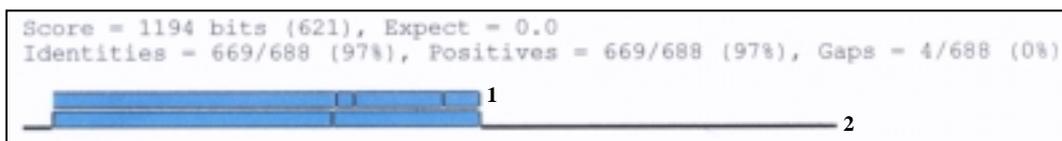


FIG33: Representação esquemática do alinhamento de confirmação dos clones seqüenciados.

O alinhamento foi realizado com o programa “BLAST 2 Sequences”. 1 – representa a EST de PRPPase da cana-de-açúcar do banco de dados do SUCEST. 2 – representa o clone 3 (RZ3035C11 – PRPP sintase) recebido para análises futuras.

Com a seqüência de nucleotídeos dos clones obteve-se a seqüência de aminoácidos com as seis fases de leitura de cada uma das proteínas em estudo. Como segue representado pela enzima PRPP sintase, cada fase de leitura foi alinhada com as ESTs depositadas no banco de dados do SUCEST, encontrando a fase de leitura correta de cada proteína selecionada (fig34).

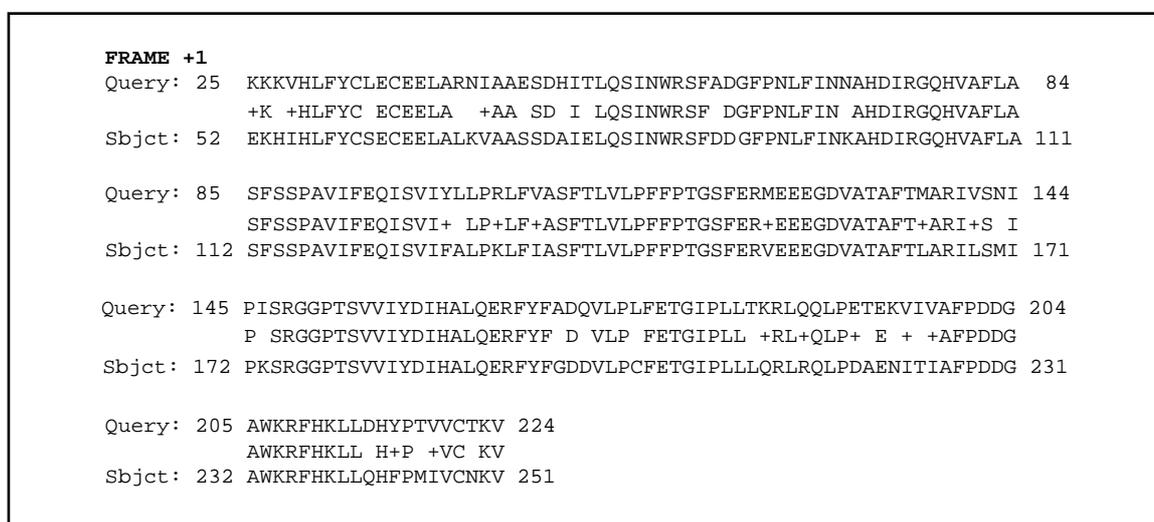


FIG34: Alinhamento representativo das fases de leitura (Frames). “Query” representa a proteína PRPP sintetase da cana-de-açúcar homóloga Arabdopsis do banco de dados do NCBI e como “Sbjct” clone 3 da PRPP sintase de cana-de-açúcar sequenciado.

A partir deste alinhamento teve-se como resultado as seguintes fases de leitura apresentadas na tabela 15. Este procedimento foi realizado com o objetivo de encontrar o códon iniciador (ATG-Formilmetionina) possibilitando a construção dos oligonucleotídeos utilizados na amplificação de cada gene. Os clones CL3140E09, ST3C04C09, RZ1003F04, CI3120E04, não apresentaram alinhamento de nenhum de seus “frames” com as proteínas esperadas, sendo descartados.

TABELA15: Resultado obtido a partir do alinhamento para obtenção da fase de leitura significativa.

CLONE	PROTEÍNA	FASE DE LEITURA
SCJLRZ1024F09	PRPPase	1
SCJLAM2093F02		2
SCACRZ3035C11		1
SCVPRT2083H04		1
SCQSAM2100E05	SAICARase	2
SCBGLR10447H09		3
SCEQRT2095D07	AIRase	1

4.3 - SEQÜENCIAMENTO DO GENE DA PRPP SINTASE (CLONE 3)

O clone 3 (SCACRZ3035C11) representante da PRPP sintase teve seu gene seqüenciado várias vezes utilizando os oligonucleotídeos T7, OligoA e OligoB com base na estratégia descrita na seção 3.2 do Material e Métodos. Os vários seqüenciamentos foram corrigidos e alinhados usando o programa SeqMan

formando um seqüência consenso, como descrito na estratégia de alinhamento da figura 35.

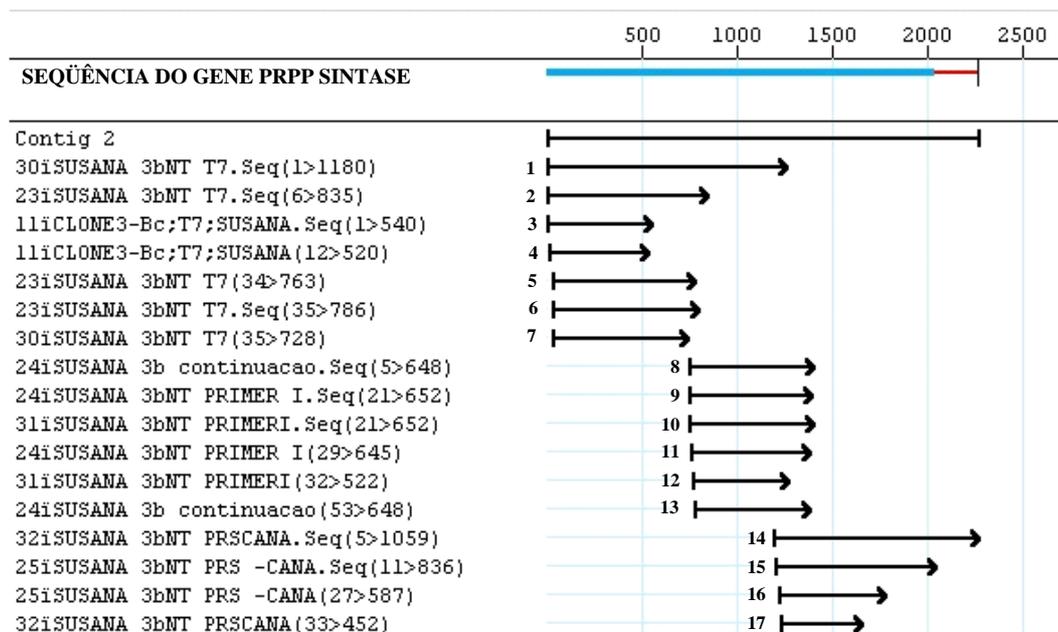


FIG35: Alinhamento múltiplo realizado através do programa SeqMan. O “contig2” representa a seqüência consenso devidamente corrigida. As setas representam o seqüenciamento do gene da extremidade 5’ a 3’ utilizando os oligonucleotídeos T7 (seta 1 a 7), OligoA (seta 8 a 13) e OligoB (seta 14 a 17). A barra em azul representa a seqüência completa do gene da PRPP sintase.

Este processo resultou na determinação da seqüência completa de nucleotídeos da fase aberta de leitura do gene *prs* que codifica para a proteína PRPP sintase (fig36).

A partir da seqüência de nucleotídeos, foi feita a dedução da seqüência de aminoácidos da enzima (fig36) e determinada sua composição (tabela 3).

01		GAGGGNNAN	09
10	NNNGGNNNTNNNNTTTGAAAACCCCTTCGCCTGCAGCCACCGGTCCGGAATT		60
61	CCCGGGTCGACCCACGCGTCCGCCGCGCCGTTCTTTTGGTTCCCTGTCC		111
112	CCGCCGCCGCGAAACCCACCGATGGAGGTCGTCGGGGCCGGAAGCAGAA		162
163	GGCGAAGAAGCACGAGAAATCGAGGTCGTCGGCCGCCGGAAGCAGAAGGCG		213
		M E V V G A A K Q K A	11
214	AAGAAGCACGAGAAGCACATACACCTTCTACTGCTCCGAATGCGAGGAG		264
12	K K H E K H I H L F Y C S E C E E		28
265	CTCGCCCTCAAGGTCGCCGCGAGCTCCGACGCCATCGAGCTCCAGTCCATC		315
29	L A L K V A A S S D A I E L Q S I		45
316	AACTGGCGGAGCTTCGACGATGGGTTCGCCAACCTGTTTCATCAACAAGGCC		366
46	N W R S F D D G F P N L F I N K A		62
367	CACGACATCCGGGGCAGCACGTCGGCGTTCCTGGCCTCCTTCAGCTCGCCG		417
63	H D I R G Q H V A F L A S F S S P		79
418	GCGGTCATATTCGAGCAGATCTCCGTCATCTTCGCGCTGCCTAAGCTCTTC		468
80	A V I F E Q I S V I F A L P K L F		96
469	ATCGCTCATTCACTCTCGTGTGCCTTCTTCCACGGGCTCCTTCGAG		519
97	I A S F T L V L P F F P T G S F E		113
520	CGCGTTGAGGAGGAGGGCGATGTCGCCACCGCCTTCACCCTCGCGCGCATC		570
114	R V E E E G D V A T A F T L A R I		130
571	CTCTCCATGATCCCCAAGTCGCGCGCGGGCCGACAGCGTCGTCATCTAC		621
131	L S M I P K S R G G P T S V V I Y		147
622	GACATCCACGCGCTCCAGGAGAGGTTCTATTTTCGGTGACGACGCTCCTGCCA		672
148	D I H A L Q E R F Y F G D D V L P		164
673	TGCTTTGAGACAGGGATTCCGCTCCTGCTGCAGCGCTCCGCCAGCTCCCT		723
165	C F E T G I P L L L Q R L R Q L P		181
724	GACGAGAAAATATCACCATTCGCTTTCAGATGATGGAGCCTGGAAGCGG		774
182	D A E N I T I A F P D D G A W K R		198
775	TTCCACAAGCTGTTGCAGCACTTCCAATGATAGTCTGTAACAAGGTTTCGT		825
826	GAAGGTGACAAGAGAATAGTTCGTATAAAGGAAGGAAATCCTGAAGGCCGT		876
199	F H K L L Q H F P M I V C N K V R		215
216	E G D K R I V R I K E G N P E G R		232
877	CATGTTGTTATTGTTGATGATTTAGTGCAATCTGGGGAACTCTTCGAGAA		927
233	H V V I V D D L V Q S G G T L R E		249
928	TGCCAGAAAGTTCTAGCTTCGCATGGCGCTGCAAAAGTTAGTGCTTATGTG		978
250	C Q K V L A S H G A A K V S A Y V		266
979	ACTCATGCTGTGTTCCCTGAGCGCTCATATGAACGTTTCATGTCGTCGATT		1029
267	T H A V F P E R S Y E R F M S S I		283
1030	TCTGCTGGGCCAGGCGACCAGTTTGCTTACTTCTGGATCACGGACTCATGC		1080
284	S A G P G D Q F A Y F W I T D S C		300
1081	CCACACACAGTAAAAGCTATTGGCCAAAGACCTCCATTTGAGGTTCTGAGC		1131
301	P H T V K A I G Q R P P F E V L S		317
1132	CTTGCTGGCTCAATTGCTGATGCTCTTCAGATCTCGATGTGTACTGTTAGA		1182
318	L A G S I A D A L Q I *		328
1183	TGGGCTTGAATCGTGGATGCCCATCAGATTGAACTTGTACTATGAGATGAG		1233
1234	CTGCTCCTGCTTTTCCATAAATGTAAGAGTTTGTTCAGTTCCCTGGAAAAG		1284
1285	CAAATAATGTGGAGTAAGTTAAGTACCTGAAGGTTCCATAAACAGCACTGT		1335
1336	GAATTGATAATAAATCAGGCATTGGATACTTGTGTGCTG-Poly A		1374

FIG.36: Sequência de nucleotídeos e aminoácidos da PRPP sintase da cana-de-açúcar. A sequência mostra 1374 pb. A região codificante da proteína dentro do gene *prp* é enumerada positivamente a partir do códon iniciador Met (ATG está em vermelho) e os resíduos de aminoácidos preditos da PRPP sintase são numerados positivamente começando da primeira Met. O códon de terminação (TAG) está indicado em vermelho na posição 1165. Sequências de nucleotídeos não resolvida na região 5'UTR estão indicadas com N. O sítio de poliadenilação está identificado na posição +1374 em azul.

A seqüência da PRPP sintase da cana-de-açúcar apresenta 1374 pares de bases. A seqüência amplificada do gene que codifica para a proteína PRPP sintase possui 984 pares de bases na fase aberta de leitura (ORF) que codifica para uma proteína de 328 aminoácidos com massa molecular de 36,6 kDa. O número de resíduos carregados positivamente (Arg e Lys) e negativamente (Asp e Glu) é 34 e 38 respectivamente (tabela 16). O ponto isoelétrico teórico foi determinado em 6,31. O códon iniciador formilmetionina (ATG) se localiza a 180 pares de bases após a região 5'UTR e o códon de terminação (TGA) está em fase no nucleotídeo 1165. Um sítio de poliadenilação é identificado na posição + 1374.

TABELA16: Composição dos aminoácidos

Ala (A)	31	9,5%	Leu (L)	27	8,2%
Arg (R)	17	5,2%	Lys (K)	17	5,2%
Asn (N)	6	1,8%	Met (M)	4	1,2%
Asp (D)	17	5,2%	Phe (F)	25	7,6%
Cys (C)	6	1,8%	Pro (P)	18	5,5%
Gln (Q)	13	4,0%	Ser (S)	24	7,3%
Glu (E)	21	6,4%	Thr (T)	11	3,4%
Gly (G)	20	6,1%	Trp	3	0,9%
His (H)	12	3,7%	Tyr (Y)	6	1,8%
Ile (I)	25	7,6%	Val (V)	25	7,6%

4.4 - ALINHAMENTO DA SEQÜÊNCIA DE AMINOÁCIDOS DA PRPP SINTASE COM SUAS HOMÓLOGAS

A Seqüência de aminoácidos deduzida da PRS da cana-de-açúcar foi comparada e alinhada com algumas proteínas PRS de plantas obtidas a partir da busca no banco de dados do NCBI (fig37).

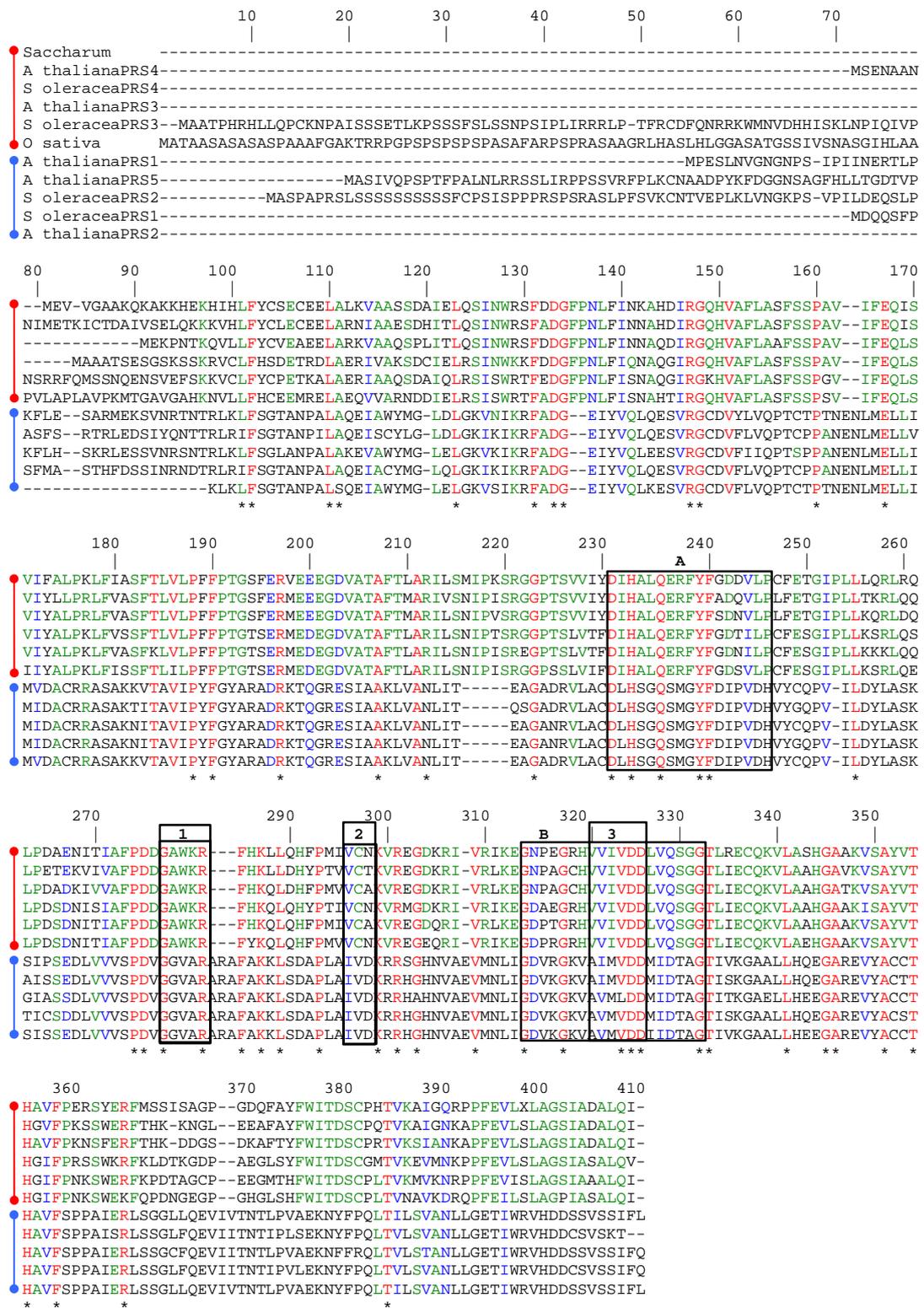


FIG37: Alinhamento múltiplo de seqüências representativas de PRS de plantas. As posições dos aminoácidos estão indicadas na parte superior do alinhamento. O consenso do alinhamento está indicado por “*” para aminoácidos idênticos. As caixas indicam os domínios ligantes preditos. CaixaA – o sítio ligante de cátion divalente Mg^{2+} . CaixaB – motivo ligante PRPP/Rib-5-Fosfato. Caixa 1, 2 e 3 – possíveis resíduos ligantes de ATP. A barra vermelha representa as PRPP sintases e a azul as PRPP sintetases.

Verifica-se através do alinhamento que a região N-terminal apresenta-se heterogênea enquanto a região C-terminal é conservada. Dois blocos conservados: um constituído das PRPP sintases representada pela barra vermelha e outro constituído das PRPP sintetases representada pela barra azul, foram identificados. Os resíduos 231 a 247 representam o sítio ligante de cátion divalente (Mg^{2+}) (CaixaA). Os cinco aminoácidos a partir do resíduo 276 (Caixa1), os aminoácidos 296 a 298 (Caixa2) juntamente com os resíduos 320 a 325 (Caixa3) representam os possíveis aminoácidos envolvidos na ligação do ATP. A caixa B representada pelos resíduos 314 a 332 corresponde ao motivo ligante PRPP/Rib-5-Fosfato. Os asteriscos (*) representam os resíduos de aminoácidos altamente conservados. Os aminoácidos em vermelho ou verde são regiões semi conservadas.

A seqüência de aminoácidos da PRPP sintase de cana-de-açúcar também foi alinhada com várias outras isoformas de outros organismos. O resultado deste alinhamento dois a dois está representado na tabela 17. O alinhamento das diferentes seqüências revela que a PRPP sintase de cana-de-açúcar possui 77% de identidade com a isoenzima PRS3 de *Arabidopsis thaliana* e *Spinacia oleracea* e com a *Oryza sativa* e 78% de identidade com a seqüência de aminoácidos da isoenzima PRS4 de *Arabidopsis thaliana* e *Spinacia oleracea*.

TABELA17: Alinhamento da PRPP sintase com PRS de outros organismos.

	CANA	A. thal1	A. thal2	A. thal3	A. thal4	A. thal5	S. Oler1	S. Oler2	S. Oler3	S. Oler4	Oryza	S. cer1	S. cer2	S. cer3	S. cer4	S. cer5	Human1	Human2	Human3	E. coli
CANA	100%	22/ 42	23/43	72/82	77/82	23/42	23/42	23/43	73/84	78/83	73/84	32/55	25/42	25/41	24/41	37/53	23/44	24/45	22/45	28/46
A. thal1	1e-09	100%	95/98	25/47	26/46	82/89	86/92	85/92	23/43	26/47	23/74	28/44	40/60	39/61	40/59	25/40	45/65	44/66	45/66	47/70
A. thal2	3e-10	e-169	100%	25/48	25/45	88/95	90/95	90/95	23/46	25/47	25/49	27/44	39/61	39/62	41/61	25/40	45/66	44/66	44/66	48/71
A. thal3	e-128	5e-14	5e-12	100%	73/83	26/48	26/48	26/48	80/88	73/84	79/88	36/55	27/47	26/42	25/47	36/57	25/45	25/44	24/44	26/44
A. thal4	e-135	2e-10	1e-09	e-131	100%	27/46	27/44	26/46	74/84	86/93	71/84	36/53	27/47	24/43	25/45	30/53	23/46	25/45	23/44	27/46
A. thal5	6e-11	e-160	e-157	2e-12	4e-10	100%	91/95	79/88	23/46	25/45	24/47	27/44	38/59	39/61	39/58	24/39	45/66	44/66	44/66	48/70
S. Oler1	1e-10	e-165	e-160	6e-13	1e-09	e-168	100%	87/93	25/46	27/45	26/49	27/45	40/61	39/61	41/61	25/40	45/66	44/66	44/66	47/70
S. Oler2	3e-10	e-172	e-160	2e-12	2e-09	e-160	e-167	100%	24/47	26/46	25/49	27/43	40/60	41/61	41/60	24/39	44/65	44/65	43/65	46/68
S. Oler3	e-129	2e-12	2e-11	e-152	e-134	2e-10	4e-12	3e-12	100%	73/83	79/87	33/48	28/45	24/41	25/45	40/57	23/43	23/43	22/42	27/45
S. Oler4	e-135	2e-12	4e-11	e-132	e-156	4e-12	1e-12	7e-12	e-133	100%	72/85	23/42	28/48	24/43	25/43	45/63	22/45	22/45	21/41	28/46
Oryza	e-130	2e-13	8e-14	e-145	e-130	6e-13	6e-13	2e-12	e-146	e-131	100%	36/56	27/48	26/42	26/46	40/61	24/45	25/45	23/44	26/45
S. cer1	4.6	6e-41	2e-40	0.53	3.7	7e-39	7e-39	5e-40	0.62	0.66	0.057	100%	36/51	34/49	36/51	29/43	36/52	37/54	36/52	30/45
S. cer2	4e-09	5e-60	3e-59	1e-11	1e-11	6e-57	2e-58	8e-59	8e-12	4e-11	2e-13	3e-64	100%	61/77	86/90	31/43	61/77	61/76	60/76	46/65
S. cer3	1e-05	9e-61	5e-61	2e-08	1e-06	1e-58	9e-60	3e-60	2e-07	4e-07	1e-08	3e-60	e-108	100%	63/79	30/44	60/78	61/78	58/76	47/65
S. cer4	9e-07	2e-62	1e-61	6e-11	2e-09	4e-58	3e-59	4e-59	1e-10	1e-08	4e-12	3e-65	e-153	e-110	100%	32/45	62/78	62/78	61/77	46/64
S. cer5	0,017	1e-40	2e-40	0.004	0.010	1e-36	1e-37	6e-38	0.003	0.007	5e-04	4e-53	3e-54	9e-54	2e-60	100%	34/45	34/46	35/46	26/41
Human1	2e-08	4e-73	2e-73	4e-11	2e-07	1e-72	9e-72	2e-71	9e-09	8e-07	2e-10	1e-70	3-105	e-106	e-107	4e-69	100%	95/98	94/96	48/68
Human2	1e-08	1e-72	6e-73	2e-11	1e-08	5e-72	4e-71	2e-70	7e-08	1e-07	9e-11	2e-70	e-105	e-106	e-107	5e-70	e-172	100%	91/96	48/68
Human3	5e-07	2e-71	2e-71	5e-10	1e-06	8e-71	7e-70	7e-70	1e-07	9e-06	2e-08	3e-69	e-101	e-101	e-104	1e-70	e-169	e-164	100%	47/68
E. coli	2e-14	4e-81	2e-81	2e-14	1e-12	5e-80	1e-80	2e-77	6e-15	3e-14	4e-14	3e-42	2e-68	1e-70	3e-69	4e-42	5e-78	7e-78	1e-75	100%
B. sub	2e-08	4e-83	3e-83	3e-05	2e-05	1e-80	5e-83	8e-84	1e-40	3e-06	9e-06	2e-39	9e-62	2e-60	1e-61	7e-32	1e-65	3e-65	1e-64	5e-75

 "e-values"

 % Identidade/ % Similaridade

Através do alinhamento da PRS da cana-de-açúcar com 25 seqüências de aminoácidos de doze organismos diferentes e com uso de métodos de reconstrução de árvores filogenéticas, obteve-se a árvore consenso descrita na figura 38. Nesta árvore foi usado como “outgroup” a PRS de *Archaeoglobus fulgidus*. Os mamíferos apresentaram uma clade monofilética, os fungos por sua vez não apresentaram uma clade monofilética, mas estão relacionados aos mamíferos. As PRS 1, 2 e 5 de *Arabidopsis thaliana* juntamente com as PRS 1 e 2 de *Spinacia oleracea* formam um grupo com uma possível localização no cloroplasto. A PRS de cana-de-açúcar forma um grupo com a PRS4 de *A. thaliana* e *S. oleracea*. A isoenzima PRS3 de *A. thaliana* e *S. oleracea*, juntamente com a *Oryza sativa* apresenta uma dicotomia dos ramos não resolvidos.

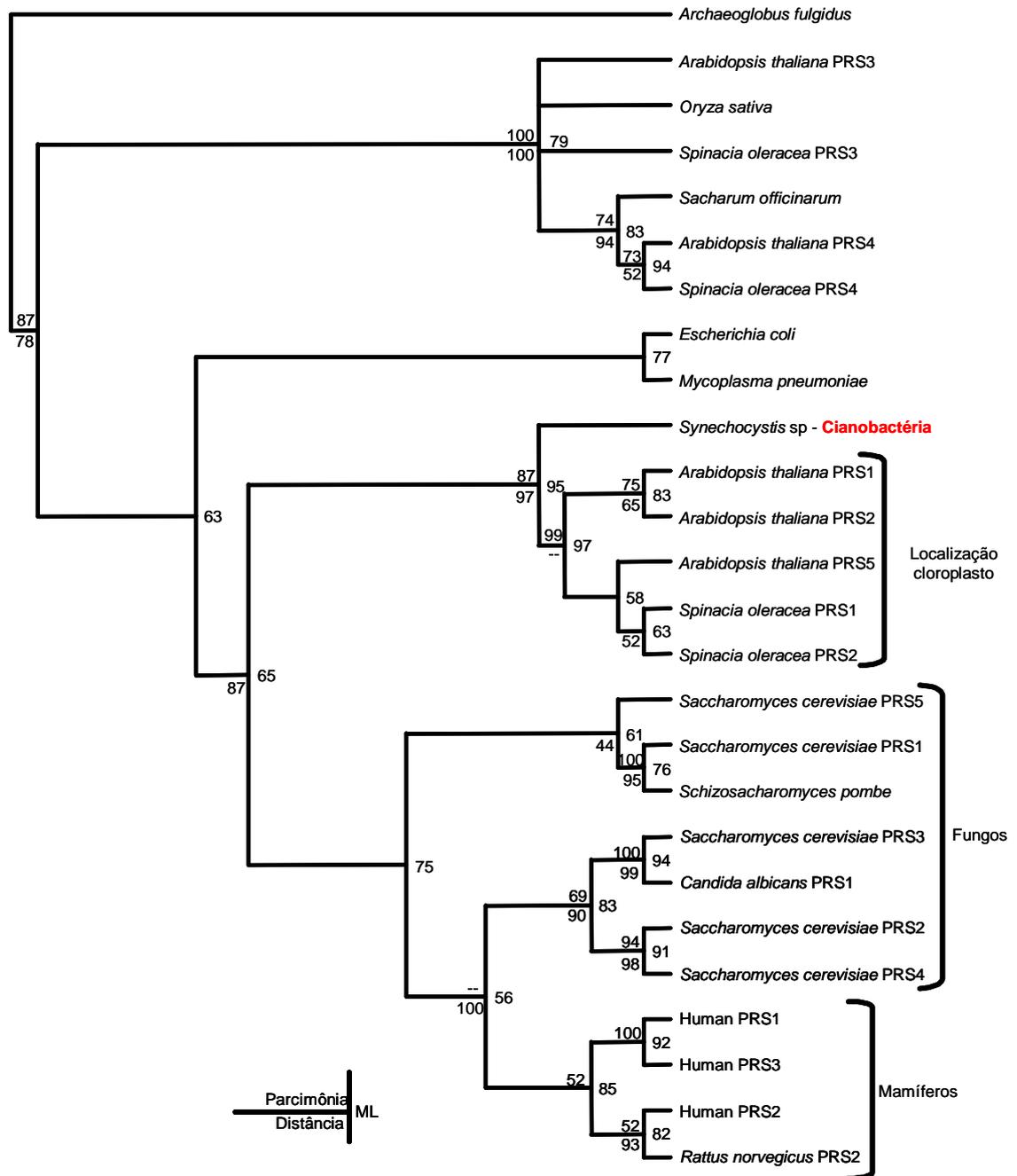


FIG38: Reconstrução filogenética de proteínas PRS representativas. A árvore apresenta raiz tendo como “outgroup” PRS de *A. fulgidus*. Os valores de fidelidade dos ramos fornecidos pela Parcimônia, Distância e ML estão indicados nos nós como valores do “bootstrap” de 1000 replicações. A seqüência de aminoácidos da enzima PRPP sintase da cana-de-açúcar (número de acesso AF536565) e PRSs homólogas a partir das isoformas 1 a 5 de *Arabidopsis thaliana* (CAA58717; CAA 63552; CAB43552; CAB43553 e AC004521 respectively [27]), isoformas 1 a 3 de *Homo sapiens* (BAA00733 [42]; P11908 [43] e AAB59463 [44]), isoformas 1 a 5 de *Saccharomyces cerevisiae* (CAA49674; CAA52436; CAA52437; CAA84888 e NP_014580 [25]), *Rattus norvegicus* (NP_036766 [45]), *Escherichia coli* (BAA36065 [28]), *Oryza sativa* (BAB89498) [46], isoformas 1 to 4 de *Spinacia oleracea* (CAB43599; CAB43600; CAB43601 e CAB43602 [36, 47]), *Mycoplasma pneumoniae* (P75044 [48]), *Synechocystis* sp (NP_442534), *Schizosaccharomyces pombe* (NP593826), e *Candida albicans* (P46585) foram alinhadas através do programa CLUSTALX [49].

4.5 - CLONAGEM E EXPRESSÃO DA PRPP SINTASE

Selecionou-se os clones 1, 3 e 9 da PRPP sintetase, o clone 6 da AIR sintetase e o clone 5 da SAICAR sintetase. Estes clones foram submetidos à amplificação e purificação do DNA (fig39).

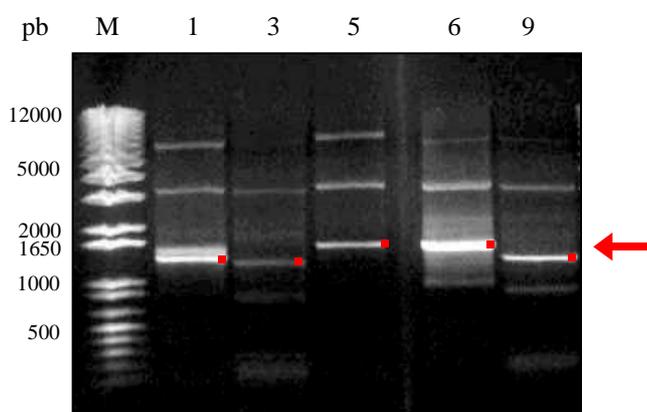


FIG39: Resultado da amplificação do DNA. Foi analisada em gel de agarose 0,8% TAE [1X]. Os números representam: 1, 3 e 9 – PRPP , 5- SAICAR e 6 AIR. M representa o marcador de massa molecular de 1kb plus. A seta e o símbolo (■) representam a banda do gel que foi retirada para ser feita a purificação do DNA amplificado (dados não mostrados).

Pelo gel de agarose verifica-se que os produtos de PCR possuem entre 1000 a 1650 pares de bases (fig39).

Os cinco clones selecionados, amplificados e purificados, foram inseridos no vetor de clonagem e expressão pCRT7 TOPO NT e CT. Para verificar se o inserto se ligou na direção correta no vetor de expressão realizou-se uma nova amplificação como representado na figura 40. Com esta estratégia foi possível amplificar o fragmento de DNA de interesse e, portanto visualizá-lo no gel de agarose (fig41).



FIG40: Representação esquemática da estratégia de amplificação. TOPO e SP6 representam os oligonucleotídeos utilizados na nova amplificação do fragmento de DNA purificado.

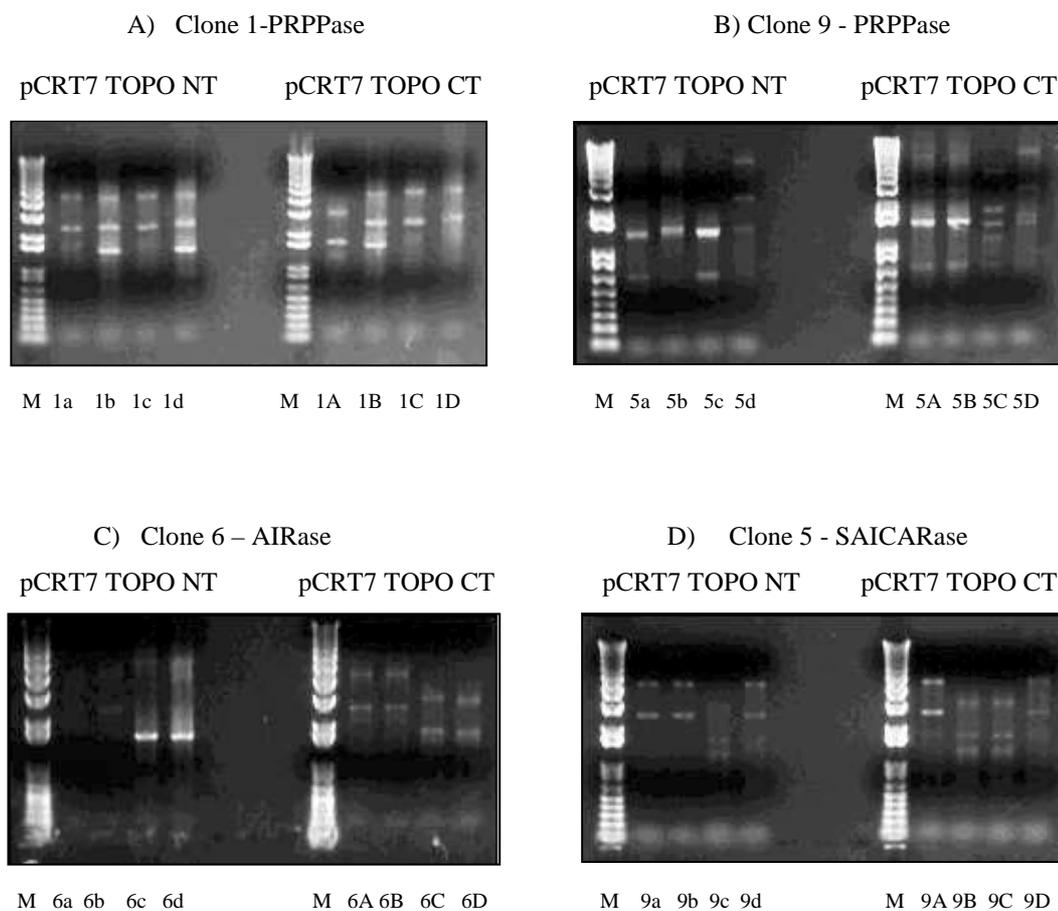


FIG41: Experimento de amplificação do DNA. Representa a checagem da orientação da construção recombinante a partir de amplificação por PCR com oligonucleotídeos TOPO e SP6. As figuras A, B, C, D, representam os clone 1 (PRPPase), 9 (PRPPase), 6 (AIRase) e 5 (SAICARase) respectivamente. **pCRT7 TOPO NT** representa a construção com vetor contendo a cauda de histidina na região N-terminal. **pCRT7 TOPO CT** representa a construção sem a cauda de histidina na região C-terminal. M representa o marcador de massa molecular de 1 kb plus. A numeração composta de letras na parte inferior do gel representa as colônias selecionadas para verificação.

Verifica-se que com o clone 5 nenhum dos transformantes foi amplificado. Com o clone 6 obteve-se apenas amplificação dos transformantes no vetor NT já os clones 1 e 9 apresentaram regiões amplificadas tanto no vetor NT quanto no CT.

O mesmo experimento foi realizado com o clone 3 (PRPP sintase), utilizando mais um tipo de construção com dois oligonucleotídeos 5' – T7 e OligoTOPO (fig42).

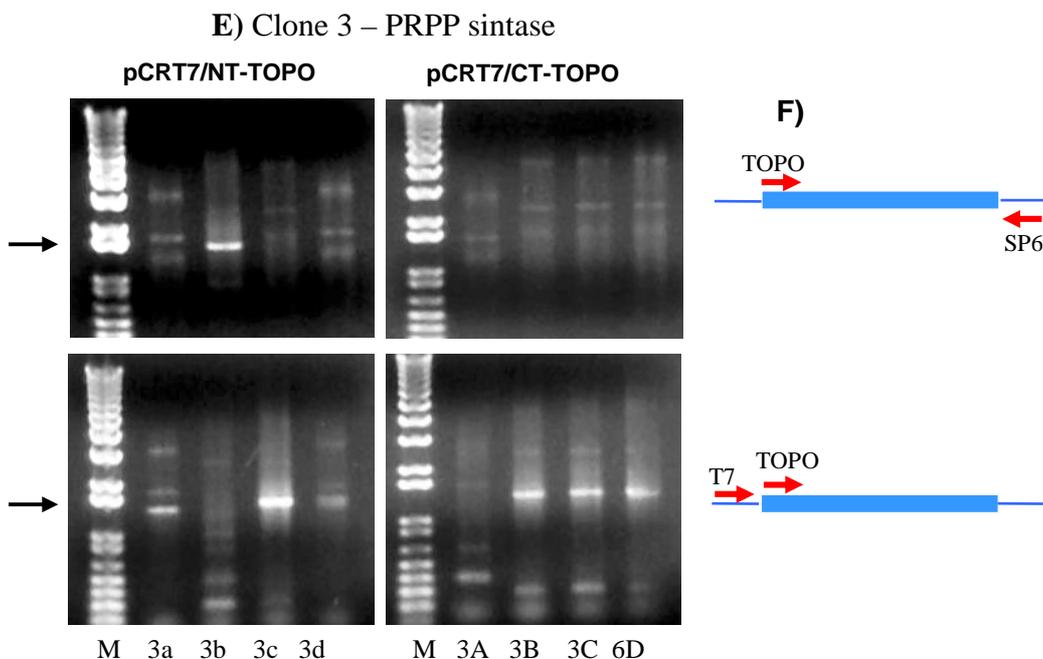


FIG42: Experimento de amplificação do gene da PRPP sintase. Representa a checagem da orientação da construção recombinante a partir de amplificação por PCR com combinação de oligonucleotídeos. **E)** representa o clone da PRPP sintase amplificado que será usado em testes de expressão. **F)** esquema da combinação dos oligonucleotídeos TOPO e SP6 e dos oligonucleotídeos T7 e TOPO. M representa o marcador de massa molecular de 1 kb plus. A numeração composta de letras na parte inferior do gel representa as colônias selecionadas para verificação.

De cada resultado apresentado nos géis (fig41 e 42) selecionou-se o produto recombinante pCRT7 TOPO CT e NT dos clones 1, 3 e 9 da PRPP sintase juntamente com um controle negativo (tabela18).

TABELA 18: Seleção das colônias resultantes da ligação do inserto ao vetor, verificadas por PCR de amplificação

CLONE PRPP	pCRT7 NT TOPO		pCRT7 CT TOPO	
	POSITIVA	NEGATIVA	POSITIVA	NEGATIVA
1	b	c	B	C
3	b	c	A	B
9	c	d	A	C

Estas construções recombinantes foram usadas para transformar células competentes de *E. coli* BL21(DE3) pLysS como descrito em Material e Métodos fazendo-se então uma expressão piloto. Apenas o clone 3b pCRT7 TOPO NT ofereceu resultado positivo (fig43). O controle negativo utilizado foi o plasmídeo religado designado 3c pCRT7 TOPO NT onde o inserto se encontra ligado na orientação oposta no vetor.

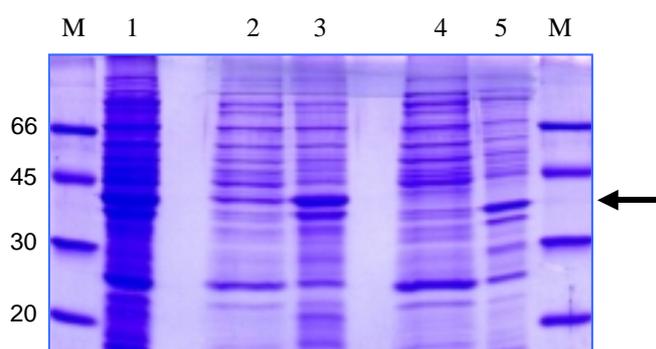


FIG43: Resultado do teste de expressão com o clone 3b da PRPP sintase. Análise feita em 15% SDS-PAGE para extrato protéico induzido e não induzido do sistema de expressão. Na coluna M foi aplicado o padrão de massa molecular em kDa. A numeração tem a seguinte representação: **Clone 3bNT:** 1 - culturas de células antes da lise celular; 2 – sobrenadante após a lise celular; 3 - precipitado após lise celular. Controle negativo Clone 3cNT – 4: sobrenadante após lise por sonicação e 5 representa o precipitado após lise das células. A seta indica a migração da proteína PRPP sintase.

4.5.1 - TESTES DE INDUÇÃO COM A PROTEÍNA PRPP SINTASE

Com base nos resultados da expressão piloto foram feitos testes de indução variando a temperatura e a concentração de IPTG. A melhor condição apresentada foi à temperatura de 37°C e 1 mM IPTG (fig44A). Partindo-se, então, para a expressão da proteína larga escala.

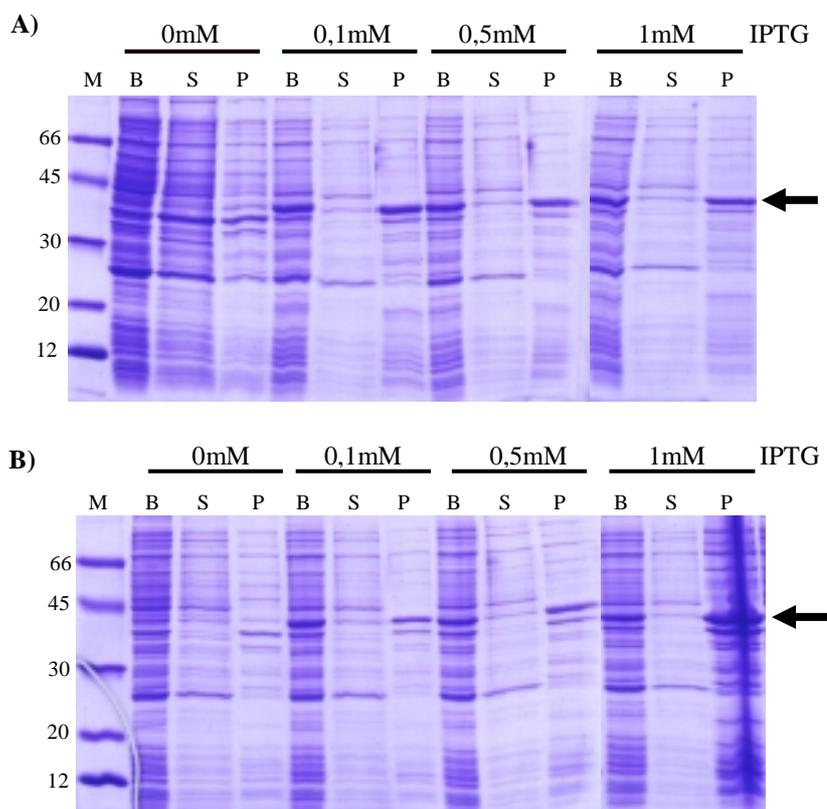


FIG44: Resultados dos testes de expressão da proteína PRPP sintase. Analisada em 15% SDS-PAGE. Na coluna M aplicou-se o padrão de massa molecular em kDa. B representa o extrato bruto não lisado. S e P – cultura de células após a lise: Sobrenadante e Precipitado respectivamente. A seta indica a migração da proteína expressa. **A)** representa o teste realizado a temperatura de 37°C. **B)** representa o teste realizado a temperatura de 20°C.

4.6 - PURIFICAÇÃO DA PROTEÍNA PRPP SINTASE

4.6.1 - PRECIPITAÇÃO COM SULFATO DE AMÔNIO

O resultado da purificação parcial através de fracionamento com sulfato de amônio foi analisado em SDS-PAGE 15% mostrando que a proteína precipita a uma concentração de 35% e 40% de sulfato de amônio como demonstrado no gel da figura 45.

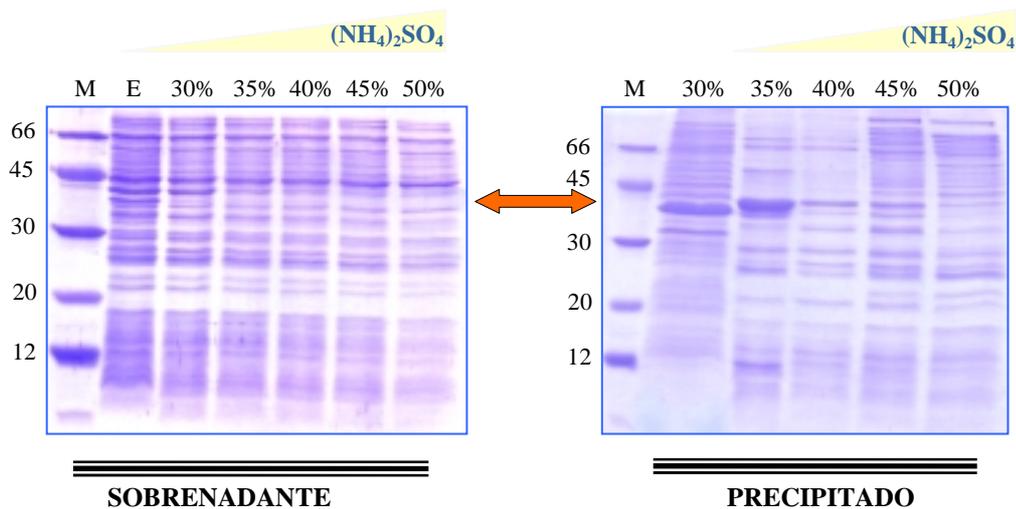


FIG45: Resultado do fracionamento com sulfato de amônio da PRPP sintase (3bNT) em 15% SDS-PAGE. M representa o marcador de massa molecular em kDa. A numeração tem a seguinte representação: E representa o sobrenadante resultante da clarificação do lisado. De 30% a 50% a concentração de sulfato de amônio utilizada na precipitação da proteína. A seta indica a migração da PRPP sintase no gel.

4.6.2 - CROMATOGRAFIA LÍQUIDA DE AFINIDADE

A proteína precipitada em 35% e 40% de sulfato de amônio foi ressuspendida em 20 mL de Tampão A, em seguida dialisada contra o mesmo tampão, sendo então submetida à cromatografia líquida de afinidade em resina de Ni-NTA sob um gradiente linear de 10 mM a 500 mM de imidazol. A proteína foi eluída na concentração 260 mM de imidazol e o resultado dessa eluição pode ser verificado no cromatograma apresentado na figura 46.

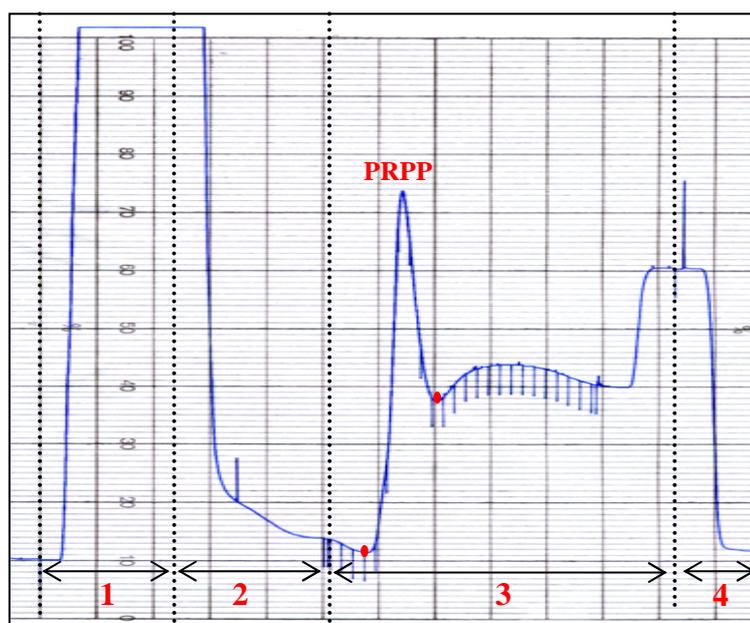


FIG46: Perfil cromatográfico da eluição da proteína em coluna de Ni-NTA. A numeração na parte inferior do cromatograma representa: 1 – aplicação da amostra; 2 – lavagem com dois volumes de coluna usando o tampão A; 3 – eluição da PRPP sintase com um gradiente linear de 10 mM a 500 mM de imidazol; 4 – lavagem da coluna com um gradiente linear de 500 mM a 10 mM de imidazol com dois volumes de coluna. A marcação (•) vermelha representa a saída da PRPP sintase.

O resultado foi analisado em 15% SDS-PAGE que apresentou as frações purificadas da proteína (fig47).

Através do gel determinou-se que a proteína recombinante contendo a cauda de histidina apresenta massa molecular de aproximadamente 42 kDa.

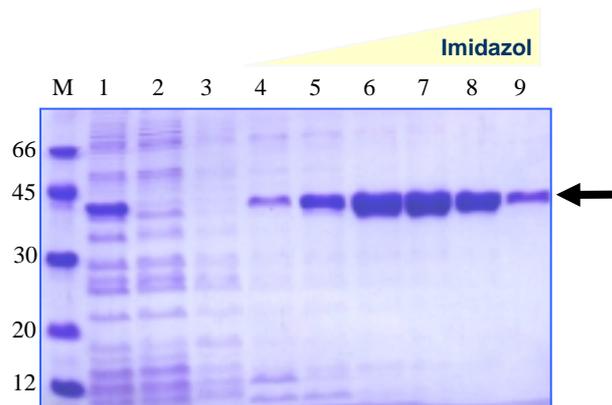


FIG47: Resultado da purificação através de cromatografia de afinidade – Ni-NTA. Na coluna M foi aplicado um padrão de massa molecular formado de proteínas de 66, 45, 30, 20 kDa. As colunas representam: 1– fração solúvel; 2 – “void”; 3 – lavagem com tampão A contendo 10 mM imidazol; 4 a 9 representam as frações eluídas com um gradiente de concentração de 10 mM a 500 mM de imidazol. A seta indica a proteína pura.

4.6.3 - PURIFICAÇÃO POR EXCLUSÃO MOLECULAR EM COLUNA SUPERDEX 200 HR 10/30

A proteína após eluída em coluna de Ni-NTA foi aplicada na coluna de exclusão molecular Superdex 200 HR 10/30, apresentando-se em um único pico com volume de eluição de 11 mL (fig48). A proteína purificada em um único pico pode ser visualizada no gel da figura 49, apresentando como valor de massa molecular 42 kDa.

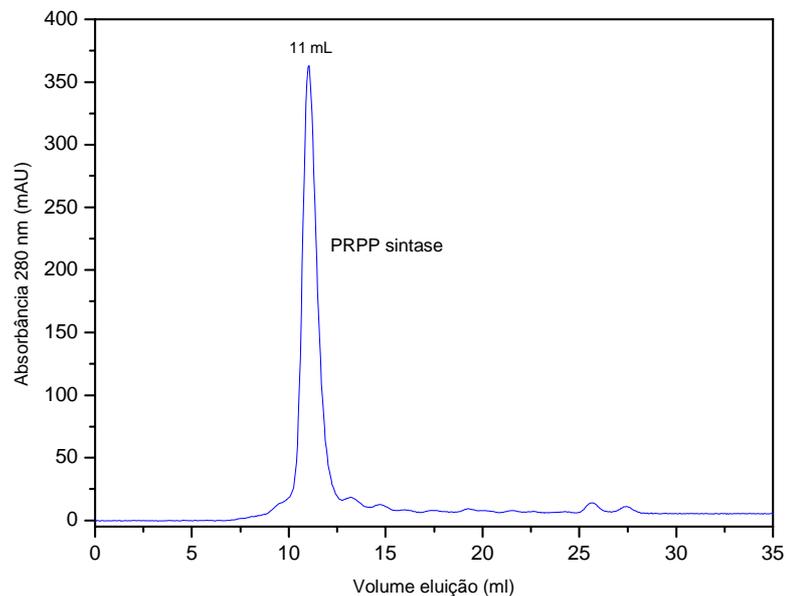


FIG48: Perfil cromatográfico da eluição da PRPP sintase na coluna Superdex 200 HR 10/30.
O pico representa a PRPP sintase com volume de eluição de 11 mL.

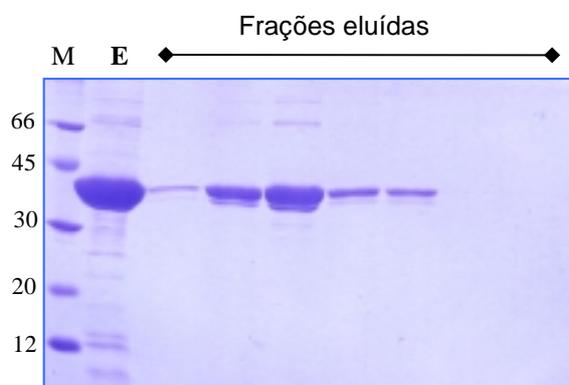


FIG49: Resultado da purificação em coluna de exclusão molecular Superdex 200 HR 10/30.
M representa o padrão de massa molecular em kDa. E representa a amostra injetada na coluna. A barra representa as frações eluídas da proteína.

A proteína purificada depois de concentrada oferece um rendimento de aproximadamente 3 mg por litro de meio de cultura.

4.7 - CARACTERIZAÇÃO DA PROTEÍNA RECOMBINANTE PRPP SINTASE

4.7.1 - ANÁLISE DA DETERMINAÇÃO DA MASSA MOLECULAR

Para verificar o estado oligomérico da proteína foi realizada uma cromatografia com vários padrões de massa molecular, como descrito no Material e Métodos. O resultado do cromatograma pode ser visualizado na figura 50.

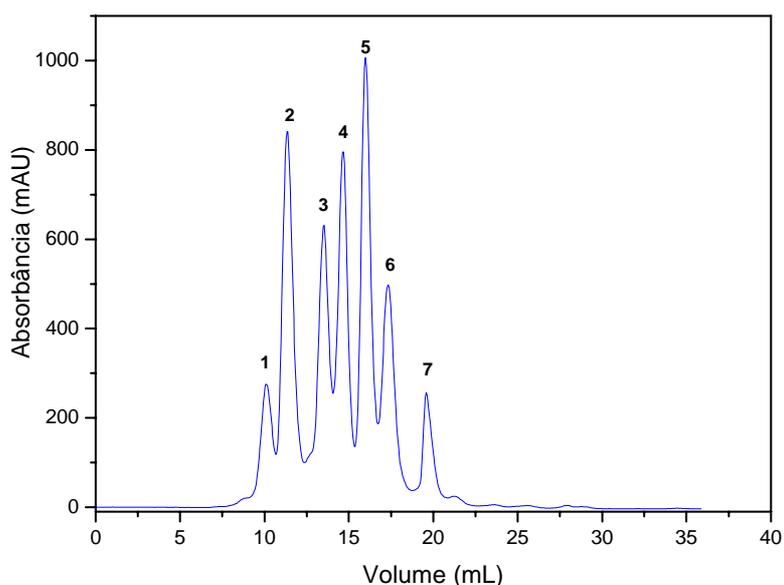


FIG50: Cromatografia de calibração da coluna Superdex 200 HR 10/30. Utilizaram-se proteínas conhecidas com diferentes massas moleculares. Os números 1, 2, 3, 4, 5, 6 e 7 na saída dos picos representam as massas moleculares de 440, 152, 66, 45, 36, 29, 13 kDa respectivamente.

A tabela 19 apresenta todos os padrões de massa molecular utilizados juntamente com seus respectivos volumes de eluição.

TABELA19: Padrões de massa molecular utilizado na coluna Superdex 200 HR 10/30 e seus respectivos volumes de eluição

Padrões de MW	MW (Da)	Ve (mL)
Ferritina	440.000	10.11
β -Amilase	152.000	11.37
Albumina sérica bovina	66.000	13.52
Ovalbumina	45.000	14.66
Gliceraldeído-3-fosfato desidrogenase	36.000	16.00
Anidrase Carbônica	29.000	17.33
Ribonuclease A	13.000	19.58

Utilizando o V_e de cada proteína padrão e o Log da massa molecular construiu-se o gráfico de regressão linear que apresenta a equação da reta descrevendo o comportamento apresentado pela curva como sendo o $\text{Log MW} = 4,04403 - 0,15583 \times V_e$ (figura 51)

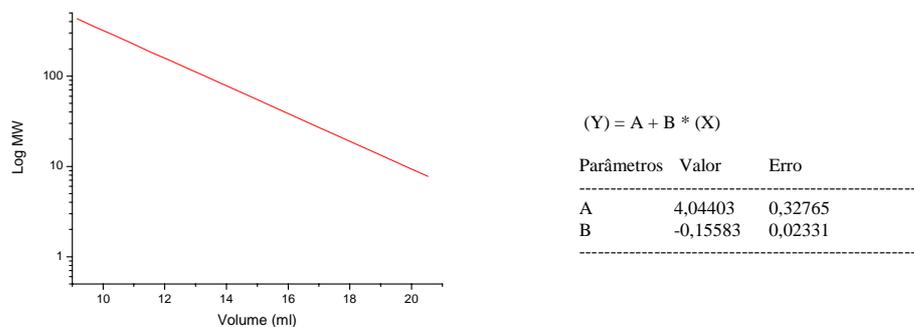


FIG51: Curva de calibração da coluna Superdex 200 HR 10/30. Utilizaram-se proteínas conhecidas com diferentes massas moleculares. A equação da reta é representada por $\text{Log MW} = 4,04403 - 0,15583 \times V_e$.

O perfil obtido no cromatograma da figura 48 forneceu como volume de eluição para a PRPP sintase o valor 11 mL que aplicado a equação da reta determinou o valor da massa molecular da proteína como sendo de 214 kDa,

tendo a proteína um monômero de massa molecular de 42 kDa pode-se dizer que a PRPP sintase apresenta-se sob a forma de um homopentâmero.

4.7.2 - EXPERIMENTO DE ESPALHAMENTO DINÂMICO DE LUZ

O resultado da massa molecular obtido por cromatografia de exclusão molecular foi confirmado pela técnica de Espalhamento Dinâmico de Luz.

A partir do histograma (fig52) obtém-se o raio hidrodinâmico da amostra de 6,04 nm podendo-se a partir deste estimar que a massa molecular da proteína é de 226 kDa.

Pelo perfil do histograma pode-se dizer que a proteína também se apresenta homogênea.

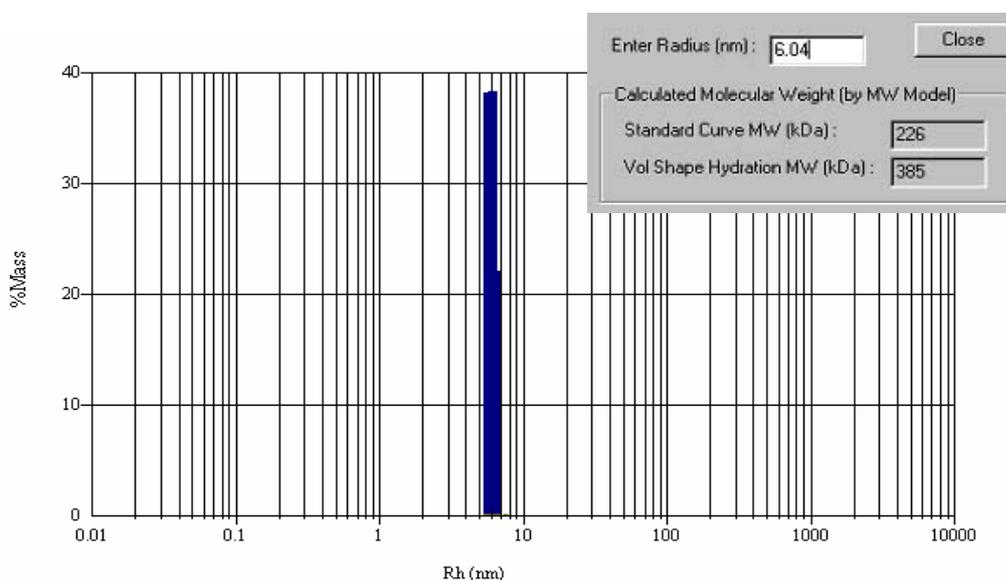


FIG52: Histograma representativo do R_H da proteína PRPP sintase. O valor do R_H apresentado é de 6,04 o que fornece uma massa molecular de 226 kDa.

4.7.3 - DETERMINAÇÃO DO PONTO ISOELÉTRICO

O ponto isoelétrico determinado experimentalmente (fig53) a partir da proteína purificada foi de 6,5.

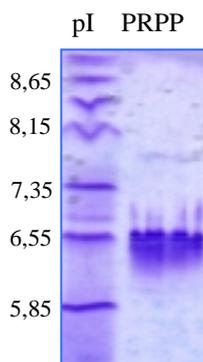


FIG53: Determinação do ponto isoelétrico da proteína PRPP sintase. pI representa o padrão de ponto isoelétrico. PRPP representa a migração da proteína até o ponto onde a somatória de suas cargas é zero ou nula.

4.7.4 - EXPERIMENTO DE DICROÍSMO CIRCULAR

O espectro da PRPP sintase a pH 7,8 pode ser visualizado na figura 54 representado pelo espectro azul. Este espectro apresenta dois mínimos, um em 210 nm e outro em 222 nm, apresenta também um máximo próximo a 194 nm.

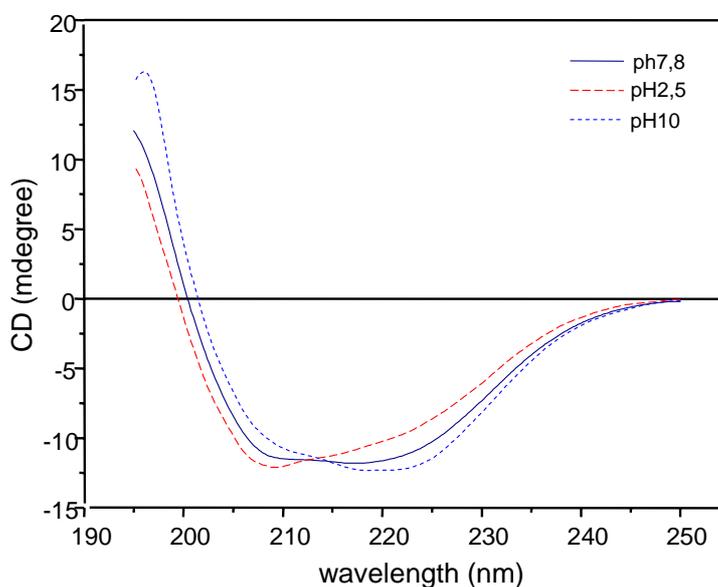


FIG54: Espectro de CD da proteína PRPP sintase. Espectro de CD da PRPP sintase em pH 7,8 (linha cheia azul), em pH 10,0 (linha pontilhada azul), em pH 2,5 (linha pontilhada vermelha) foram realizados a partir de 195 a 250 nm em uma cubeta com caminho óptico de 1 mm com uma média de 16 medidas a 25°C.

A PRPP sintase foi testada em diferentes pHs. A figura 54 mostra o espectro de CD da proteína em pH 2,5 e pH 10. No intervalo de pH de 3,5 a pH 9,0 o espectro de CD da proteína foi o mesmo apresentado pela proteína no pH 7,8. No pH 2,5 e pH 10 verifica-se uma variação na estrutura secundária da proteína que pode ser verificado a partir da desconvolução dos espectros (tabela 20) o qual fornece os componentes de estrutura secundária da proteína.

TABELA 20: Desconvolução dos espectros de CD através do programa SELCON 2.

pH	7,8	2,5	10,0
Hélice α	21%	16%	21%
Folhas β	30%	36%	30%
Desordenadas	49%	48%	49%

4.7.5 - CLIVAGEM COM A ENZIMA ENTEROQUINASE_{Max} (EK_{Max})

Para clivagem da cauda de histidina e resíduos de aminoácidos adicionais derivados do vetor de expressão, fez-se uma clivagem proteolítica com EK_{Max} a duas temperaturas, 18°C e 4°C que foram analisadas em 15% SDS-PAGE representados na figura 55A e 55B respectivamente.

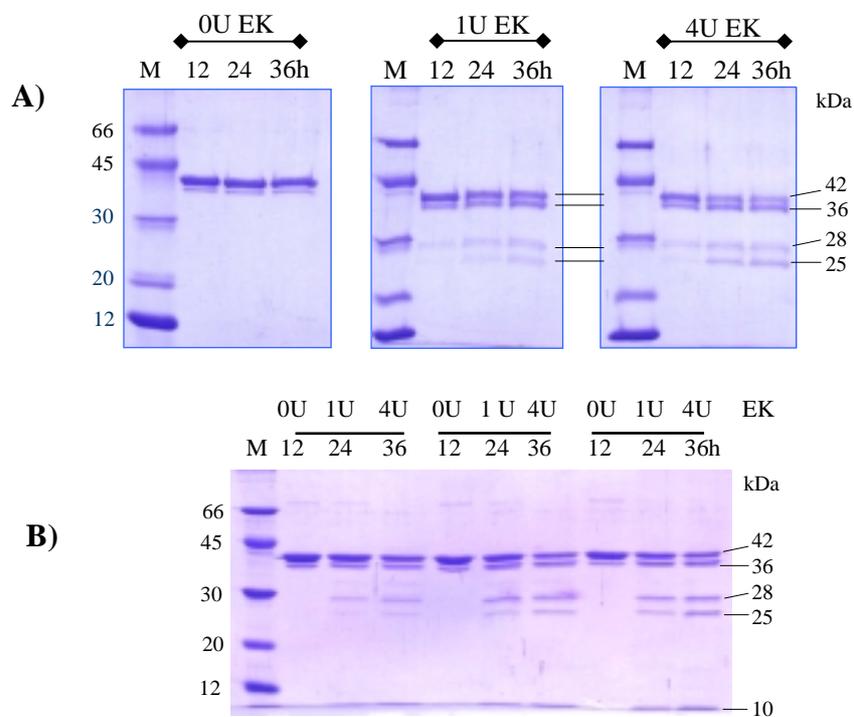


FIG55: Experimento de clivagem da proteína PRPP sintase com a enzima EK_{Max}. M representa o padrão de massa molecular em kDa. Os valores acima do gel representam a quantidade de EK_{Max} utilizada na digestão. 12h, 24h e 36h representam o tempo de incubação em horas. Os valores à direita do gel representam a massa molecular dos produtos de clivagem. **A)** e **B)** representam os resultados da incubação na temperatura de 18°C e de 4°C respectivamente.

A clivagem apresentou como produto quatro bandas com diferentes massas moleculares (fig55) que foram submetidas ao seqüenciamento N-terminal. A melhor condição de clivagem foi estabelecida como sendo a 18°C, por 36 horas de incubação utilizando 1U da enzima EKMax em solução contendo 10 µg da proteína PRPP sintase.

4.7.6 - DETERMINAÇÃO N-TERMINAL

O produto da clivagem da proteína com EKMax, foi submetido a seqüenciamento N-terminal, cuja seqüência de aminoácidos de cada produto individual está descrito na figura 56.

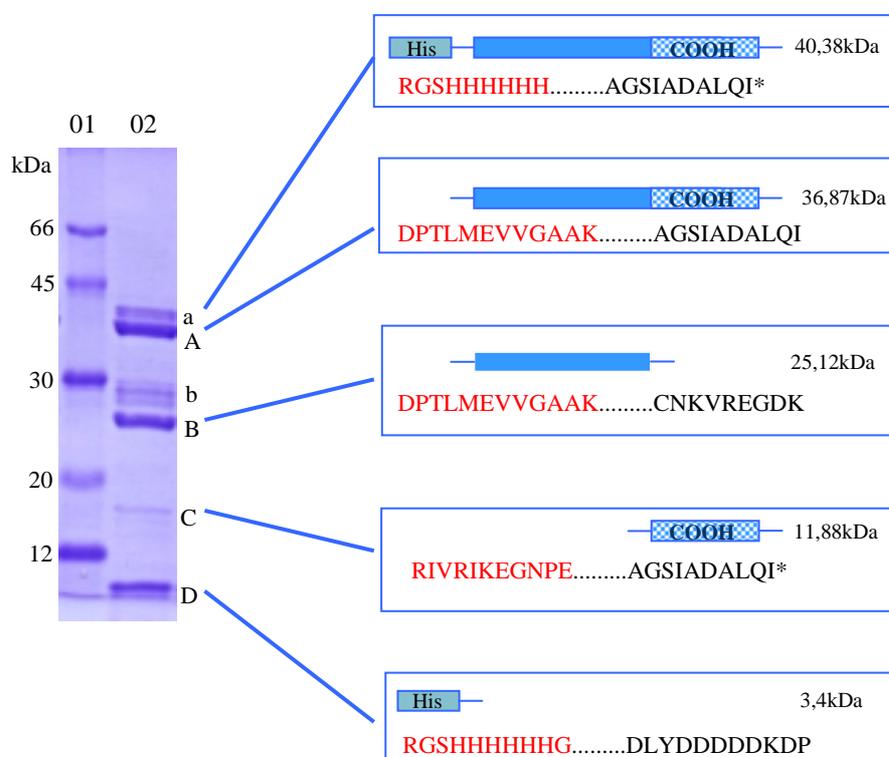


FIG56: Produto de digestão da proteína com a enzima EKMax submetido a seqüenciamento N-terminal. 01 – padrão de massa molecular. 02 – resultado da clivagem da proteína

recombinante. As caixas representam um esquema da clivagem juntamente com a seqüência de aminoácidos obtida através seqüenciamento N-terminal da proteína pelo método de Degradação de Edman representadas em vermelho, e a massa molecular de cada produto da clivagem está representada. (*) representa o códon de terminação. (.....) representa os aminoácidos internos.

A seqüência N-terminal da proteína determinada foi a MEVVGAAK que condiz com o N-terminal esperado.

As bandas a e b não foram sujeitas ao seqüenciamento N-terminal. A banda **a** refere-se à proteína não digerida constituída pela seqüência completa da proteína com massa molecular de 40,38 kDa. A banda **b** um possível contaminante. A banda **A** corresponde a proteína inteira sem a cauda de histidina com massa molecular de 36,87 kDa contendo o N-terminal (MEVVGAAK). A banda **B** com massa molecular de 25,12 kDa representa a proteína sem a cauda de histidina contendo um sítio de clivagem interno, portanto contendo a seqüência N-terminal com a região C-terminal clivada (RIVRIKEGNPE). A banda **C** de 11,88 kDa determinou ser a proteína clivada internamente apresentando apenas a região carboxi terminal. Por último tem-se um banda **D** com massa molecular de 3,4 kDa representando a cauda de histidina clivada.

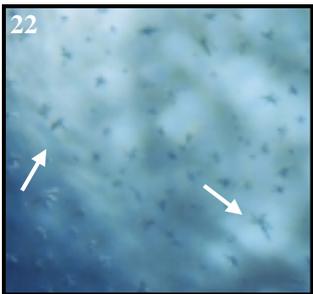
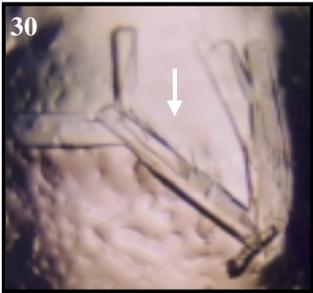
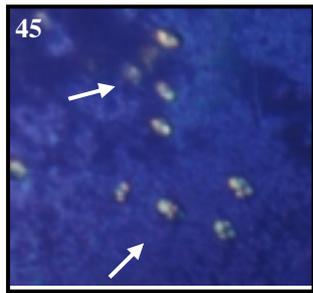
4.8 - ENSAIOS DE CRISTALIZAÇÃO

Com a proteína purificada e concentrada (9 mg/mL) foram realizados testes cristalização utilizando o Fatorial I (“Crystal Screen I” - 50 condições) e Fatorial II (“Crystal Screen I” 48 condições) {Hampton Research} a temperatura de 18°C. Na maioria das gotas observou-se a presença de um precipitado amorfo. Por outro lado várias outras condições apresentaram a formação de cristais junto

com precipitados (tabela 21 e 22). O cristal da condição 5 do fatorial II foi o que apresentou melhor resultado difratando a $3,7\text{\AA}$ de resolução podendo-se determinar seu sistema cristalino como sendo um sistema tetragonal.

TABELA21: Ensaios de cristalização com Fatorial I da Hampton Research

CONDIÇÃO	SAL	TAMPÃO	PRECIPITANTE
03	-	-	Fosfato desidrogenado de amônio 0,4M
16	-	Hepes de sódio 0,1 M Ph 7,5	Sulfato monohidratado de lítio 1,5 M
36	-	Tris/HCl 0,1M pH 8,5	8% PEG 8.000

TABELA22: Ensaios de cristalização com Fatorial II da Hampton Research	
	
	

CONDIÇÃO	SAL	TAMPÃO	PRECIPITANTE
05	Sulfato de amônio 2,0 M	-	5% (v/v) de Isopropanol
22	-	Mes 0,1 M pH 6,5	12% PEG 20.000
30	-	Mes 0,1 M pH 7,5	10% PEG 6.000, 5% MPD
41	Cloreto hexaidratado de níquel 0,01 M	Tris 0,1 M pH 8,5	Sulfato de lítio 1,0 M

DISCUSSÃO E CONCLUSÕES

5 - DISCUSSÃO E CONCLUSÕES

Para desenvolvimento deste trabalho foram identificados a partir do banco de dados do SUCEST (projeto Genoma da Cana-de-açúcar) treze clones (tabela 2, p. 32) envolvidos no metabolismo de purinonucleotídeos (via da síntese *de novo*) que é de central importância aos organismos vivos e tem sido investigada como um possível alvo para o planejamento de drogas que combatam agentes fitopatogênicos, a exemplo do que se tem feito com muitos parasitas. Enzimas desta via metabólica em plantas têm sido pouco estudadas, por isso o interesse em caracterizar uma enzima desta via presente em cana-de-açúcar e devido ao fato desta planta ser uma das mais importantes culturas cultivadas no país.

De todos os clones de cDNA selecionados apenas o clone 3 (SCACRZ3035C11) foi caracterizado. Este clone representa a enzima Fosforribosil pirofosfato sintase (PRPP sintase) que é a iniciadora da cascata enzimática da síntese *de novo* de purinas (fig16, p. 21) e cuja ausência impede o funcionamento de toda a via metabólica. A PRPP sintase é importante em outras diferentes vias metabólicas como a via fosfato pentose, via de recuperação de purina e pirimidina nucleotídeos, na biossíntese de coenzimas e na síntese dos aminoácidos histidina e triptofano.

O grande papel desta enzima foi o principal motivo de seleção deste clone além do fato de ter apresentado resultados positivos, do alinhamento a expressão da proteína em células de *E. coli*.

Um fragmento de 1194 pb foi amplificado a partir do clone de cDNA identificado no banco de dados do SUCEST como contendo a sequência completa

do gene *prs* da PRPP sintase. A estratégia de amplificação empregou um oligonucleotídeo específico 5' e um oligonucleotídeo específico promotor SP6 do vetor como oligonucleotídeo 3'. Esta estratégia permitiu assim clonar a fase aberta de leitura ("ORF") diretamente para o vetor de expressão pCRT7 NT TOPO. Esta amplificação foi possível, pois o gene foi obtido a partir de um clone de cDNA que foi clonado direcionalmente dentro do vetor pSPORT [6] que permitiu o uso da seqüência conhecida 5' a partir da fase aberta de leitura do gene e possibilitou também o uso do oligonucleotídeo SP6 da região promotora do vetor pSPORT [6].

A seqüência amplificada contém uma "ORF" de 984 pb, que codifica para uma proteína de 328 aminoácidos com massa molecular de 36,6 kDa (fig36, p. 74). O ponto isoelétrico determinado experimentalmente da proteína foi de 6,55 que é consistente com o pI de 6,31 predito a partir da seqüência de aminoácidos (fig53, p. 93). Estes valores de MW e pI condizem com os valores encontrados para outros organismos já estudados cujas massas moleculares dos monômeros variam de 33 a 57 kDa [69, 70].

O códon iniciador ATG do gene *prs* da cana-de-açúcar é precedido por 180 nucleotídeos da seqüência 5' – UTR. Um códon de terminação TGA em fase está presente na posição 1165. Um sítio de poliadenilação com 209 nucleotídeos a partir do códon de terminação TGA é observado. Estes resultados indicam um mRNA transcrito de 1374 nucleotídeos (fig36, p. 74).

O alinhamento dois a dois da seqüência de aminoácidos da PRS da cana-de-açúcar com proteínas PRS de *A. thaliana*, *O. sativa* e *S. oleracea*, *S. cerevisiae*, *H. sapiens*, *E. coli*) (tabela 17, p. 78) mostrou que a proteína PRS da cana-de-

açúcar apresenta 77% e 78% de identidade com a isoenzima PRS4 de *A. thaliana* e *S. oleracea* respectivamente, podendo-se dizer que a PRS da cana-de-açúcar é uma isoenzima do tipo PRS4 devido a grande identidade encontrada. A seqüência de *O. sativa* não tem sido identificada como isoenzima, entretanto a partir da comparação feita neste trabalho (tabela 17, p. 78) pode-se identificá-la como a isoforma PRS3, pois apresenta 79% de identidade com as isoenzimas PRS3 de *A. thaliana* e *S. oleracea*. Já com a PRS da cana-de-açúcar a PRS de *O. sativa* apresenta 73% de identidade. Analisando o alinhamento com os demais organismos verifica-se que estes valores de identidade são no máximo de 37%.

Analisando o alinhamento da PRPP sintase de cana-de-açúcar com PRS de plantas homólogas (fig37, p. 76) verifica-se a formação de dois blocos de alinhamento. O primeiro bloco constituído pelas PRS 4 e 3 de *A. thaliana* e *S. oleracea* e a PRS de *O. sativa* que fazem parte das PRPP sintases das quais pode-se sugerir serem enzimas independentes de fosfato inorgânico (P_i) para sua atividade. Isso se explica através de experimentos realizados com a isoenzima PRS1 e 2 de *S. oleracea* que mostram que estas são dependentes da presença do P_i e que suas atividades são inibidas pelo ADP, o que não acontece com as isoenzimas PRS3 e 4 que independem do P_i e são insensíveis à inibição por ADP nas mesmas condições, apresentando desta maneira uma nova classe de isoenzimas PRS3 e 4 [36 e 47]. Esta conclusão também pode ser verificada através de trabalhos desenvolvidos com as isoenzimas PRS3 e 4 de *A. thaliana* confirmando esta independência ao P_i [27]. O segundo bloco apresentado pelas PRS1, 2 e 5 de *A. thaliana* e *S. oleracea* pertencem a classe das isoenzimas dependentes de P_i . A verificação destas duas classes formando dois blocos

distintos deve-se também pela presença nítida de “gaps” que podem ser visualizados na figura 37, p. 76. A inspeção do alinhamento das seqüências revela uma alta divergência destas na região N-terminal enquanto a região C-terminal apresenta-se bastante conservada.

Essas conclusões quanto à identidade das isoformas é consistente com a reconstrução filogenética (fig38, p. 80). É interessante a identificação de um sinal de direcionamento no cloroplasto na seqüência N-terminal das isoenzimas PRS1 e PRS2 de *S. oleracea* e isoenzimas PRS1, 2 e 5 de *A. thaliana* esta observação consiste em dados experimentais obtidos com *S. oleracea* PRS1 e PRS2 sendo esta última sintetizada *in vitro*, importada e processada em cloroplastos de ervilha (*Pisum sativum*) [36].

A relação filogenética da PRS da cana-de-açúcar com 25 seqüências adicionais de aminoácidos de PRPP sintetases a partir de 12 organismos diferentes foi analisada (fig38, p. 80). As isoenzimas PRS1 e 2 de *S. oleracea* dependentes de fosfato formam um grupo com as isoenzimas PRS1 e 2 de *A. thaliana* juntamente com a PRPP sintetase de *Synechocystis* sp. indicando uma estreita relação filogenética, consistente com estas isoenzimas de plantas que possuem uma possível localização em cloroplasto já que *Synechocystis* sp é uma cianobactéria ancestral de plantas e algas verdes.

Um outro grupo é observado sendo constituído pelas isoenzimas PRS de fungos que se relaciona com o grupo constituído pela isoenzimas de mamíferos, que condiz com a evolução filogenética.

Em adição, um grupo de isoenzimas PRS de plantas é evidente a partir de sua filogenia. O grupo contém as isoenzimas PRS3 e PRS4 de *A. thaliana* e *S.*

oleracea independentes de fosfato inorgânico e a PRS de arroz formando um grupo relacionado com a isoenzima PRS4 de cana-de-açúcar. Estas seqüências agrupadas sugerem que a *A. thaliana*, *S. oleracea*, *O. sativa* e a enzima PRS de cana-de-açúcar são também isoenzimas independentes de fosfato constituindo a nova classe das PRPP sintases.

Os 984 pares de bases do gene *prs* contidas no fragmento de PCR foi subclonado do vetor pCRT7 NT TOPO para expressão em larga escala. A construção do plasmídeo resultante foi seqüenciada para confirmar a fidelidade da seqüência e verificar a orientação correta com relação à região promotora. O clone 3 selecionado foi usado na transformação de células competentes *E. coli* BL21(DE3) pLysS. As células de *E. coli* transformadas, induzidas como descrito na figura 44, p. 85, produziram um alto nível da enzima PRPP sintase insolúvel, apresentando-se também

satisfatoriamente na fração solúvel. A expressão em larga escala da proteína PRS foi realizada e a PRS foi purificada a homogeneidade inicialmente por fracionamento com sulfato de amônio (fig45, p. 86) seguido pela cromatografia de afinidade em coluna Ni-NTA eluída com um gradiente contínuo de imidazol (fig46, p. 87). A purificação da PRS de cana-de-açúcar foi visualizada em várias frações separadas em SDS-PAGE corado com “Coomassie blue” (fig47, p. 88). Como mostrado a PRS expressa é eluída em 260 mM de imidazol e migra como uma proteína de 42 kDa em SDS-PAGE. Esta massa é a esperada para seqüência PRS de cana-de-açúcar contendo a cauda de hexahistidinas no N-terminal derivado do vetor de expressão. A opção de clonar do vetor pCRT7 NT TOPO é que este contém a cauda 6xHis na região N-terminal que facilita a purificação e a

torna mais rápida, apresentando também a possibilidade de clivagem desta cauda por digestão com EKMax obtendo a proteína nativa purificada.

Para clivar a cauda de hexahistidinas N-terminal e resíduos de aminoácidos adicionais originados da construção do vetor de expressão, a digestão proteolítica com EKMax foi realizada (fig55, p. 95). A análise da seqüência primária da proteína PRPP sintase de cana-de-açúcar não indicou a presença de uma seqüência de reconhecimento interna para a enzima EKMax. Mas como mostrado na figura 55, o resultado do padrão de clivagem foi mais complexo do que se esperava. Um total de 4 bandas foram visualizadas em SDS-PAGE corado com “Coomassie”. Para determinar a identidade de cada banda fez-se necessário o seqüenciamento N-terminal de cada produto individual. As bandas marcadas em **a** e **b** na figura 56 (p. 96) são a proteína PRS não digerida e um contaminante não identificado, respectivamente. A Banda **A** é a enzima PRS com sua seqüência completa depois que a seqüência N-terminal do vetor foi removida resultando em uma seqüência N-terminal DPTLMEVVGAAK e uma proteína de 36,87 kDa. A banda **B** tem a seqüência N-terminal da enzima PRS esperada clivada corretamente (DPTLMEVVGAAK), mas migra em SDS-PAGE como uma proteína de 25,12 kDa. Esta migração corresponde à ocorrência de uma clivagem interna entre os aminoácidos 219-220 e revela uma PRS C-terminal com um fragmento 11,88 kDa. A banda **C** é consistente com o fragmento C-terminal clivado de 11,88 kDa e a banda **B** a porção N-terminal resultante a partir de um sítio de clivagem interno. Banda **D** no qual não é visível em toda a preparação devido a uma pequena massa molecular foi identificada como sendo a seqüência do vetor (RGSHHHHHHG) contendo a cauda hexahistidina e resíduos de

aminoácidos adicionais com massa molecular de 3,4 kDa (fig56, p. 96). Esta banda migra na frente do gel e sua massa não pode ser verificada desta forma. Banda **C** aparece consistentemente como uma banda fraca, podendo ser resultado de uma pobre ligação do corante “Coomassie” devido a composição da seqüência de aminoácidos do fragmento.

Embora a presença não esperada de um sítio interno de clivagem, a enzima PRPP sintase de cana-de-açúcar sem a seqüência do vetor N-terminal foi fracionada por separação cromatográfica em coluna Superose 12 HR 10/30 com o objetivo de separar os demais fragmentos (dados não mostrados), mas não se obteve o resultado esperado apresentando um perfil cromatográfico com sua eluição visualizado em SDS-PAGE mantendo-se a proteína agrupada com os fragmentos não desejados. Provavelmente a repetição do experimento com fracionamento por troca iônica poderá separar esses fragmentos uma vez que o pI de cada fragmento é de: **a** – 6,22; **A** – 6,3; **B** – 5,91; **C** – 7,08; **D** – 6,27.

A proteína purificada por cromatografia de exclusão molecular (fig49, p. 89) mostra a PRS sob a forma de um homopentâmero de aproximadamente 214 kDa, consistente com as observações obtidas para outras enzimas PRS previamente isoladas. A PRS presente em seringueiras e a PRS de *Giardia intestinalis* apresentam-se sob a forma de um tetrâmero de massa molecular de 200 kDa e 150 kDa com subunidades de 57 kDa e 38 kDa respectivamente [70 e 34]. A PRS de *Bacillus subtilis* com massa molecular de 280 kDa e subunidade de 34 kDa foi consistente com uma estrutura quaternária octamérica [35] e estudos recentes determinaram pela primeira vez a estrutura tridimensional de uma PRS como sendo um hexâmero determinando também seus domínios ligantes [33, 71].

Sabendo-se a posição do sítio ligante do cátion divalente, do motivo ligante da PRPP/Rib-Fosfato e os possíveis aminoácidos ligantes de ATP pode-se determinar as seqüências de aminoácidos correspondentes a estes domínios na seqüência de aminoácidos da PRPP sintase através do alinhamento desta com a seqüência de aminoácidos do *Bacillus subtilis*. Estes domínios podem ser visualizados na figura 37, p. 76.

A análise da estrutura quaternária da proteína foi confirmada pelo experimento de espalhamento dinâmico de luz que forneceu o raio hidrodinâmico de 6,04 nm. A massa molecular estimada a partir deste raio foi de 226 kDa consistente com a forma de um pentâmero apresentado pela cromatografia de exclusão molecular (fig52, p. 92).

A técnica de Dicroísmo Circular (CD) foi usada para a análise da distribuição dos componentes de estrutura secundária da proteína PRS de cana-de-açúcar (fig54, p. 94). A PRPP sintase teve verificado a sua composição de estrutura secundária em pH 7,8 apresentando um espectro caracterizado por dois mínimos um em 210 nm e outro a 222 nm e um máximo próximo a 194 nm que pode ser visualizado na figura 54. A estabilidade da proteína em função da variação de pH de 2,5 a pH 10,0 também foi analisada, apresentando uma pequena variação nos componentes de estrutura secundária principalmente em pH 2,5 que tiveram seus mínimos e máximos modificados se comparado com o espectro da proteína em pH 7,8.

A desconvolução do perfil em pH 7,8 mostrou 21% de hélice α , 30% de componentes β (folhas β paralela e antiparalela) e 49% de outras contribuições (não ordenadas e “turns”), com um desvio médio quadrático (r.m.s.d.) de 1%. A

desconvolução do espectro gerado pelo pH 10,0 mostrou valores de componentes de estrutura secundária semelhantes ao espectro desconvolvido observado em pH 7,8, ambos são compatíveis com a composição de estrutura secundária de PRS de *Bacillus subtilis* [33].

No extremo de pH 2,5 os componentes de estrutura secundária obtidos foram 16% hélice α , 36% de componentes β (folhas β paralela e antiparalela) e 48% de outras contribuições e o desvio médio quadrático foi de 12%. Este alto valor do r.m.s.d. indica uma diferença estrutural significativa induzida por condições extremas de pH. Através da análise de CD pode-se observar que alterações estruturais estão limitadas ao pH extremo e ocorrem abruptamente em transições de pH 3,5 para pH 2,5 e de pH 9,0 para pH 10,0. Estes resultados mostram uma interessante estabilidade da PRPP sintase de cana-de-açúcar no intervalo de pH de 3,5 a pH 9,0.

Pelas propriedades apresentadas da PRPP sintase de cana-de-açúcar incluindo sua expressão solúvel, fácil purificação e rendimento médio de 3 mg por litro de meio de cultura sugere que esta seja uma boa candidata à cristalização e determinação da estrutura por análises de difração de raios X.

Partindo-se deste princípio a proteína foi utilizada para ensaios de cristalização testando-a nos fatoriais I e II. Obteve-se a formação de precipitados amorfos em algumas condições e em outras se verificou a formação de cristais como mostrado nas tabelas 21 e 22 (p. 98). A obtenção destes cristais nos fornece um passo na resolução da estrutura tridimensional da enzima PRPP sintase de cana-de-açúcar por cristalografia de raios-X. As condições nas quais houve

crescimento de cristais estão sujeitas a refinamento dos parâmetros de cristalização.

Os resultados descritos são importantes no avanço da caracterização de proteínas PRS de plantas que atualmente são pouco descritas, incluindo a possibilidade da realização de estudos estruturais em tais enzimas. Também, temos fortes evidências filogenéticas indicando PRS homólogas aqui descrita como uma nova classe de isoenzimas independente de fosfato inorgânico para sua atividade, representando uma contribuição no estudo desta família de enzimas. Futuramente a partir da cristalização pode-se ter a primeira PRPP sintase de planta com estrutura tridimensional resolvida por cristalografia de raios-X facilitando o desenvolvimento de drogas que combatam doenças em plantas, realizar melhoramento genético e principalmente servir de estudo comparativo em pesquisas com plantas homólogas, compreendendo-se melhor como a PRPP sintase atua sobre o mecanismo da via metabólica da síntese *de novo* de purinonucleotídeos.

**PERSPECTIVAS
FUTURAS**



6 - PERSPECTIVAS FUTURAS

- ✚ Caracterização dos parâmetros cinéticos da proteína PRPP sintase.
- ✚ Otimizar os experimentos de clivagem da cauda hexahistidina N-terminal da proteína.
- ✚ Realizar experimentos de difração de raios-X com os cristais obtidos.
- ✚ Refinar as condições de cristalização da proteína e realizar experimentos de cristalização com a proteína nativa.
- ✚ Determinar a estrutura tridimensional da proteína PRPP sintase e refiná-la.

**REFERÊNCIAS
BIBLIOGRÁFICAS**

7 - REFERÊNCIAS BIBLIOGRÁFICAS

[1] Sterky, F. and Lundeberg, J. (2000). Sequence analysis of genes and genomes. *J. Biotech.* 76: 1-31.

[2] International Human Genome Sequencing Consortium (2001). Initial Sequencing and Analysis of the Human Genome. *Nature.* 409: 860-921.

[3] The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis (2000). The Genome Sequence of the Plant Pathogen *Xylella fastidiosa*. *Nature.* 406: 151-157.

[4] Vettore A. L.; Silva F.R.; Kemper E. L. and Arruda P. (2001). The libraries that made SUCEST. *Genetics and Mol. Biol.* 24(1-4): 1-7.

[5] Telles, G. P., and Silva, F. R. (2001). Trimming and Clustering Sugarcane ESTs. *Genetics and Mol. Biol.* 24(1-4): 17-23.

[6] Jancso M. A.; Sculaccio S. A. and Thiemann O. H. (2001). Identification of sugarcane genes involved in the purine synthesis pathway. *Genetics and Mol. Biol.* 24(1-4): 251-255.

[7] Stryer, L. (1992). Bioquímica. 3ª ed. Rio de Janeiro/RJ. Ed. Guanabara Koogan S.A. 495.

[8] Nelson, D. L. and Cox, M. M. (2000). Principles of Biochemistry. 3ª ed. New York. Worth Publishers. 848 -849.

[9] Kornberg, A., Lieberman J., and Simms E. S. (1954). *JACS* 76: 2027. (1955). *J. Biol. Chem.* 215-389.

[10] Berens, R. L., Krug, E. C., and Marr, J. J. (1995). Purine and pyrimidine metabolism: In Biochemistry of Parasitic Organisms and its Molecular Foundations. J. J. Marr and M. Muller, eds. London: Academic Press.

- [11] Hassan, H. F. and Coombs, G. H. (1986). A comparative Study of the Purine and Pyrimidine-metabolising Enzymes of a Range of Trypanosomatids. *Comp. Biochem. Physiol.* 84B: 217-223.
- [12] Henderson, J. F. and Khoo, M. Y. (1965). *J. Biol. Chem.* 240: 2349.
- [13] White, M. N., Olszowy, J., and Switzer, R. L. (1971). *J. Bacteriol.* 108, 122.
- [14] Switzer, R. L. (1969). *J. Biol. Chem.* 244, 2854.
- [15] Wong P. C. L., and Murray, A. W. (1969). *Biochemistry.* 8: 1608.
- [16] Fox, I. H., and Kelley, W. N. (1971). *J. Biol. Chem.* 246: 5739.
- [17] Fernandes, J. F., Pereira, J. P. M., and Nussenzweig, R. S. (1959). *Rev. Brasil. Biol.* 19: 331.
- [18] Hager, S. E., and Jones, M. E. (1967). *J. Biol. Chem* 242:5674.
- [19] Hoffmeyer, J., and Neuhard, J. (1971). *J. Bacteriol.* 106: 14.
- [20] Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E. V., Davis, R. W. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science.* 282: 754-759.
- [21] Anderson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C. M., Podowski, R. M., Näslund, A. K., Eriksson, A. S., Winkler, H. H. Kurland, C. G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature.* 396: 133-140.
- [22] Taira, M., Ishijima, S., Kita, K., Yamada, K., Iizasa, T., Tatibana, M. (1987). Nucleotide and deduced amino acid sequences of two distinct cDNAs for rat phosphoribosylpyrophosphate synthetase. *J. Biol. Chem.* 262: 14867-14870.

- [23] Taira, M., Iizasa, T., Shimada, H., Kudoh, J., Shimizu, N., Tatibana, M. (1990). A human testis-specific mRNA for phosphoribosylpyrophosphate synthetase that initiates from a non-AUG codon. *J. Biol. Chem.* 265: 16491-16497.
- [24] Hernando, Y., Parr, A., Schweizer, M. (1998). *PRS5*, the fifth member of the phosphoribosyl pyrophosphate synthetase gene family in *Saccharomyces cerevisiae*, is essential for cell viability in the absence of either *PRS1* or *PRS3*. *J. Bacteriol.* 180: 6404-6407.
- [25] Carter, A. T., Narbad, A., Pearson, B. M., Beck, K. F., Logghe, M., Contreras, R., Schweizer, M. (1994). Phosphoribosylpyrophosphate synthetase (*PRS*): a new gene family in *Saccharomyces cerevisiae*. *Yeast.* 10: 1031-1044.
- [26] Carter, A. T., Beiche, F., Hove-Jensen, B., Narbad, A., Barker, P. J., Schweizer, L. M., Schweizer, M. (1997). *PRS1* is a key member of the family encoding Phosphoribosylpyrophosphate synthetase in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 254: 148-156.
- [27] Krath B. N, Eriksen T. A., Poulsen T. S, Hove-Jensen B. (1999). Cloning and sequencing of cDNAs specifying a novel class of phosphoribosyl diphosphate synthase in *Arabidopsis thaliana*. *Biochim. Biophys. Acta.* 1430(2): 403-8.
- [28] Hove-Jensen B., Harlow K. W., King C. J. & Switzer R. L. Phosphoribosylpyrophosphate synthetase of *Escherichia coli* (1986). Properties of the purified enzyme and primary structure of the *prs* gene. *J. Biol. Chem.* 261: 6765-6771.
- [29] Switzer, R. L. (1969). Regulation and mechanism of phosphoribosylpyrophosphate synthetase I: Purification and properties of the enzyme from *Salmonella typhimurium*. *J. Biol. Chem.* 244: 2854-2863.
- [30] Hove-Jensen, B. (1989). Phosphoribosylpyrophosphate (PRPP)-less mutants of *Escherichia coli*. *Mol. Microbiology* 3(11): 1487-1492.

- [31] Hove-Jensen, B., Harlow, K. W., King, C. J., Switzer, R. L. (1986). Phosphosibosylpyrophosphate synthetase of *Escherichia coli*: Properties of the purified enzyme and primary structure of the *prs* gene. *J. Biol. Chem.* 261: 6765-6771.
- [32] Becker, M. A., Raivio, K. O., and Seegmiller, J. E. (1979). *Adv. Enzymol. Relat. Areas Mol. Biol.* 49: 281-306.
- [33] Eriksen T. A., Kadziola A., Bentsen A. K., Harlow K. W. and Larsen S. (2000). Structural basis for the function of *Bacillus subtilis* phosphoribosylpyrophosphate synthetase. *Nature Struc. Biol.* 7: 303-308.
- [34] Lee, C. S., Asai, T., and O'Sullivan, W. J. (1992). Studies on phosphoribosylpyrophosphate synthetase from *Giardia intestinalis*. *International Journal for Parasitology.* 22: 29-33.
- [35] Arnvig, K., Hove-Jensen, B., and Switzer, R. L. (1990). Purification and properties of phosphoribosyl-diphosphate synthetase from *Bacillus subtilis*. *Eur. J. Biochem.* 192: 195-200.
- [36] Krath B. N. and Hove-Jensen B. (1999). Organellar and cytosolic localization of four phosphoribosyl diphosphate synthase isozymes in Spinach. *Plant. Phys.* 119: 497-505.
- [37] Smith, P. M. C., and Atkins, C. A. (2002). Purine Biosynthesis. Big in Cell Division, Even Bigger in Nitrogen Assimilation. *Plant. Physiology.* 128: 793-802
- [38] Laemmli U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227: 680-685.
- [39] Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., Petersen, G. B. (1977). DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA.* 74: 5463-5467.

- [40] Nelson, D. L. and Cox, M. M. (2000). Principles of Biochemistry. 3^a ed. New York. *Worth Publishers*. 353.
- [41] Innis, M., and Gelfand, D. Optimization of PCR: Conservations between Michael and David. PCR Applications. *Academic Press*. 03-21.
- [42] Sonoda T., Taira M., Ishijima S., Ishizuka T., Iiazasa T. and Tatibana M. (1991). Complete nucleotide sequence of human phosphoribosyl pyrophosphate synthethase subunit I (PRS I) cDNA and a comparison with human and rat PRPS gene families. *J. Biochem.* 109: 361-364.
- [43] Iizasa T., Taira M., Ishijima S., Shimada H. and Tatibana M. (1989). Molecular cloning and sequencing of human cDNA for phosphoribosyl pyrophosphate synthethase subunit II. *FEBS Lett.* 244(1): 47-50.
- [44] Iizasa T., Taira M., Shimada H., Kudoh J., Shimizu N. and Tatibana M. (1990). A human testis-specific mRNA for phosphoribosylpyrophosphate synthethase that initiates from a non-AUG codon. *J. Biol. Chem.* 265: 16491-16497.
- [45] Ishijima S., Taira M. and Tatibana M. (1989). Complete cDNA sequence of rat phosphoribosylpyrophosphate synthethase subunit II (PRS II). *Nucleic Acids Res.* 17 (21): 8859.
- [46] Sasaki T., Matsumoto T. and Yamamoto K.. *Oryza sativa nipponbare* (GA3) genomic DNA, chromosome 1, PAC clone. Published only in database (2001).
- [47] Krath, B. N., and Hove-Jensen, B. (2001). Class II Recombinant Phosphoribosyl Diphosphate Synthase from Spinach. Phosphate Independence and Diphosphoryl Donor Specificity. *J. Biol. Chem.*. 276(21): 17851-17856.
- [48] Himmelreich R., Hilbert H., Plagens H., Pirkl E., Bi-Chen Li and Herrmann R. (1996). Complete sequence analysis of the genome of bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24: 4420-4449.

- [49] Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., Higgins D. G. (1997). The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
- [50] Castreana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol and Evolution.* 17: 540-552.
- [51] Strimmer K. and Haeseler A. von. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13: 964-969.
- [52] Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, university of Washington, Seattle.
- [53] Strimmer K. and Haeseler A. von. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *PNAS (USA).* 94: 6815-6819.
- [54] Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struthl, K. Short Protocols in Molecular Biology. 3^a ed. USA, Ed. John Wiley & Sons, Inc.
- [55] De Robertis, E. M. F., and Hib, J. (2001). Bases da Biologia Celular e Molecular. 3^a ed. Ed. Guanabara. 398.
- [56] Nelson, D. L. and Cox, M. M. (2000). Principles of Biochemistry. 3^a ed. New York. Worth Publishers. 130-133.
- [57] England, S., and Seifter, S. (1990). Methods in Enzymology, vol 182. Precipitation Techniques. *Academic Press. Inc.* 285-296

- [58] Santos, C. N., and Castanho, A. R. B. M. (1996). Teaching Light Scattering Spectroscopy: the dimension and shape of Tobacco Mosaic Virus. *Biophysical Journal*. 71: 1641-1646.
- [59] Garfin, D. E. (1990). Methods in Enzymology, vol 182. Isoelectric Focusing. *Academic Press, Inc.* 459-467.
- [60] Campana, P. T. (1998). Desnaturação e Reenovelamento da Frutalina, uma Lectina Ligante de D-Galactose. Tese de Mestrado, Instituto de Física da Universidade de São Paulo – São Carlos.
- [61] Bortoleto, R. K. (2001). Transformações Estruturais e Função de Fosfolipase. Tese de Doutorado, Departamento de Física, IBRCE, UNESP, Brasil.
- [62] Sreerama N. and Woody R. W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* 209: 32-44.
- [63] Allen, G. (1981). Sequencing of protein and peptides. (T. S. Work and R. H. Burdon, eds) *Elsevier*, Amsterdam, New York.
- [64] Matsudaira, P. (1990). Limited N-terminal Sequence Analysis. Methods in Enzymology, vol 182. *Academic Press, Inc.* 602-626.
- [63] Matsudaira P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* 262(21): 10035-8.
- [64] Ollis, D., and White, S. (1990). Protein Crystallization. Methods in Enzymology, vol 182. *Academic Press, Inc.* 646-659.
- [65] Branden, C., and Tooze, J. (1991). Introduction to Protein Structure. Determination of Protein Structures. New York. *Garland Publishing, Inc.* 269-273

- [66] Bradford M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254.
- [67] Becker, M. A., Meyer, L. J., Huisman, W. H., Lazar, C., and Adams, W. B. (1997). Human Erythrocyte Phosphoribosylpyrophosphate Synthetase. Subunit Analysis and States of Subunit Association. *J. Biol. Chem.* 252(11): 3911-3918.
- [68] Gallois, R., Prévôt, J.- C., Clément, A., and Jacob, J.-L. (1997). Purification and Characterization of phosphoribosylpyrophosphate Synthetase from Rubber Tree Latex. *Plant Physiol.* 115: 847-852.
- [69] Bentsen, A-K., Larsen, T. A., Kadziola, A., Larsen, S., and Harlow, K. W. (1996). Overexpression of *Bacillus subtilis* Phosphoribosylpyrophosphate Synthetase and Crystallization and Preliminary X-Ray Characterization of free Enzyme and Its Substrate-Effector Complexes. *Proteins: Structure, Function, and Genetics.* 24: 238-246.
- [70] Campana, P. T. (1998). Desnaturação e Reenovelamento da Frutalina, uma Lectina Ligante de D-Galactose. Tese de Mestrado, Instituto de Física da Universidade de São Paulo – São Carlos.
- [71] Bortoleto, R. K. (2001). Transformações Estruturais e Função de Fosfolipase. Tese de Doutorado, Departamento de Física, IBRCE, UNESP, Brasil.

ANEXO I

ANEXO

MEIO LB (Luria Bertane)	
Extrato de levedura	5 g/L
Bacto-triptona	10 g/L
Cloreto de Sódio	10 g/L
pH 7,5	
Para LB sólido adicionar 15 g/L de Agar	

TAMPÃO TAE [50X]	
Tris Base pH 8,5	242 g
Ácido acético glacial	51,1 mL
Na ₂ EDTA	37,2 g
Água q.s.p 1 L	

TAMPÃO TBE [10X]		
Tris Base	890 mM	108 g/L
Ácido Bórico	890 mM	55 g/L
Na ₂ EDTA – 2H ₂ O	20 mM	7,44 g/L

GEL DE AGAROSE 0,8%	
Tampão TAE	100 mL
Agarose	0,8 g
Brometo de Etídio	4µL

TAMPÃO EKMA _x [10X]		
Tris/HCl pH 8,0	50 mM	60,5 g/L
CaCl ₂	10 mM	1,47 g/L
Tween-20 (v/v)	1%	10 mL
Ajustar o volume para 1 litro com água q.s.p.		

TAMPÃO TA [3X]

Tris/HCl 2M pH 6,8	2,35 mL	0,188 M
SDS	5 g	6%
Glicerol 87%	8,6 mL	30%
β -Mercaptoetanol	3,75 mL	15%
Azul de Bromofenol 0,5%	0,5 mL	0,01%
Água q.s.p	25 mL	

TAMPÃO DE LISE

Fosfato de potássio pH 7,8	500 mM
NaCl	400 mM
KCl	100 mM
Glicerol	10%
Triton X-100	0,5%
Imidazol	10 mM

TAMPÃO A

Fosfato de potássio pH 7,8	50 mM
NaCl	400 mM
Imidazol	10 mM

TAMPÃO B

Fosfato de potássio pH 7,8	50 mM
NaCl	150 mM

ANEXO II

Identification of sugarcane genes involved in the purine synthesis pathway

Mario A. Jancso¹, Susana A. Sculaccio^{1,2} and Otavio H. Thiemann^{1*}

Abstract

Nucleotide synthesis is of central importance to all cells. In most organisms, the purine nucleotides are synthesized *de novo* from non-nucleotide precursors such as amino acids, ammonia and carbon dioxide. An understanding of the enzymes involved in sugarcane purine synthesis opens the possibility of using these enzymes as targets for chemicals which may be effective in combating phytopathogen. Such an approach has already been applied to several parasites and types of cancer. The strategy described in this paper was applied to identify sugarcane clusters for each step of the *de novo* purine synthesis pathway. Representative sequences of this pathway were chosen from the National Center for Biotechnology Information (NCBI) database and used to search the translated sugarcane expressed sequence tag (SUCEST) database using the available basic local alignment search tool (BLAST) facility. Retrieved clusters were further tested for the statistical significance of the alignment by an implementation (PRSS3) of the Monte Carlo shuffling algorithm calibrated using known protein sequences of divergent taxa along the phylogenetic tree. The sequences were compared to each other and to the sugarcane clusters selected using BLAST analysis, with the resulting table of p-values indicating the degree of divergence of each enzyme within different taxa and in relation to the sugarcane clusters. The results obtained by this strategy allowed us to identify the sugarcane proteins participating in the purine synthesis pathway.

INTRODUCTION

The purine nucleotide synthesis and recycling (or salvage) pathways are of central importance to all living organisms (Marr, 1991) because they provide the purines (adenine and guanine) necessary for nearly all biochemical processes such as DNA and RNA metabolism, the biosynthesis intermediates ATP, GTP and coenzymes as well as being an important metabolic regulator in cell signaling (Stryer, 1995). Due to their importance, purine synthesis pathways have been investigated as potential targets for chemotherapy in several different scenarios (Marr, 1991; Ullman and Carter, 1997; Marr and Ullman, 1995). Cancer cells are dependent on the purine salvage pathway to provide the large quantity of purine nucleotides for RNA synthesis because of the increased intracellular synthesis rates of cancerous cells. Due to differences between host and pathogens, the purine synthesis enzymes have been proposed as potential anti-parasitic targets since several parasites are purine auxotrophs, lacking the *de novo* synthesis pathway and relying entirely on the salvage pathway.

The *de novo* purine synthesis pathway is composed of eleven enzymatic steps leading to the formation of inosinate (Figure 1). The pathway initiates with the synthesis of 5-phosphoribosyl-1-pyrophosphate (PRPP) from ribose-5-phosphate and ATP, a reaction catalyzed by PRPP synthetases (PRS; EC: 2.7.6.1). PRS is an important enzyme involved in the salvage as well as *de novo* pathway

because PRPP is a substrate for the major salvage enzymes adenine-phosphoribosyl-transferase (APRT) and hypoxanthine-guanine-phosphoribosyl-transferase (HGPRT).

Due to its cellular importance and potential as a chemotherapeutic target, the purine synthesis and salvage pathways have been extensively studied in several organisms. Most of the genes encoding for the proteins involved in these pathways have been cloned from several organisms. Interestingly, the genetic organization varies between organisms across diverse taxa and some remarkable gene fusions have been described for several of the enzymes participating on this pathway.

In this paper we describe the identification of the eleven enzymes from the sugarcane expressed sequence (SUCEST) tag project homologous to known *de novo* purine synthesis proteins. Several genes are represented by full-length expressed sequence tags (ESTs) and will allow further sequence and functional analysis. Better understanding of the enzymes involved in sugarcane purine synthesis will help elucidate the genetic components of this important pathway in sugarcane and may open the possibility of using such enzymes as a target for chemicals which may be effective in combating phytopathogens (Kubinyi, 1998).

¹Laboratory of Protein Crystallography and Structural Biology, Physics Institute of São Carlos, University of São Paulo - USP, Av. Trabalhador São-carlense 400, 13566-590 São Carlos, SP, Brazil.

²Department of Genetics and Evolution, Federal University of São Carlos, São Carlos, Brazil.

Corresponding author: Otavio H. Thiemann. E-mail: thiemann@if.sc.usp.br.

MATERIAL AND METHODS

The search strategy employed is summarized in Figure 2. Representative protein sequences for each member of the purine *de novo* synthesis pathway (Figure 1) were chosen from the NCBI database (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov). The first approach was to select protein sequences of organisms as closely related to sugarcane in the phylogenetic tree as possible, to avoid possible sequence divergence effects in our search methodology. In three cases (Glutamine-PRPP-amidotransferase, Adenylosuccinate lyase and 5-aminoimidazole-4-carboxamide ribonucleotide transformylase/inosine monophosphate cyclohydrolyase bifunctional enzyme) sequences from chicken, protozoa and rat had to be used as representatives due to the lack of reliable full-length plant homologues. In those cases, sequence divergence was estimated to be low by comparison of the available homologues with the Monte Carlo shuffling strategy implemented by the PRSS3 program (<http://fasta.bioch.virginia.edu/fasta/prss.htm>).

The selected peptide sequences were used to search the entire translated SUCEST database with the available BLAST facility (tBLASTn) (Altschul *et al.* 1990; Madden *et al.* 1996; Zhang and Madden, 1997). Sugarcane clusters retrieved with the BLAST program were further tested for the statistical significance of the alignment by the PRSS3 program. To calibrate the Monte Carlo analysis, known

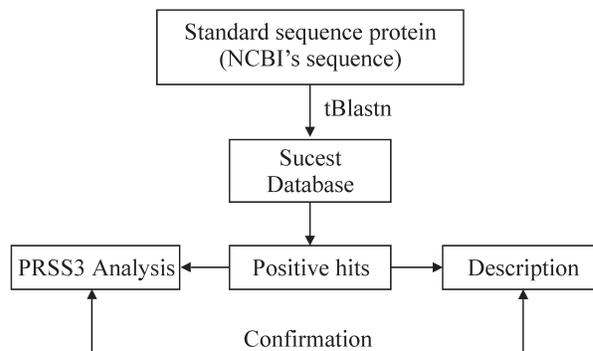


Figure 2 - Schematic representation of the search strategy. Sequences representative of each of the eleven proteins participating in the purine synthesis pathway were selected from NCBI (National Center for Biotechnology Information) database. Those peptide sequences were used to screen the sugarcane expressed sequence tag (SUCEST) database using the tBLASTn program. The positive hits were further confirmed by the PRSS3 implementation of the Monte Carlo shuffling algorithm, resulting in a table of statistically significance values.

protein sequences of divergent taxa along the phylogenetic tree were used. Those sequences were compared to each other and to the clusters using the PRSS3 program. The resulting table of p-values aided to estimate the degree of divergence of each enzyme between different taxa and in relation to the sugarcane clusters. The alignments and the results from the PRSS3 program analysis were taken into

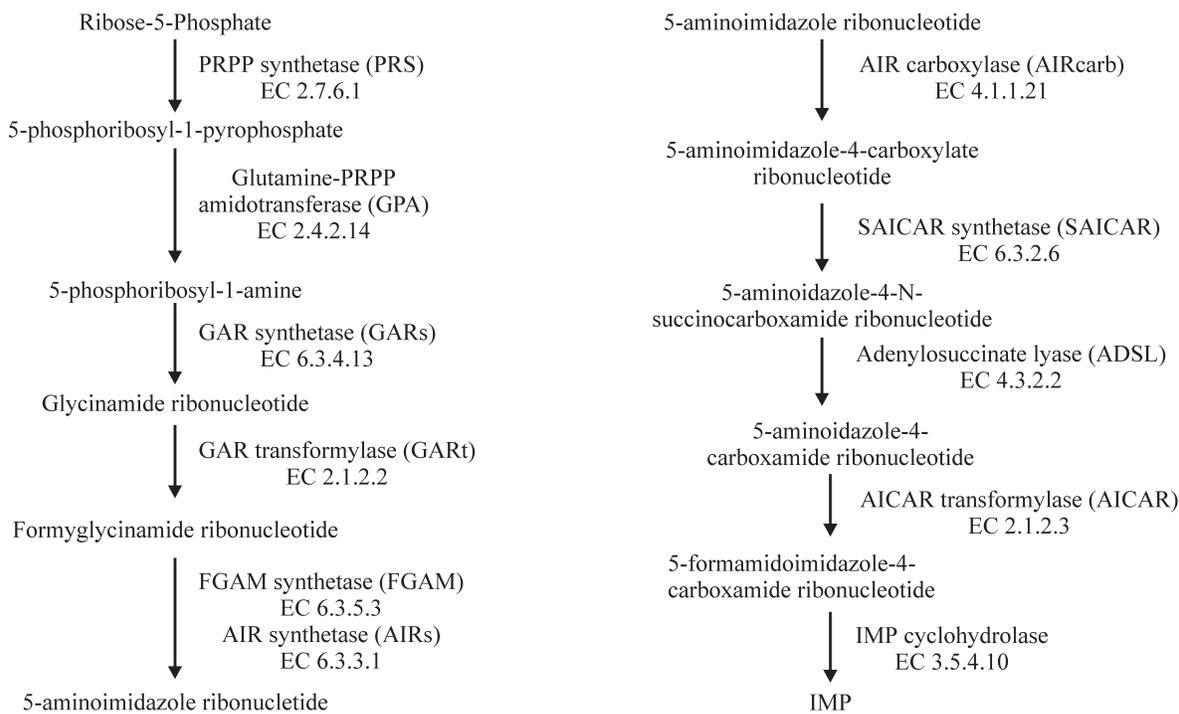


Figure 1 - The purine synthesis pathway showing the eleven enzymes participating in the metabolic cascade and the pathway intermediates. The EC numbers are also indicated below the enzyme name. Formylglycineamide ribotide (FGAM) synthetase and aminoimidazole ribotide (AIR) synthetase participate in the same enzymatic step and are so indicated. Sugarcane aminoimidazole carboxamide ribotide (AICAR) transformylase and inosine monophosphate (IMP) cyclohydrolyase have been identified as a single cluster representing a bifunctional enzyme.

consideration to confirm a selected cluster as a homologue to a known purine synthesis pathway member as described in Figure 2. The PRSS3 comparison tables are not presented here due to lack of space but they are available at the sugarcane data-mining site (<http://sucest.lad.ic.unicamp.br/cgi-bin/prod/mining-reports/mining-reports.pl>).

The representative enzyme sequences for each member in the *de novo* pathway and the corresponding accession numbers are: PRPP synthetase (*Arabidopsis thaliana*, CAA63552) (Krath *et al.*, 1999), Glutamine-PRPP-amidotransferase (*Rattus norvegicus*, A46088) (Iwahana *et al.*, 1993), glycinamide ribonucleotide (GAR) synthetase (*Arabidopsis thaliana*, P52420) (Schnorr, *et al.*, 1994), GAR transformylase (*Arabidopsis thaliana*, P52422) (Schnorr, *et al.*, 1994), Formylglycinamide ribonucleotide (FGAM) synthetase (*Glycine max*, AAF21596, unpublished direct submission), 5-aminoimidazole ribonucleotide (AIR) synthetase (*Arabidopsis thaliana*, Q05728) (Senecoff and Meagher, 1993), AIR carboxylase (*Arabidopsis thaliana*, T02535, unpublished direct submission), 5-aminoimidazole-4-N-succinocarboxamide ribonucleotide (SAICAR) synthetase (*Arabidopsis thaliana*, P38025, unpublished direct submission), adenylosuccinate lyase (*Leishmania major*, CAC22697) (Ivens *et al.* 1998), with AICAR transformylase and IMP cyclohydrolase bifunctional enzymes (*Gallus gallus*, P31335) (Ni *et al.*, 1991).

The clusters selected using the above strategy were further aligned to representative sequences of each enzyme of the pathway. The presence of a full-length EST sequence within a cluster is identified by a significant alignment in the translated N-terminal sequence of that EST, taking into consideration the natural heterogeneity of those regions.

This strategy allowed us to identify at least one potential case of polymorphism in sugarcane, the PRPP synthetase protein, an essential enzyme of the *de novo* purine synthesis pathway. This case is being further analyzed by the cloning and functional characterization of the sugarcane genes.

RESULTS

The search strategy adopted (Figure 2) allowed the estimation of a statistical support for the alignment of representative sequences to the sugarcane clusters using the Monte Carlo shuffling strategy. The initial analysis of several clusters identified as representatives for each member of the purine synthesis pathway indicated the degree of divergence within those enzymes. This result was compared to the statistical score of the Sugarcane cluster to each of the different representative enzyme sequences retrieved from the NCBI (National Center for Biotechnology Information) database. With such a strategy we were able to identify clusters representing each enzyme of the sugarcane purine synthesis pathway with a high degree of confidence. As shown in Table I, the analysis of the results obtained in each of the different libraries indicates that the purine synthesis pathway is present in every tissue of the plant. Fifty-three EST clones forming different clusters have been identified as unambiguously belonging to the purine synthesis pathway. In four libraries (normalized tissue, seeds, stem bark and leaves) we did not identify any EST sequence or cluster belonging to the purine synthesis pathway. However, these libraries represent the ones with the least number of ESTs clones sequenced in the project. Interestingly, the homologues of glutamine-PRPP-amidotransferase (GPA), AIR carboxylase (AIRcarb) and SAICAR synthetase were iden-

Table I - Frequency of purine synthesis homologue EST clones per library.

Libraries	PRS	GPA	GARs	GARt	FGAM	AIRs	AIRcarb	SAICAR	ADSL	AICAR & IMPcyc	Total
Callus	1	0	0	2	0	2	0	2	3	0	10
Sugarcane flower	0	0	2	2	1	0	0	0	1	1	7
Leaf Roll	1	2	0	1	0	0	3	0	0	0	7
Root	5	0	2	0	0	0	0	0	0	0	7
Apical Meristem	3	0	1	1	1	0	0	0	0	0	6
Leaf-root transition zone	5	0	1	1	1	0	0	0	1	1	5
Stem	0	0	0	0	0	1	0	0	0	0	1
Lateral Bud	0	0	0	1	0	0	0	0	0	0	1
Stem Bark (large insert)	0	0	0	0	0	0	0	0	0	0	0
Normalized Library	0	0	0	0	0	0	0	0	0	0	0
Seeds	0	0	0	0	0	0	0	0	0	0	0
Leaves	0	0	0	0	0	0	0	0	0	0	0
<i>Acetobacter diazotrophicans</i>	0	0	1	0	0	0	0	0	0	1	2
<i>Herbaspirillum rubrisubalbicans</i>	1	0	1	0	0	0	0	0	0	0	2
Total	16	2	8	8	3	3	3	2	5	3	53

tified in only one library (Table I). The EST clone identified as having the highest frequency was the homologue to PRPP synthetase (PRS), this being consistent with this enzyme being central to several metabolic pathways.

The presence of EST clones within clusters representing full-length sequences is of great importance for the further investigation of the genes involved in this pathway. Such EST clones will allow the fast analysis of the entire sequence and its sub-cloning into appropriate vectors for functional investigation of the expressed protein. The alignment of the individual EST clone sequence and the reference sequences revealed the frequency of full length clones (Table II). From the 53 EST clones identified and distributed in the different clusters (Table I), 13 are full length ESTs and 40 represent partial length EST clones. The sequence for the bifunctional enzyme AICAR transformylase/IMP cyclohydrolase was identified in one instance as a full-length EST clone.

DISCUSSION

The investigation of central metabolic pathways is of great interest for the advancement of the knowledge of any organism. The purine *de novo* synthesis pathway (Figure 1) represents a central pathway affecting several metabolic routes of great importance and has been conserved phylogenetically in several taxa. However, some organisms (*e.g.* protozoan parasites) are purine auxotrophs and are dependent on the salvage pathway for providing the purine nucleotides necessary for intracellular metabolism. Differences in metabolic pathways or in the protein sequences that participate in such pathways can be exploited as potential targets for the development of inhibitory compounds, an approach which is being aggressively undertaken in various rational drug design programs.

We have applied a stringent search strategy (Figure 2) for the identification of homologue genes from the sugarcane clusters which participate in the purine synthesis path-

way. The use of Monte Carlo shuffling analysis in our screening strategy was important to avoid the subjectivity normally involved in the evaluation of the alignment results of divergent sequences. Our results (Table I) show that all the enzymes necessary for the purine *de novo* synthesis pathway are present in sugarcane.

The identification of some clusters with a higher frequency of EST clones than others clusters (depending on the tissue or library analyzed) may be a result of the different number of sequences deposited in the SUCEST database for each library. At this point we can not conclude that these results reflect differential expression or mRNA stability for each gene. However, it is interesting to notice the high frequency of PRPP synthetase (PRS) ESTs identified because the product of PRS (PRPP) is a substrate for diverse metabolic pathways and its abundance in the sugarcane libraries may reflect the importance of this enzymes. In several organisms, including mammals, isoforms of the PRS gene have been identified. In mammals at least three forms are known, PRS-I, PRS-II and PRS-III which act as a multienzyme complex. It is likely that in sugarcane the PRS gene is also present as a set of isoforms. We will further investigate such hypothesis.

The last enzyme in the metabolic cascade leading from Ribose-5-phosphate to IMP has also been identified in sugarcane as the bifunctional enzyme AICAR transformylase /IMP cyclohydrolase, but there is no evidence of any other multifunctional gene in this pathway in sugarcane. Four EST clones have been identified from the libraries of infected plants (*Gluconacetobacter diazotrophicans* and *Herbaspirillum rubrisubalbicans*) as shown in Table I. These sequences have been identified as belonging to sugarcane genes and not the bacterial homologues, however only the full-length sequence will identify their origin with certainty.

The alignment of the identified sequence clusters with representatives of the purine metabolic pathway has allowed us to verify the presence of 13 full-length EST clones including one full-length bifunctional AICAR/IMPcyc homologue. These findings do not directly reflect the expression of the sugarcane genes because they are not the result of direct mRNA quantification.

The present study identified all the genes of the purine *de novo* synthesis pathway in sugarcane. Several full-length EST clones are presently being investigated for their enzymatic characteristics. Such research may lead to the rational design and future development of anti-pathogenic agents and also provide insights into the evolutionary origin of this conserved metabolic pathway.

RESUMO

A via de síntese de purino nucleotídeos é considerada uma via de central importância para todas as células. Na maioria dos organismos, os purino nucleotídeos são sintetizados "de novo" a partir de precursores não-nucleotídicos

Table II - Frequency of full-length EST clones.

Enzyme	Full-length	Partial-length
PRPP synthetase	7	9
Glutamine-PRPP-amidotransformylase	0	2
GAR synthetase	2	6
GAR transformylase	1	7
FGAM synthetase	0	3
AIR synthetase	1	2
AIR carboxylase	0	3
SAICAR synthetase	1	1
Adenylosuccinate lyase	0	5
AICAR transformylase – IMP cyclohydrolase	1	2
Total	13	40

como amino ácidos, amônia e dióxido de carbono. O conhecimento das enzimas envolvidas na via de síntese de purinas da cana-de-açúcar vai abrir a possibilidade do uso dessas enzimas como alvos no desenho racional de inibidores no combate a agentes fitopatogênicos, como esta sendo feita com diversos parasitos e células cancerosas. A seguinte estratégia esta sendo utilizada na identificação de genes de cana-de-açúcar para cada membro da via de síntese de purinas: Seqüências representativas dos genes que compõem a via foram escolhidas do banco de dados NCBI. Essas seqüências de peptídeos estão sendo utilizadas em buscas ao banco de dados gerado pelo SUCEST pelo programa BLAST (implementação tBLASTn). Alinhamentos com os clusters de cana-de-açúcar são posteriormente analisados para sua significância estatística pela implementação PRSS3 do algoritmo conhecido como Monte Carlo shuffling. Para calibrar a análise dos resultados de PRSS3, foram empregadas seqüências conhecidas de diferentes taxas ao longo da árvore filogenética. Essas seqüências são comparadas duas a duas e com o cluster da cana-de-açúcar. A tabela de valores-p resultante indica o grau estatístico de similaridade e divergência entre as seqüências já descritas e entre essas e os clusters de cana-de-açúcar. Os resultados obtidos dessas análises estão descritos neste artigo.

ACKNOWLEDGMENTS

This work was supported by a research grant 99/02874-9 to O.H. Thiemann from the Fundação de Amparo a Pesquisa no Estado de São Paulo (FAPESP) as part of the ONSA network. We would like to thank the members of the Protein Crystallography and Structural Biology Group (IFSC-USP) for helpful discussions in the course of this work.

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and Smith, D.F. (1998). A physical map of the *Leishmania major* Friedlin genome. *Genome Res.* 8 (2): 135-145.
- Iwahana, H., Yamaoka, T., Mizutani, M., Mizusawa, N., Ii, S., Yoshimoto, K. and Itakura, M. (1993). Molecular cloning of rat amidophosphoribosyltransferase. *J. Biol. Chem.* 268 (10): 7225-7237.
- Karlin S, Altschul SF. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. U.S.A.* 87 (6): 2264-8.
- Krath, B.N., Eriksen, T.A., Poulsen, T.S. and Hove-Jensen, B. (1999). Cloning and sequencing of cDNAs specifying a novel class of phosphoribosyl diphosphate synthase in *Arabidopsis thaliana*. *Biochim. Biophys. Acta* 1430 (2): 403-408.
- Kubinyi, H. (1998). Structure-based drug design *Curr. Opin. Drug. Disc. Dev.* 1: 4-15.
- Madden, T.L., Tatusov, R.L. and Zhang, J. (1996). Applications of network BLAST server *Meth. Enzymol.* 266: 131-141.
- Marr, J. J. and Ullman, B. (1995). *Biochemistry and Molecular Biology of Parasites* (Marr, J.J. and Muller, M., eds) 323-336.
- Marr, J. J. (1991). Purine analogs as chemotherapeutic agents in Leishmaniasis and American Trypanosomiasis *J. Lab. Clin. Med.* 118: 111-119.
- Ni, L., Guan, K., Zalkin, H. and Dixon, J.E. (1991). De novo purine nucleotide biosynthesis: cloning, sequencing and expression of a chicken PurH cDNA encoding 5-aminoimidazole-4-carboxamide-ribonucleotide transformylase-IMP cyclohydrolase. *Gene* 106 (2): 197-205.
- Pearson, W.R and Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* 85 (8): 2444-8.
- Pearson, W.R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* 183: 63-98.
- Schnorr, K.M., Nygaard, P. and Laloue, M. (1994). Molecular characterization of *Arabidopsis thaliana* cDNAs encoding three purine biosynthetic enzymes. *Plant J.* 6 (1): 113-121.
- Senecoff, J.F. and Meagher, R.B. (1993). Isolating the *Arabidopsis thaliana* genes for de novo purine synthesis by suppression of *Escherichia coli* mutants. I. 5'-Phosphoribosyl-5-aminoimidazole synthetase. *Plant Physiol.* 102 (2): 387-399.
- Stryer, L. (1995). *Biochemistry*. 4th edition: pp. 739-743.
- Ullman, B. and Carter, D. (1997). Molecular and biochemical studies on the hypoxanthine-guanine phosphoribosyltransferases of the pathogenic haemoflagellates. *Int. J. Parasitol.* 27: 203-213.
- Zhang, J. and Madden, T.L. (1997). PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res.* 7: 649-656.