



**Transportadores Iônicos:  
Importância da Alça Citoplasmática Maior na  
Associação entre Subunidades Catalíticas de P-  
ATPases e Identificação de Transportadores em  
*Gluconacetobacter diazotrophicus* PAL5**

*Wellington Inácio de Almeida*

Tese submetida ao Instituto de Bioquímica Médica da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Biológicas (Química Biológica).

*Dezembro 2007*

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**Transportadores Iônicos: Importância da Alça Citoplasmática Maior na Associação entre Subunidades Catalíticas de P-ATPases e Identificação de Transportadores em *Gluconacetobacter diazotrophicus* PAL5**

*Wellington Inácio de Almeida*

Tese submetida ao Instituto de Bioquímica Médica da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Biológicas (Química Biológica).

Banca examinadora:

.....  
Prof. Julio Alberto Mignaco  
Professor Adjunto do Instituto de Bioquímica Médica/UFRJ

.....  
Prof<sup>a</sup>. Selma Gomes Ferreira Leite  
Professor Titular da Escola de Química/UFRJ

.....  
Prof. Antônio Ferreira Pereira  
Professor Associado do Instituto de Microbiologia Prof. Paulo de Góes/UFRJ

.....  
Prof<sup>a</sup>. Gabriela de Oliveira Paiva e Silva  
Professor Adjunto do Instituto de Bioquímica Médica/UFRJ  
(Revisora – Suplente Interno)

.....  
Prof. Orlando Bonifácio Martins  
Professor Adjunto do Instituto de Bioquímica Médica/UFRJ  
(Orientador)

.....  
Prof. Paulo Cesar de Carvalho-Alves  
Professor Adjunto do Instituto de Bioquímica Médica/UFRJ  
(Co-orientador)

.....  
Prof. Vânia Lúcia Muniz de Pádua  
Professor do Centro Universitário Estadual da Zona Oeste/UEZO  
(Suplente Externo)

**Transportadores Iônicos: Importância da Alça Citoplasmática Maior na  
Associação entre Subunidades Catalíticas de P-ATPases  
e Identificação de Transportadores em  
*Gluconacetobacter diazotrophicus* PAL5**

Almeida, Welington Inácio de.

.- Welington Inácio de Almeida. Rio de Janeiro: UFRJ/ IBqM, 2007.

x, f.: il.; 31 cm.

Transportadores Iônicos: Importância da Alça Citoplasmática na Associação entre Subunidades Catalíticas de P-ATPases, e Identificação de Transportadores em *Gluconacetobacter diazotrophicus* PAL5

Tese (doutorado) – UFRJ/ Instituto de Bioquímica Médica/

Programa de Pós-Graduação em Química Biológica, 2007.

Referências Bibliográficas: f. .

1. Bioinformática. 2. Expressão Heteróloga. 3. Estrutura de Proteínas. 4. Dimerização. 5. Modelagem Molecular. I. Carvalho-Alves, Paulo Cesar; Martins, Orlando Bonifácio. II. Universidade Federal do Rio de Janeiro, Instituto de Bioquímica Médica, Programa de Pós-Graduação em Química Biológica. III. Título.

Esta tese foi realizada no Laboratório de Biologia Molecular, do Instituto de Bioquímica Médica da Universidade Federal do Rio de Janeiro, sob a orientação do Professor Orlando Bonifácio Martins e co-orientação do Professor Paulo Cesar de Carvalho Alves, com auxílios financeiros da Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), e do Laboratório Nacional de Luz Síncrotron (LNLS).

## QUE É? O QUE É?

Gonzaguinha

Viver e não ter a vergonha de ser feliz  
Cantar, e cantar, e cantar  
A beleza de ser um eterno aprendiz  
Ai meu Deus, eu sei (eu sei)  
Que a vida devia ser bem melhor  
E será, mas isso não impede que eu repita  
É bonita, é bonita, é bonita  
E a vida, e a vida o que é diga lá meu irmão  
Ela é a batida de um coração  
Ela é uma doce ilusão, ê ô  
E a vida, ela é maravilha ou é sofrimento  
Ela é alegria ou lamento  
O que é, o que é meu irmão  
A quem fale que a vida da gente  
É um nada no mundo  
É uma gota, é um tempo que nem dá um segundo  
Há quem fale que é um divino mistério profundo  
É o sopro do criador numa atitude repleta de amor  
Você diz que é luta e prazer  
Ele diz que a vida é viver  
Ela diz que melhor é morrer  
Pois amada não é, e o verbo é sofrer  
Eu só sei que confio na moça  
E na moça eu ponho a força da fé  
Somos nós que fazemos a vida  
Como der ou puder ou quiser  
Sempre desejada  
Por mais que esteja errada  
Ninguém quer a morte  
Só saúde e sorte  
A pergunta roda  
E a cabeça agita  
Eu fico com a pureza da resposta das crianças  
É a vida, é bonita, é bonita  
Viver...

## AGRADECIMENTOS

Primeiramente gostaria de agradecer à minha família, principalmente aos meus pais, por servirem de exemplo, demonstrando compreensão, paciência, e por sempre me incentivar.

Ao Professor Orlando Bonifácio Martins por criar um ambiente de trabalho e estudo diversificado, congregando diferentes projetos e reflexão nos vários momentos na “luta contra as moléculas”, e ao Professor Paulo Cesar de Carvalho Alves, por ter demonstrado confiança em mim, pelo incentivo. Obrigado, pela oportunidade de aprendizagem e pela amizade...

À Professora Verônica Salerno Pinto, por sua amizade e por me indicar como estudante de IC ao Professor Paulinho depois do seu regresso da USP em 1999, e à Professora Martha M. Sorenson por sua paciência na revisão dos artigos.

Às estudantes de Iniciação Científica que me ajudaram durante o período de doutorado: Rebeca G. Costa e Caroline B. S. Oliveira, pelo convívio e por compartilharem frustrações e alegrias.

Aos amigos: Alexander MC, Sylvia Alqueres, Rodrigo “Gaúcho” e Ariane Larentis, pelas discussões e colaborações nos trabalhos que desenvolvemos e na realização desta tese.

Aos demais colegas de LBM e colaboradores: Dr. Ricardo Pilz Vieira, Dra. Maysa Mandetta, Professora Carla Polycarpo, Cynthia Silveira, Aline Turque, Marcio Loureiro, Marcelo Betarlan, e do Laboratório de Contração Muscular: Marcelo Sant’Anna e Joyce, pelas discussões e sugestões.

Aos professores do Laboratório de Artrópodes Hematófagos: Pedro Oliveira, Marcos H. Sorgine, Marcus e Aurélio, e aos demais colegas desse



Laboratório: Ana Caroline, Christine, Kathy, Renata, José Henrique, Maria Clara, Fabiana, Milane e Carlo Donato pelo convívio estimulante de discussão nos *journals* e pelo aprendizado.

Aos professores do Laboratório de Biofísica Química de Proteínas, hoje Laboratório de Doenças Neurodegenerativas: Sérgio Ferreira, Fernanda de Felice e Margareth Magdesian, pelas sugestões e pelo aprendizado.

Ao pessoal do Laboratório de Doenças Neurodegenerativas, principalmente, à Samantha pela ajuda, incentivo e pelos bate-papos da antiga salinha ao novo laboratório; também ao Adriano Sebolella e Matthias Gralle pelo auxílio no HPLC e na análise dos espectros de CD.

Ao Leonardo Mega e à Dra. Vânia Lúcia (Núcleo de Estudos de Genomas Johanna Döbenheiner/UFRJ), à Flavia Alvin e ao Professor Rodolpho Albano (Laboratório de Genômica, UERJ) pelos seqüenciamentos; aos demais integrantes do RioGene Virtual Institute, e demais colaboradores: Alessandra Gonzalez e Dr. Rodolfo Paranhos (IB/UFRJ), Prof. João Alfredo Medeiros (IQ/UFRJ), Karen Silva e ao Professor Ulysses Lins (IMPPG/UFRJ), ao Prof. Paulo M. Bisch e Shaila Rossle (IBCCF). Ao Ranlig Medeiros, Arlan e ao Professor Pedro Pascutti, pela atenção e discussão em torno das simulações de dinâmica molecular do complexo dimérico de P-ATPases.

O apoio do Laboratório Nacional de Luz Síncrotron: Drs. Íris Torriani e Cristiano Oliveira, da linha de SAXS.

Aos novos colegas da Divisão de Biotecnologia (DIBIOTEC/DIQUIM II) do INPI, especialmente ao grupo de “novíssimos”: Juliana Manasfi, Giany Mello de Oliveira, Alessandra Costa, Paula Salles e Debora Gomes; e, também aos “supervisores”: Rodrigo Ferraro, a minha ex-professa de Biologia Molecular Claudia Magioli, Vânia Lúcia Linhares, e a Margareth Maia, pelo convívio tranquilo no âmbito de serviço de Exame de Patentes.

À Professora Gabriela Paiva e Silva, por sua atenção e pela revisão criteriosa dessa tese. Aos membros da banca, por aceitarem participar desse momento.

A ajuda de muitos pesquisadores e estudantes de todos os laboratórios do nosso Departamento, agora Instituto de Bioquímica Médica que foram fundamentais para esta tese, como também foi a ajuda de outros laboratórios, principalmente aos Professores Adalberto Vyeira (IBCCF) e Luiz Maurício Trambaiolli.

## SUMÁRIO

Abreviaturas	9
Índice de Figuras	10
Índice de Tabelas	11
Resumo	12
Abstract	14
I. Introdução	16
1. Mudança de Paradigmas em Ciências Biológicas	17
1.1 Células, Genes e Proteínas	17
1.2 (Bio)Diversidade, Conhecimento e Tecnologia	19
2. Funcionalidade Biológica - as proteínas	22
3. As Membranas Biológicas e o fluxo iônico	23
3.1 Transporte Ativo, ATPases tipo F e ATPases tipo V	28
3.2 ATPases tipo P	30
4. <i>Gluconacetobacter diazotrophicus</i> e o seqüenciamento do seu genoma	45
4.1 Anotação do genoma de <i>G. diazotrophicus</i> : Sistema de Transporte	48
II. Objetivos	54
III. Materiais e Métodos	56
IV. Resultados, Discussão e Conclusão	72
P-ATPases	74
Manuscrito 1	74
Proposição e Análise de Interface de Dimerização	85
Arranjo Tridimensional do Complexo dimérico do LCD/SERCA	96
Genoma	102
V. Referências Bibliográficas	118
Apêndices: Outros Projetos e Publicações	130
Parte 1: Prospecção Microbiana	132
Manuscrito 1	138
Parte 2: Modelagem Molecular	167
Manuscrito 2	167
Manuscrito 3	178
Manuscrito 4	188
Parte 3: Archaea e Biotecnologia	198
Manuscrito 5	200
Manuscrito 6	208

**ABREVIATURAS**

3D – tridimensional  
ADP – Adenosina 5'-difosfato  
ATP – Adenosina 5'-trifostato  
BLAST – Basic Local Alignment Sequence Tool  
BSA – albumina de soro bovina  
CD – dicroísmo circular  
CDB – Convenção da Diversidade Biológica  
cDNA – DNA complementar  
Da – Dáltons  
DNA – Ácido Desoxiribonucléico  
DTT – ditionitrito  
EDTA – ácido etileno diamino tetracético  
EP – fosfoenzima  
FITC – isotiocianato de fluoresceína  
FBN – Fixação biológica de nitrogênio  
HPLC – cromatografia de alta eficiência  
IPTG – isopropil- $\beta$ -D-galactosídeo  
LCD – alça citoplasmática maior  
MD – dinâmica molecular  
MW – peso molecular  
NCBI – National Center of Biotechnology Information  
Ni-NTA – níquel-ácido nitriloacético  
O.D. – densidade ótica  
ORF – seqüências abertas de leitura  
pb – pares de bases  
PCR – reação em cadeia da polimerase  
Pi – fosfato inorgânico  
PIC – coquetel de inibidores de proteases  
PMA1 ou PMA – H<sup>+</sup>-ATPase de membrana plasmática de *Saccharomyces cerevisiae*  
PMSF – fluoreto de fenil-metil sulfonila  
SAXS – espalhamento de raios-X a baixo ângulo  
SERCA – Ca<sup>2+</sup>-ATPase de retículo sarco(endo)plasmático

## INDICE DE TABELAS

<b>TABELA I – Análise de interação entre resíduos</b>	91
<b>TABELA II – Parâmetros Estruturais.</b>	100
<b>Tabela III – Descrição das ORFs identificadas com pertencentes ou relacionadas com o sistema de transportes no genoma da <i>G. diazotrophicus</i></b>	103
<b>Tabela IV – Comparação entre o sistema de transportes no genoma da <i>G. diazotrophicus</i> com outros genomas publicados</b>	113
<b>Tabela V – Principais classes de transportadores anotadas no genoma de <i>G. diazotrophicus</i></b>	114

## INDICE DE FIGURAS

FIGURA 1. Gradientes iônicos.	25
FIGURA 2. Representação esquemática das ATPases tipo F e V	29
FIGURA 3. Representação esquemática das ATPases tipo P	31
FIGURA 4 – Esquema simplificado do ciclo catalítico da SERCA	35
FIGURA 5. Representação das ATPases tipo P	38
FIGURA 6. Estrutura da Ca <sup>2+</sup> -ATPase de retículo sarco(endoplasmático) (SERCA) com alta resolução (2,6 Å)	41
FIGURA 7. Diferenças estruturais na Ca <sup>2+</sup> -ATPase de retículo sarco(endoplasmático) (SERCA)	42
Figura 8. Microscopia de varredura de <i>Gluconacetobacter diazotrophicus</i>	47
Figura 9. Modelo hipotético para o dímero da SERCA	86
Figura 10. Sistema para simulação de Dinâmica Molecular	86
Figura 11. Análise do sistema dimérico	87
Figura 12. Novo modelo para o dímero da SERCA	88
Figura 13. Gráfico do RMSD	89
Figura 14. Gráfico de Pontes de Hidrogênio	90
Figura 15. Eventos de Dimerização	96
Figura 16. Modelos Teóricos de LCD Dimerização	97
Figura 17. Função de distribuição (Pr) para o LCD/SERCA	98
Figura 18. Comparação dos dados de SAXS	99

## RESUMO

Apesar dos recentes avanços no conhecimento da estrutura tridimensional das P-ATPases, a exata unidade funcional residente nas membranas biológicas ainda é controversa.

Este trabalho apresenta novas evidências estruturais para a relevância da alça citoplasmática maior (LCD) na auto-associação de P-ATPases e, em particular, o papel fundamental do domínio P para a oligomerização. Arranjos oligoméricos de P-ATPases foram propostos por um longo tempo, a partir das observações da dependência bifásica à concentração de ATP na atividade enzimática. No entanto, o resultado mais importante aqui é a constatação de que LCDh reenovelado mostrou dois picos fluorescentes principais, quando analisados por HPLC, calculados com pesos moleculares de 95 e 39 kDa, compatíveis com as formas dimérica e monomérica, respectivamente. Estes dados sugerem que, como anteriormente observado para o LCD/SERCA isolado, LCDh reenovelado existe em um equilíbrio monômero-dímero. Por outro lado, a cromatografia de exclusão por tamanho da LCD $\Delta$ h mostrou um único pico, correspondente a um valor estimado de peso molecular de 29 kDa, e nenhum outro pico em pesos moleculares mais elevados, o que indica que esse fragmento é essencialmente monomérico, e que a presença do domínio P é essencial para a dimerização. Além disso, os parâmetros estruturais (raio de giro e diâmetro máximo) do LCD/SERCA obtidos por SAXS corroboram o modelo em que a porção carboxi-terminal do domínio P pode estar envolvida na auto-associação e funcional relevância para P-ATPases em membranas nativas.

De acordo com estes dados, propomos um modelo dimérico geral para a maioria dos P - ATPases (pelo menos para as subclasses II e III), tendo em conta que a hipótese de movimentos dos domínios N e A durante o mecanismo de translocação iônica não estão restringidos pela associação dos domínios P. Finalmente, o nosso modelo proposto não se opõe à associação de outras proteínas reguladoras, como o phospholamban, a qual propostamente

associaria com dois monômeros de  $\text{Ca}^{2+}$ -ATPase, ligando-o às Lys 393 ou Lys 400 do domínio N.

Paralelamente, contribuimos no projeto de seqüenciamento de *G. diazotrophicus* com a identificação de ORFs relacionados ao sistemas de transporte. Nós identificamos 130 genes pertencentes ao transportadores ABC e 48 à superfamília da MFS, além de uma grande quantidade de genes de exportação de metais pesados e outros compostos tóxicos. Além disso, podemos encontrar 40 genes da família RND e 41 genes de secreção tipo IV que estão relacionados com a exportação de lipooligosacarídeos e com sistemas de conjugação (transporte de proteínas, DNA e outras macromoléculas), o que talvez tenha sido permitido pela aquisição de uma grande quantidade de genes por transferência laterais, respectivamente; como também a presença de 38 genes responsáveis pela produção de glicosiltransferases, que são genes requeridos para a biossíntese de polissacarídeos de superfície celular (principalmente succinoglicanos). Acredita-se que estes apresentem importante papel de reconhecimento celular, no decurso de mecanismos de interação planta-bactéria, os quais correspondem a requisitos necessários ao reconhecimento planta-bactéria, e a conseqüente ocorrência do processo de fixação biológica de nitrogênio.

Ressalta-se ainda a ocorrência de 23 genes relacionados com o transporte de ferro (TonB) que foram encontrados no genoma de *G. diazotrophicus*. A presença destas proteínas, associada a genes responsáveis pela síntese de bacteriocinas e antibióticos, pode conferir à *G. diazotrophicus* uma vantagem competitiva em relação a outros endosimbiontes vegetais.



## ABSTRACT

In spite of the recent advances in the knowledge of the three-dimensional structure of P-ATPases, its exact functional unit resident in the biological membranes is still controversial.

This work provides new structural evidence for the relevance of Large Cytoplasmatic Domain (LCD) in self-association of P-ATPases, and in particular, the fundamental role of the P site for oligomerization. Oligomeric arrangements of P-ATPases have been proposed for a long time, starting by early observations of biphasic dependence on ATP concentration for enzyme activity. However, the most important result here is the finding that refolded LCDh showed two main fluorescent elution peaks, when analyzed by size-exclusion HPLC, with calculated molecular weights of 95 and 39 kDa, compatible with dimeric and monomeric forms, respectively. These data suggest that, as previously observed for isolated LCD/SERCA, refolded LCDh exists in a monomer–dimer equilibrium. On the other hand, size-exclusion chromatography of LCD $\Delta$ h showed a major peak corresponding to an estimated molecular weight of 29 kDa, and no other peak at higher molecular weights, indicating that this fragment is essentially monomeric and that the presence of the P site is essential for dimerization. In addition, the structural parameters (radius of gyration and maximum diameter) of LCD/SERCA obtained by SAXS corroborate to the model that the carboxyl-terminus of P domain may be involved in the self-association and the functional relevance of P-ATPases in the native membrane.

In agreement with these data, we propose this general dimeric model for most of P-ATPases (at least for the II and III subclasses), taking into account that the proposed movements of N and A domains during the ion translocation mechanism are not constrained by such P-P site association. Finally, our proposed model does not preclude association of other regulatory proteins such as phospholamban, which has also been proposed to associate with two Ca<sup>2+</sup>-ATPase monomers, and to bind to Lys 393 or Lys 400 in the N domain.

In parallel, we contribute in the sequencing project of *G. diazotrophicus* with the identification of ORFs related to the transport systems. Genes encoding transport systems constitute the largest (14.1%) class of genes in the *G. diazotrophicus* genome. They are related to transport of amino acids, sugars, ammonia, nitrate, sulphate and other ions. We identified 130 genes belonging to the ATP-Binding Cassette (ABC) and 48 Major Facilitator superfamilies (MFS). A great amount of genes for heavy metals and other toxic compounds export (multidrug-resistance homologs) are present in *G. diazotrophicus* transport systems. RND family proteins (40 genes) are also related to export lipooligosaccharides concerned with plant nodulation for the purpose of symbiotic nitrogen fixation. Moreover, we can find 41 genes for type IV secretion systems. They are generally involved in conjugation systems (transport of proteins, DNA and other macromolecules) and perhaps have allowed the acquisition of a great amount of genes by lateral transfer. The presence of these proteins, associated with genes responsible for the synthesis of bacteriocins and antibiotics can give to *G. diazotrophicus* a competitive advantage in relation to others plant endosymbionts. Recently, it has been shown that association with *G. diazotrophicus* makes sugarcane more resistant to pathogen infections.

## **INTRODUÇÃO**

# I. INTRODUÇÃO

---

*“Nós somente podemos aspirar  
por aquilo que já encontramos”.*

*Erwin Chargaff*

## 1. Mudança de Paradigmas em Ciências Biológicas

### 1.1. Células, Genes e Proteínas

De modo geral, a Ciência busca reconhecer na natureza, e explicitar, em linguagem adequada, padrões de “comportamento” que independam do observador e que possam ser generalizados sem limitações de tempo e de espaço. Segundo essa perspectiva da compreensão humana sobre os fenômenos naturais, as Ciências Biológicas vêm contribuindo significativamente no estabelecimento de padrões, sobretudo a partir da segunda metade do século XIX (Rose, 2002; Kafatos, 2002, Woese, 2004).

De forma ilustrativa, destacamos a seguir três marcos científicos: Primeiramente, o desmoronar da idéia da geração espontânea, com os experimentos do francês Louis Pasteur (1822-1895). O segundo exemplo, que na realidade compreende um período longo de tempo (não especificamente um pesquisador ou experimento), inicia-se em 1865 com os experimentos do monge Gregor Mendel (1822-1884) e estende-se até 1953, e estabelece a unificação das diferentes áreas das Ciências Biológicas (Zoologia, Botânica, Microbiologia, Histologia, Anatomia, Genética, etc.) em torno de uma entidade física e universal a todos os seres vivos: a molécula de ácido desoxirribonucléico (DNA)<sup>1</sup> (Avery, MacLeod e McCarty, 1944; Chargaff *et al.*, 1949, Hershey e Chase, 1952, Watson e Crick, 1953a,b).

---

<sup>1</sup>O DNA está presente nos diferentes organismos vivos na superfície do planeta Terra como a molécula portadora das informações genéticas (hereditárias), e em certos grupos de vírus. Outros grupos de vírus apresentam o RNA (ácido ribonucléico) como a molécula de informação.

E por fim, notadamente, uma vez estabelecida a base experimental da vida e a teoria cromossômica aceita, o caminho estava aberto para um rápido avanço na compreensão dos fenômenos biológicos, no processo de suscitar a utilização da informação gênica do DNA e, desse modo, permitir a existência, manutenção e evolução de sistemas biológicos mais complexos (Jacob and Monod, 1961; Nirenberg e Matthei, 1961; Crick, 1970; Sanger e Coulson, 1975).

Desse modo não é difícil justificar os mais de 3.000 projetos de seqüenciamento de genomas completos, desde 1995, representando uma nova perspectiva na compreensão das diferentes formas de vida, desde arqueas ao homem (Kyrpides et al., 1999; <http://www.genomesonline.org>).

Apesar da pesquisa genômica ser considerada uma pesquisa de ponta, ela é, sem dúvida, a pesquisa mais básica de todas, isso porque o projeto de seqüenciamento tem como objetivo primário decifrar todas as bases nitrogenadas de um genoma<sup>2</sup> (Santos, 2005). Por outro lado, o potencial máximo da genômica será alcançado quando as funções biológicas das proteínas forem determinadas e compreendidas, visto que sua função é determinada tanto por sua conformação tridimensional como pela sua interação com outras moléculas (Sanchez & Sali, 1999), o que não é possível deduzir somente a partir da seqüência primária dos aminoácidos.

Não obstante, um recorrente problema é a disparidade/discrepância entre os bancos de dados de seqüências, tal como o GenBank, e de estruturas de macromoléculas, como PDB (Protein Data Bank), devido, principalmente, às dificuldades técnicas na elucidação de estruturas tridimensionais (3D) por difração de raios-X ou ressonância magnética nuclear (NMR), o que concorre para a pouca informação estrutural da grande maioria das seqüências de proteínas conhecidas.

Assim, na era pós-genômica, diversas abordagens computacionais (ferramentas de Bioinformática) têm sido aplicadas para responder questões biológicas relacionadas à coleta, organização e análise de dados, como

---

<sup>2</sup> Genoma de um organismo seja ele unicelular ou pluricelular, representa toda a sua carga de DNA. Nos eucariotos o genoma encontra-se no núcleo e em organelas como mitocôndria e cloroplastos. Nos procariotos o genoma encontra-se disperso no citoplasma.

também para decifrar informações funcionais contidas nas seqüências protéicas (Sanchez & Sali, 1999). Tais ferramentas computacionais variam desde programas e/ou algoritmos matemáticos para a identificação de seqüências abertas de leitura – “open reading frames” (ORFs), identificação de similaridade e conservação de bases nitrogenadas nas seqüências de ácidos nucléicos e/ou de aminoácidos nas proteínas, predição de domínios funcionais e do enovelamento protéico, estabelecimento de relações de filogênese, e até a análise comparativa entre dados de diferentes genomas. Além disso, atendem à crescente necessidade de estabelecer padrões de interações moleculares que governam os mais diversos sistemas biológicos, e sua relação em eventos patológicos (Prosdocimi *et al.*, 2002).

Além disso, o desenvolvimento de métodos computacionais mais sofisticados (como a Modelagem Molecular e simulações de Dinâmica Molecular) tem permitido um avanço sem precedentes na compreensão de processos biológicos em nível atômico-molecular, aliando uma enorme economia de tempo em ensaios experimentais com a possibilidade de uma análise detalhada das propriedades físico-químicas do reconhecimento molecular.

## **1.2. (Bio)Diversidade, Conhecimento e Tecnologia**

Desde seus primórdios, a humanidade depende para sua sobrevivência dos recursos naturais, sejam os biológicos (animais, plantas e microorganismos), sejam abióticos (água, ar solo e recursos minerais). E ao longo dos séculos, o uso desses recursos tem sofrido modificações, ocasionando o início do desenvolvimento econômico e do progresso social (Garcia, 1995), e recentemente, extinções e modificações climáticas (REF)>

No decorrer da história, os recursos biológicos passaram a ser considerados mercadorias, objeto de troca e comércio. Para exemplificar, estima-se ainda que a utilização dos compostos da biodiversidade seja responsável por cerca de 45% do PIB mundial, e respondem por cerca de 30% das exportações brasileiras (Carneiro, 2007).

Por outro lado, após o desenvolvimento científico e tecnológico experimentado a partir da Revolução Industrial (século XX), nas últimas décadas, o conceito ecológico vem se ampliando, dentro de um modelo de desenvolvimento que busca uma relação de equilíbrio dinâmico. Tal modelo, resgata uma nova ética na relação do homem com a natureza pela interrelação evidente dos recursos naturais biológicos e não biológicos, concorrendo às primeiras discussões acerca das estratégias para a conservação dos recursos naturais que tem resultado em diversos tratados internacionais e na conservação do meio ambiente (Garcia, 1995).

Portanto, a questão ecológica tende a ocupar posição prioritária nos esforços mundiais para a conservação da biodiversidade<sup>3</sup>, a partir da adoção de um projeto de desenvolvimento sustentável, integrando o conhecimento científico às necessidades sociais (Garcia, 1995).

Com a realização da Convenção da Diversidade Biológica (CDB) em 1992 (ECO 92), verificou-se um verdadeiro divisor de águas quanto às questões ligadas a biodiversidade, uma vez que essa, antes considerada patrimônio da humanidade, passou a ser considerada como parte integrante da soberania do país no qual as espécies se encontram. No entanto, também suscitou conflitos de interesses polarizados entre os países desenvolvidos, possuidores de tecnologia, e os países em desenvolvimento, detentores da biodiversidade, e constantes discussões sobre o acesso aos recursos genéticos, aos conhecimentos tradicionais associados, e à repartição de benefícios.

Nesse cenário, alinha-se o conhecimento científico e desenvolvimento sócio-econômico à necessidade da conservação da diversidade biológica. Primeiramente, o Brasil abriga entre 10 a 20% do número de espécies conhecidas pela ciência (MMA, 1998). Cerca de 200.000 espécies já foram descritas no Brasil e estima-se que a biodiversidade do país seja composta por um número 6 a 10 vezes maior, ou seja, aproximadamente dois milhões de formas de vida, mas para muitos grupos de invertebrados e, mais ainda, para

---

<sup>3</sup> Biodiversidade ou Diversidade Biológica significa a variabilidade de organismos vivos de todas as origens, compreendendo, dentre outros, os ecossistemas terrestres, marinhos e outros ecossistemas aquáticos e os complexos ecológicos de que fazem parte, compreendendo ainda a diversidade dentro de espécies, entre espécies e ecossistemas, segundo a Convenção sobre Diversidade Biológica (CDB) promulgada durante a Conferência das Nações Unidas sobre o Meio Ambiente e Desenvolvimento, a Rio-92.

os microrganismos, a informação é demasiadamente incompleta (Lewinsohn & Prado, 2005). Além disso, o Brasil destaca-se pela participação de grupos de pesquisa brasileiros em projetos de seqüenciamento de genomas completos, como o que levou ao primeiro seqüenciamento de um fitopatógeno (Simpson *et al.*, 2000), pela definição de mecanismos econômicos para os países em desenvolvimento de uma parcela justa e eqüitativa do lucro econômico proveniente do uso comercial dos recursos biológicos e dos genes neles contidos, e pela manipulação do material genético.

Assim, em resumo, as Ciências Biológicas, diante desses paradigmas, avançam com novas interfaces (numa visão de interdisciplinaridade) com as Ciências da Computação, a Química, a Matemática, a Medicina, a Economia e a Física, a fim de permitir uma melhor compreensão dos sistemas biológicos em suas complexidades organizacionais, uma estabilidade dos ecossistemas, e reconhecida utilidade dos recursos naturais para a humanidade.

Sendo assim, tendo em vista agregar os fatos acima destacados, o presente trabalho, que abrange uma visão de interdisciplinaridade a experimentação bioquímica (*em bancada*) e a instrumentalização da bioinformática (*in silico*) aplicadas a questões biológicas clássicas, será apresentado em dois grupos de dados/resultados separadamente, a saber:

- 1) As interações moleculares em proteínas, em especial para a subfamília das P-ATPases;
- 2) A identificação de transportadores, utilizando a anotação do genoma da bactéria *Gluconacetobacter diazotrophicus*.

Embora os dois temas não se relacionem diretamente, representam vertentes que segui durante a pós-graduação, as quais se transformaram na base da minha formação de Doutorado.



## 2. Funcionalidade Biológica – as proteínas

Ao contrário de muitos outros polímeros artificiais e naturais, que surgem da repetição de um único bloco ou da mistura randômica de diferentes blocos, as seqüências de aminoácidos das proteínas são perfeitamente definidas pelo código genético. No entanto, as proteínas têm propriedades emergentes que as tornam muito diferentes de moléculas pequenas (como os aminoácidos) e também de outras macromoléculas, e podem ter um grupo de conformações especialmente favoráveis, que costumamos chamar de “estado nativo”, o que torna difícil experimentá-las em escala de tempo de observação e de relevância biológica (Karplus *et al.*, 1987; Frauenfelder *et al.*, 1991; Shortle *et al.*, 1998).

Portanto, para entender as funções biológicas de uma proteína, os parâmetros globais derivados da análise conformacional estatística de cada resíduo ao longo da cadeia polipeptídica não bastam, e precisamos descrever a estrutura local. Contudo, mesmo para aquelas proteínas que foram cristalizadas, as estruturas cristalográficas aparentemente fixas escondem variações substanciais e importantes (DePristo *et al.*, 2004), uma vez que as proteínas, mesmo sem estímulo externo, não são estáticas, mas se apresentam constantemente em movimento.

Além disso, muitas vezes a funcionalidade biológica das proteínas depende da energia de interação destas moléculas com ligantes ou substratos, e a aquisição da estrutura biologicamente ativa por um domínio pode depender da presença de outro domínio, mesmo se os dois domínios estiverem bastante afastados na seqüência de aminoácidos (Fedorov e Baldwin, 1999; Jansens *et al.*, 2002). Ela pode depender, ainda, da participação de outra(s) proteína(s) (Szepanski *et al.*, 1994). Desta forma, os diferentes domínios e monômeros colaboram para criar proteínas com funções novas.

Desse modo, o conhecimento da seqüência primária e/ou da estrutura tridimensional (3D) das proteínas não nos permite uma determinação de todas as propriedades emergentes das proteínas e, portanto, diferentes metodologias teóricas e experimentais devem ser aplicadas e somadas para o avanço do nosso conhecimento a respeito de sua relevância e de funcionalidade biológica. Especialmente, métodos de determinação estrutural

em solução, por ressonância magnética ou espalhamento de raios-X, oferecem a possibilidade de modelar explicitamente a dinâmica funcional da proteína, não revelada em estados cristalográficos (Arai *et al.*, 2004; Lindorff-Larsen *et al.*, 2005).

### 3. As Membranas Biológicas e o Fluxo Iônico

*“É difícil pensar em uma definição para a vida que não implique uma separação dos fluídos intracelular e extracelular.”*  
Simkiss, 1998

O surgimento de um sistema de compartimentação durante o desenvolvimento da vida no planeta colaborou efetivamente para a existência, manutenção e evolução de sistemas biológicos mais complexos (Deamer, 1997). Assim, a vida explorou, essencialmente, a estratégia de compartimentalização de ambientes aquosos por meio de uma camada hidrofóbica (bicamada lipídica), por esta respeitar as seguintes leis termodinâmicas: i) a estrutura da bicamada lipídica é viável, ocorre naturalmente; ii) sua flexibilidade permite o crescimento e movimento, como também a inserção e operação da maquinaria protéica; e, iii) sua estrutura tem uma constante dielétrica baixa, propriedade esta utilizada em mecanismos de sinalização, transporte e transdução de energia (Njus, 2000a).

Assim sendo, o conceito, de uma barreira física explícito no modelo de Davson e Danielli (1943) evoluiu progressivamente, nos últimos trinta anos, até um sistema dinâmico regulatório, como proposto por Singer e Nicholson (1972) (Simkiss, 1998; Lee, 2002).

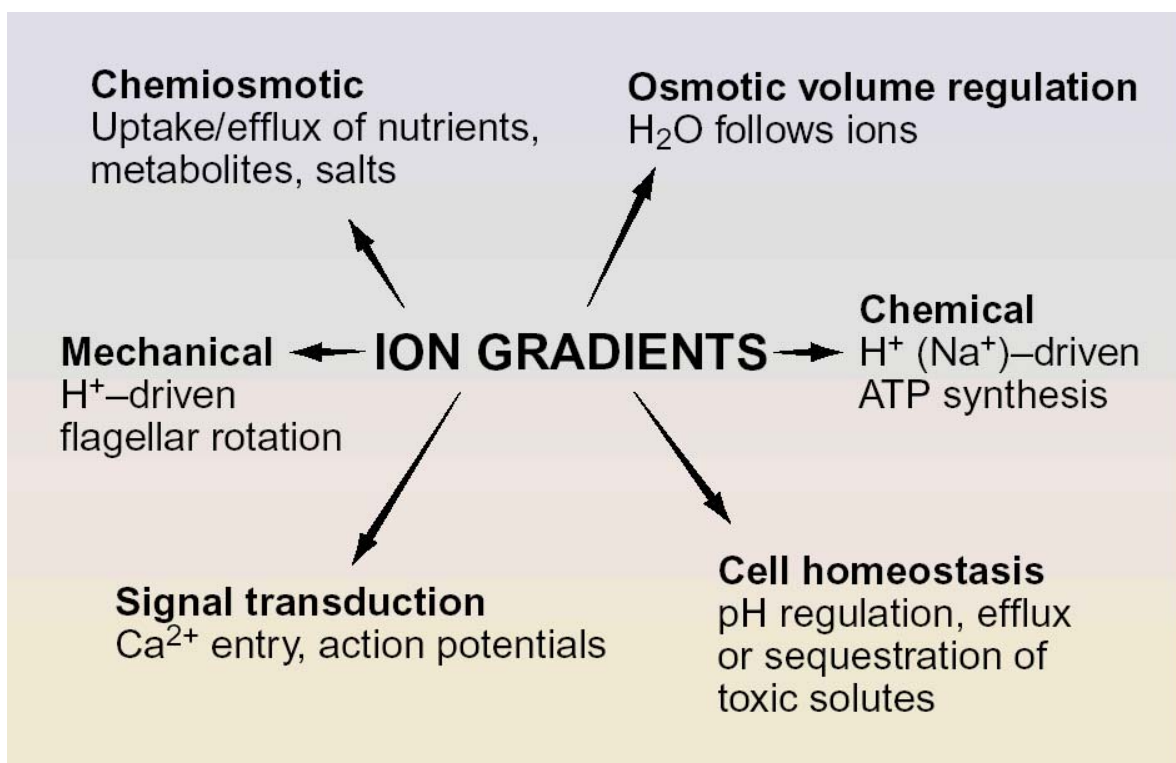
O estudo das membranas biológicas, assim como das proteínas de membranas, que representam cerca de 30% de todos genes (Saraste e

Walker, 1998; Werten et al., 2002), é extremamente importante, uma vez que essas cumprem funções vitais como interface com o meio exterior, interface entre células e compartimentalização intracelular (Werten et al., 2002). As membranas biológicas estão direta ou indiretamente relacionadas a numerosas doenças, como hiperinsulinemia, diabetes insipidus nefrogênica, falência congestiva cardíaca, cirrose hepática, hiper- e hipotensão, edema pulmonar, epilepsia e catarata (Blaustein e Lederer, 1999; Féraillé e Doucet, 2001; Werten et al., 2002). Porém, somente 30 estruturas unitárias de proteínas de membranas foram resolvidas em nível atômico, comparadas às mais de 3.000 estruturas unitárias de proteínas solúveis, dada a dificuldade de produzir análises tridimensionais (3D) satisfatórias por difração de raios  $\text{-X}$  de proteínas de membranas solubilizadas em detergentes (Werten et al., 2002).

A composição química das células varia notoriamente quando comparada ao ambiente. Para manter estes desequilíbrios, as membranas biológicas usam mecanismos de transporte ativo e passivo de nutrientes, inclusive íons, e com isso garantem a manutenção da constância da composição química do meio intracelular (Stryer, 1998). Funcionam, portanto, como uma barreira perfeita, de modo a manter gradientes químicos e/ou elétricos entre os compartimentos, além de possuir inúmeros mecanismos, operados especialmente por proteínas nelas incorporadas, como: i) gerar gradientes químicos e elétricos, e a conversão deles em trabalho útil; ii) permitir o transporte específico de uma variedade de moléculas que conduzem a uma troca controlada entre o interior celular e o ambiente; e, iii) perceber sinais químicos ou comprimentos apropriados de luz e transduzir por redes

sofisticadas numa variedade de respostas celulares (Dimroth, 1997; Maier, Jr. 2000).

Dentre as mais diversas moléculas que circulam através das membranas, os íons desempenham papel fundamental na energética celular (Harold, 1986, apud Sze et al., 1999). Realmente, o fluxo de íons é de fundamental importância na captação de energia durante a respiração e fotossíntese, e também media a interconversão das formas de energias química, osmótica e elétrica em biológica (Sze et al., 1999; Rees e Howard, 1999). Este fluxo apoia uma gama de funções fisiológicas, como a regulação do volume celular, comportamento natatório de organismos unicelulares, abertura dos estômatos nas plantas, contração muscular, secreção endócrina, e a geração da excitabilidade neuronal nos animais (Ranganathan, 1994) (Figura 1).



**FIGURA 1. Gradientes iônicos.** Os diferentes papéis que os gradientes iônicos desempenham nos sistemas biológicos.

A via de permeabilização responsável pela passagem de íons através do ambiente hidrofóbico das membranas é provido por três classes distintas de proteínas integrais de membranas: i. canais; ii. proteínas trocadoras; e, iii. proteínas transportadoras (Wolfersberger, 1994).

Os canais são, tipicamente, poros íon-específicos, que abrem e fecham transitoriamente, de maneira regulada. Quando um canal está aberto, os íons rapidamente atravessam a membrana pelo canal, dirigidos por gradiente elétrico e de concentração. Este movimento de íons pelos canais controla o potencial elétrico de membranas, e sinais elétricos em membranas excitáveis de células nervosas e musculares (Ranganathan, 1994; Wolfersberger, 1994).

As proteínas trocadoras, são “*enzyme-like proteins*”, pois promovem o transporte ativo secundário de solutos através das membranas contra gradientes de concentrações, utilizando-se de gradientes iônicos como uma fonte de energia para executar trabalho (Wolfersberger, 1994).

As proteínas transportadoras também são chamadas de bombas primárias que são catalisadores vetoriais, pois podem transduzir diretamente a energia química ou luminosa para o transporte de íons ou outros solutos através das membranas contra gradiente de concentrações (Nelson, 1994).

A maioria das proteínas íon-transportadoras conhecidas biologicamente bombeiam cátions. Embora elas pudessem da mesma maneira translocar ânions, há 3 bilhões de anos os cátions foram selecionados durante a evolução pelas células primitivas. Acredita-se que estas proteínas evoluíram independentemente desde o seu surgimento. Atualmente, suas estruturas, mecanismos de translocação do íon, e a fonte de energia variam consideravelmente (Nelson, 1988; Palmgren, 1998). Consequentemente, várias

proteínas íon-transportadoras não relacionadas, podem ser “energizadas” pela energia luminosa, energia de oxirredução, hidrólise de pirofosfato ou ATP (Nelson, 1994; Axelsen e Palmgren, 1998).

Este sistema de bombas primárias pode ser agrupado em duas categorias distintas, segundo Dimroth (1997). As bombas da primeira categoria servem diretamente às células e não tem nenhuma função na transdução de energia. Exemplos são os exportadores de substâncias tóxicas às células, que incluem os transportadores de arsênio nas bactérias (Carlin et al., 1995; Rosen et al., 1995; Rensing et al., 1999) ou os transportadores de resistência múltipla a drogas, que são responsáveis pela resistência das células tumorais ao tratamento quimio-terapêutico. Algumas ATPases tipo P também podem ser agrupadas nesta categoria, pois servem a necessidade nutricional, acumulando ou expulsando íons ( $K^+$ ,  $Cu^+$ ,  $Mg^{2+}$ ,  $Cd^{2+}$ ,  $Ag^+$ ), acidificando o lúmen do estômago, estabilizando gradientes iônicos para sinalização ( $Ca^{2+}$ ) ou promovendo a excitação elétrica ( $Na^+$ ,  $K^+$ ) em células nervosas (Axelsen e Palmgren, 1997; Scarborough, 1999).

A função fisiológica da segunda categoria de bombas primárias é criar um estado energizado nas membranas, na forma de um gradiente eletroquímico, que pode ser usado através de outros sistemas de trabalho acoplados a membrana. Este trabalho pode ser mecânico, como o movimento do flagelo bacteriano; químico, como a síntese de ATP; ou osmótico, como a captação de nutrientes pelos numerosos sistemas de transporte secundário (Dimroth, 1997).

### **3.1 Transporte Ativo, ATPases tipo F e ATPases tipo V**

O transporte ativo é usualmente definido como transporte de moléculas ou íons de uma região de baixa para uma região de alta concentração (isto é, transporte contra um gradiente de concentração)(Stryer, 1998; Njus, 2000b). Porém, definiremos em um senso restrito como transporte ativo aquele que requer energia química ou fotoquímica. Por esta definição, somente um pequeno número de enzimas é capaz de catalisar o transporte ativo e, com poucas exceções, elas podem transportar cátions (Saier, Jr., 1998; Njus, 2000b).

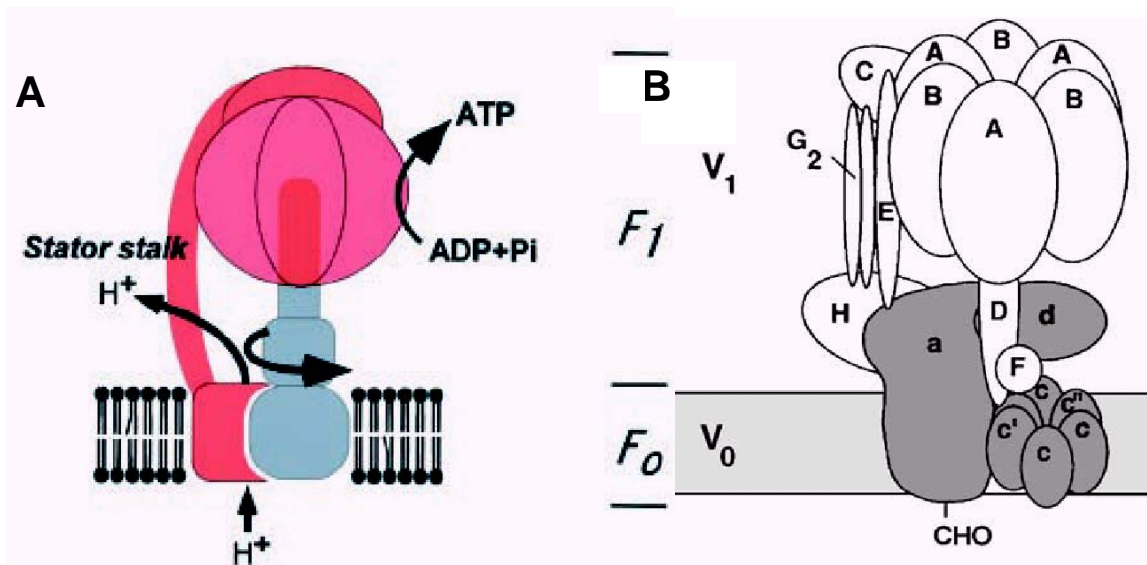
Uma das classes majoritárias de proteínas transportadoras, são as ATPases íon-transportadoras. As primeiras concepções a respeito das ATPases íon-transportadoras foram feitas por Dean (1941) (*apud* Skou, 1957, 1989), sendo identificadas posteriormente por Skou (1957). Dada a importância esta descoberta foi novamente publicada (Skou, 1989), e contemplada com o Prêmio Nobel em 1997.

Desde então, as ATPases que transportam ativamente cátions, foram extensivamente estudadas por diversos grupos de pesquisadores, e categorizadas por Pedersen e Carafoli (1987) em três classes: ATPases tipo F (F-ATPases), ATPases tipo V (V-ATPases) e ATPases tipo P (P-ATPases).

As F-ATPases estão localizadas na membrana plasmática bacteriana, na membrana interna mitocondrial e membrana dos cloroplastos, e operam *in vivo* como ATPases reversas, ou ATP-sintases. Estas sintetizam ATP a partir de ADP e fosfato inorgânico (Pi) acoplada ao fluxo de prótons movido por um gradiente eletroquímico favorável através da membrana (Amzel e Pedersen, 1983; Yoshida e Noji, 2000).

As V-ATPases utilizam a hidrólise de ATP para bombear prótons, gerando um gradiente eletroquímico. Elas são amplamente distribuídas em compartimentos intracelulares de células eucarióticas, incluindo endossomos, lisossomos, vesículas secretórias (incluindo vesículas sinápticas), vâcuolos de fungos e plantas, como também nas membranas plasmáticas de vários tipos celulares animais, tais como osteoclastos e macrófagos (Harvey, 1992; Forgac, 2000).

As classes F- e V-ATPases são constituídas por unidades estruturais complexas; morfologicamente, tais enzimas têm três componentes: i. um setor ancorado a membrana, (F<sub>0</sub>/V<sub>0</sub>) (Figura 2), que consiste em um canal iônico; ii. uma parte “conectora” central; e, iii. uma parte globulosa, (F<sub>1</sub>/V<sub>1</sub>) (Figura 2) que contém os sítios catalíticos (Muller et al., 2001).



**FIGURA 2.** Representação esquemática das ATPases tipo F e V. Em A, uma ATPase tipo F (Reproduzido de Yoshida e Noji, 2000); Em B, uma ATPase tipo V (Reproduzido de Forgac, 2000).

### 3.2 ATPases tipo P

#### 3.2.1 Considerações Gerais e Classificação

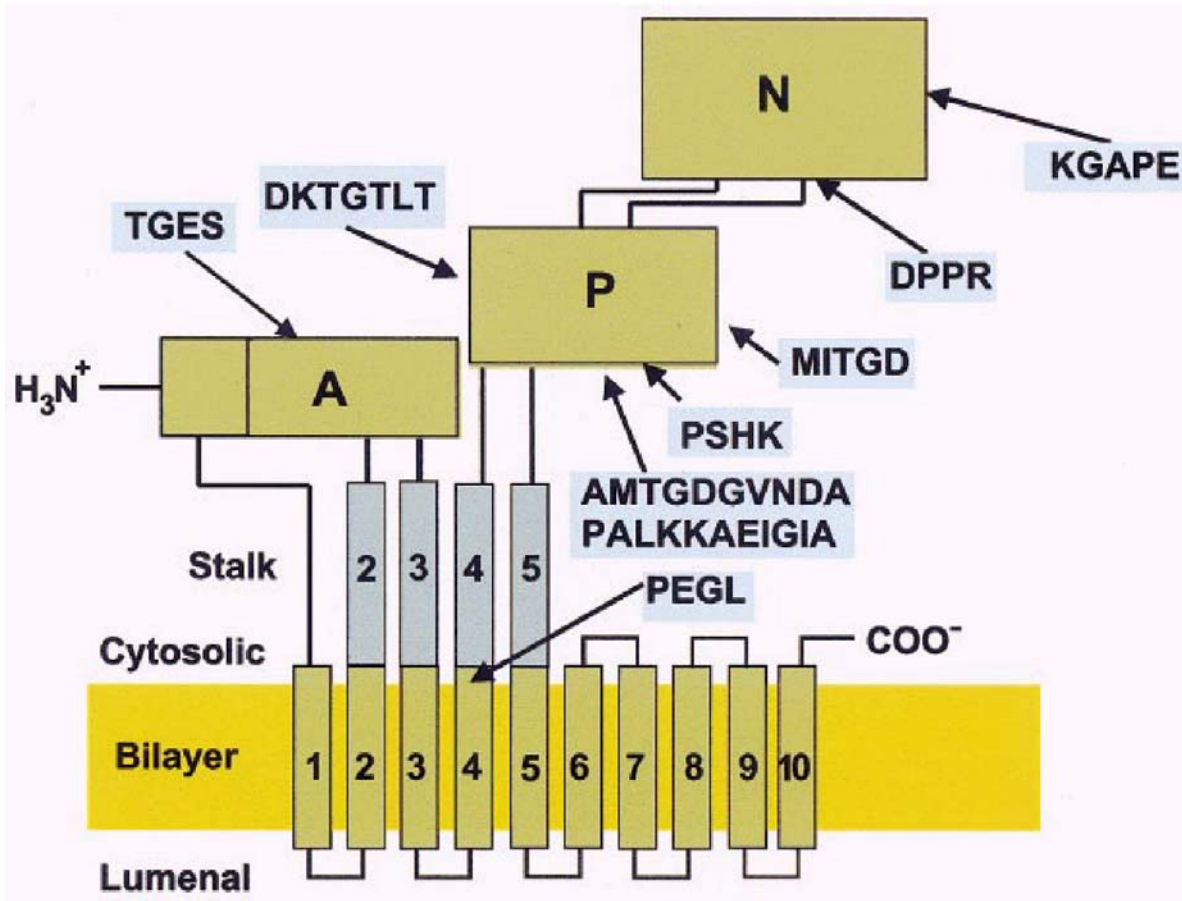


ATPases tipo P compreendem uma classe de proteínas integrais de membranas que estão envolvidas no transporte ativo de solutos carregados através das membranas biológicas (Moller et al., 1996). São encontradas em todas formas de vida, desde arqueobactérias aos animais, com poucas exceções como a bactéria *Borrelia burgdorferi* (Serrano, 1989; Lutsenko e Kaplan, 1995; Palmgren e Axelsen, 1998; Scarborough, 1999).

As ATPases tipo P são responsáveis por vários processos celulares. Em animais, por exemplo, estão envolvidas na manutenção de potenciais de ação nos tecidos nervosos, secreção e reabsorção de solutos nos rins, acidificação do estômago, absorção de nutrientes nos intestinos, contração/relaxamento muscular, transdução de sinal (cálcio dependente), crescimento e diferenciação celular, entre outros (Axelsen e Palmgren, 1997; Scarborough, 1999).

Uma das características marcantes dessas ATPases é a presença de algumas regiões altamente conservadas (seqüências padrão/motivos) facilmente identificadas nas estruturas primárias e que persistiram, de forma ligeiramente variável, ao longo da evolução (Moller et al., 1996). O mais eminente desses motivos é DKTG(I,L)T (Figura 3), seqüência característica do sítio de fosforilação; e uma longa seqüência "hinge" XMX GDGXNDXPX (Figura 3), que conectaria o domínio de ligação de ATP ao segmento transmembrana envolvido na ligação do íon e sua translocação (Moller et al., 1996). Na Figura 3 podem ser vistos também outras seqüências menores, de 3-5 resíduos de aminoácidos de comprimento, altamente conservadas, como o motivo PEG(L), que é considerado de crucial importância na transdução de energia; o padrão KGAXE, característico encontrado na alça citoplasmática maior de ATPases

em eucariontes, que contém o resíduo de lisina responsável pela ligação com o fluoróforo isotiocianato de fluoresceína (FITC); e, o motivo TGES/A, localizado no “loop” flexível (alça citoplasmática menor) (Lutsenko e Kaplan, 1995; Moller et al., 1996; Lee e East, 2001).



**FIGURA 3. Representação esquemática das ATPases tipo P, identificando os vários padrões característicos (setas), baseada no modelo estrutural de Toyoshima et al., 2000.** Observam-se três domínios citoplasmáticos denominados de A (atuador), com o motivo TGES (TGES/A) localizado na alça citoplasmática menor; P (fosforilação), com os motivos DKTGTLT (DKTG(I,L)T), MITGD, PSHK, e AMTG DGVNDA PALKKAEIGIA (XM X GDGXNDXPX); e, N (ligação de nucleotídeo), com os motivos KGAPE (KGAXE) e DPPR – todos localizados na alça citoplasmática maior; o motivo PEGL, no segmento transmembrana 4; bem como os 10 segmentos transmembranares, e as porções amino e carboxi-terminais. Os domínios A, P e N serão melhor discutidos posteriormente no corpo do texto. (Extraído de Lee e East, 2001).

Além da estrutura, outras duas características importantes distinguem as P-ATPases das ATPases tipo F e V. Estas características foram identificadas por estudos cinéticos anteriores ao conhecimento das estruturas primárias e

tridimensionais, e são: i) a inibição por vanadato; e, ii) a formação de um intermediário fosforilado (um aspartil-fosfato) durante o ciclo catalítico (*apud* Pedersen e Carafoli, 1987; Stokes e Wagenknecht, 2000).

Vários grupos de pesquisadores especularam a relação entre as várias P-ATPases, quanto a organização estrutural, diversidade, íon transportado e sua relação evolutiva (Serrano, 1988; Fagan e Saier, 1994; Lutsenko e Kaplan, 1995; Moller et al., 1996). Contudo, o número de proteínas incluído nestas análises foi relativamente baixo, o que dificultou a elaboração de um sistema de classificação confiável. Após 1994, pelo menos uma nova classe de P-ATPases, a classe IV, foi descoberta (Tang et al., 1996), e com isso a classificação ficou incompleta. Isto sem mencionar o sequenciamento completo de vários genomas (Fleischmann et al., 1995; Fraser et al., 1995; Bult et al., 1996; Goffeau et al., 1997), e a substancial quantidade de informações de seqüências de vários outros organismos, o que favoreceu finalmente a proposição de uma classificação mais confiável.

Tradicionalmente, as P-ATPases eram divididas em famílias de acordo com alguns critérios, tais como especificidade por íons, origem bacteriana ou eucariótica (Carafoli et al., 1991), ou pelo número de segmentos transmembranares e íon transportado (Lutsenko e Kaplan, 1995). Atualmente, as ATPases tipo P estão classificadas em pelo menos 5 famílias, baseadas em suas seqüências, substratos específicos, e presença de vários domínios regulatórios, cada uma podendo ser subdividida em duas ou mais subfamílias (Axelsen e Palmgren, 1998; Palmgren e Axelsen, 1998).

As ATPases tipo IA compreendem um complexo de quatro proteínas de membranas: KdpF, KdpA, KdpB, KdpC, a Kdp-ATPase. O complexo tem uma

quantidade equimolar das três subunidades maiores: KdpA, KdpB e KdpC, e parece trabalhar como um oligômero. Então, sua estequiometria mínima é A<sub>2</sub>B<sub>2</sub>C<sub>2</sub>, mas pouco se conhece sobre o arranjo das subunidades no complexo KdpFABC. A KdpB é homóloga à subunidade catalítica de outras ATPases tipo P, contendo as estruturas (seqüências) fundamentais comuns desta classe de ATPases, como o sítio de ligação de ATP e a seqüência altamente conservada DKTGT (Gabel et al., 1998).

As ATPases tipo IB transportam metais como zinco, prata, cobre, cobalto e chumbo. Funcionam na manutenção de homeostase desses metais essenciais; conferem resistência para concentrações tóxicas de chumbo, cádmio, cobre, zinco e prata. Em humanos, estão relacionadas às doenças de Menke e de Wilson, que são causadas por mutações em duas Cu<sup>+</sup>-ATPases, ATP7A e ATP7B, respectivamente (Vulpe et al., 1993; Tanzi et al., 1993; Bull et al., 1994). Uma característica eminente dessas ATPases é a presença de 1-6 repetições de um motivo conservado, ligador de metal, GTMCXXCXXC ou M/HXXMDHS/GXM, na porção amino terminal da cadeia polipeptídica (Silver et al., 1993; Bull e Cox, 1994; Solioz e Vulpe, 1996; Gatti et al., 2000).

No grupo II, estão agrupadas as enzimas ATPases tipo P melhor caracterizadas com respeito a cinética de reação, modificação química e mutagênese sítio-dirigida (Vasilets, e Schwarz, 1993; Moller et al., 1996; MacLennan et al., 1997): as Ca<sup>2+</sup>-ATPases de retículo sarco(endo)plasmático (SERCA) e de membrana plasmática (PMCA)– tipos IIA e IIB, respectivamente; Na<sup>+</sup>, K<sup>+</sup>-ATPase, a bomba sódio-potássio, e a H<sup>+</sup>, K<sup>+</sup>-ATPase de membrana plasmática da mucosa gástrica – tipo IIC.

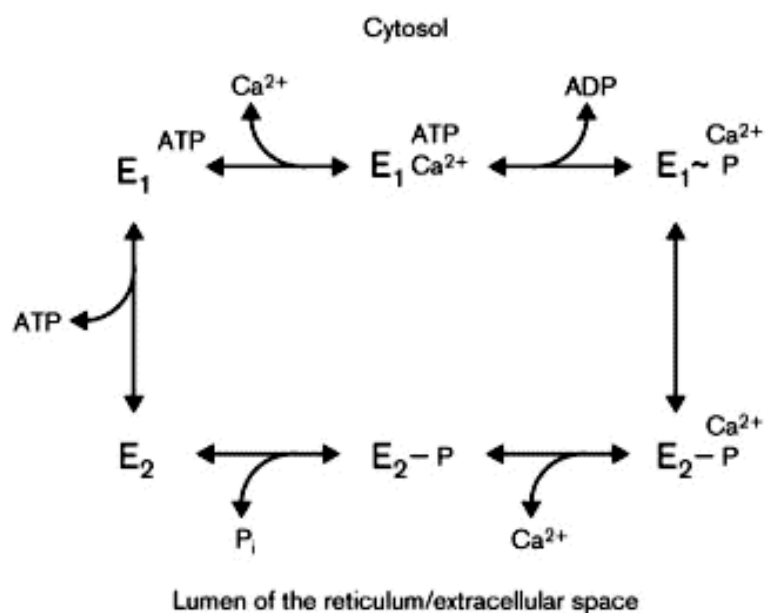
No grupo III encontra-se as H<sup>+</sup>-ATPases de membrana plasmática de fungos e plantas, que desempenham papel crucial na fisiologia desses organismos (Bowman e Bowman, 1986; Gaber, 1992; Portillo, 2000; Persomme e Boutry, 2000).

Os dois outros grupos – IV e V – foram recentemente identificados, como resultado dos projetos de sequenciamento de genomas e “*expressed sequence tags*” – EST (Tang et al., 1996; Goffeau et al., 1997; Halleck et al., 1998). As P-ATPases tipo V parecem ser abundantes em células eucarióticas, mas a sua função ainda é desconhecida (Axelsen e Palmgren, 1998).

Análises filogenéticas demonstraram que uma simples ATPase em um organismo é, freqüentemente, mais homóloga a uma ATPase da mesma classe de outro organismo distante evolutivamente que a outra ATPase no mesmo organismo. Então, parece que, pelo menos, a especificidade de íons evoluiu muito cedo, e antes da divisão de Eubacteria, Archaea e Eucarya (Palmgren e Axelsen, 1998). P-ATPases tipo IA são encontradas somente em bactérias e, com base em sua estrutura primitiva, podem estar relacionadas a origem ancestral das proteínas ATPases tipo P. P-ATPases tipo IB e IIA são encontradas em todos os domínios, apareceram cedo durante a evolução. P-ATPases tipo III são encontradas em arqueas, plantas e fungos, mas não em bactérias, e provavelmente evoluíram depois. P-ATPases tipos IV, IIB, e IID são encontrados somente em eucariontes e provavelmente evoluíram depois da divisão entre Archaea e Eucarya (Palmgren e Axelsen, 1998).

### **3.2.2 Aspectos Cinéticos e Estruturais**

As  $\text{Na}^+$ ,  $\text{K}^+$ -ATPase,  $\text{H}^+$ ,  $\text{K}^+$ -ATPase e  $\text{Ca}^{2+}$ -ATPase de retículo sarco(endo)plasmático (SERCA) foram intensamente estudadas do ponto de vista cinético e mecanístico (veja como revisões: Moller et al., 1996; MacLennan et al., 1997). A descrição do mecanismo de transporte das P-ATPases é baseado na hipótese proposta, no final da década de 60, por Albers (1967) e Post e colaboradores (1969) para a  $\text{Na}^+$ ,  $\text{K}^+$ -ATPase (apud Vasilets, e Schwarz, 1993), e por Makinose (1973) para a  $\text{Ca}^{2+}$ -ATPase (apud Minitz e Guillain, 1997), no qual a enzima pode adotar duas conformações durante o ciclo de transporte, denominadas de E1 e E2. O modelo proposto por de Meis e Vianna (1979) é representado de forma simplificada na Figura 4.



**FIGURA 4 – Esquema simplificado do ciclo catalítico da SERCA.** Ilustra os principais intermediários e a seqüência de ligação de íons e principais substratos. Baseado no modelo proposto por de Meis e Vianna (1979).

De acordo com a Figura 4, assumindo como exemplo a  $\text{Ca}^{2+}$ -ATPase de retículo sarco(endo)plasmático (SERCA), o estado E1 tem dois sítios de alta afinidade para  $\text{Ca}^{2+}$  do lado citosólico. Então, E1 é capaz de ligar  $\text{Ca}^{2+}$  do citosol em baixas concentrações e, depois dos dois sítios serem ocupados por

$\text{Ca}^{2+}$  em presença de  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}\text{E}$  é fosforilada por ATP originando  $\text{Ca}^{2+}\text{E}\sim\text{P}$  (uma fosfoenzima) (Shigekawa et al., 1978; Dupont, 1980; Kubo et al., 1990; Lewis e Thomas, 1992; Lacapere e Guillain, 1993). Esta sofre uma mudança conformacional ( $\text{E2Ca}^{2+}\sim\text{P}$ ) que exibem sítios de baixa afinidade para  $\text{Ca}^{2+}$  no sentido do lúmen, liberando  $\text{Ca}^{2+}$  (Froehlich e Heller, 1985; Hanel e Jecks, 1991; Orłowski e Champeil, 1991) e passando a  $\text{E2}\sim\text{P}$ . Finalmente, esta fosfoenzima é hidrolisada liberando o fosfato inorgânico ( $\text{Pi}$ ), gerando  $\text{E2}$ , sendo capaz de reiniciar novamente o ciclo, no estado  $\text{E1}$  (Kanazawa e Boyer, 1973; Froehlich e Taylor, 1975; 1976).

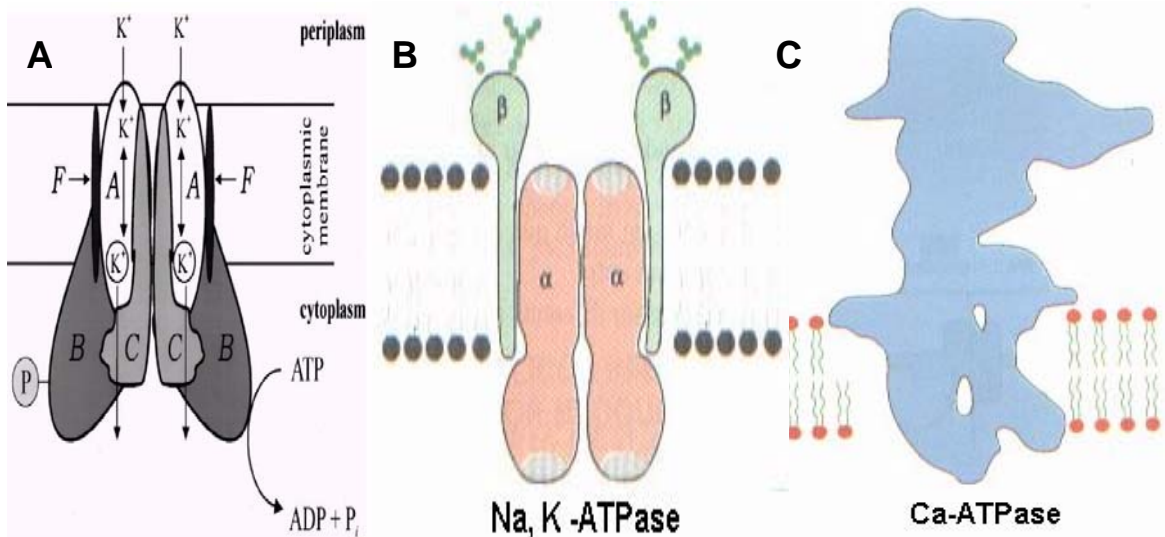
O ciclo é totalmente reversível, o que pode ser demonstrado de duas formas diferentes: i) em presença de cálcio, vesículas de retículo sarco(endo)plasmático (SR) previamente carregadas com cálcio, podem sintetizar ATP a partir de ADP e  $\text{Pi}$ , com efluxo de dois  $\text{Ca}^{2+}$  por ATP sintetizado; ii) em ausência de cálcio, vesículas SR “mal vedadas” – com ionóforos, por exemplo – podem sintetizar ATP pela reversão direta da ATPase, etapa por etapa, iniciando-se pela fosforilação espontânea da proteína por  $\text{Pi}$  em presença de magnésio e ausência de cálcio (Masuda e de Meis, 1973) com a formação de  $\text{E}\sim\text{P}$ . Numa nova etapa, com concentrações milimolares de  $\text{Ca}^{2+}$  e ADP adicionados juntos, o  $\text{Ca}^{2+}$  liga-se nos sítios internos induzindo a formação de espécies  $\text{Ca}^{2+}\text{E}\sim\text{P}$ , que conseqüentemente ligam ADP, e transferem o fosfato ao ADP, para sintetizar uma molécula de ATP.

Modelos similares foram propostos para outras ATPases tipo P, que formam um intermediário fosforilado (fosfoenzima) e sofrem mudanças

conformacionais durante o ciclo catalítico (transporte do substrato e hidrólise de ATP).

A natureza molecular das ATPases é variável. Podem compreender formas complexas com hetero-subunidades, como: i) no complexo da Kdp de bactérias (Figura 5A; Gabel et al., 1998); ii) na Na<sup>+</sup>, K<sup>+</sup>-ATPase (Figura 5B), e a H<sup>+</sup>, K<sup>+</sup>-ATPase constituídas das subunidades  $\alpha$  (catalítica) e  $\beta$  (glicosilada), ou, iii) somente com uma subunidade, como nas Ca<sup>2+</sup>-ATPases de retículo sarco(endo)plasmático (SERCA) (Figura 5C), de membrana plasmática (PMCA), e nas H<sup>+</sup>-ATPases de membrana plasmática de fungos e plantas. Podem também ser observadas “subunidades adicionais” (com propriedades regulatórias), como a “subunidade  $\gamma$ ” na Na<sup>+</sup>, K<sup>+</sup>-ATPase nos rins (Blanco e Mercer, 1998; Jorgensen et al., 1998), o “phospholemman” (Chen et al., 1998), e o “phospholamban” (Simmernan e Jones, 1998).





**FIGURA 5. Representação das ATPases tipo P, natureza molecular diversificada.** Em **A**, o complexo Kdp, uma ATPase transportadora de potássio, com as subunidades: A, “canal de íons”; B, subunidade catalítica para hidrólise de ATP; C e F, responsáveis pela estabilização do complexo (Modificada de Gabel et al., 1998). Em **B**, bomba de sódio-potássio (Na<sup>+</sup>, K<sup>+</sup>-ATPase), com as subunidades catalíticas  $\alpha$  e glicosilada  $\beta$ , evidenciadas. Em **C**, Ca<sup>2+</sup>-ATPase de retículo sarco(endo)plasmático (SERCA), constituída de uma cadeia polipeptídica [B e C - Modificadas de Bioquímica (Lubert Stryer, 4ª edição – 1998, Editora Guanabara Koogan S.A., Rio de Janeiro; páginas 291 e 295, respectivamente)].

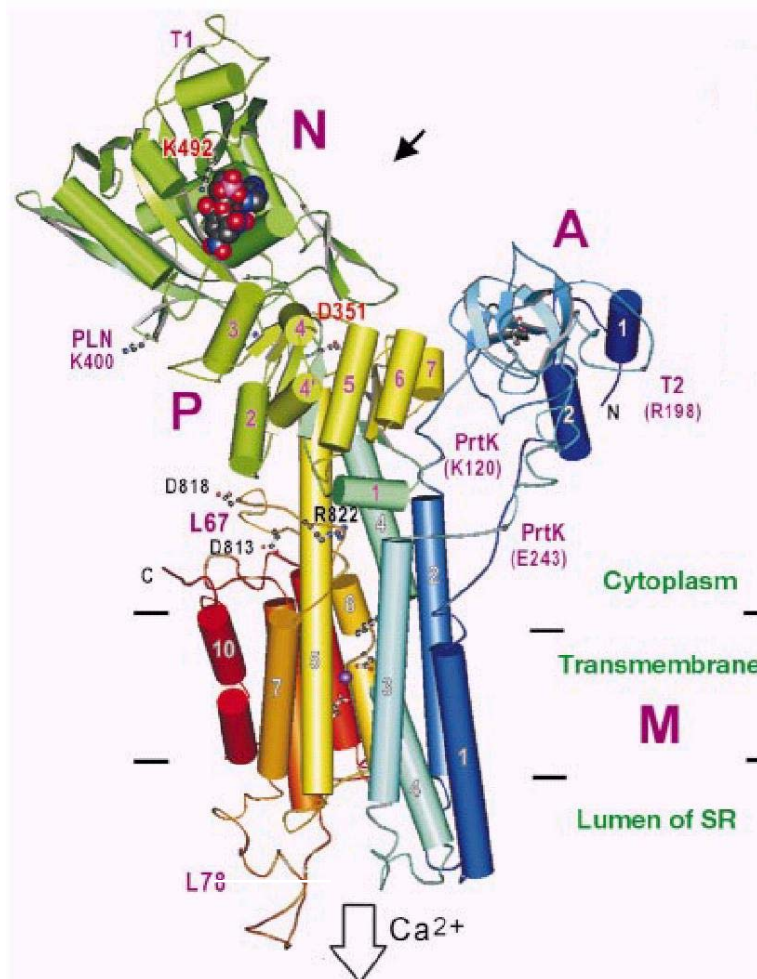
Contudo, evidências experimentais sugeriram que grandes mudanças conformacionais ocorriam desde a formação da fosfoenzima à translocação do(s) íon(s) pela membrana, e seriam igualmente compartilhados pelos membros dessa família. Tais evidências incluem clivagem proteolítica (Jorgensen, 1975), medidas espectroscópicas (Karlsh e Yates, 1978; Karlsh, 1980; Pick e Karlsh, 1980; Yamamoto et al., 1989) e “cross-linking” (McIntosh, 1992) (ver como revisão Bigelow e Inesi, 1992; Martonosi, 1995). Mas, muito pouco se sabia sobre os mecanismos moleculares durante o transporte iônico e a interconversão de energia (Andersen e Vilsen, 1995; Moller et al., 1996; MacLennan et al., 1997), principalmente devido à falta de informação estrutural de alta resolução destas bombas (Jahn et al., 2001).

Os estudos estruturais, por muito tempo, foram limitados a análises de eletromicroscopia. Cristais bidimensionais da  $\text{Ca}^{2+}$ -ATPase foram descritos na década de 80, em duas formas diferentes. Em ausência de cálcio, em meio contendo  $\text{Mg}^{2+}$  e decavanadato, ocorreu a cristalização de dímeros da  $\text{Ca}^{2+}$ -ATPase. Em presença de cálcio, os cristais apresentam monômeros (Dux e Martinosi, 1983; Dux et al., 1987; Stokes e Green, 1990; Martonosi, 1995). Estes trabalhos conduziram aos primeiros mapas de isodensidade, mas como foram conduzidos depois de revelação negativa, a parte principal ou o domínio de membrana não foram revelados (Mintz e Guillain, 1997). Posteriormente, novos experimentos de crio-eletromicroscopia permitiram: i) avaliar mais precisamente as dimensões e distribuição de massa da  $\text{Ca}^{2+}$ -ATPase, 120Å (altura total), 70% da massa em domínio de citoplasma, 25% na membrana e 5% no lúmen (Toyoshima et al., 1993); ii) determinar os domínios transmembranares; e iii) revelar rearranjos dos domínios citoplasmáticos pelas pequenas mudanças entre as 10  $\alpha$ -hélices transmembranares, comparando as estruturas das  $\text{Ca}^{2+}$ -ATPase (Zhang et al., 1998) e  $\text{H}^{+}$ -ATPase de *Neurospora crassa* (Auer et al., 1998), em diferentes estados conformacionais (Stokes et al., 1999). Restava ainda, portanto, compreender a intimidade da comunicação entre os domínios citoplasmáticos e os segmentos de membrana.

Assim sendo, vários grupos realizaram estudos de caracterização dos domínios citoplasmáticos de várias ATPases (Capieaux et al., 1994; Moutin et al., 1994, 1998; Falson et al., 1997; Gatto et al., 1998; Champeil et al., 1998; Tran e Farley, 1999; McIntosh et al., 1999; Carvalho-Alves et al., 2000), bem como inferiram a estrutura terciária do domínio catalítico (Smirnova et al., 1998), identificavam do sítio de ligação de  $\text{Mg}^{2+}$  (Rider e Dijkstra, 1999) e

modelaram o enovelamento da dehalogenase (Stokes e Green, 2000), baseado na homologia da superfamília haloácido dehalogenases (HADs) (Aravind et al., 1998).

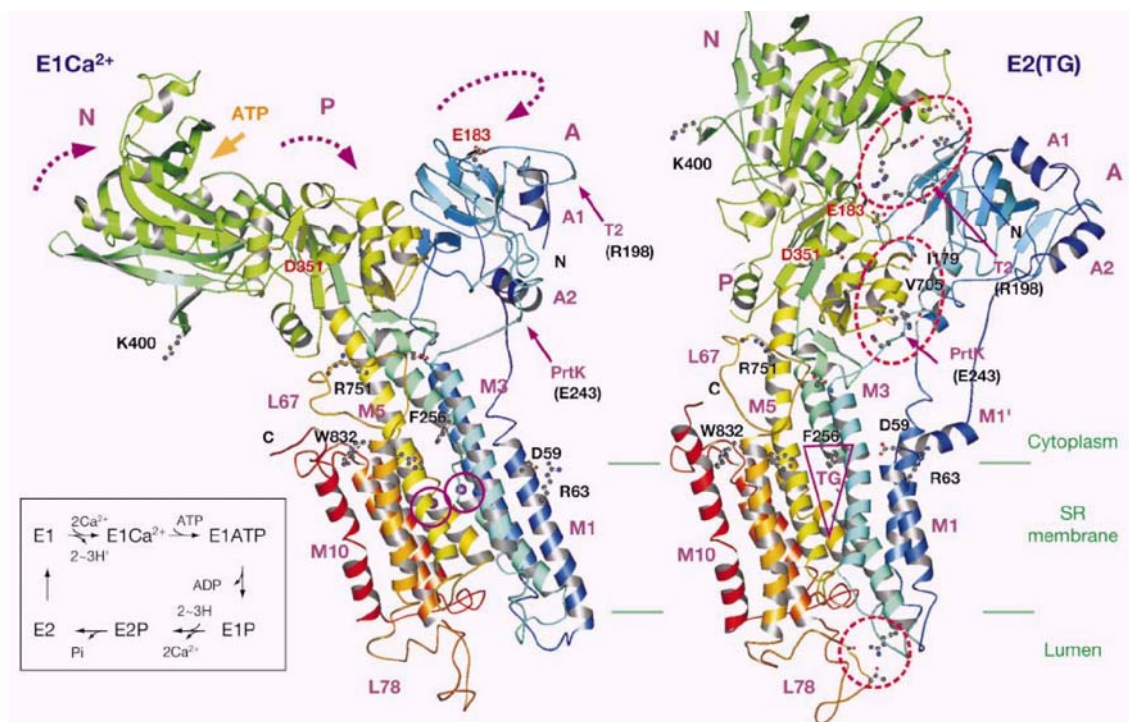
Em 2000, Toyoshima e colaboradores reportaram a estrutura 3D da  $\text{Ca}^{2+}$ -ATPase ligado a cálcio com alta resolução (2,6 Å) (Figura 6; Toyoshima et al., 2000). Neste trabalho, eles confirmaram o arranjo das 10  $\alpha$  hélices transmembranas e a similaridade do domínio catalítico da haloácido dehalogenase (Aravind et al., 1998) no domínio P, e dividiram a porção citoplasmática em três domínios funcionais: domínio A (atuador/ativação), formado pela porção amino-terminal e alça citoplasmática menor, localizada entre as hélices M2-M3, responsável pelas principais mudanças conformacionais; domínio N (ligação de nucleotídeo), cuja estrutura característica “Rossmann fold” (Lee e East, 2001) é responsável pela ligação de ATP; e o domínio P (fosforilação) por 7 (sete) fitas  $\beta$ -paralelas que formam uma única folha e oito pequenas  $\alpha$  hélices. Além disso, confirmaram os sítios de ligação dos íons  $\text{Ca}^{2+}$ , entre as hélices M4-M6 e M8 (Clarke et al, 1989).



**FIGURA 6 – Estrutura da  $\text{Ca}^{2+}$ -ATPase de retículo sarco(endo)plasmático (SERCA).** Coloração muda gradualmente da região amino terminal (azul) ao carboxi terminal (vermelho). Visão frontal, identificação dos 10 segmentos transmembranas (numerados de 1-10); os domínios funcionais: A (em azul); P (em amarelo); e, N (em verde); os íons cálcio (em roxo). Em destaque os resíduos D351, que é fosforilado no domínio P; K492, que liga o nucleotídeo (na figura um análogo de ATP, TNP-AMP) (Extraído de Toyoshima et al., 2000).

Os três domínios funcionais (domínios A, N e P) que estão separados na estrutura da  $\text{Ca}^{2+}$ -ATPase (Figura 6; Toyoshima et al., 2000), teriam que sofrer grandes mudanças conformacionais, em ausência de  $\text{Ca}^{2+}$ , conforme proposto por Green e MacLennan (2000) – em concordância com a estrutura tubular com decavanadato e ausência de  $\text{Ca}^{2+}$ . De fato, mudanças conformacionais ocorrem e foram elucidadas por estudos estruturais de eletromicroscopia (Xu et al., 2002), ou por cristalografia (Figura 7; Toyoshima e Nomura, 2002) e,

interessantemente, podem ser monitoradas ao longo do ciclo catalítico (Danko et al., 2001).



**FIGURA 7. Diferenças estruturais na Ca<sup>2+</sup>-ATPase de retículo sarco(endo)plasmático (SERCA).** Coloração muda gradualmente da região amino-terminal (azul) ao carboxi-terminal (vermelho). Na direita (E1Ca<sup>2+</sup>), SERCA no estado conformacional E1, com íons cálcio ligados (esferas roxas, circuladas). Na esquerda (E2(TG)), SERCA no estado conformacional E2, com taspigargina (TG), círculos pontilhados vermelhos indicam pontes de hidrogênio extras na E2TG. Setas largas em E1Ca<sup>2+</sup> indicam a direção dos movimentos dos domínios citoplasmáticos durante o ciclo E1Ca<sup>2+</sup> a E2(TG). Em destaque, um esquema simplificado do ciclo catalítico (Extraído de Toyoshima e Nomura, 2002).

### 3.2.3 P-ATPases: monômeros ou oligômeros?

A manutenção das concentrações iônicas dentro da normalidade nos diversos compartimentos dos seres vivos é fundamental para a realização das diferentes características de cada célula (Moller *et al.*, 1996). Desta forma, as ATPases do tipo P (P-ATPases) desempenham importantes papéis na fisiologia celular, como o transporte de íons e estabelecimento de gradientes eletroquímicos através das membranas biológicas, e por isso, têm sido alvo de vários estudos físico-químicos.

Desde sua descoberta há aproximadamente 50 anos (Skou, 1957), vários progressos têm sido feitos nas análises estrutural e funcional dessas ATPases, promovendo com isso um modelo cinético geral para o transporte iônico e seu relacionamento estrutural, dada pela seqüência de aminoácidos, predição da estrutura secundária, mutagêneses, eletromicroscopia, difração de raios-X e/ou dados de espectroscopia (Bigelow e Inesi, 1992; Vasilets, e Schwarz, 1993; Martonosi, 1995; MacLennan *et al.*, 1997).

Muito embora a elucidação da estrutura tridimensional (3D) da  $\text{Ca}^{+2}$ -ATPase de retículo sarco(endo)plasmático – SERCA (Toyoshima *et al.*, 2000) seja útil na interpretação de muitos dados cinéticos e de mutagênese, alguns fatos permanecem sem resposta (Lee, 2002); dentre eles se destaca a questão da oligomerização. Ainda é controverso se as unidades funcionais das P-ATPases nas membranas existiriam como estruturas monoméricas ou oligoméricas.

Apesar de ter sido demonstrado que unidades monoméricas de ATPases sejam capazes de realizar o ciclo completo de hidrólise de ATP e transporte (Goormaghtigh *et al.*, 1986; Andersen, 1989) há inúmeras evidências de estados oligoméricos em várias P-ATPases. Estas evidências incluem diferentes abordagens experimentais, como: a ativação da atividade ATPásica pelo excesso de ATP (Moller *et al.*, 1980), dados estruturais de eletromicroscopia ou difração de raios-X (Dux e Martonosi, 1983; Jahn *et al.*, 2001), medidas espectroscópicas (Coelho-Sampaio *et al.*, 1991; Levi *et al.*, 2000; 2002), inativação por radiação (Cavieries, 1984; Briskin e Reynolds-

Niesman, 1989), imunquímica (Blanco *et al.*, 1994; Koster *et al.*, 1995; Maguiere e Ohlendieck, 1996), tanto para ATPases na membrana nativa quanto para proteínas reconstituídas em lipossomos.

Estudos cinéticos também sugerem que estados oligoméricos sejam importantes no mecanismo de transporte (Froehlich & Taylor, 1976; Wang, 1986; Mahaney *et al.*, 1995) e que mudanças dinâmicas nas interações proteína-proteína são cruciais durante o ciclo catalítico (Karon & Thomas, 1993; Merino *et al.*, 1999). Isso sem mencionar o papel que o estado oligomérico tem na cooperatividade cinética e na estabilização enzimática (Blanco *et al.*, 1994; Maguire & Ohlendieck, 1996; Levi *et al.*, 2002).

Porém, em muitos casos a existência destes complexos oligoméricos poderia ser devido ao modo de preparação das membranas, e não refletir interações específicas entre subunidades catalíticas. Além disso, a análise da contribuição de domínios hidrofílicos ou hidrofóbicos na “auto-associação” não havia sido investigada, provavelmente, por causa da dificuldade de gerar proteínas recombinantes com alto grau de pureza e em quantidade suficiente para estudos físico-químicos (Carvalho-Alves *et al.*, 2000).

Em 1994, foi demonstrada a primeira evidência de haver uma interação específica entre subunidades  $\alpha$  de  $\text{Na}^+, \text{K}^+$ -ATPase (Blanco *et al.*, 1994). Em seguida, foi demonstrado que uma região de 150 resíduos de aminoácidos, localizada dentro da alça citoplasmática maior, seria responsável por este processo de associação das subunidades  $\alpha$  (Koster *et al.*, 1995).

Posteriormente, em 2000, foi observado que a alça (ou domínio) citoplasmática(o) maior da  $\text{Ca}^{+2}$ -ATPase de retículo sarco(endo)plasmático (LCD/SERCA), quando expressa heterologicamente, formava uma estrutura dimérica, baseado na medida do raio de Stokes e no raio de giro (Carvalho-Alves *et al.*, 2000). Mais interessante ainda foi a verificação de que o tratamento da LCD/SERCA com a proteinase K gerava um fragmento de 30 kDa (p30) resistente à digestão, similar ao da proteína nativa (Champeil *et al.*, 1998), e que este fragmento (p30) comportava-se como um monômero (Carvalho-Alves *et al.*, 2000). Esta observação nos levou a propor que a extensa região conservada, chamada de “hinge domain” (Clarke *et al.*, 1989)

contida no carboxi-terminal liberada pelo tratamento com proteinase K, contribuiria para a dimerização da enzima (Carvalho-Alves *et al.*, 2000).

Mais recentemente, também foi demonstrado que a LCD/Na<sup>+</sup>,K<sup>+</sup>-ATPase tende a se auto-associar e, interessantemente, interage com a enzima nativa, com dependência do substrato – Mg<sup>+</sup>-ATP (Costa *et al.*, 2003).

O atual projeto visa esclarecer o papel da alça citoplasmática maior (LCD) de P-ATPases, em particular o sítio de fosforilação (P domain), na interação entre cadeias  $\alpha$  para a formação de dímeros, e se esse seria um mecanismo geral, e assim propor um modelo estrutural para este dímero.

#### **4. *Gluconacetobacter diazotrophicus* e o seqüenciamento de seu genoma**

O Brasil vem se destacando com uma crescente participação mundial de genomas, desde a publicação do genoma da *Xylella fastidiosa* (Simpson *et al.*, 2000), uma vez que o seqüenciamento de genomas completos fornece informações detalhadas sobre uma espécie que não podem facilmente ser gerada por meios alternativos. Estas informações incluem genes únicos, proteínas hipotéticas conservadas, composição de GC, o índice de variações totais e elementos repetitivos e/ou de inserção, além de permitir uma reconstrução detalhada do perfil fisiológico e metabólico dos organismos, e uma análise funcional subsequente dos genes, dos quais nenhuma informação precedente estava disponível.

Segundo Santos (2005) cerca de 50% dos projetos de seqüenciamento de genomas bacterianos tem algum interesse biotecnológico e, devido ao fato de grande parte da nossa economia ser baseada na atividade agropecuária, quase a totalidade dos diferentes projetos de seqüenciamentos de genomas desenvolvidos em território nacional são voltados para esse setor da economia (Carraro & Kitajima, 2002). Então, seguindo essa tendência nacional, em 2000, a Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de



Janeiro (FAPERJ) lançou o projeto de seqüenciamento do genoma completo da bactéria *Gluconacetobacter diazotrophicus* (<http://www.lncc.br/~riogene>).

A *G. diazotrophicus* (*syn.*, *Acetobacter diazotrophicus*), que é um bacilo Gram negativo, reto e com extremidades arredondadas, que possui flagelos laterais ou peritríquios, de metabolismo microaeróbico, pertencente ao subgrupo das  $\alpha$ -proteobactérias, apresenta a capacidade de fixar nitrogênio atmosférico e produzir ácido acético a partir de etanol, além da capacidade de redução de acetileno (Cavalcante e Döbereiner, 1988). Este microrganismo é encontrado no espaço intercelular (Figura 8), realizando associação endofítica<sup>4</sup> em todas as estruturas da cana de açúcar, e com os mais variados genótipos deste vegetal, especialmente em áreas agrícolas onde a adição de fertilizantes nitrogenados nos cultivares é baixa (Gillis *et al.*, 1989; Reis, Olivares e Döbereiner, 1994; Fuentes-Ramírez *et al.*, 1999; Steenhoudt e Vanderleyden, 2000; Fischer e Newton, 2005). Desta forma, contribui com uma parcela do nitrogênio necessário para o desenvolvimento da planta hospedeira, além de produzir diversas auxinas (como o ácido 3-indol acético, IAA) e citocinas que podem exercer efeitos diretos sobre a fisiologia da planta, influenciando no crescimento vegetal.

De fato, tem sido largamente documentado que *G. diazotrophicus* pode aumentar a taxa de crescimento da cana de açúcar, e este benefício tem sido associado, em parte, com a transferência do nitrogênio bacteriologicamente fixado ao vegetal, e ainda com a produção de fatores estimulantes do crescimento vegetal, destacando-se a produção dos fitormônios auxina e giberilina, que têm sido amplamente detectados em culturas deste endófito (Lee *et al.*, 2000; Muthukumarasamy *et al.*, 2002; Muñoz-Rojas e Caballero-Mellado, 2003; Lee *et al.*, 2004; Muthukumarasamy *et al.*, 2006).

Desse modo, essa bactéria torna-se um modelo de estudo de organismos fixadores de nitrogênio, tanto pela compreensão da fixação biológica de nitrogênio (FBN) propriamente dita quanto pelo seu potencial biotecnológico como, por exemplo, pela possibilidade de substituição de fertilizantes nitrogenados em diferentes culturas de importância econômica,

---

<sup>4</sup> Endofíticos - são aqueles microrganismos que habitam o interior das plantas, sendo encontrados em órgãos e tecidos vegetais como as folhas, ramos e raízes.

principalmente na cana-de-açúcar. Nos últimos anos, a cultura da cana vem sendo considerada estratégica no desenvolvimento brasileiro, graças ao Pró-Álcool (Programa Nacional do Álcool), que é uma importante iniciativa ao uso alternativo a combustíveis fósseis, e o Brasil é o único país a implementar um programa de larga escala do uso de um combustível alternativo ao petróleo (Zanin *et al.*, 2000).

Outro exemplo de importância biotecnológica está relacionado a processos de controle biológico, devido a uma atividade antagonista contra *Xanthomonas albilineans*, como resultado da produção de uma bacteriocina similar à lisozima, a qual é capaz de inibir o crescimento deste fitopatógeno responsável pela doença da folha escaudada na cana de açúcar, possivelmente através da destruição da parede celular da cepa patogênica (Piñon *et al.*, 2002; Blanco *et al.*, 2005).

Outro aspecto que deve ser destacado também é que, diferentemente da interação rizóbio/leguminosa, na associação *G. diazotrophicus*-planta não há formação de nódulos ou qualquer outra estrutura formada ou induzida pela bactéria, o que sugere uma interação muito bem sucedida no interior da planta (Figura 8) promovendo o crescimento vegetal sem causar nenhum sintoma de doença, cujo mecanismo molecular precisa ser melhor entendido (Baldani & Baldani, 2005).

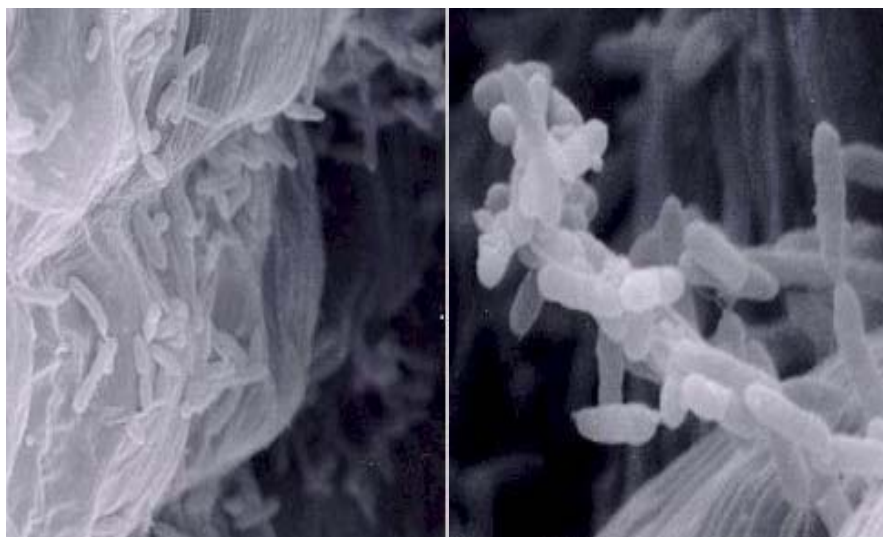


Figura 8. Microscopia de varredura de *Gluconacetobacter diazotrophicus*.

Estudos de expressão gênica de cana de açúcar durante a associação com *G. diazotrophicus*, utilizando diferentes abordagens tais como cDNA-AFLP fingerprinting, perfis transcricionais gerados pelo SUCEST (Sugarcane EST Sequencing Project) e microarranjos, sugerem uma participação efetiva da cana de açúcar, pela decodificação de sinais bacterianos e por respondê-los adequadamente, com a mudança nos perfis de expressão gênica, favorecendo o seu sucesso na associação com o vegetal (Nogueira *et al.*, 2001; Vargas *et al.*, 2003).

Recentemente foi evidenciada, através de preparações de glicoproteínas purificadas de cana de açúcar e marcadas com isotiocianato de fluoresceína (FITC), uma interação específica dessas glicoproteínas com receptores presentes na parede celular bacteriana (Blanco *et al.*, 2005), que por sua vez sugerem a primeira etapa da discriminação biológica de um endófito simbiote compatível pelo vegetal no processo de colonização, e a participação de estruturas e mecanismos moleculares da superfície dessa bactéria nesse fenômeno.

#### **4.1 Anotação do Genoma de *G. diazotrophicus*: Sistema de Transporte**

Em projetos de genomas de procariotos, normalmente realiza-se a quebra do DNA inteiro do organismo desejado em fragmentos pequenos (através da técnica de “shotgun”) que são clonados em vetores plasmidiais que serão seqüenciados em suas extremidades.

Após uma primeira etapa de montagem desse genoma, fragmentos maiores são clonados em cosmídeos e seqüenciados. Essa segunda etapa é importante para a montagem do genoma completo do organismo, já que a primeira normalmente produz uma seqüência incompleta, apresentando alguns buracos de seqüência (gaps).

Uma vez obtidos os dados do seqüenciamento das moléculas de DNA é preciso saber o que representa cada uma das seqüências nucleotídicas produzidas. A anotação consiste simplesmente no processo de identificação

dessas seqüências.

A partir da anotação de seqüências nucleotídicas procura-se, primeiramente, identificar a natureza de uma determinada seqüência. Então, pretendemos descobrir se tal seqüência está inserida em uma região gênica, se representa uma molécula de RNA transportador ou RNA ribossômico, se pertence a algum tipo de região repetitiva já descrita, ou se apresenta algum marcador genético conhecido em seu interior.

O principal objetivo dessa etapa é construir um mapa do genoma do organismo, posicionando cada um dos possíveis genes e caracterizando as regiões não-gênicas. Nesta fase, alguns programas de predição gênica são usados para a localização de possíveis genes nas seqüências de DNA. A procura por elementos como o códon de iniciação de proteínas (a trinca de nucleotídeos ATG, por exemplo) e códons de terminação na mesma fase de leitura são utilizados por alguns desses programas, como o Glimmer. O tamanho delimitado por esta janela de leitura é freqüentemente utilizado para definir uma determinada região como sendo gênica ou não.

Mapeados os genes, a etapa seguinte consiste em identificar quais proteínas são codificadas, e nisso consiste o processo de anotação das seqüências protéicas. Nessa etapa, procura-se montar um catálogo dos genes presentes no organismo estudado, dando-lhes nomes e associando-os a prováveis funções. No caso do trabalho descrito nesta tese, buscamos por proteínas de membranas associadas aos sistemas de transporte.

A análise das seqüências biológicas de um projeto de seqüenciamento genômico, através da utilização das mais variadas ferramentas de bioinformática e a comparação com os mais variados bancos de dados, compreende uma etapa final conhecida como anotação, que consiste na conversão dos dados de uma seqüência biológica, em informações biologicamente relevantes. A anotação consiste em uma meta-informação, ou seja, uma descrição de características em mais alto nível da seqüência biológica (Lemos, 2004).

Dentre as principais ferramentas de bioinformática, utilizadas para execução do processo de anotação, destacam-se: o BLAST (Basic Local Alignment Sequence Tool), que consiste numa ferramenta utilizada para

verificação de homologia de uma determinada seqüência, com seqüências estudadas previamente por outros pesquisadores (Altschul, 1990; Altschul, 1998; WU-BLAST, 2006); o RBSFinder, que localiza sítios de ligação ribossomal (TIGR, 2006); programas de predição de genes como o GLIMMER (Delcher et al., 1999) e o ORFFinder (NCBI, 2006a); algoritmos para predição de estruturas de proteínas como o Modeller (Sali, 2006) e o Threader (Jones, 2006); pacotes de análise filogenética como o PHYLIP (PHYLogeny Inference Package) (Felsenstein, 2006); e programas projetados para comparação de vários genomas como o Alfresco (Sanger, 2006).

Com relação aos bancos de dados, podemos destacar os que armazenam seqüências de nucleotídeos, tais como o Genbank-NT (NCBI, 2006b), GSDB (Harger et al., 1998), GDB (Letovsky et al., 1998) e o EMBL (Cochrane et al., 2006); os que armazenam seqüências de aminoácidos como o SWISS-PROT (Boeckmann, 2003), PIR (Wu et al., 2002) e o Genbank-NR (NCBI, 2006b); bancos de dados de estruturas terciárias de proteínas como o PDB - Protein Data Bank (Westbrook et al. 2002); padrões de seqüências proteicas conservadas associadas com funções específicas num organismo, tais como o PROSITE (Falquet et al., 2002); e bibliotecas que oferecem padrões de seqüências de estruturas de domínios protéicos mais longos, como o PFAM (Bateman et al., 2004).

O processo de anotação apresenta 4 etapas distintas, onde a primeira consiste na anotação em nível de nucleotídeos, cujo principal objetivo é a localização física de ORFs (fases de leitura aberta) nas seqüências de DNA, assim como elementos repetitivos, promotores e sítios de ligação ribossomais. A segunda etapa da anotação ocorre em nível de proteínas, cujos objetivos residem na determinação de identidades e prováveis funções dos genes, visando identificar genes presentes (e eventualmente ausentes) num determinado genoma. A terceira etapa ocorre em nível de processo, onde os grupos de pesquisa buscam realizar a identificação de diversas vias metabólicas. A quarta etapa consiste na análise das seqüências reguladoras presentes no genoma, visando implementar uma anotação funcional eficiente, e aprimorar a compreensão das diversas vias metabólicas presentes no organismo, em especial aquelas que apresentam importância econômica,

científica, tecnológica e acadêmica (Koonin, Mushegian et al., 1996; Downs e Escalante-Semerena, 2000; Schwartz, 2000; Sistema Bionotes 2006b).

Em linhas gerais, podemos afirmar que a anotação de um genoma consiste num processo interativo, que requer a utilização de diversas ferramentas de bioinformática, com finalidade de refinar os dados brutos gerados a partir de um projeto de seqüenciamento genômico (Sistema Bionotes, 2006b), através da utilização de ferramentas que possibilitem a localização de regiões codificantes numa seqüência de DNA (Delcher et al., 1999; NCBI, 2006a), a predição de função e estrutura da proteína codificada por um determinado gene (Jones, 2006; Sali, 2006), assim como a determinação de relações filogenéticas entre diferentes organismos (Felsenstein, 2006; Sanger, 2006). Desta forma, são elucidadas questões relacionadas com organização genômica, aspectos bioquímicos, fisiológicos e evolutivos do organismo em estudo, e de espécies relacionadas e/ou espécies que ocorrem em associação (Koonin, Mushegian et al., 1996; Moszer, 1998).

Assim, os diferentes grupos que integram o RioGene ficaram responsáveis pela anotação e identificação de genes de sua proficiência e interesse. Desse modo, devido a minha experiência com transportadores iônicos, contribuí no projeto de seqüenciamento da *G. diazotrophicus* com a identificação de ORFs relacionadas com os sistemas de transporte em membranas.

Todos os estudantes das Ciências Biológicas hão de concordar em um ponto essencial: a Vida envolve tanto interconversões químicas quanto fluxos através das membranas biológicas, que referimos como metabolismo e transporte, respectivamente, e nenhuma descrição de vida é completa sem uma compreensão inclusa de ambos fenômenos (Saier, 2000). Assim, os sistemas de transporte em membranas têm papéis fundamentais no metabolismo e atividade celular, por catalisarem a translocação de solutos por distintos mecanismos moleculares, e funcionam na fisiologia celular em processos distintos, tais como a aquisição de nutrientes orgânicos, e a manutenção da homeostase iônica, a extrusão de compostos tóxicos, a sinalização e comunicação celular e ambiental, por exemplo (Saier, 2000).

Os vários sistemas de transporte existentes diferem em relação a sua topologia na estrutura das membranas biológicas, mecanismos de acoplamento energético e transporte, e especificidade por substrato (Paulsen *et al.*, 1998). Contudo, análises filogenéticas entre os diferentes grupos de transporte demonstram que uma classe é, freqüentemente, mais homóloga a uma proteína da mesma classe, em outro organismo distante evolutivamente, que a outra classe de transportador no mesmo organismo. Então, parece que pelo menos a especificidade de acoplamento de energia e por substrato evoluiu muito cedo, e antes da divisão dos três domínios da vida: Bacteria, Archaea e Eucarya (Palmgren & Axelsen, 1998).

Assim cada transportador é classificado pelo TCDB (*Transporter Classification System*; <http://www.tcdb.org>) em famílias e subfamílias, de acordo com sua função, filogenia e/ou especificidade de substrato, sendo que mecanismos de acoplamento energético e transporte servem como base preliminar para a classificação, por causa de suas características relativamente estáveis (Saier, 1999).

Há, portanto, cinco classes principais de transportadores no sistema TCDB: canais, transportadores (ativos) primários, transportadores secundários, translocadores, e uma classe distinta, onde são incluídos transportadores cujo mecanismo energético ou de função são desconhecidos (Saier, 1999).

Desse modo, o estudo das membranas biológicas, bem como das proteínas de membranas, que representam cerca de 30% de todos os genes, é extremamente importante, uma vez que estas cumprem as funções vitais, como interface com o meio exterior, interface entre células, e como limites de compartimentos intracelulares (Werten *et al.*, 2002). Além disso, entre 2% a 16% das ORFs, tanto de genomas procarióticos quanto eucarióticos, foram preditos a codificar proteínas de transporte da membrana, fato esse que enfatiza a importância dos transportadores no estilo de vida de todas as espécies (Ren & Paulsen, 2005).

Apesar das proteínas de transportes serem vitais à compreensão da potencialidade metabólica de organismos, poucos grupos têm se dedicado à análise sistemática dos sistemas de transportes dos microorganismos, em especial daqueles em que os genomas completos foram publicados, e/ou

ainda uma análise comparativa dos sistemas de transportes entre os mais diferentes microorganismos (Ren & Paulsen, 2005).

Por esta razão, e tendo em vista uma análise mais detalhada do genoma da *G. diazotrophicus*, a condução de um inventário de todas as proteínas de transporte é uma etapa chave para entender as estratégias desenvolvidas por este microorganismo para adaptar-se às circunstâncias ambientais, e um interesse adicional para o estudo da identificação das proteínas de transporte das diferentes classes que estejam envolvidas no processo de associação com o vegetal.



## **OBJETIVOS**

## II. OBJETIVOS

---

Tendo em vista agregar os fatos acima destacados, o presente trabalho, que abrange em uma visão de interdisciplinaridade a experimentação bioquímica (*em bancada*) e a instrumentalização da bioinformática (*in silico*) aplicadas a questões biológicas clássicas, será apresentado em dois grupos separadamente, a saber:

- 1) As interações moleculares em proteínas, em especial para a subfamília das P-ATPases;
- 2) A identificação de ORFs relacionadas com os sistemas de transporte em membranas, no projeto de seqüenciamento da *G. diazotrophicus*;

Objetivos específicos:

- a. Expressão heteróloga da alça citoplasmática maior (LCD) da H<sup>+</sup>-ATPase de membrana plasmática de *Saccharomyces cerevisiae*;
- b. Caracterização da participação do enovelamento nativo e, da participação de resíduos pertencentes à porção C-terminal do sítio de fosforilação no fenômeno de auto-associação entre P-ATPases;
- c. Proposição de um modelo geral dimérico para P-ATPases;

E, como um objetivo secundário, devido a minha experiência com transportadores iônicos:

- d. Identificação de ORFs relacionadas com o sistema de transportadores em membrana, durante o processo de anotação, no projeto genoma *G. diazotrophicus*.

## **MATERIAIS E MÉTODOS**

### III. MATERIAIS E MÉTODOS

---

#### 3.1 Cepas Utilizadas e Condições de Cultivo

Dentre os vários sistemas de expressão disponíveis para produção de proteínas heterólogas, a bactéria gram-negativa *Escherichia coli* é um dos mais utilizados e atrativos, devido a sua habilidade de crescer rapidamente, com altas concentrações celulares e em substratos baratos, além sua genética ser bem conhecida e haver disponibilidade de muitos vetores de clonagem e cepas mutantes (Baneyx, 1999). Os microorganismos utilizados neste trabalho foram as cepas XL1-Blue, DH5, M15[pREP-4] (Qiagen), e BL21-SI de *Escherichia coli*.

Todas as cepas de *Escherichia coli* foram cultivadas por um período de 18 a 24 horas a 37°C sob agitação constante (150 rotações por minuto - rpm) no meio de cultura LB (Luria Bertani - pH 7,5); composto de triptona 1% (p/v); extrato de levedura 0,5% (p/v) e cloreto de sódio 0,5% (p/v) (Sambrook *et al.*, 1989). A cepa BL21-SI foi crescida em meio de cultura LB modificado - pH 7,5, sem adição de cloreto de sódio 0,5%, sendo denominado então de LBON (Donahue e Bebee, 1999). Os meios de cultura foram esterilizados por autoclavagem a 120°C por 15 minutos (Sambrook *et al.*, 1989).

#### 3.2 Primers e Reação em Cadeia da Polimerase (PCR)

O gene codificante da H<sup>+</sup>-ATPase de membrana plasmática de *Saccharomyces cerevisiae*, PMA1 (Serrano *et al.*, 1986), havia sido subclonado em vetor pBlueScript (Stragene) obtido com o grupo da Dra. Caroline Slayman (Yale University), sendo denominado pBlue/PMA1.

Os oligonucleotídeos (primers) para a clonagem e expressão do LCD/PMA1 foram desenhados de forma que anelassem em regiões correspondentes ao LCD/SERCA1 (identificadas através do alinhamento das duas seqüências utilizando o BLAST (Basic Local Alignment Search Tool) 2 SEQUENCES – NCBI (Tatusova e Madden, 1999)), el compreenderia a região

codificante entre os resíduos Leu356 ao Iso674 (SERCA - resíduos Met326 ao Iso743) contendo a região do “Hinge Domain”; e, a região codificante entre os resíduos Leu356 ao Tre544 (SERCA – resíduos Met326 ao Ser610), porção similar ao clivado pela proteinase K (Champeil *et al.*, 1998).

O isolamento das regiões da alça citoplasmática maior, com e sem “hinge domain”, denominados LCDh e LCD $\Delta$ h, respectivamente; foram amplificadas por PCR a partir do plasmídeo pBlue/PMA1 (20 ng), em um volume final de 25  $\mu$ l, usando 200  $\mu$ M de dNTP; 2,5 U de Taq DNA polimerase, tampão T (Tris-Cl 10 mM; MgCl<sub>2</sub> 1,5 mM e KCl 50 mM) e 1,5  $\mu$ M de cada oligonucleotídeo.

O gene foi amplificado utilizando a seguinte programação: 94°C por 10 minutos, 29 ciclos de 1 minuto a 94°C, 1 minuto a 50°C e 1 minuto a 72°C. Os fragmentos amplificados foram então analisados em gel de agarose (Sambrook *et al.*, 1989).

### **3.3 Eletroforese de DNA**

Os produtos das reações de PCR e as mini-preparações de DNA plasmidial e digestões com enzimas de restrição foram eletroforeticamente separados em gel de agarose 1% (p/v) preparado em tampão TAE (Tris-acetato 40 mM, EDTA 1 mM), contendo brometo de etídio na concentração final de 0,5  $\mu$ g/ml.

### **3.4 Clonagem e Seleção de Clones**

O pGEM-T Easy Vector é linear e suas extremidades 5' possuem timidinas, as quais viabilizam a ligação de produtos de PCR diretamente, pois a enzima Taq polimerase acrescenta uma adenina na extremidade 3' dos produtos de PCR, que pareiam com as timidinas do vetor, possibilitando (otimizando) a ligação. Além disso, este vetor traz o gene lacZ da  $\beta$ -galactosidase, possibilitando a identificação das colônias que apresentam o plasmídeo recombinante de interesse. Quando o produto do PCR é ligado ao

vetor o gene da  $\beta$ -galactosidase fica interrompido e a enzima, por isso, fica inativa gerando colônias brancas. Do contrário o gene fica contínuo, e a enzima funcional, de forma que quando expressa catalisa a reação de hidrólise do X-Gal (5-bromo-4-cloro-3-idoil- $\beta$ -D-galactosídeo), um análogo cromogênico da lactose, formando colônias azuis (Dower et al., 1988; Sambrook *et al.*, 1989).

Os clones recombinantes foram confirmados através de PCR, onde colônias isoladas crescidas em LB acrescido de agar 1,5% (w/v) e ampicilina a 50  $\mu$ g/ml, foram colocadas em um tubo contendo 100  $\mu$ l de água destilada autoclavada. Posteriormente esta mistura foi aquecida a 90°C por 10 minutos e centrifugada a 12.000 x g por 5 minutos a 4°C. O sobrenadante foi então transferido para um novo tubo e 5 $\mu$ l usados para a reação de PCR. Foram utilizados os seguintes oligonucleotídeos na clonagem no vetor pGEM-T Easy Vector (Promega) M13F (5'TAATACGACTCACTATAGGGCGA3') e M13R (5'ATTTAGGTGACACTATAGAATA3').

### **3.5 Sistema de Clonagem de DNA por Recombinação Sítio-Específica (Gateway Cloning Technology)**

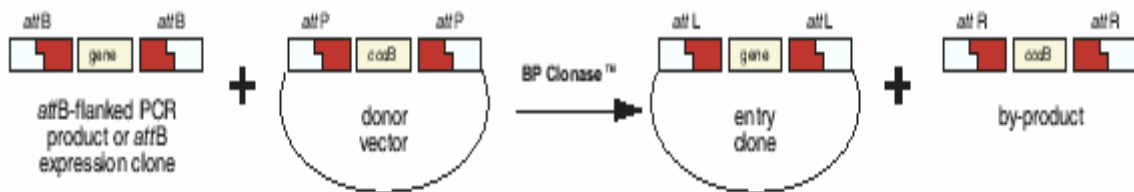
O sistema por recombinação sítio-específico é uma técnica de clonagem *in vitro* baseada no sistema de recombinação sítio-específico do bacteriófago lambda (Landy, 1989; Hartley et al., 2000) na transição entre as vias lítica e lisogênica (Ptashne, 1992).

A recombinação é catalisada por uma mistura de enzimas que ligam-se em seqüências específicas (sítios *att*), reúne estes sítios, cliva-os e liga covalentemente o DNA (gene de interesse). As proteínas de recombinação envolvidas dentro da reação diferem dependendo de quais vias o bacteriófago lambda utiliza: a via lítica ou lisogênica.

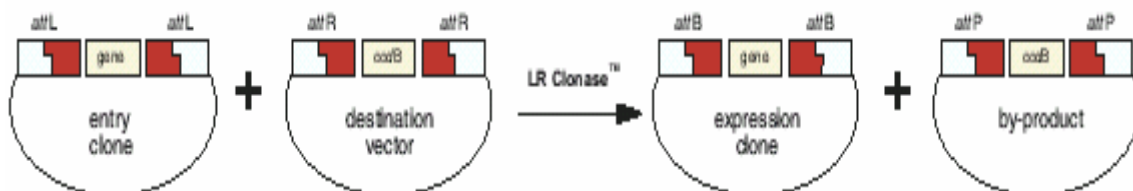
A via lisogênica é catalisada pelas proteínas IHF (Integration Host Factor) e  $\lambda$  Int (Integrase), encontrada no mix BP Clonase, enquanto na via lítica é catalisada pelo mix LR Clonase, constituído pelas proteínas  $\lambda$  Int (Integrase) e Excisionase (Xis), e IHF (Integration Host Factor). Duas reações

de recombinação sítio-específica, portanto constituem a base da *Tecnologia de Clonagem Gateway*.

- i) **BP Reação** - Recombinação, catalisada pelo mix BP Clonase, de um substrato de *attB* (produto de *attB*-PCR ou um vetor linearizado *attB*) com um substrato de *attP* (vetor doador – *Donor Vector*) para criar um Clone de Entrada (*Entry Clone*) - contendo sítios *attL*.



- ii) **LR Reação** - Recombinação, catalisada pelo mix LR Clonase, de um substrato de *attL* (Clone de Entrada) com um substrato de *attR* (Vetor de Destino – *Destination Vector*) criando um Clone de Expressão (*Expression Clone*) - contendo sítios *attB*.



Estas reações de recombinação são altamente específicas, isto é, os sítios *attB1* recombinam somente com *attP1*, mas não com *attP2*; assim como sítios *attL1* recombinam somente com *attR1*, mas não com *attL2*.

Pode-se reparar também, nos esquemas acima, que os resultados das reações de LR e BP são o vetor com o qual se deseja trabalhar, e um vetor contendo o gene *ccdB*. A proteína, *ccdB*, tem a peculiaridade de ser altamente tóxica para organismos como bactérias e leveduras, de forma que, caso a recombinação não ocorra, e portanto o fragmento que codifica para a proteína

*CcdB* não seja eliminado, o organismo não consegue se desenvolver, sendo esse sistema uma maneira de selecionar transformantes.

Este sistema, por excelência, foi desenvolvido para facilitar os processos de clonagem de genes, pois elimina etapas de digestão com endonuclease de restrição, ligações, géis e amplificações; permitindo a inserção de um gene simultaneamente em múltiplos vetores em paralelo (vetores de *E. coli*, leveduras, células de mamíferos, baculovírus, como proteínas de fusão – His-tag, GST-tag, etc.), com eficiência maiores que 90% em reações bastante rápidas (~60 minutos) em temperatura ambiente e, sempre mantendo a orientação do segmento de DNA durante a recombinação e sem mutações (i.e., sem ganhos ou perdas de nucleotídeos) (Gateway Technology Cloning, Invitrogen).

A formação do Clone de Entrada pode ser feita a partir de várias opções: *i.* inserção do gene em Vetor de Entrada (Entry Vector) pelo uso de enzimas de restrição e ligases; *ii.* utilização de clone de entrada de uma biblioteca de cDNA (com sítios *attL* terminais); *iii.* transformação de um Clone de Expressão (com sítios *attB*) em um Clone de Entrada a partir de um vetor que contenha sítios *attP* por um reação com sistema BP Clonase; e, *iv.* Reação de um produto de PCR com sítios terminais *attB* (formado a partir de “primers” com seqüências *attB* de 25pb + 4G terminais) com um vetor doador que contenha sítios *attP*.

Para tanto, desenhamos novo conjunto de “primers”, contendo sítios *attB*. As seqüências estão descritas, a seguir. Em itálico, 4 G na porção terminal; em negrito, sítios *attB* para recombinação sítio-específica; sublinhado, seqüência do gene PMA1: “primers GWLSR e GWSSR, para fragmentos LDh e LCD $\Delta$ h (com e sem “hinge domain, respectivamente); em itálico e azul, códon de terminação.

1.N-terminal(GWNSF):

5'GGGG**ACAAGTTTGTACAAAAAGCAGGCTTC**TTGGCTAAGAAACAAGCCATTG 3'

**sítio *attB*1**

PMA1



2.C-terminal (GWCLSR):

5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAATAGCAGATAGACC 3'

**sítio attB2**

PMA1

3.C-terminal $\Delta$ h (GWCSSR):

5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAGTTTGAGCAGTATC 3'

**sítio attB2**

PMA1

Uma vez os fragmentos amplificados por PCR, estes foram eluídos do gel de agarose (conforme já descrito anteriormente) e, 2  $\mu$ l destes foram incubados com 1  $\mu$ l (70 ng) do Vetor Doador (pDONR-201) contendo o sítio *attP* para recombinação, em presença de 1  $\mu$ l mix BP Clonase, Tampão TE (10 mM Tris-Cl, pH 7,5; 1 mM EDTA) q.s.p 5  $\mu$ l final por 16h a 25°C (caracterizando a reação BP).

Após as 16h, as misturas foram utilizadas para transformar por eletroporação células *E. coli* DH5, e estas foram selecionadas por resistência a canamicina, digestão com enzima de restrição BanII e PCR de colônias recombinantes. Os clones positivos, Vetores de Entrada (pENTR/LCDh e pENTR/LCD $\Delta$ h) foram submetidos a extração de DNA plasmidial, para serem utilizados na reação de LR: 1  $\mu$ l de DNA plasmidial foi incubado com 1  $\mu$ l (70ng) do Vetor de Destino (pDEST-17; Figura 15) contendo o sítio *attL* para recombinação, em presença de 1  $\mu$ l mix LR Clonase, Tampão TE q.s.p 5  $\mu$ l final (Tris-Cl 10 mM, pH 7,5; EDTA 1 mM) por 16h a 25°C. Passadas as 16h, então, o DNA foi utilizado para transformar células DH5 (por eletroporação).

### **3.6 Transformação de *Escherichia coli***

As diferentes cepas de *E. coli* foram transformadas utilizando-se os métodos de eletroporação.

Quarenta mililitros de células eletrocompetentes de *E. coli* XL1-Blue preparadas de acordo com Sambrook *et al* (1986), foram misturadas a 1,5 ml de pGEM-T/LCDh e pGEM-T/LCD $\Delta$ h , então transformadas no eletroporador Gene-Pulser II (Bio Rad) a 25 mF, 200 W e 2,5 KV.

Imediatamente, 1 ml de meio LB, foi acrescentado ao cultivo de bactérias eletroporadas e a cultura foi crescida por 1 h a 37°C em shaker. 100 ml da cultura assim crescida foram plaqueadas em meio LB sólido com ampicilina (100 mg/ml), IPTG (10 ml de uma solução 200 mg/ml distribuído sobre a placa) e X-Gal (40 ml de uma solução 20 mg/ml distribuído sobre a placa) para identificar-se os clones positivos.

Os plasmídeos pENTR/LCDh e pENTR/LCD $\Delta$ h foram submetidos ao mesmo procedimento, contudo foram utilizados 5 ml do meio reacional da ligação e plaqueados 200 ml de LB sólido com canamicina 25 µg/ml (recombinantes pENTR), desta vez sem X-Gal e IPTG.

### **3.7 Seqüenciamento dos Clones Positivos**

As reações de sequenciamento foram realizadas no seqüenciador automático “ABI 377 Sequencer” (Applied Biosystems Perkin Elmer) de acordo com as indicações do fabricante. Foi utilizada a seguinte programação: 5 minutos a 95°C; 25 ciclos de 10 segundos a 95°C; 5 segundos a 50°C e 4 minutos a 60°C.

O sequenciamento foi efetuado no Núcleo de Estudos de Genoma Johanna Döbehreiner da Universidade Federal do Rio de Janeiro.

### **3.8 Análise das Seqüências de DNA**

Todas as análises foram realizadas utilizando o conjunto de programas BLAST (Basic Local Alignment Search Tool), disponibilizado na internet pelo NCBI (National Center for Biotechnology and Information) - <http://www.ncbi.nlm.nih.gov>. (Altschul, et al., 1997).

### **3.9 Expressão e Preparação dos Extratos Proteicos**

Uma colônia isolada da bactéria contendo o plasmídeo de interesse, crescida em meio LB acrescido de ágar a 1,5% e ampicilina a 100 µg/ml, foi inoculada em 3 ml de meio líquido LB contendo ampicilina a 100 µg/ml. Este pré-inóculo foi incubado a 37°C, sob agitação moderada, por aproximadamente 16 horas. Após este tempo, inoculou-se 100 µl da cultura em 100 ml de meio LB contendo 100 µg/ml de ampicilina. O cultivo foi incubado a 37°C sob agitação moderada até atingir uma D.O. a 600 nm de aproximadamente 0,5~0,6 (cerca de 2 horas), quando então se adicionou NaCl (concentração final: 1 mM). Antes da adição de IPTG, uma alíquota (sem indução) de 1,5 ml da cultura foi armazenada a -20°C. Alíquotas de 1 ml após 3 horas de indução foram centrifugadas a 5.000 x g por 2 minutos. O sedimento foi ressuscitado em 100 µl de solução de lise (NaH<sub>2</sub>PO<sub>4</sub> 50 mM a pH 8,0; EDTA 1 mM; β-mercaptoetanol 1 mM; fluoreto de fenilmetilsulfonila – PMSF 1 mM; azida 1 mM). Adicionou-se então 10 µl de lisozima (10 µg/µl), misturou-se e incubou-se esta solução por 30 minutos no gelo. Após este tempo, as amostras foram submetidas ao ultrassom (Branson/250) a 80 W de potência em 6 ciclos de 10 segundos. Após centrifugação a 5.000 x g por 5 minutos, o sobrenadante foi transferido para um novo tubo, sendo o sedimento ressuscitado em 100 µl de uréia 6 M.

No caso da expressão da proteína em células BL21-SI, em vez de 1 mM IPTG e meio LB fora utilizado NaCl 300 mM e meio LBON.

### **3.10 Eletroforese desnaturante em Gel de Poliacrilamida (SDS-PAGE)**

As eletroforeses em gel de poliacrilamida em presença de SDS foram realizadas segundo Laemmli (1970), utilizando géis de corrida (6,5 x 7,0 cm e 1 mm de espessura) contendo 10% de poliacrilamida em Tris-Cl 375 mM pH 8,8 e SDS 0,1%. Os géis de concentração tinham 5% de poliacrilamida em Tris-Cl pH 6,8 156,3 mM e SDS 0,1%. As corridas foram realizadas utilizando um

sistema BIO-RAD a 20 mA em tampão Tris base 25 mM, glicina 5 mM e SDS 0,1% por 80 min a temperatura ambiente.

As amostras foram diluídas em tampão de amostra, contendo Tris-Cl (31,2 mM), glicose 30% (p/v),  $\beta$ -mercaptoetanol 1,3%, SDS 1,0% e azul de bromofenol 0,01%. Os géis foram revelados por azul Coomassie R-250.

### **3.11 Purificação das LCDs recombinantes em coluna de afinidade (Ni-NTA)**

Uma colônia isolada da bactéria contendo o plasmídeo de interesse, crescida em meio LBON acrescido de ágar a 1,5% e ampicilina a 100  $\mu$ g/ml, foi inoculada em 3 ml de meio líquido LBON contendo ampicilina a 100  $\mu$ g/ml. Este pré-inóculo foi incubado a 37°C, sob agitação moderada por aproximadamente 16 horas. Após este tempo, esta cultura foi usada como inóculo em 500 ml de meio LBON contendo 100  $\mu$ g/ml de ampicilina. O cultivo foi incubado a 37°C sob agitação moderada até atingir uma D.O.a a 600 nm de aproximadamente 0,5~0,6 (cerca de 2 horas) quando então se adicionou NaCl (concentração final: 300  $\mu$ M). Antes da adição de NaCl, uma alíquota (sem indução) de 1,5 ml da cultura foi armazenada a -20°C. A indução prosseguiu por 3 horas, também a 37°C e sob agitação de 200 rpm; após este período as células foram centrifugadas a 10.000 x g por 10 minutos a 4°C. Depois foram ressuspensas em 20 ml de tampão de lise (NaH<sub>2</sub>PO<sub>4</sub>- 50 mM a pH 8,0, EDTA 1 mM, beta-mercaptoetanol 1 mM, Fluoreto de fenilmetilsulfonila – PMSF 1 mM, azida 1 mM, 1X PIC – coquetel de inibidores de proteases). Adicionou-se então 100  $\mu$ l de lisozima (10  $\mu$ g/ml), misturou-se, e incubou-se esta solução por 30 minutos no gelo. Após este tempo, as amostras foram submetidas ao ultrassom (Branson/250) a 80 W de potência em 6 ciclos de 10 segundos. A extração da proteína de interesse foi realizada seguindo um protocolo de desnaturação em presença de 6M de Uréia. O precipitado do material lisado foi ressuspenso em 10 ml de tampão de extração (uréia 6 M, NaH<sub>2</sub>PO<sub>4</sub>-25 mM a pH 7,5, sulfato de amônio 200 mM, beta-mercaptoetano 10 mM). Ciclos de congelamento e descongelamento mantiveram a alça maior na porção solúvel. Após uma etapa

de centrifugação de 10 minutos a  $10.000 \times g$ , a  $4^{\circ}\text{C}$ , o sobrenadante pôde ser aplicado em uma coluna de cromatografia de afinidade à níquel (Ni-NTA).

Para purificação da proteína recombinante de interesse utilizou-se a resina “Chelating Sepharose Fast Flow” (Pharmacia). Colocou-se 3 ml da resina, previamente ativada com níquel, numa coluna de cromatografia. Inicialmente, 30 ml de tampão de extração foram usados para equilibrar a coluna num fluxo de 1 ml/min. A coluna foi então carregada com 10 ml da porção solúvel em uréia. Adicionalmente, lavou-se a coluna com 50 ml de tampão de extração. A proteína recombinante de interesse foi então eluída com a passagem de 20 ml de tampão de extração com 20 mM de imidazol. Esta eluição foi coletada em frações em 0,5 ml para identificação em gel de poliacrilamida.

A eliminação da uréia, visando a renaturação da proteína, foi iniciada por diálise da proteína contra o tampão D ( $\text{NaH}_2\text{PO}_4$  25 mM a pH 7,5, sulfato de amônio 200 mM,  $\beta$ -mercaptoetanol 10 mM, DTT 3 mM) contendo concentrações decrescentes – de 4 a 0 M – de uréia, para garantir a completa remoção da uréia; a diálise com o tampão sem uréia foi repetida, realizada por aproximadamente 12 horas, usando membrana de poro de exclusão de 12.000-14.000 Da.

### **3.12 Western Blot**

Após SDS-PAGE, as proteínas do gel de poliacrilamida foram transferidas eletroforéticamente (SemiDry, Bio-Rad) para uma membrana de PVDF (Imobilon-P, Millipore), empregando-se 150 mA por 60 minutos, utilizando o tampão CAPS (10 mM CAPS e 10% metanol). Após o término da transferência a membrana foi lavada com 25 mL de solução TBS (Tris-HCl 2 mM a pH 7,5, NaCl 50 mM) por 10 minutos. A membrana foi então bloqueada com 5% (p/v) de leite em pó desnatado (dissolvido em TBS contendo azida sódica 0,01%) por 24 horas. Após este período, a membrana foi lavada três vezes por 5 minutos com 25 mL de TBS, duas vezes por 5 minutos com 25 mL de TTBS (TBS com 0,05% de Tween-20) e finalmente incubada por 24 horas com 10 mL de anticorpo monoclonal anti-HisTag (Amersham Biosciences/GE

Healthcare), o qual reconhece a cauda de seis histidinas que os vetores de expressão adicionam à proteína, diluído 1:1000 em tampão de anticorpos (leite desnatado 3% (p/v) em TTBS com 0,01% de azida sódica). A membrana foi então lavada três vezes com 25 mL de TTBS por 10 minutos e incubada por 2 horas com 10 mL do anticorpo secundário anticamundongo conjugado com a peroxidase. A revelação foi conduzida com o sistema ECL (Amersham Biosciences/GE Helthcare) de detecção, seguindo as instruções dos fabricantes.

### **3.13 Espalhamento de raios-X em baixos ângulos (SAXS)**

Experimentos de SAXS foram realizados no Laboratório Nacional de Luz Sincrotron (LNLS), Campinas, Brasil, com a assistência técnica do Dr. Cristiano Oliveira e supervisão da Dra. Iris Torriani. As aquisições foram feitas em soluções de 4 mg / ml de LCD/SERCA em tampão S ( $\text{NaH}_2\text{PO}_4$  25 mM a pH 7,5; sulfato de amônio 200 mM,  $\beta$ -mercaptoetanol 10 mM, DTT 200 mM) em células de 1 mm de percurso, na ausência ou presença de  $\text{MgCl}_2$  50 mM. Baixas concentrações de proteína foram testadas e os resultados obtidos foram semelhantes. As amostras foram expostas ao feixe de raios-X a 1,6 Å por 10 a 20 minutos em temperatura ambiente. A intensidade dos dados foi corrigida i.e., a aquisições coletadas em condições idênticas, mas sem proteína.

Os dados brutos obtidos são dispostos no gráfico  $I(q)$ , onde a intensidade  $I$  da radiação espalhada é plotada (geralmente em escala logarítmica) em função do ângulo de espalhamento, representado por  $q$ , onde  $q = 4 \pi \sin(\theta)/\lambda$ , e  $2\theta$  é o ângulo entre os raios transmitidos e espalhados. As informações básicas fornecidas a partir destes dados são o peso molecular, o raio de giro e a dimensão máxima. O peso molecular pode ser calculado a partir da intensidade  $I(0)$  no ângulo de espalhamento zero, que precisa ser obtida por extrapolação das intensidades medidas, pois não é acessível experimentalmente. O raio de giro pode ser calculado a partir da curva de Guinier, uma aproximação linear do espalhamento medido nos ângulos mais baixos, ou a partir de um ajuste da curva de espalhamento inteiro. Para isso, os dados experimentais são transformados em uma distribuição  $P(r)$  das

distâncias  $r$  entre os átomos da proteína. O raio de giro é uma média ponderada destas distâncias, enquanto a dimensão máxima é a maior distância interatômica encontrada na curva.

### 3.14 Mutagêneses sítio-dirigida

Oligonucleotídeos foram desenhados de forma que anelassem em regiões correspondentes do resíduo Asp660 ao resíduo Leu663 do LCD/SERCA (na tétrede Asp660-Leu661-Pro662-Leu663), sendo substituídos por resíduos de Alanina (Ala ou A), utilizando o QuikChange® Multi Site-Directed Mutagenesis Kit (Stratagene). Criamos, assim 04 (quatro) peptídeos LCD/SERCA mutantes, dos quais dois são mutantes simples (i.e., alteração em somente um resíduo de aminoácido – sendo eles: D660A e L663A, respectivamente); um com dupla mutação (D660A, L663A), e um mutante para os quatro resíduos (D660A-L661A-P662A-L663A), de acordo com os primers a saber:

1. Mutante1 (D600A):

Senso: 5'GAGTTTGACGCCCTGCCCTGGCCGAG3'

Antisenso: 5'CTCGGCCAGGGCAGGGCGTCAAACCTG3'

2. Mutante 2 (L663A):

Senso: 5'GAGTTTGACGACCTGCCCGCCGCCGAG3'

Antisenso: 5'CTCGGCGGCGGGCAGGTTCGTCAAACTC3'

3. Mutante 3 (D660A, L663A):

Senso: 5'GAGTTTGACGCCCTGCCCGCCGCCGAG3'

Antisenso: 5'CTCGGCGGCGGGCAGGGCGTCAAACCTC3'

4. Mutante 4 (D660A-L661A-P662A-L663A):

Senso: 5'GAGTTTGACGCCCGCCGCCGCCGAG3'

Antisenso: 5'CTCGGCGGCGGGCAGGGCGTCAAACCTC3'

Os plasmídeos mutantes resultantes foram inicialmente utilizados na transformação de *E. coli* XL1-Blue e, posteriormente em *E. coli* M-15[pREP-4] (Qiagen).

A confirmação dos clones contendo os plasmídeos mutantes foi realizada através de PCR das colônias recombinantes, onde colônias isoladas crescidas em LB acrescido de agar 1,5% (w/v) e ampicilina a 50 µg/ml, foram colocadas em um tubo contendo 100 µl de água destilada autoclavada. Adicionalmente, os plasmídeos recombinantes contendo as mutações pretendidas foram submetidos a sequenciamento de DNA (dados não mostrados), onde confirmamos o sucesso desse processo de mutagêneses sítio-dirigida.

### **3.15 Identificação de ORFs relacionadas ao sistema de transportadores no genoma da *Gluconacetobacter diazotrophicus* PAL5**

A maioria dos projetos de seqüenciamento de genoma completo é geralmente dividido em três fases: o seqüenciamento de “shotgun” (WGS), o seqüenciamento de finalização e a anotação. Conforme os “reads” são incluídos na montagem do genoma, os “contigs” mudam de uma forma muito dinâmica. Apesar da estratégia do WGS ser um processo de seqüenciamento aleatório, a maior parte do genoma é seqüenciado, sobrando apenas algumas regiões problemáticas, difíceis de serem clonadas ou seqüenciadas, criando assim regiões não seqüenciadas no genoma, ou buracos. Desse modo, muito freqüentemente, a anotação do genoma é realizada quando terminadas as duas primeiras etapas.

Todo o processo de anotação do genoma do *Gluconacetobacter diazotrophicus* PAL5, foi executado no sistema Bionotes (Lemos, Seibel & Casanova, 2003; Lemos, 2004), pelos membros integrantes do Projeto Riogene. Foi realizado um processo de anotação manual, onde o primeiro passo consistiu na validação das ORFs, pois o programa Glimmer (Delcher et al., 1999) utilizado para predição de ORFs é considerado um programa que gera um grande número de falsos positivos, onde diversas ORFs podem apresentar sobreposições completas.



Nestas situações, a seleção das ORFs corretas foi baseada nos resultados de Blast (Altschul et al., 1990; Altschul, 1998), onde dentre as ORFs sobrepostas, as que apresentaram melhores alinhamentos em relação aos bancos de dados foram validadas. Dentre as variantes de Blast (WU-BLAST, 2006) utilizadas no processo de anotação, o Blastp foi utilizado como ferramenta inicial, pois compara seqüências de aminoácidos, contra seqüências de proteínas depositadas nos bancos de dados públicos, realizando uma análise menos restrigente que outras ferramentas, tais como o Blastn (utilizado para comparação de seqüências de nucleotídeos), devido ao fato de que os mais variados organismos podem utilizar códons diferenciados, para codificar o mesmo aminoácido.

A inferência de prováveis funções das proteínas codificadas pelo genoma de *G. diazotrophicus* PAL5, com base nos resultados de Blastp obedeceu os seguintes critérios: i) o alinhamento deve ter e-value inferior a  $10^{-5}$ ; ii) porcentagens de identidade superiores a 30%; e iii) alinhamentos de 60% ou mais da seqüência da ORF foram considerados suficientes para caracterizar homologia com seqüências de proteínas depositadas em bancos de dados. Quando não houve detecção de sobreposição de seqüências de ORFs, e mesmo assim o resultado de Blast indicou uma similaridade muito baixa ou nenhuma em relação aos bancos de dados, estas seqüências foram classificadas como seqüências de proteínas hipotéticas.

Posteriormente, realizou-se uma análise na base de dados PFAM (Bateman, 2004), que apresenta uma coleção de famílias e domínios protéicos. O PFAM possui seqüências conservadas de cerca de 6200 famílias protéicas, das quais aproximadamente 75% das proteínas encontradas no SWISS-PROT (Boeckmann et al., 2003), possuem pelo menos um domínio do PFAM, e cada resultado relatado no PFAM acompanha um comentário sobre o domínio encontrado, facilitando a identificação da possível função da ORF. Após a predição da possível função das ORFs, estas foram validadas e as suas respectivas categorias e subcategorias funcionais identificadas, assim como o nome do gene, número de EC (Enzyme Category), número de COG (Clusters of Orthologous Groups), organismos relacionados, e comentários adicionais dos pesquisadores acerca da ORF.

Adicionalmente, na tentativa de determinar prováveis funções para ORFs de algumas proteínas hipotéticas, foram analisados os alinhamentos de seqüências do *G. diazotrophicus* PAL5 (Blastp), que apresentaram alinhamentos com e-values inferiores a  $10^{-3}$ , e alinhavam com domínios protéicos conservados depositados no PFAM. As proteínas hipotéticas com valores de e-value superiores a  $10^{-3}$  nos seus alinhamentos continuaram a receber esta classificação.

A maioria, se não todas, as proteínas que estão localizadas na membrana externa das bactérias, por meio de uma via secretora, contem um peptídeo de sinal. O programa SignalP analisa a seqüência de aminoácido de uma ORF para um tal peptídeo de sinal, e os primeiros 60 aminoácidos N-terminais de cada ORF foram analisados por SignalP, usando a base de dados de bactérias Gram negativas.

Outro programa utilizado foi o Psort, que determina a localização de uma proteína com base na sua seqüência de sinal, no reconhecimento de segmentos transmembranares e na análise de sua composição de aminoácidos. Considera-se que um produto de uma ORF é uma proteína de membrana interna/externa se os dados do resultado prevêm com um valor de segurança de 0,5 ou maior.

Adicionalmente, realizou-se análises de alinhamentos comparativas com bancos de dados para proteínas de transportes, que foram gentilmente disponibilizadas pelos grupos: i) Dr. Milton Saier Jr. (University of California, San Diego, UCSD/USA) para utilizar o banco de dados do TCDB (Saier, et al., 2006; Transporter Classification System – também disponível em: <http://www.tcdb.org/>) ; ii) Drs. Ian T. Paulsen e Ren Qun do Insititute J.Craig Venter e The Institute for Genomic Research – TIGR para anotação automática e comparativa do sistema de transporte de *G. diazotrophicus* com outros genomas seqüenciados e disponíveis no TransportDB – Genomic Comparisons of Membrane Transport System: <http://www.membranetransport.org/>; e, iii) Dr. Yves Quentin da Universite´ Paul SABATIER-CNRS/France para anotação automática no ABCdb (Archaeal and Bacterial ABC transporter database; disponível em : <http://www-abcdb.biotoul.fr/>; Fichant et al., 2006).

## **RESULTADOS, DISCUSSÃO E CONCLUSÕES**

#### **IV. RESULTADOS, DISCUSSÃO E CONCLUSÃO**

---

Nesta seção serão apresentados e discutidos os resultados que foram obtidos durante o doutoramento. Os resultados serão apresentados tanto em forma livre quanto em forma de manuscritos.

Os resultados serão divididos em duas seções distintas. A primeira parte, denominada: “*P-ATPases*”, compreenderá os estudos e a caracterização molecular do estado agregacional de P-ATPases, tanto para a H<sup>+</sup>-ATPase de membrana plasmática de levedura quanto para a SERCA de coelho.

Na segunda parte, intitulada: “*Genoma*”, será apresentada a análise da anotação dos Sistemas de Transporte da *G. diazotrophicus*.

## ***P-ATPases***

**Manuscrito 1:** “*Self-association of isolated large cytoplasmic domain of plasma membrane H<sup>+</sup>-ATPase from Saccharomyces cerevisiae: Role of the phosphorylation domain in a general dimeric model for P-ATPases*”.

Muito embora a elucidação da estrutura tridimensional (3D) da SERCA tivesse sido obtida, ainda era controverso se as unidades funcionais das P-ATPases nas membranas existiriam como estruturas monoméricas ou oligoméricas.

Neste trabalho, demonstramos haver uma interação específica entre duas cadeias de LCD, e a existência de uma região de 150 resíduos de aminoácidos localizada na região C-terminal desse domínio citoplasmático maior como responsável nesse processo.

Em adição, realizamos uma série de estudos *in silico*, pelo alinhamento das seqüências de várias P-ATPases, buscando identificar se a oligomerização é algo realmente conservado entre as mais diversas P-ATPases ou somente existente (compartilhado) entre alguns subgrupos dessas enzimas. Com a análise preliminar por alinhamento das seqüências primárias e a modelagem comparativa de diferentes P-ATPases foi possível observar a presença de muitos resíduos de aminoácidos com caráter hidrofóbico ou carregado, expostos ao solvente (na superfície da enzima, utilizando como modelo a SERCA), o que poderia favorecer interações entre cadeias. Em contraposição, a existência de alças ("*loops*"), devido a inserções e/ou deleções de aminoácidos, poderia ser responsáveis pela criação de interações específicas entre cadeias homólogas, especialmente em membranas biológicas onde diferentes P-ATPases são encontradas, o que evitaria a formação de heterodímeros não-funcionais.

Particularmente, para a SERCA1a, análises de cristais tubulares formados em membranas nativas revelam uma estrutura dimérica, com as unidades polarizadas em direções opostas, associadas como uma fita adornando de uma forma espiralada em torno dos túbulos, onde uma ATPase

monômerica de uma fita está em contato com o outro monômero através da região hidrofílica citoplasmática. Estas observações estão de acordo com nossos resultados que sugerem o envolvimento do domínio de fosforilação (Domínio P, encontrado justamente na região C-terminal do LCD) na dimerização, tanto para a SERCA1a quanto para a PMA1, como também não interferem nos movimentos propostos dos domínios de ligação de nucleotídeo (N) e de atuação (compreendida na alça citoplasmática menor - domínio A ) durante o mecanismo de transporte, corroborando a idéia que essa interação entre duas cadeias de LCD seja de relevância funcional para grande parte de P-ATPases na membrana nativa.

# Self-association of isolated large cytoplasmic domain of plasma membrane H<sup>+</sup>-ATPase from *Saccharomyces cerevisiae*: Role of the phosphorylation domain in a general dimeric model for P-ATPases

W.I. Almeida\*, O.B. Martins, P.C. Carvalho-Alves

*Instituto de Bioquímica Médica, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-590, Brazil*

Received 10 March 2006; received in revised form 21 August 2006; accepted 22 August 2006

Available online 1 September 2006

## Abstract

Large cytoplasmic domain (LCD) plasma membrane H<sup>+</sup>-ATPase from *S. cerevisiae* was expressed as two fusion polypeptides in *E. coli*: a DNA sequence coding for Leu353–Ileu674 (LCDh), comprising both nucleotide (N) and phosphorylation (P) domains, and a DNA sequence coding for Leu353–Thr543 (LCDΔh, lacking the C-terminus of P domain), were inserted in expression vectors pDEST-17, yielding the respective recombinant plasmids. Overexpressed fusion polypeptides were solubilized with 6 M urea and purified on affinity columns, and urea was removed by dialysis. Their predicted secondary structure contents were confirmed by CD spectra. In addition, both recombinant polypeptides exhibited high-affinity 2',3'-O-(2,4,6-trinitrophenyl)adenosine-5'-triphosphate (TNP-ATP) binding ( $K_d=1.9 \mu\text{M}$  and  $2.9 \mu\text{M}$  for LCDh and LCDΔh, respectively), suggesting that they have native-like folding. The gel filtration profile (HPLC) of purified LCDh showed two main peaks, with molecular weights of 95 kDa and 39 kDa, compatible with dimeric and monomeric forms, respectively. However, a single elution peak was observed for purified LCDΔh, with an estimated molecular weight of 29 kDa, as expected for a monomer. Together, these data suggest that LCDh exist in monomer–dimer equilibrium, and that the C-terminus of P domain is necessary for self-association. We propose that such association is due to interaction between vicinal P domains, which may be of functional relevance for H<sup>+</sup>-ATPase in native membranes. We discuss a general dimeric model for P-ATPases with interacting P domains, based on published crystallography and cryo-electron microscopy evidence.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** P-type; ATPase; Oligomerization; Yeast; Protein structure; Self-association

## 1. Introduction

P-type ATPases constitute an important ubiquitous family of enzymes that transport ions or charged molecules across biological membranes, using energy released from ATP hydrolysis [1,2]. Since the first member of this family was

discovered about five decades ago [3], the great number of studies dealing with P-ATPases has led to significant progress in knowledge of these enzymes' functional features, such that a general kinetic model for active ionic transport has been developed (reviewed in [4–6]). Although the recent determination of sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase (SERCA) tertiary structure [7,8] provides the framework for interpreting kinetic and mutagenesis data in structural terms, but left many fundamental questions remain unanswered [9]. Among them, protein oligomerization and the exact functional unit resident in biological membranes are still controversial subjects.

Dimeric models that explain kinetic behavior of these enzymes are supported by several physicochemical and immunochemical studies dealing with different P-ATPases [10,11]. Even though monomeric P-ATPase can carry out active

*Abbreviations:* ATP, adenosine triphosphate; ATPase, adenosine-triphosphatase; bp, base pairs; CD, circular dichroism; DTT, dithiothreitol; HPLC, high performance liquid chromatography; LCD, Large cytoplasmic domain; PCR, polymerase chain reaction; PMSF, phenylmethylsulfonyl fluoride; SERCA, sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; TNP-ATP, 2',3'-O-(2,4,6-trinitrophenyl)adenosine-5'-triphosphate; MOPS, 3-N-morpholinopropanesulfonic acid

\* Corresponding author. Tel./fax: +55 21 2562 6789.

E-mail address: [welington@bioqmed.ufrj.br](mailto:welington@bioqmed.ufrj.br) (W.I. Almeida).

ion transport [12,13], much evidence points to functional multimeric states for several P-ATPases in their native membranes. For example, kinetic studies suggest that oligomeric states are important for transport [14,15], and that dynamic changes in protein–protein interaction occur during catalytic cycle [16], explaining both cooperative kinetic behavior and enzymatic stability [17–19]. In addition, P-ATPase oligomeric structures have been identified by using proteins in native membranes or by proteins reconstituted in phospholipids/detergent micelles, based on different experimental approaches, including X-rays crystallography or microscopy of tubular crystals [20–28], spectroscopic data [19,29,30], radiation inactivation [31–33], immunochemical detection [18], low-angle laser light scattering photometry [34], co-precipitation [17,35], thermal denaturation [36] and others [37]. However, in some cases, the oligomeric complexes observed were conceivably produced by changes in membrane composition or protein conformation due to the choice of purification method or the experimental conditions, and may not reflect real subunit interactions in native forms. Detailed quaternary structure of these enzymes is still unknown and is a subject of much debate.

The Large Cytoplasmic Domain (LCD) of P-ATPases represents 40–45% of the polypeptide  $\alpha$ -chain, and contains most of the enzyme's hydrophilic portion, comprising both nucleotide-binding (N domain) and phosphorylation (P domain) sites. Recently, elucidation of SERCA three-dimensional structures has shown that the P domain assumes the Rossman fold motif, formed by the LCD N-terminal beta-sheet containing Asp 351 (the residue that is phosphorylated by ATP) surrounded by 50 C-terminal amino acids, before re-entering the membrane [7,8], as suggested by homology modeling with haloacid dehalogenase [38].

Evidence suggesting that LCD plays a critical role in protein oligomerization came first from studies with  $\text{Na}^+, \text{K}^+$ -ATPase [17], and it has been proposed that a 150 amino-acid segment at LCD C-terminus is essential for association of  $\alpha$ -subunits [35]. Recently, we have shown that isolated LCD from SERCA (LCD/SERCA), heterologously expressed in *E. coli* as a fusion protein, is unable to form dimers when the equivalent C-terminal portion was removed by proteinase-K treatment [39]. The proteolyzed segment was the P domain and the monomeric segment was identified as the N domain [7]. Thus, expressing and characterizing LCD from other P-ATPase family members is of interest as to address the question of whether P domain is important for protein–protein interaction in this ATPase class as a general rule. Accordingly, we describe cloning, heterologous expression, purification, and structural characterization of both LCDh, a fragment comprising the entire N and P domains of plasma membrane  $\text{H}^+$ -ATPase from *S. cerevisiae* (*pma1*); and LCD $\Delta$ h, a truncated LCD fragment lacking the P domain C-terminus.

Our data indicate that P domain is necessary for dimerization of LCDh, as previously shown for LCD/SERCA [39], supporting the hypothesis that such interaction may be of functional relevance for P-ATPases in native membranes. We discuss our data in terms of a general dimeric model for P-ATPases based on crystallography and cryo-electron microscopy.

## 2. Materials and methods

### 2.1. Subcloning of LCDh and LCD $\Delta$ h and construction of expression vectors

The segments corresponding to plasma membrane  $\text{H}^+$ -ATPase Large Cytoplasmic Domain with (LCDh) and without (LCD $\Delta$ h) its C-terminus were amplified by PCR with *Pfx*-polymerase, using *pma1* gene as template (kindly supplied by Dr. Carolyn Slayman, Yale University, USA), and three oligonucleotides (primers): 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTGGCTAAGAAACAAGCCATTG-3' and either 5'-GGGGACCACTTTGTACAAGAAAGCTG GGTCCCTAAATAGCAGATAGACC-3' for LCDh or 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCCCTAAGTTTGAGCAGTATC-3' for LCD $\Delta$ h. These primers were designed on the basis of GenBank entry 1168544, *pma1* gene [40] with the added site-specific recombination (underlines) of phage lambda [41,42] (Invitrogen).

PCR products were subcloned into *attP* sites of pDONR-201 (Invitrogen), yielding either entry clone LCDh (pENTR/LCDh) or entry clone LCD $\Delta$ h (pENTR/LCD $\Delta$ h). Direction of insertion was checked by restriction enzyme digestion. DNAs of entry clones were used to transfer LCDh and LCD $\Delta$ h into destination vector pDEST-17 (which contained a phage T7 promoter; Invitrogen), yielding pDEST-17/LCDh and pDEST-17/LCD $\Delta$ h, respectively. This vector introduces an amino-terminal hexahistidine tag suitable for  $\text{Ni}^{2+}$  affinity purification. Recombinant plasmids were transformed into *E. coli* DH5 $\alpha$  for DNA cloning and sequencing and BL21-SI (salt inducible) for protein expression [43] (Invitrogen). Gene sequence was confirmed by automated DNA sequence analysis.

### 2.2. Overexpression, purification and renaturation procedures

Engineered pDEST-17 was used to transform BL21-SI strain. *E. coli* transformants were selected on LB agar without sodium chloride, dubbed LBON [44] which contain ampicillin (100  $\mu\text{g}/\text{ml}$ ). An overnight culture of 5 ml was prepared and used to inoculate 500 ml of LBON medium containing ampicillin (100  $\mu\text{g}/\text{ml}$ ). Expression was induced by adding 0.3 M NaCl at an optical density of 0.6–0.7 at 600 nm. After further growth in 150 min, bacteria were collected, suspended in 10 ml of lysis buffer (50 mM sodium phosphate, 1 mM PMSF, pH 7.0), and sonicated at 80 W using 6 pulses of 30 s (Branson/250) on ice. The lysate was then centrifuged, and the pellet was resuspended in 10 ml of extraction buffer (6 M urea, 25 mM sodium phosphate, 200 mM ammonium sulfate, pH 7.5 and 10 mM  $\beta$ -mercaptoethanol) and centrifuged (5,000 $\times g$  for 15 min at 4  $^{\circ}\text{C}$ ). The supernatant was collected (10 ml containing 10 mg protein) and used to load the affinity column.

To purify the engineered polypeptides LCDh and LCD $\Delta$ h, we used 3 ml Ni-NTA resin pre-equilibrated with extraction buffer. The 6 M urea extract was applied and the column was washed twice with 10 ml of extraction buffer. Fusion polypeptides were eluted with 10 ml of extraction buffer supplemented with 25 mM imidazole for LCDh or 120 mM imidazole for LCD $\Delta$ h. All protein determinations were performed in accordance with Bradford [45]. Aliquots containing 10–25  $\mu\text{g}$  of LCD polypeptides were used for SDS-PAGE. Protein expressions and purification were examined on 10% SDS-PAGE [46]. Protein bands were observed by staining the gels with Coomassie Brilliant Blue. For Western blotting, the polypeptides were electro-transferred from gels to PVDF membranes (Immobilon-P, Millipore) using 150 mA for 60 min, using the mouse anti-hexahistidine antibody (Sigma) diluted 1/4000. Antibodies were detected using the ECF kit (Amersham, UK).

Refolding of engineered proteins was initiated by a progressive elimination of urea. For this purpose, LCDh and LCD $\Delta$ h polypeptides were dialyzed for 12 h at 4  $^{\circ}\text{C}$  against decreasing concentrations of urea (4 M, 2 M or no urea) and 25 mM sodium phosphate, pH 7.5, 200 mM ammonium sulfate, 10 mM  $\beta$ -mercaptoethanol, 3 mM DTT. After dialysis the protein was stored on ice. Alternatively, the protein was concentrated 10-fold using an Ultrafree-15 centrifugal filter device (Millipore, Bedford, MA, USA) and the buffer was changed to 20 mM MOPS at the same pH and in the presence of the same salt and reducing agents.

### 2.3. Circular dichroism

CD measurements were carried out on a J-715 spectropolarimeter (Jasco Corp., Japan), using a 0.1-cm path-length quartz cell. Spectra were obtained in



25 mM sodium phosphate buffer containing 100 mM ammonium sulfate, 10 mM  $\beta$ -mercaptoethanol, and 0.3 mM DTT at pH 7.0, and spectra from buffer alone were subtracted. Three scans from 190 to 260 nm were performed to obtain each spectrum. The units are expressed as molar ellipticity per residue ( $\text{degree cm}^2 \text{dmol}^{-1}$ ).

Three different algorithms were used to deconvolute CD spectra of proteins using CDPro package [47]. The algorithms used were Selcon, containing 33 proteins in the data base, Contin with 16 proteins in the data base; and CDSSTR. The values found were compared with predicted contents for LCDh and LCD $\Delta$ h polypeptides, using SOPMA [48].

#### 2.4. TNP-ATP fluorescence measurements

TNP-ATP fluorescence measurements were performed at room temperature in a medium containing 20 mM MOPS pH 7.25, 100 mM KCl, and 0.2 M ammonium sulfate, in a  $1 \times 1$  cm fluorescence cuvette under continuous stirring using a photocounting spectrofluorimeter. The excitation wavelength was 410 nm (bandwidth=5 nm) and the emission wavelength was 540 nm (bandwidth=20 nm). Nucleotide binding was measured from the difference of fluorescence intensity in the presence and absence of  $1 \mu\text{M}$  of either LCDh or LCD $\Delta$ h after correcting for inner filter effect.

#### 2.5. Gel filtration

Gel filtration experiments were carried out by HPLC (Shimadzu, Japan) using a GP 100 column. The flow rate was 0.5 ml/min with eluting buffer (25 mM sodium phosphate, 100 mM ammonium sulfate, pH 7.0). The column was calibrated for molecular mass with  $\alpha$ -amylase, albumin, ovalbumin and carbonic anhydrase [49]. Protein elution was monitored by detecting fluorescence emission at 320 nm.

### 3. Results

The overall aim of our study is to investigate the role of Large Cytoplasmic Domain (LCD) of P-type ATPases in self-association, and in particular, the relevance of C-terminal residues, which compose most of the P domain. For this purpose we subcloned and overexpressed two constructions of LCD from *S. cerevisiae* plasma membrane  $\text{H}^+$ -ATPase: LCDh (from

Leu 353 to Ile 674), which contains both N and P domains; and LCD $\Delta$ h (from Leu 353 to Thr 543), which lacks the C-terminal portion of the P domain.

#### 3.1. Construction of the *E. coli* expression vectors

Recombinant plasmids encoding LCDh (about 1000 bp) and the C-terminus-deleted LCD $\Delta$ h (about 700 bp) were checked by restriction-enzyme digestion and confirmed by DNA sequencing. pENTR clones generated by BP reaction (i.e., recombinant plasmids for insertion of PCR products with *attB* into pDONR-201) were checked with *Ban*II, and recombinant pDEST-17 plasmids, generated by LR reaction were checked with *Eco*RI (data not shown).

#### 3.2. Overexpression and purification of the LCD from $\text{H}^+$ -ATPase polypeptides in *E. coli*

Expression in *E. coli* BL21-SI was induced by 0.3 M sodium chloride. The expected polypeptides bands were observed in SDS-PAGE (35,705 Da for LCDh and 20,994 Da for LCD $\Delta$ h, Fig. 1A). In both cases, after induced cells were disrupted by sonication, most of the LCD polypeptides appeared as inclusion bodies, which were then solubilized with extraction buffer containing 6 M urea. After centrifugation, to discard the remaining aggregates, the fusion polypeptides were purified on a  $\text{Ni}^{2+}$  affinity resin (Fig. 1B). Then they were refolded by stepwise dialysis and recovered in soluble form (around 3–5 mg/500 ml of cultured cells).

#### 3.3. CD spectra and binding of TNP-ATP

CD spectroscopy was employed in order to characterize LCD secondary structure (Fig. 2A and B). Table 1 summarizes secondary structure content of each construction based on CD

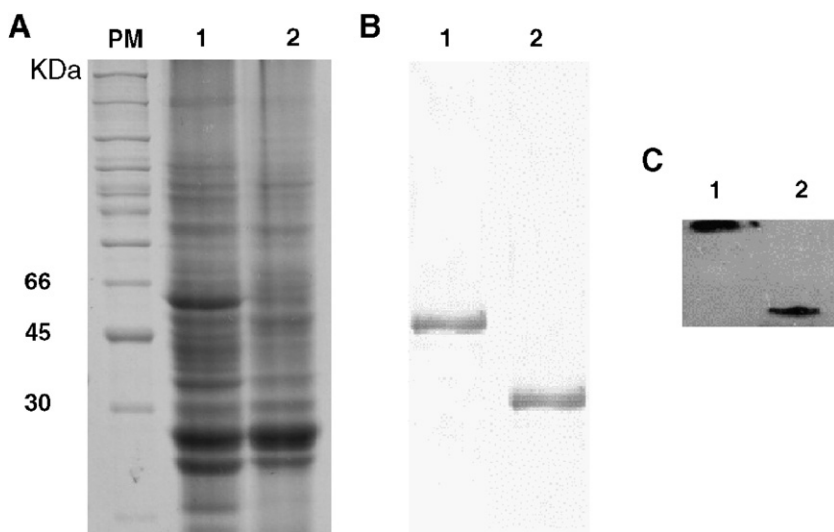


Fig. 1. Purification of LCDs from PMA1. SDS-PAGE of fractions obtained during purification stained with Coomassie Blue. PM, molecular weight standards (molecular masses in kDa, indicated at left); 1, crude extract of the induced LCDh; 2, crude extract of the induced LCD $\Delta$ h; (B) 1, affinity-column elution of purified LCD $\Delta$ h (about 21 kDa) with extraction buffer supplemented with 120 mM imidazole; 2, affinity-column elution of purified LCDh (about 36 kDa) with extraction buffer supplemented with 25 mM imidazole; (C) Western blot of LCD polypeptides with anti-hexa-His amino-tag antibody. 1, purified LCD $\Delta$ h; 2, purified LCDh.

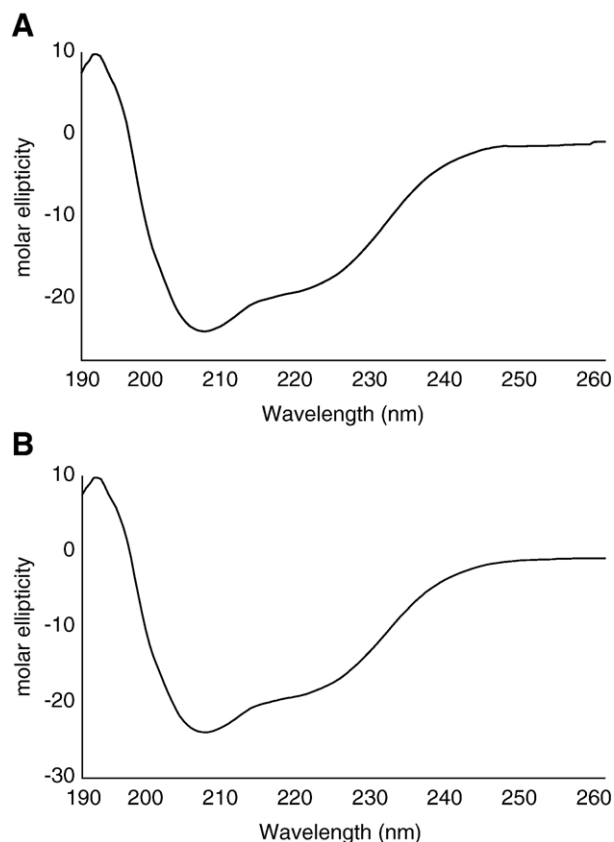


Fig. 2. CD spectra of LCDs from PMA1. (A) Spectrum of LCDh. (B) Spectrum of LCD $\Delta$ h. Both spectra were obtained in 25 mM sodium phosphate buffer containing 100 mM ammonium sulfate,  $\beta$ -mercaptoethanol and 0.3 mM DTT at pH 7.0. CD measurements were carried out on a Jasco J-715 spectropolarimeter, using an 0.1-cm path-length quartz cell. Molar ellipticity is in  $\text{deg cm}^2 \text{dmol}^{-1}$ . Resultant spectra are average of 3 scans.

spectrum deconvolution using three different algorithms. These values can be compared with hypothetical contents predicted from LCD polypeptide primary structure using SOPMA. For LCDs from PMA1, both  $\alpha$ -helix and  $\beta$ -strand contents, obtained from the three deconvolution algorithms were similar to those predicted from primary sequence (44% and 30%, respectively).

TNP-ATP binding affinity was measured by the difference in fluorescence intensity ( $F$ ) in the presence as opposed to the absence of the polypeptides (Fig. 3). LCDh and LCD $\Delta$ h bound TNP-ATP with high affinity ( $K_d=1.9\pm 0.1 \mu\text{M}$  and  $2.9\pm 1.1 \mu\text{M}$ , respectively). The apparent affinity ( $K_d$ ) and maximal fluorescence ( $F_{\text{max}}$ ) were obtained by a non-linear fitting of the

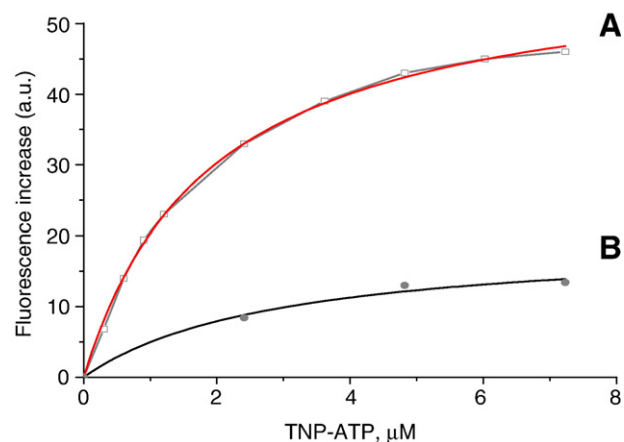


Fig. 3. Binding of TNP-ATP to LCDs from PMA1. TNP-ATP binding was determined by measuring the difference between TNP-ATP fluorescence in the presence of 1  $\mu\text{M}$  LCDh (A) or LCD $\Delta$ h (B), and in the absence of protein (not shown). Fluorescence increase ( $F$ ) is represented on an arbitrary scale. The solid lines represent a non-linear fit of data, from where the dissociation constant ( $K_d$ ) and the maximal fluorescence at saturation ( $F_{\text{max}}$ ) were calculated.

data from experiments performed in the presence of each polypeptides, assuming a hyperbolic dependence of  $F$  on TNP-ATP concentration. These  $K_d$  values are very similar to those previously described for the recombinant LCD/SERCA [39] or for the recombinant LCD/PMA [50]. The value of  $F_{\text{max}}$  calculated for LCDh was 3-fold higher than that calculated for LCD $\Delta$ h. Together these data suggest that our recombinant LCD polypeptides have a three-dimensional structure very similar to that found in the native  $\text{H}^+$ -ATPase.

#### 3.4. Association state of LCD polypeptides

The association state of each LCD polypeptide was examined by size-exclusion HPLC. Intrinsic fluorescence emission (at 320 nm) was employed to monitor protein elution. LCDh exhibited two main peaks, with relative molecular masses of 95 and 39 kDa, compatible with dimeric and monomeric forms, respectively (Fig. 4A). On the other hand, LCD $\Delta$ h showed a monodisperse profile, with a single peak at 29 kDa estimated molecular weight, as expected for monomers (Fig. 4B). The same amount of protein was applied in each case. These data indicate that the C-terminal portion of LCD from PMA1, which is absent in the LCD $\Delta$ h construction, is necessary to observe protein association. These data are in line with those obtained previously using isolated LCD/SERCA [39].

Table 1  
Secondary structure content of LCDs obtained by CD spectra deconvolution

Secondary structure motif	CONTINLL		SELCON 3		CDSSTR		Predicted*	
	LCDh	LCD $\Delta$ h	LCDh	LCD $\Delta$ h	LCDh	LCD $\Delta$ h	LCDh	LCD $\Delta$ h
Alpha-helix (%)	39	38	43	40	49.5	45	44.3	41.4
Beta-strand (%)	30	30	30	30	29.2	30	27.1	30.4
Other (%)	31	32	27	30	21.3	25	28.6	28.2

Algorithms shown along the top were used for deconvolution of Fig. 2 CD spectra (CDPro package); \* predicted refers to fold predicted for LCD peptides using SOPMA.

#### 4. Discussion

This work provides new structural evidence for the relevance of LCD in self-association of P-ATPases, and in particular, the fundamental role of the P site for oligomerization. Oligomeric arrangements of P-ATPases have been proposed for a long time, starting early observations of biphasic dependence on ATP concentration for enzyme activity [10,11]. Before publication of Toyoshima's model [7], attempts to obtain higher resolution structural data resulted in several reports describing over-expression of soluble domains of P-ATPases in *E. coli*, as an alternative approach [50–54]. Those data demonstrated that isolated LCDs are able to bind nucleotides, but crystals suitable for X-ray analysis have never been obtained, nor have the LCD oligomeric states been properly investigated. The first evidence suggesting that LCD is important for P-ATPase self-association came from the observation that  $\text{Na}^+, \text{K}^+$ -ATPase oligomeric complex depended on interactions between  $\alpha$ -subunits [17];

more specifically, on a sequence of 150 amino acids located within the LCD C-terminal portion [35]. Later, it was demonstrated that isolated LCD/SERCA, when expressed in *E. coli* as a fusion protein, has a high tendency to form dimers [39]. Interestingly, LCD/SERCA treated with proteinase K generated a monomeric fragment, similar to that produced upon digestion of native ATPase by the same protease [55]. This monomeric peptide core was identified, in the light of Toyoshima's model, as the whole nucleotide binding site (N), devoid of N- and C-terminal sequences that compose the phosphorylation site (P). Thus, LCD/SERCA self-association needs at least part of the P site, as the N site alone does not form dimers [39].

We decided to extend the study with isolated LCD, as fusion protein, to the  $\text{H}^+$ -ATPase from *S. cerevisiae*, in order to investigate whether it also self-associates and, in this case, whether such association would also be dependent on C-terminal residues of the P site. For this purpose, we have constructed and expressed vectors containing the gene sequence coding for N and P sites (LCDh), or an alternative deleted version without all P site C-terminal amino-acid residues (LCD $\Delta$ h). As expected, purified LCDh and LCD $\Delta$ h showed in SDS-PAGE mainly bands of about 36.0 and 21.0 kDa (accounting for about 97% of total protein), respectively (Fig. 1B). CD spectra deconvolution of both purified proteins, after stepwise dialysis (Fig. 2A and B) were consistent with predicted secondary structure contents (Table 1), and therefore we believe that they have folded similarly to native protein. From these data, LCDh would be quite helical (44%), in contrast to either LCD/SERCA, which contains 20–23%  $\alpha$ -helix and 21–25%  $\beta$ -sheet [39] or LCD/ $\text{Na}^+, \text{K}^+$ -ATPase, which contains 23%  $\alpha$ -helix and 27%  $\beta$ -sheet [52]. Such structural difference between LCDh and other LCDs may be attributed to deletions in the N domain, and may account for PMA1's high specificity for ATP [56], in contrast to SERCA and  $\text{Na}^+, \text{K}^+$ -ATPase, which are known to hydrolyze many different phosphorylated compounds. In addition, we showed that both LCD polypeptides bind TNP-ATP (Fig. 3) with high affinity, a property which is similar to those described for other P-type LCD recombinant constructions [39,50]. These data support the conclusion that the polypeptides have folded into a biologically relevant conformation.

However, the most important result here is the finding that refolded LCDh showed two main fluorescent elution peaks, when analyzed by size-exclusion HPLC, with calculated molecular weights of 95 and 39 kDa, compatible with dimeric and monomeric forms, respectively (Fig. 4A). These data suggest that, as previously observed for isolated LCD/SERCA [39], refolded LCDh exists in a monomer–dimer equilibrium. On the other hand, size-exclusion chromatography of LCD $\Delta$ h showed a major peak corresponding to an estimated molecular weight of 29 kDa, and no other peak at higher molecular weights, indicating that this fragment is essentially monomeric (Fig. 4B) and that the presence of the P site is essential for dimerization. These results are also in line with the studies with  $\text{Na}^+, \text{K}^+$ -ATPase described above [17,35], and altogether point to a fundamental role of P domain in self-association of P-ATPases in their native membranes. In addition, expressed LCD/ $\text{Na}^+, \text{K}^+$ -ATPase has recently been demonstrated to

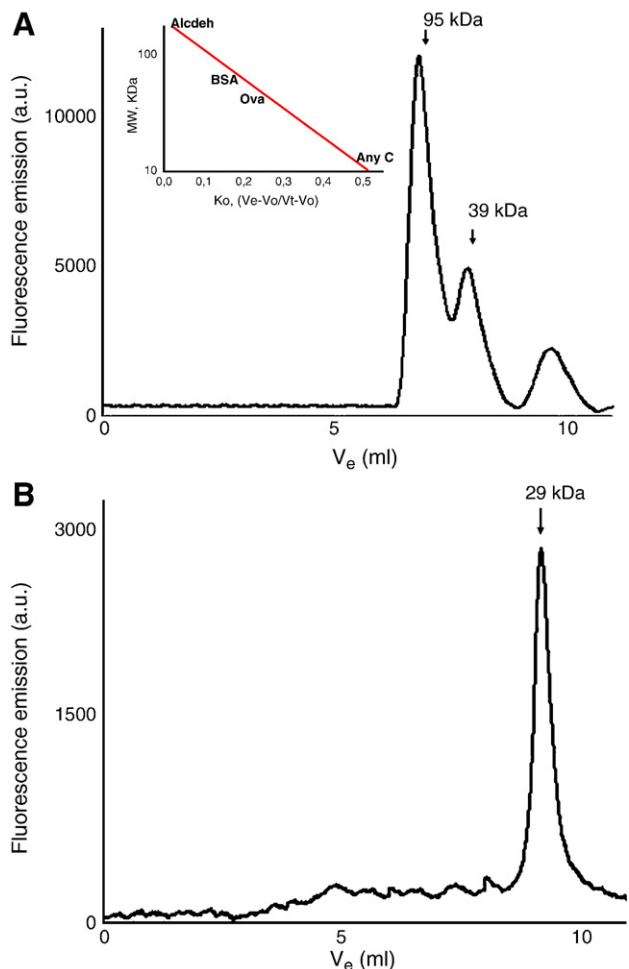


Fig. 4. Size-exclusion HPLC of LCDs. LCDh elution profile (A) shows 2 main peaks (arrows) corresponding to MWs of 95 kDa and 39 kDa, as expected for dimeric and monomeric forms, respectively; while LCD $\Delta$ h elution profile (B) shows only one peak (arrow) corresponding to MW of 29 kDa, as expected for monomeric form. Elution performed in 25 mM sodium phosphate, 100 mM ammonium sulfate buffer (pH 7.0). The apparent molecular weights of the main elution peaks were estimated by referring to a calibration curve (Inset).

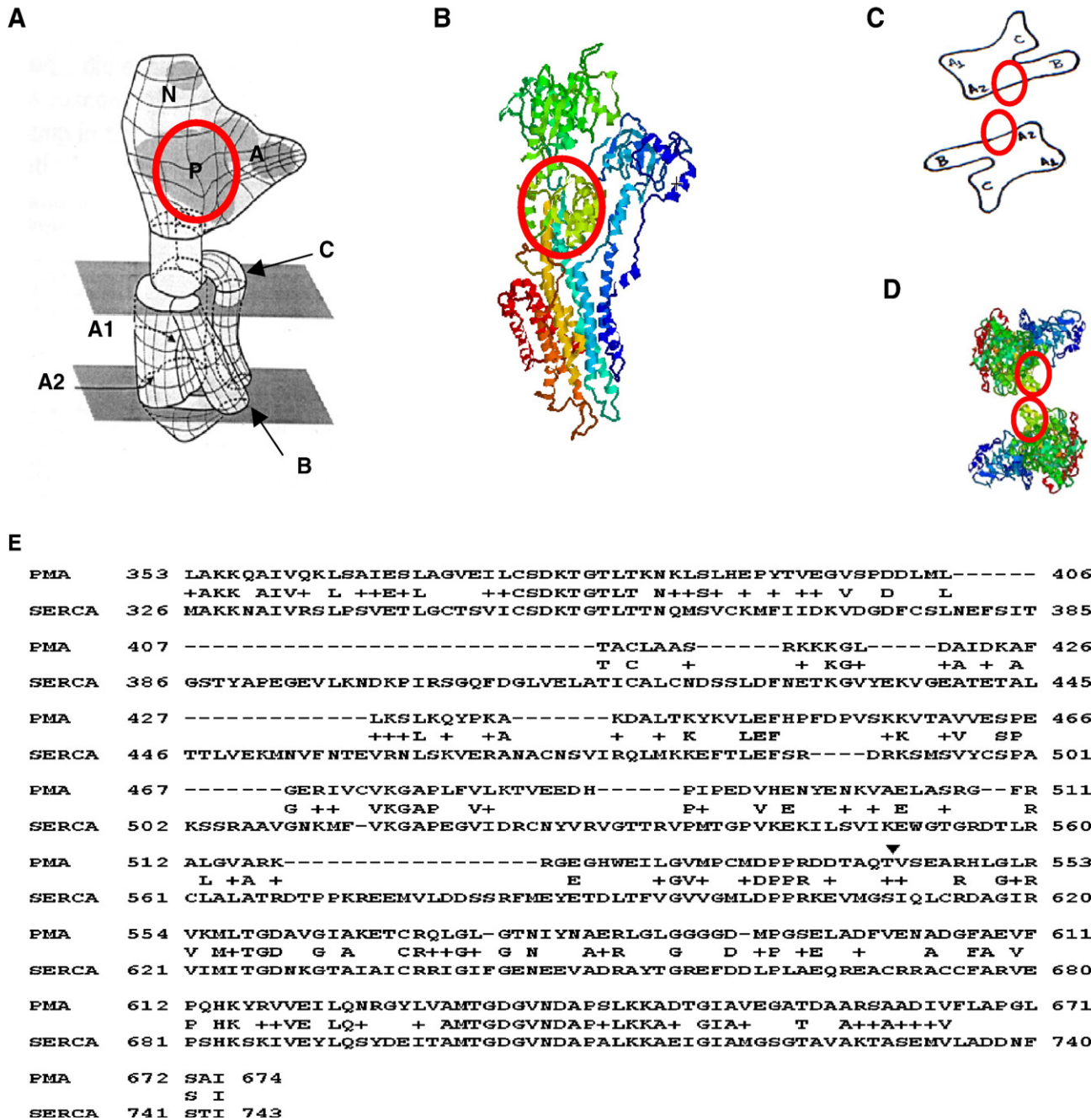


Fig. 5. Assigning the SERCA three-dimensional structure in the tubular crystal model. (A) Side view of the tubular crystal SERCA monomer model (from the opposite strand of the dimer ribbon) identifying N domain (N), A domain (A) and C-terminal portion of P domain from Asp 650 to Glu 680 (P, red circles) and the transmembrane segments, named A1, A2, B and C (arrows; scheme modified from [22]); (B) Equivalent side view using the SERCA three-dimensional structure model (1IWO); (C) Transverse cross-section of the SERCA dimer, at the lower plane in (A), identifying transmembrane segments and putative projections of the C-terminal portion of P domain of each monomer (red circles; scheme modified from [22]); (D) three-dimensional structure of (B) assigned to the dimer model of (C); top view of the cytoplasmic projection showing contact between P domain C-terminal portion of each monomer (red circles); (E) alignment between LCDh and LCD/SERCA sequences using BLAST. The arrow indicates the sequence to LCD $\Delta$ h.

associate with detergent-solubilized native  $\alpha$ -subunit [57]. As the association was stronger in the presence of MgATP for both LCD–LCD and LCD– $\alpha$ -subunit, N site occupancy may play a critical role in stabilizing the self-association during the catalytic cycle.

Although recent determinations of SERCA structures in two different conformations [7,8,58] showed single nucleotide-81

binding and phosphorylation sites, then do not exclude the possibility of ATPase–ATPase interactions, nor do they clarify the molecular basis of the regulatory site [59–61]. In fact, dimeric organizations have been described for different P-ATPase crystals in membranes [25–28]. In particular for SERCA, cryomicroscopy analysis of tubular crystals formed in native membranes revealed a dimer ribbon structure with two strands of

polarized unit, pointing in opposite directions, where the ATPase monomer from one strand is in face-to-face contact with the other through the head-shaped hydrophilic domain [22,62]. In addition, contact between adjacent P sites can also be identified in tubular crystals formed in both phosphorylated and unphosphorylated state [28]. At the time this interface could not be properly identified, due to the low resolution of the images. However, by assigning the recent high-resolution structure (E2-Tg, pdb 1IWO) with the vanadate tubular crystals, at transverse membrane level cut (Fig 5A and B), and then projecting to the cytoplasmic region, it is possible to observe that the dimer interface is formed by interaction between the C-terminal portions of P domains (Fig. 5C and D, red circles). In fact, during the final revision of the present manuscript, Hinsén and collaborators published fitting of 1IWO atomic structures into the 8 Å model [63], using a method for flexible docking of high-resolution atomic structures into electron microscopy lower-resolution densities [64]. Their results strongly corroborate the proposition that the sequence from Asp 650 to Glu 680 is positioned at the dimeric interface in such tubular crystal preparations. Comparing E1 (1SU4) and E2-Tg (1IWO) structures, it can be seen that this region remains accessible for interaction in both conformations. In agreement with these data, we propose this general dimeric model for most of P-ATPases (at least for these subclasses), taking into account that the proposed movements of N and A domains during the ion translocation mechanism are not constrained by such P-P site association. In fact, attempts to associate movements of N and A regions during the reaction cycle suggest that these two domains behave as rigid bodies, while the P domain is flexible, coordinating movements of associated transmembrane helices [65]. Thus, it is even possible to attribute subunit interplay to this plastic region of the enzyme structure.

By modeling the P site C-terminal region of several P-ATPases, it is possible to observe the presence of many hydrophobic and charged amino-acid residues at the surface (data not shown), which would favor inter-chain interactions. On the other hand, badly matched loops, due to insertion and/or deletion of amino-acid residues, might lead to specific interactions between homologous chains, especially in membranes where different ATPases are found, thus avoiding formation of non-functional heterodimers, as proposed by studies with chimeric H<sup>+</sup>,K<sup>+</sup>-ATPase and Na<sup>+</sup>,K<sup>+</sup>-ATPase  $\alpha$ -subunits [35]. In addition, our model does not exclude formation of higher oligomeric structures, such as the hexameric structures observed with another H<sup>+</sup>-ATPase, from *Neurospora crassa* [23,66–69]. It is possible that the regulatory domain (the C-terminal portion of H<sup>+</sup>-ATPase chains) can alter the interaction between P-sites described here, by linking one monomer to the N domain of an adjacent chain, as proposed by Kühlbrandt et al. [68]. This effect could be dependent on phosphorylation level of the regulatory domain. Finally, our proposed model does not preclude association of other regulatory proteins such as phospholamban, which has also been proposed to associate with two Ca<sup>2+</sup>-ATPase monomers [25], and to bind to Lys 393 or Lys 400 in the N domain [70,71].

## Acknowledgments

We would like to thank Drs Helena Maria Scofano and Sergio Teixeira Ferreira for their valuable criticism. The paper has been kindly revised by Dr. Martha M Sorenson.

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ).

## References

- [1] S. Lutsenko, J.H. Kaplan, Organization of P-type ATPases: significance of structural diversity, *Biochemistry* 34 (1995) 15607–15612.
- [2] J.V. Moller, B. Jull, M. le Maire, Structural organization, ion transport and energy transduction of P-type ATPases, *Biochim. Biophys. Acta* 1286 (1996) 1–51.
- [3] J.C. Skou, The influence of some cations on an adenosine triphosphate from peripheral nerves, *Biochim. Biophys. Acta* 23 (1957) 394–401.
- [4] D.J. Bigelow, G. Inesi, Contributions of chemical derivation and spectroscopic studies to the characterization of the Ca<sup>2+</sup> transport ATPase of sarcoplasmic reticulum, *Biochim. Biophys. Acta* 1113 (1992) 323–338.
- [5] A.N. Martonosi, The structure and interactions of Ca<sup>2+</sup>-ATPase, *Biosci. Rep.* 15 (1995) 263–281.
- [6] D.H. MacLennan, W.J. Rice, N.M. Green, The mechanism of Ca<sup>2+</sup> transport by sarco(endo)plasmic reticulum Ca<sup>2+</sup>-ATPases, *J. Biol. Chem.* 272 (1997) 28815–28818.
- [7] C. Toyoshima, M. Nakasto, H. Nomura, H. Ogawa, Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution, *Nature* 405 (2000) 647–655.
- [8] C. Toyoshima, H. Nomura, Structural changes in the calcium pump accompanying the dissociation of calcium, *Nature* 418 (2002) 605–611.
- [9] A.G. Lee, A calcium pump made visible, *Curr. Opin. Struct. Biol.* 12 (2002) 547–554.
- [10] G. Inesi, S. Watanabe, Temperature dependence of ATP hydrolysis and calcium uptake by fragmented sarcoplasmic membranes, *Arch. Biochem. Biophys.* 121 (1967) 665–671.
- [11] J.V. Moller, K.E. Lind, J.P. Andersen, Enzyme kinetics and substrate stabilization of detergent-solubilized and membranous (Ca<sup>2+</sup>-Mg<sup>2+</sup>)-activated ATPase from sarcoplasmic reticulum, *J. Biol. Chem.* 255 (1980) 1912–1920.
- [12] E. Goormaghtigh, C. Chadwick, G.A. Scarborough, Monomers of Neurospora plasma membrane H<sup>+</sup>-ATPase catalyze efficient proton translocation, *J. Biol. Chem.* 261 (1986) 7466–7471.
- [13] J.P. Andersen, Monomer-oligomer Ca<sup>2+</sup>-ATPase and the role of the subunit interaction in the Ca<sup>2+</sup> pump mechanism, *Biochim. Biophys. Acta* 988 (1989) 47–72.
- [14] J.E. Mahaney, J.P. Froehlich, D.D. Thomas, Conformation transitions of the sarcoplasmic Ca<sup>2+</sup>-ATPase studied by time-resolved EPR and quenched-flow kinetics, *Biochemistry* 34 (1995) 4864–4879.
- [15] M. Morii, Y. Hayata, K. Mizoguchi, N. Takeguchi, Oligomeric regulation of gastric H<sup>+</sup>, K<sup>+</sup>-ATPase, *J. Biol. Chem.* 271 (1996) 4068–4072.
- [16] J.M. Merino, C. Gutiérrez-Merino, F. Henaó, Plausible stoichiometry of the interaction nucleotide-binding in the Ca<sup>2+</sup>-ATPase from sarcoplasmic reticulum membranes, *Arch. Biochem. Biophys.* 368 (1999) 298–302.
- [17] G. Blanco, J.C. Koster, R.W. Mercer, The  $\alpha$  subunit of the Na,K-ATPase specifically and stably associates into oligomers, *J. Biol. Chem.* 91 (1994) 8542–8546.
- [18] P. Maguire, K. Ohlendieck, Oligomerization of sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase from rabbit skeletal muscle, *FEBS Lett.* 396 (1996) 115–118.
- [19] V. Levi, J.P.F.C. Rossi, P.R. Castello, L.G. Flecha, Structural significance of plasma membrane calcium pump oligomeric, *Biophys. J.* 82 (2002) 437–446.
- [20] L. Dux, A.N. Martonosi, Two-dimensional arrays of proteins in

- sarcoplasmic reticulum and purified  $\text{Ca}^{2+}$ -ATPase vesicles treated with vanadate, *J. Biol. Chem.* 258 (1983) 2599–2603.
- [21] V.N. Kalabokis, J.J. Bozzola, L. Castelli, P.M.D. Hardwicke, A possible role for the dimer ribbon state of scallop sarcoplasmic reticulum. Dimmer ribbons are associated with stabilization of the  $\text{Ca}(2+)$ -free  $\text{Ca}$ -ATPase, *J. Biol. Chem.* 266 (1991) 22044–22050.
- [22] C. Toyoshima, H. Sasabe, D.L. Stokes, Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane, *Nature* 362 (1993) 469–471.
- [23] M. Auer, G.A. Scarborough, W. Kühlbrandt, Three-dimensional map of the plasma membrane  $\text{H}^{+}$ -ATPase in the open conformation, *Nature* 392 (1998) 840–843.
- [24] W.J. Rice, H.S. Young, D.W. Martin, J.R. Sachs, D.L. Stokes, Structure of  $\text{Na}^{+}$ ,  $\text{K}^{+}$ -ATPase at 11-Å resolution: comparison with  $\text{Ca}^{2+}$ -ATPase in E1 and E2 states, *Biophys. J.* 80 (2001) 2187–2197.
- [25] H.S. Young, L.R. Jones, D.L. Stokes, Locating phospholamban in co-crystals with  $\text{Ca}(2+)$ -ATPase by cryoelectron microscopy, *Biophys. J.* 81 (2001) 884–894.
- [26] T. Jahn, J. Dietrich, B. Andersen, B. Leidvik, C. Otter, C. Briving, W. Kühlbrandt, M.G. Palmgren, Large scale expression, purification and 2D crystallization of recombinant plant plasma membrane  $\text{H}^{+}$ -ATPase, *J. Mol. Biol.* 309 (2001) 465–476.
- [27] P. Morsomme, M. Chami, S. Marco, J. Nader, K.A. Jetchum, A. Goffeau, J.-L. Rigaud, Characterization of a hyperthermophilic P-type ATPase from *Methanococcus jannaschii* expressed in yeast, *J. Biol. Chem.* 277 (2002) 29608–29616.
- [28] F. Delavoie, D. McIntosh, F. Henao, G. Pernazi, P. Champeil, D.L. Stokes, J.-J. Lacapère, Projection map of covalently phosphorylated  $\text{Ca}$ -ATPase from tubular crystals, *Ann. N. Y. Acad. Sci.* 986 (2003) 17–19.
- [29] T. Coelho-Sampaio, S.F. Ferreira, G. Benaim, A. Vyeira, Dissociation of purified erythrocyte  $\text{Ca}^{2+}$ -ATPase by pressure, *J. Biol. Chem.* 266 (1991) 22266–22272.
- [30] V. Levi, J.P.F.C. Rossi, P.R. Castello, L.G. Flecha, Oligomerization of the plasma membrane calcium pump involves two regions with different thermal stability, *FEBS Lett.* 483 (2000) 99–103.
- [31] J.D. Cavières, Calmodulin and the target size of the  $(\text{Ca}^{2+}$ ,  $\text{Mg}^{2+})$ -ATPase dependent of human red-cell ghosts, *Biochim. Biophys. Acta* 771 (1984) 241–244.
- [32] B.J. Bowman, C.J. Berenski, C.Y. Jung, Size of the plasma membrane  $\text{H}^{+}$ -ATPase from *Neurospora crassa* determined by radiation inactivation and comparison with the sarcoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase from skeletal muscle, *J. Biol. Chem.* 15 (1985) 8726–8730.
- [33] D.P. Brisikin, I. Reynolds-Niesman, Change in target molecular size of the red beet plasma membrane ATPase during solubilization and reconstitution, *Plant Physiol.* 90 (1989) 394–397.
- [34] Y. Hayashi, K. Kameyama, T. Kobayashi, E. Hagiwara, N. Shinji, T. Takagi, Oligomeric structure of solubilized  $\text{Na}^{+}/\text{K}^{+}$ -ATPase linked to E1/E2 conformation, *Ann. N. Y. Acad. Sci.* 834 (1997) 19–29.
- [35] J.C. Koster, G. Blanco, R.W. Mercer, A cytoplasmic region of the  $\text{Na}$ ,  $\text{K}$ -ATPase  $\alpha$ -subunit is necessary for specific  $\alpha/\alpha$  association, *J. Biol. Chem.* 270 (1995) 14332–14339.
- [36] C. Donnet, E. Arystarkhova, K. Sweadner, Thermal denaturation of the  $\text{Na}$ ,  $\text{K}$ -ATPase provides evidence for  $\alpha$ - $\alpha$  oligomeric interaction and  $\gamma$  subunit association with the C-terminal domain, *J. Biol. Chem.* 276 (2001) 7357–7365.
- [37] S. Kaya, K. Abe, K. Taniguchi, M. Yazawa, T. Katoh, M. Kikumoto, K. Oiwa, Y. Hayashi, Oligomeric structure of P-type ATPases observed by single molecule detection technique, *Ann. N. Y. Acad. Sci.* 968 (2003) 278–280.
- [38] L. Aravind, M.Y. Galperin, E.V. Koonin, The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold, *TIBS* 23 (1998) 127–129.
- [39] P.C. Carvalho-Alves, V.R. Hering, J.M.S. Oliveira, R.K. Salinas, S. Verjovski-Almeida, Requirement of the hinge domain for dimerization of  $\text{Ca}^{2+}$ -ATPase large cytoplasmic portion expressed in bacteria, *Biochim. Biophys. Acta* 1467 (2000) 73–84.
- [40] R. Serrano, M.C. Kiehlbrandt, G.R. Fink, Yeast plasma membrane ATPase is essential for growth and has homology with  $(\text{Na}^{+}, \text{K}^{+})$ ,  $\text{K}^{+}$ - and  $\text{Ca}^{2+}$ -ATPases, *Nature* 319 (1986) 689–693.
- [41] A. Landy, Dynamic, structural and regulatory aspects of lambda site-specific recombination, *Annu. Rev. Biochem.* 58 (1989) 913–949.
- [42] J.L. Hartley, G.F. Temple, M.A. Brasch, DNA cloning using in vitro site-specific recombination, *Gen. Res.* 10 (2000) 1788–1795.
- [43] P. Bhandari, J. Gowrishankar, Na *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer, *J. Bacteriol.* 164 (1997) 4403–4406.
- [44] R.A. Dohaue, R.L. Bebee, BL21-SI competent cells for protein expression in *E. coli*, *Focus* 21 (1999) 49–51.
- [45] M.M. Bradford, A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248–254.
- [46] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 193 (1970) 680–685.
- [47] N. Sreerama, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expand reference set, *Anal. Biochem.* 287 (2000) 252–260.
- [48] C. Geourjon, G. Deléage, SOPMA: significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments, *Comput. Appl. Biosci.* 11 (1995) 681–684.
- [49] V.N. Uversky, Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule, *Biochemistry* 32 (1993) 13288–13298.
- [50] E. Capieaux, C. Rapin, D. Thinés, Y. Dupont, A. Goffeau, Overexpression in *Escherichia coli* and purification of an ATP-binding peptide from yeast plasma membrane  $\text{H}^{+}$ -ATPase, *J. Biol. Chem.* 268 (1993) 21895–21900.
- [51] M.-J. Moutin, M. Cuillel, C. Rapin, R. Miras, M. Anger, A.-M. Lompré, Y. Dupont, Measurements of ATP binding on the large cytoplasmic loop of the sarcoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase overexpressed in *Escherichia coli*, *J. Biol. Chem.* 269 (1994) 11147–11154.
- [52] C. Gatto, A.X. Wang, J.H. Kaplan, The M4M5 cytoplasmic loop of the  $\text{Na}$ ,  $\text{K}$ -ATPase, overexpressed in *Escherichia coli*, binds nucleoside triphosphates with the same selectivity as the intact native protein, *J. Biol. Chem.* 273 (1998) 10578–10585.
- [53] T. Obsil, F. Merola, A. Lewit-Bentley, E. Amler, The isolated H4–H5 cytoplasmic loop of  $\text{Na}, \text{K}$ -ATPase overexpressed in *Escherichia coli* retains its ability to bind ATP, *Gen. Physiol. Biophys.* 17 (1998) 52–55.
- [54] C.M. Tran, R.A. Farley, Catalytic activity of an isolated domain of  $\text{Na}$ ,  $\text{K}$ -ATPase expressed in *Escherichia coli*, *Biophys. J.* 77 (1999) 258–266.
- [55] P. Champeil, T. Menguy, S. Soulie, B. Jull, A. Gomez de Garcia, F. Rusconi, P. Falson, L. Denoroy, F. Henao, M. le Maire, J.V. Moller, Characterization of a protease-resistant domain of the cytosolic portion of sarcoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase, *J. Biol. Chem.* 273 (1998) 6619–6631.
- [56] B.J. Bowman, E.J. Bowman,  $\text{H}^{+}$ -ATPases from mitochondria, plasma membranes, and vacuoles of fungal cells, *J. Membr. Biol.* 94 (1986) 83–97.
- [57] C.J. Costa, C. Gatto, J.H. Kaplan, Interactions between  $\text{Na}, \text{K}$ -ATPase  $\alpha$ -subunit ATP-binding domains, *J. Biol. Chem.* 278 (2003) 9176–9184.
- [58] C. Toyoshima, T. Mizutani, Crystal structure of the calcium pump with a bound ATP analogue, *Nature* 430 (2004) 529–535.
- [59] Y. Dupont, R. Pougeois, M. Ronjat, S. Verjovski-Almeida, Two distinct classes of nucleotide binding sites in sarcoplasmic reticulum  $\text{Ca}$ -ATPase revealed by 2',3'-O-(2,4,6-trinitrocyclohexadienylidene)-ATP, *J. Biol. Chem.* 260 (1985) 7241–7249.
- [60] P.C. Carvalho-Alves, C.R.G. Oliveira, S. Verjovski-Almeida, Stoichiometric photolabeling of two distinct low and high affinity nucleotide sites in sarcoplasmic reticulum ATPase, *J. Biol. Chem.* 260 (1985) 4282–4287.
- [61] J.A. Mignaco, O.H. Lupi, F.T. Santos, H. Barrabin, H.M. Scofano, Two simultaneous binding sites for nucleotide analogs are kinetically distinguishable on the sarcoplasmic reticulum  $\text{Ca}(2+)$ -ATPase, *Biochemistry* 35 (1996) 3886–3891.
- [62] L. Castellani, P.M. Hardwicke, P. Vilbert, Dimer ribbons in the three-dimensional structure of sarcoplasmic reticulum, *J. Mol. Biol.* 185 (1985) 579–594.

- [63] P. Zhang, C. Toyoshima, K. Yonekura, N.M. Green, D.L. Stokes, Structure of the calcium pump from sarcoplasmic reticulum at 8 Å resolution, *Nature* 392 (1998) 835–839.
- [64] K. Hinsin, N. Reuter, J. Navaza, D.L. Stokes, J.J. Lacapere, Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase, *Biophys. J.* 88 (2005) 818–827.
- [65] N. Reuter, K. Hinsin, J.J. Lacapere, Transconformations of the SERCA1 Ca-ATPase: a normal mode study, *Biophys. J.* 85 (2003) 2186–2197.
- [66] C.C. Cyrklaff, E. Goormaghtigh, G.A. Scarborough, 2-D structure of the *Neurospora crassa* plasma membrane ATPase as determined by electron cryomicroscopy, *EMBO J.* 14 (1995) 1854–1857.
- [67] M. Auer, G.A. Scarborough, W. Kühlbrandt, Surface crystallisation of the plasma membrane H<sup>+</sup>-ATPase on a carbon support film for electron crystallography, *J. Mol. Biol.* 287 (1999) 961–968.
- [68] W. Kühlbrandt, J. Zeelen, J. Dietrich, Structure, mechanism, and regulation of the *Neurospora* plasma membrane H<sup>+</sup>-ATPase, *Science* 297 (2002) 1692–1696.
- [69] K.-H. Rhee, G.A. Scarborough, R. Henderson, Domain movements of plasma membrane H<sup>(+)</sup>-ATPase: 3D structures of two states by electron cryo-microscopy, *EMBO J.* 21 (2002) 3582–3589.
- [70] C. Toyoshima, M. Asahi, Y. Sugita, R. Khanna, T. Tsuda, D.H. MacLennan, Modeling of the inhibitory interaction of phospholamban with the Ca<sup>2+</sup> ATPase, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 467–472.
- [71] H.S. Young, D.L. Stokes, The mechanics of calcium transport, *J. Membr. Biol.* 198 (2004) 55–63.

## Proposição e Análise de Interface de Dimerização

Apesar dos recentes avanços no conhecimento da estrutura tridimensional de P-ATPases, a verdadeira unidade funcional presente nas membranas biológicas ainda é controversa.

Por outro lado, o domínio de fosforilação, localizado na alça citoplasmática maior de P-ATPases, tem sido identificado como uma importante superfície para auto-associação, ao menos para o subgrupo II de P-ATPase que compreende a Na<sup>+</sup>,K<sup>+</sup>-ATPase (Blanco et al., 1994; Koster et al., 1995; Costa et al., 2003) e a SERCA (Carvalho-Alves et al., 2000). Para esse último foi demonstrado que o LCD/SERCA expressa heterologicamente em *E. coli* tem forte tendência a formar um complexo dimérico.

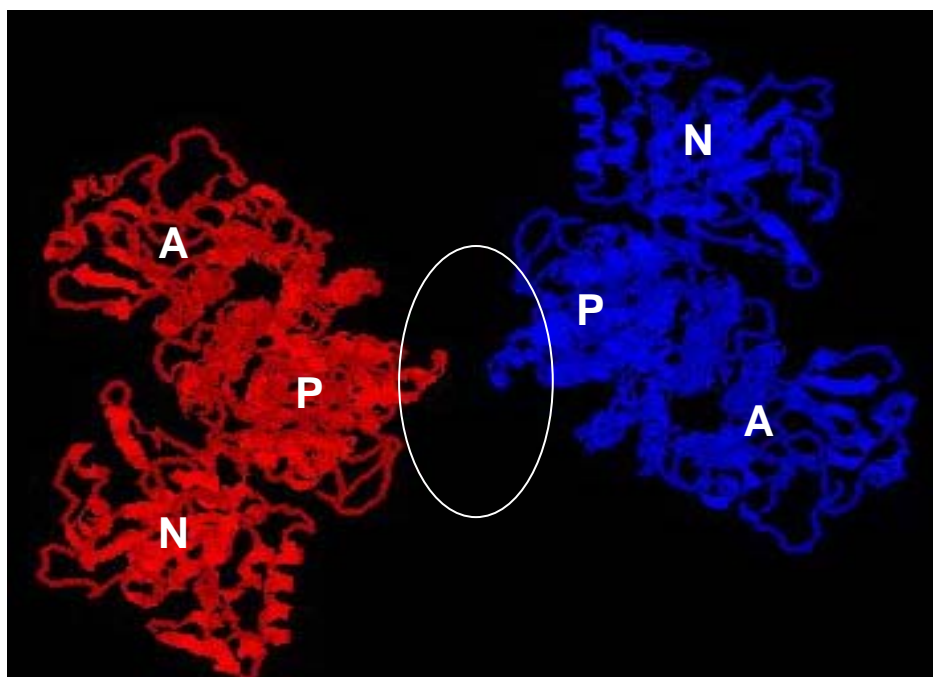
Assim um modelo teórico para o dímero da SERCA foi criado, tanto a partir do modelo cristalográfico como também de outros dados da literatura (Toyoshima et al., 1993; Morsomme et al., 2002). Esse modelo (Figura 9) foi submetido a técnicas computacionais de simulação por dinâmica molecular (MD) a fim de ter uma melhor caracterização estrutural e uma maior compreensão da dinâmica dos domínios hidrofílicos, pois a aplicabilidade de Modelagem Molecular de estruturas de biomoléculas em computadores abrange ao menos três aspectos de interesse biológico:

I. Estrutural – concepção gráfica tridimensional virtual de estrutura conhecida ou proposta, possibilitando a identificação e análise detalhada de regiões ligantes e de funcionalidade biológica;

II. Interações Moleculares – a utilização de campos de forças (Mecânica Molecular), descreve as interações intra e inter-moleculares, e estados conformacionais; e,

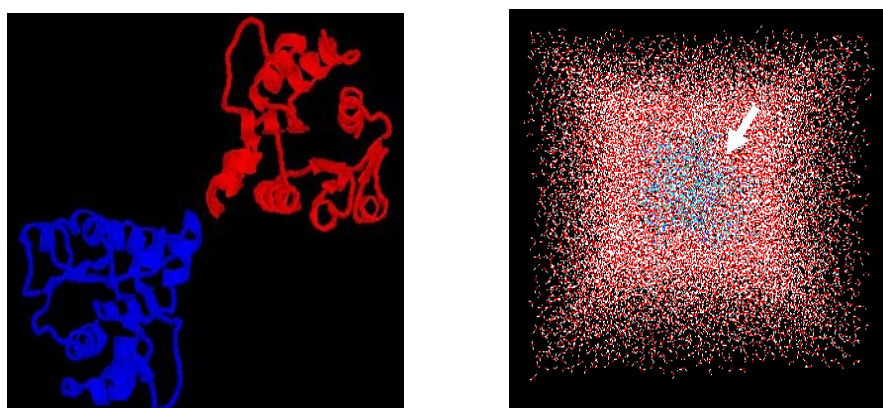
III. Evolução Conformacional – por simulações de Dinâmica Molecular pode-se caracterizar as variações conformacionais sofridas pelo sistema, como por exemplo efeitos de temperatura e pressão, ao longo do tempo.





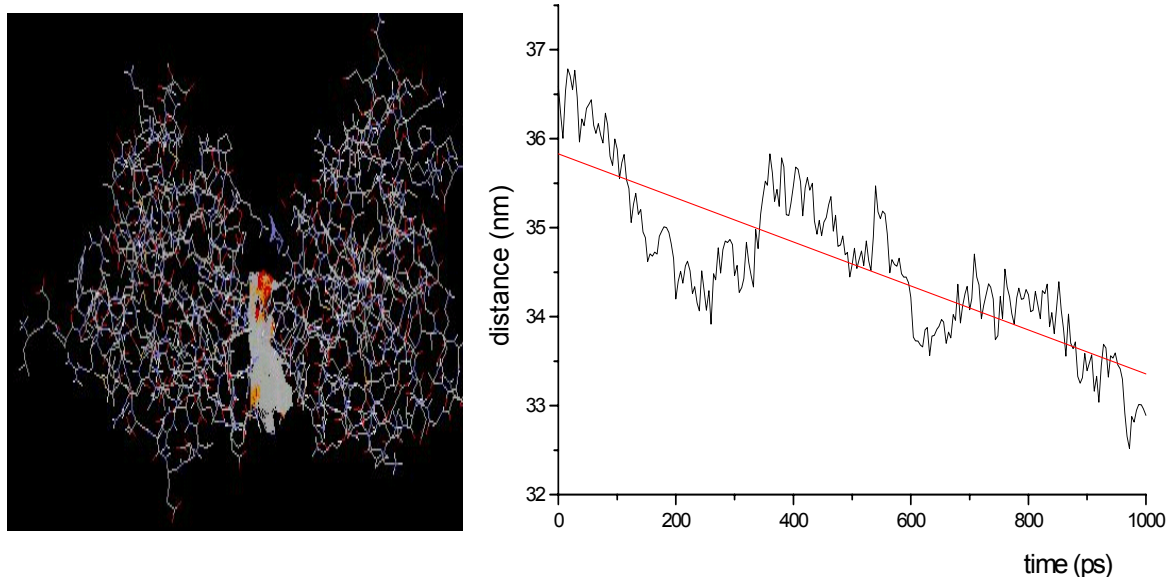
**Figura 9. Modelo hipotético para o dímero da SERCA**, destacando-se os domínios funcionais (A, N e P) e a interface de interação entre os monômeros (círculo).

O sistema simulado compreende somente a porção C-terminal do domínio P (sítio de fosforilação), a partir do resíduo de Leu610 (Figura 10). As minimizações das energias por mecânica molecular e as simulações de dinâmica molecular (MD) foram feitas com o programa THOR, utilizando campos de forças do GROMOS96, e os parâmetros de carga escolhidos para os resíduos foram padrões com representações de carga mais prováveis dos resíduos de aminoácidos em pH fisiológico.



**Figura 10. Sistema para simulação de Dinâmica Molecular.** Na esquerda, a porção C-terminal do sítio de fosforilação da SERCA; à direita, o sistema simulado em caixa d'água (no interior complexo proteico simulado (seta) e ao redor moléculas de água).

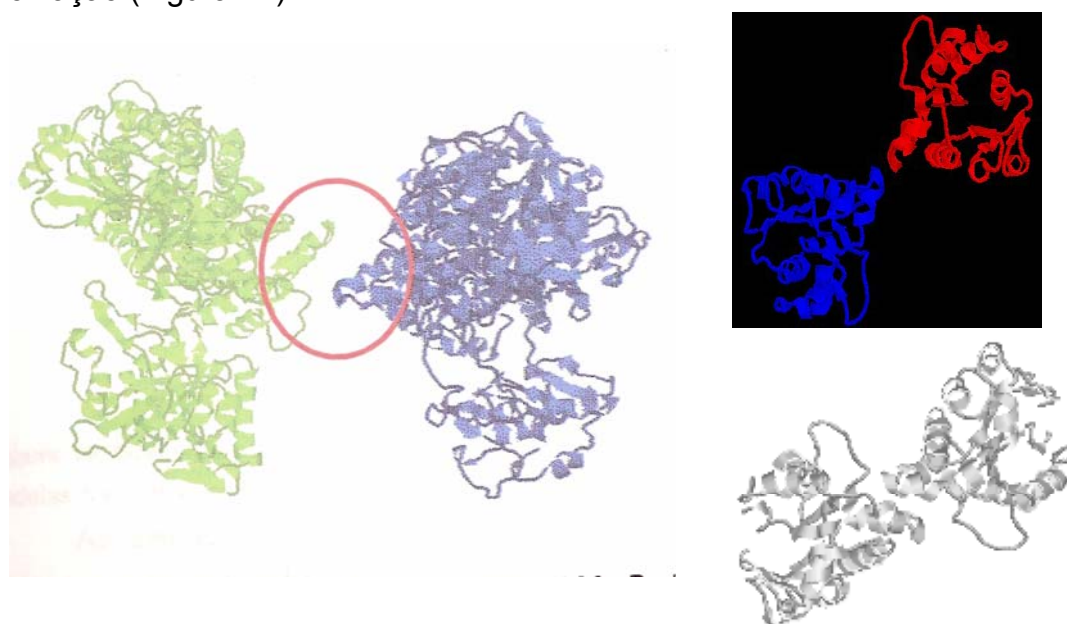
Realizamos análises para avaliar a área da superfície do sítio ativo da proteína acessível ao ligante, calculando as áreas médias de contato de cada resíduo de aminoácido com o respectivo ligante. Nessa análise fica caracterizada a importância de cada resíduo de aminoácido para a interação proteína-proteína (ou proteína-ligante). A partir dos resultados das simulações, foi então possível quantificar os contatos de van der Waals entre proteína-proteína, aplicando o algoritmo de Connolly, onde observamos que, sob essas condições, a interação entre essas regiões é de caráter hidrofóbico, e também uma diminuição no centro de massa do sistema simulado (Figura 11, respectivamente) – o que indica a aproximação entre as cadeias peptídicas simuladas. Contudo, observamos um colapso do sistema (i.e., deformação na estruturação secundária dos peptídeos simulados, dados não mostrados), o que compromete a assertiva da efetiva interação entre as cadeias e a contribuição de forças hidrofóbicas nesse fenômeno de associação.



**Figura 11. Análise do sistema dimérico.** Na esquerda, observa-se forte interação (em cinza) entre os resíduos da interface dos dois polipeptídeos, onde a interação é de caráter hidrofóbico; À direita, análise da distância entre os centros-de-massa dos polipeptídeos; observa-se uma diminuição, sugerindo aproximação entre as duas cadeias.

Então, a continuação de MD para um novo modelo similar ao modelo anterior, diferindo somente pela inserção da porção N-terminal do sítio de fosforilação, faz-se extremamente necessário. Desse modo, verificamos que, provavelmente, com a inserção da região N-terminal do sítio P teríamos uma melhora na estabilização dos peptídeos simulados, evitando o colapso evidenciado em nossos estudos de MD preliminares supracitados. Desse modo, sem a distorção da estrutura secundária dos polipeptídeos, poderíamos avaliar as forças de interação entre as cadeias de LCD e efetivamente demonstrar a tendência de auto-associação de P-ATPases através da interface do domínio P.

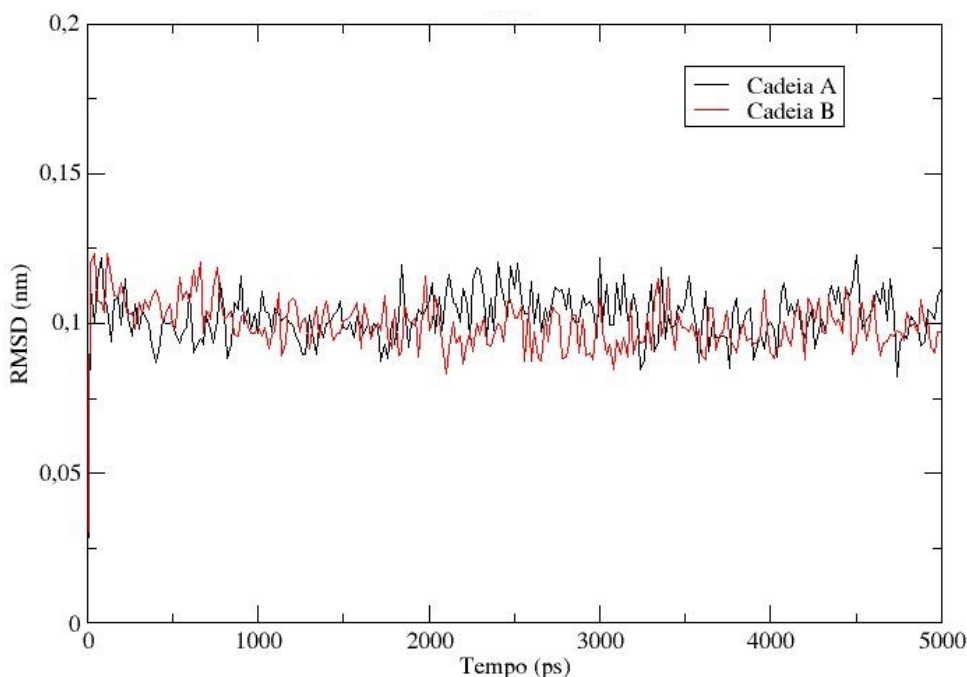
Adicionalmente, utilizando como base o modelo hipotético descrito no manuscrito presente nessa tese, como também outros dados da literatura (Toyoshima et al., 1993; Morsomme et al., 2002), aperfeiçoamos o modelo dimérico da simulação por dinâmica molecular, onde deslocamos ligeiramente as cadeias polipeptídicas para o interior (uma em relação a outra) de modo que houvesse uma maior interface de associação a partir do C-terminal do sítio de fosforilação (Figura 12).



**Figura 12. Novo modelo para o dímero da SERCA.** Destacando-se a maior interface de interação entre os monômeros (círculo), quando comparado com o modelo anterior na parte superior à direita. Na parte inferior a direita, o sistema simulado com a inserção da porção N-terminal do sítio de fosforilação.

Esse “aperfeiçoamento” do modelo foi feito com base em dados dos cristais tubulares (Toyoshima et al., 1993), pelo cálculo aproximado das distâncias entre as cadeias da SERCA nas membranas (dados não mostrados).

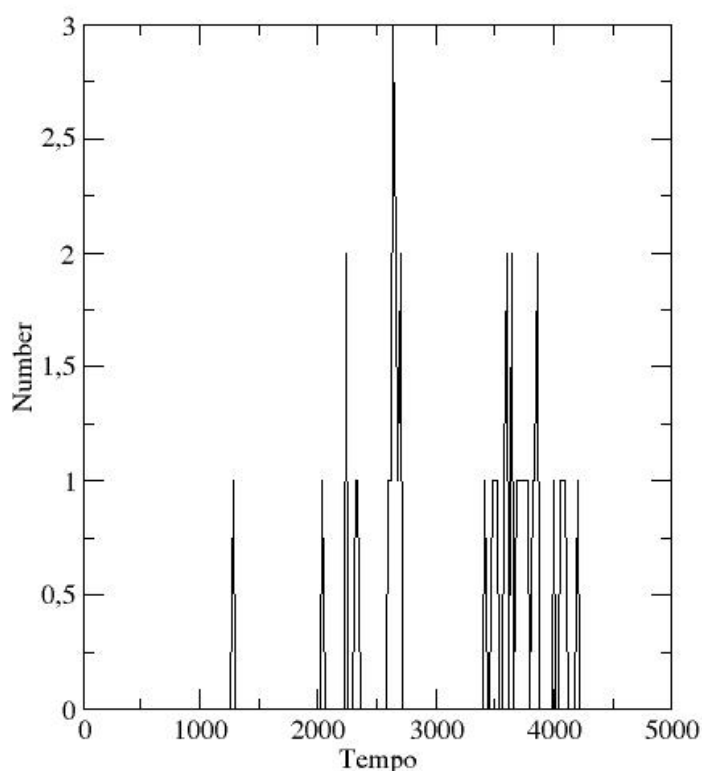
Os dados de MD para esse novo sistema demonstram que a inserção do N-terminal do sítio de fosforilação tornou-o mais estável, havendo pouca flutuação tanto do esqueleto peptídico durante toda a simulação, conforme mostrado do desvio do RMSD – que é uma análise das variações espaciais das posições dos átomos ao longo da dinâmica (Figura 13), quando no RMSF – que é uma análise das variações espaciais por resíduo (dados não mostrados). Estes dados sugerem, portanto, que o colapso observado no primeiro modelo teórico foi realmente devido à ausência da folha-beta no interior do sítio de fosforilação.



**Figura 13. Gráfico do RMS.** A Flutuação do esqueleto peptídico (backbone) na simulação das cadeias A e B do dímero do sítio de fosforilação da SERCA durante 5,0 ns.

Assim, iniciamos para o estudo das possíveis forças que manteriam ou colaboram para a associação dos polipeptídeos durante a dinâmica molecular. Como já é sabido, as interações não covalentes são essenciais para a manutenção da estrutura da proteína, para os processos de reconhecimento, e para as interações proteína/proteína ou proteína/ligantes que são

intermediadas pela água, e onde quase um quarto de todas as ligações de hidrogênio detectadas em todas as simulações. As ligações de hidrogênio são um tipo especial de interação entre átomos não-ligados. Na Figura 14 demonstramos a formação de pontes de hidrogênio nas cadeias durante os 5 ns de simulação. Com isso, pôde-se verificar poucas interações de ponte de hidrogênio no sistema, corroborando a idéia da participação de interações hidrofóbicas e/ou eletrostáticas na interface.



**Figura 14. Gráfico de Pontes de Hidrogênio.** Número de pontes de hidrogênio formadas na interface dos dímeros (cadeia A e B) durante 5,0 ns.

Sendo assim, passamos a verificar a existência de resíduos de caráter hidrofóbico voltados para fora da estrutura 3D da SERCA (Toyoshima et al., 2000), em especial na região C-terminal do domínio de fosforilação, e a contribuição de cada resíduo de aminoácido na interação proteína-proteína (Tabela I).

**TABELA I – Análise de interação entre resíduos.** Interação dos resíduos de aminoácidos da SERCA entre si (i.e., interação intramolecular) separados pela característica físico-química da interação. Nota: São mostrados somente os resíduos localizados na porção C-terminal do sítio de fosforilação da SERCA. O Total refere-se à soma do valor calculado da interação de cada resíduo participante na dada interação, e só são considerados quando maiores ou iguais a 2. Nenhum dos resíduos não realizam interações atrativas nessa região.

HIDROFÓBICA (Razão : 2 - 3,8)								
Valor	Posição1	Resíduo1	Atomo1	Atomo2	Resíduo2	Posição2	Valor	TOTAL
0,69	600	LEU	C	CD	PRO	602	0,05	0,74
0,28	603	PRO	CG	CG	MET	608	0,68	0,96
0,28	603	PRO	CB	CG2	ILE	639	0,48	0,76
0,04	604	ARG	C	CG2	VAL	607	0,27	0,31
0,04	604	ARG	CB	CG2	VAL	607	0,27	0,31
0,04	604	ARG	CA	CG2	VAL	607	0,27	0,31
0,04	604	ARG	CZ	CA	ASN	739	0,23	0,27
0,04	604	ARG	CZ	CB	ASN	739	0,23	0,27
0,26	610	SER	CB	CG2	VAL	744	0,36	0,62
0,26	611	ILE	CD1	CD1	ILE	639	0,48	0,74
0,26	611	ILE	CG1	CE1	PHE	740	0,08	0,34
0,26	614	CYS	CB	CG2	VAL	621	0,25	0,51
0,17	617	ALA	CB	CE	MET	817	0,22	0,39
0,19	619	ILE	CD1	CG1	VAL	747	0,26	0,45
0,62	620	ARG	CZ	C	ARG	671	0,82	1,44
0,25	621	VAL	CG2	CZ	PHE	740	0,08	0,33
0,56	622	ILE	CG2	CD1	PHE	676	0,4	0,96
0,12	623	MET	CE	CG2	ILE	635	0,1	0,22
0,12	623	MET	CE	CD1	ILE	639	0,48	0,6
0,01	625	THR	CG2	CB	ALA	632	0,03	0,04
0,45	628	ASN	CA	CZ	ARG	678	0,06	0,51
0,45	628	ASN	CB	CZ	ARG	678	0,06	0,51
0,64	629	LYS	CB	CG2	THR	654	0,61	1,25
0,7	633	ILE	CG2	CG2	VAL	648	0,72	1,42
0,24	636	CYS	CB	CD1	PHE	642	0,65	0,89
0,48	639	ILE	CD1	CD1	ILE	641	0,29	0,77
0,65	642	PHE	CE2	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CZ	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CZ	CB	VAL	648	0,72	1,37
0,65	642	PHE	CE1	CB	VAL	648	0,72	1,37
0,65	642	PHE	CD1	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CG	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CD2	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CZ	CB	CYS	674	0,79	1,44
0,65	642	PHE	CE2	CB	CYS	674	0,79	1,44
0,66	653	TYR	CE2	CB	ALA	669	0,68	1,34
0,2	655	GLY	CA	CE1	HIS	683	0,45	0,65
0,41	658	PHE	CZ	CB	GLN	666	0,61	1,02
0,41	658	PHE	CE1	CD	GLN	666	0,61	1,02
0,41	658	PHE	CE2	CA	GLN	666	0,61	1,02
0,41	658	PHE	CD2	CE2	PHE	676	0,4	0,81
0,51	661	LEU	CB	CB	GLU	665	0,55	1,06
0,6	662	PRO	CD	CG	GLU	665	0,55	1,15
0,6	662	PRO	CD	CD	GLU	665	0,55	1,15

0,6	662	PRO	CG	CG	GLU	665	0,55	1,15
0,55	665	GLU	C	CB	GLU	668	0,68	1,23
0,61	666	GLN	CB	CE2	TYR	690	0,69	1,3
0,76	667	ARG	CB	CD1	TYR	690	0,69	1,45
0,76	667	ARG	CA	CG	TYR	690	0,69	1,45
0,76	667	ARG	CA	CD1	TYR	690	0,69	1,45
0,76	667	ARG	CZ	CZ	TYR	694	0,79	1,55
0,76	667	ARG	CG	CE2	TYR	694	0,79	1,55
0,76	667	ARG	CG	CZ	TYR	694	0,79	1,55
0,68	668	GLU	CD	CD	ARG	671	0,82	1,5
0,73	670	CYS	CB	CB	TYR	690	0,69	1,42
0,82	671	ARG	CB	CZ	TYR	694	0,79	1,61
0,82	671	ARG	CD	CZ	TYR	694	0,79	1,61
0,82	671	ARG	CB	CE2	TYR	694	0,79	1,61
0,76	673	ALA	CB	CZ	PHE	676	0,4	1,16
0,76	673	ALA	CB	CE1	PHE	676	0,4	1,16
0,49	679	VAL	CG2	CE1	HIS	683	0,45	0,94
0,49	679	VAL	CG2	CB	HIS	683	0,45	0,94
0,49	679	VAL	CG2	CG	HIS	683	0,45	0,94
0,51	680	GLU	CG	CE1	HIS	683	0,45	0,96
0,51	680	GLU	CB	CE1	HIS	683	0,45	0,96

**PONTES DE H: Cadeia - Cadeia (Razão : 2 - 3,2)**

Valor	Posição1	Resíduo1	Atomo1	Atomo2	Resíduo2	Posição2	Valor	TOTAL
0,69	600	LEU	C	CD	PRO	602	0,05	0,74
0,28	603	PRO	CG	CG	MET	608	0,68	0,96
0,28	603	PRO	CB	CG2	ILE	639	0,48	0,76
0,04	604	ARG	C	CG2	VAL	607	0,27	0,31
0,04	604	ARG	CB	CG2	VAL	607	0,27	0,31
0,04	604	ARG	CA	CG2	VAL	607	0,27	0,31
0,04	604	ARG	CZ	CA	ASN	739	0,23	0,27
0,04	604	ARG	CZ	CB	ASN	739	0,23	0,27
0,26	610	SER	CB	CG2	VAL	744	0,36	0,62
0,26	611	ILE	CD1	CD1	ILE	639	0,48	0,74
0,26	611	ILE	CG1	CE1	PHE	740	0,08	0,34
0,26	614	CYS	CB	CG2	VAL	621	0,25	0,51
0,17	617	ALA	CB	CE	MET	817	0,22	0,39
0,19	619	ILE	CD1	CG1	VAL	747	0,26	0,45
0,62	620	ARG	CZ	C	ARG	671	0,82	1,44
0,25	621	VAL	CG2	CZ	PHE	740	0,08	0,33
0,56	622	ILE	CG2	CD1	PHE	676	0,4	0,96
0,12	623	MET	CE	CG2	ILE	635	0,1	0,22
0,12	623	MET	CE	CD1	ILE	639	0,48	0,6
0,01	625	THR	CG2	CB	ALA	632	0,03	0,04
0,45	628	ASN	CA	CZ	ARG	678	0,06	0,51
0,45	628	ASN	CB	CZ	ARG	678	0,06	0,51
0,64	629	LYS	CB	CG2	THR	654	0,61	1,25
0,7	633	ILE	CG2	CG2	VAL	648	0,72	1,42
0,24	636	CYS	CB	CD1	PHE	642	0,65	0,89
0,48	639	ILE	CD1	CD1	ILE	641	0,29	0,77
0,65	642	PHE	CE2	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CZ	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CZ	CB	VAL	648	0,72	1,37
0,65	642	PHE	CE1	CB	VAL	648	0,72	1,37
0,65	642	PHE	CD1	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CG	CG1	VAL	648	0,72	1,37

0,65	642	PHE	CD2	CG1	VAL	648	0,72	1,37
0,65	642	PHE	CZ	CB	CYS	674	0,79	1,44
0,65	642	PHE	CE2	CB	CYS	674	0,79	1,44
0,66	653	TYR	CE2	CB	ALA	669	0,68	1,34
0,2	655	GLY	CA	CE1	HIS	683	0,45	0,65
0,41	658	PHE	CZ	CB	GLN	666	0,61	1,02
0,41	658	PHE	CE1	CD	GLN	666	0,61	1,02
0,41	658	PHE	CE2	CA	GLN	666	0,61	1,02
0,41	658	PHE	CD2	CE2	PHE	676	0,4	0,81
0,51	661	LEU	CB	CB	GLU	665	0,55	1,06
0,6	662	PRO	CD	CG	GLU	665	0,55	1,15
0,6	662	PRO	CD	CD	GLU	665	0,55	1,15
0,6	662	PRO	CG	CG	GLU	665	0,55	1,15
0,55	665	GLU	C	CB	GLU	668	0,68	1,23
0,61	666	GLN	CB	CE2	TYR	690	0,69	1,3
0,76	667	ARG	CB	CD1	TYR	690	0,69	1,45
0,76	667	ARG	CA	CG	TYR	690	0,69	1,45
0,76	667	ARG	CA	CD1	TYR	690	0,69	1,45
0,76	667	ARG	CZ	CZ	TYR	694	0,79	1,55
0,76	667	ARG	CG	CE2	TYR	694	0,79	1,55
0,76	667	ARG	CG	CZ	TYR	694	0,79	1,55
0,68	668	GLU	CD	CD	ARG	671	0,82	1,5
0,73	670	CYS	CB	CB	TYR	690	0,69	1,42
0,82	671	ARG	CB	CZ	TYR	694	0,79	1,61
0,82	671	ARG	CD	CZ	TYR	694	0,79	1,61
0,82	671	ARG	CB	CE2	TYR	694	0,79	1,61
0,76	673	ALA	CB	CZ	PHE	676	0,4	1,16
0,76	673	ALA	CB	CE1	PHE	676	0,4	1,16
0,49	679	VAL	CG2	CE1	HIS	683	0,45	0,94
0,49	679	VAL	CG2	CB	HIS	683	0,45	0,94
0,49	679	VAL	CG2	CG	HIS	683	0,45	0,94
0,51	680	GLU	CG	CE1	HIS	683	0,45	0,96
0,51	680	GLU	CB	CE1	HIS	683	0,45	0,96

**CARREGADA : Atrativa (Razão : 2 - 6)**

Valor	Posição1	Resíduo1	Atomo1	Atomo2	Resíduo2	Posição2	Valor	TOTAL
0	601	ASP	OD2	NZ	LYS	352	0	0
0	601	ASP	OD1	NZ	LYS	352	0	0
0	601	ASP	OD1	NH1	ARG	638	0,68	0,68
0,01	627	ASP	OD2	NZ	LYS	352	0	0,01
0,01	627	ASP	OD1	NZ	LYS	352	0	0,01
0,77	646	GLU	OE2	NH1	ARG	637	0,6	1,37
0,77	646	GLU	OE2	NH1	ARG	651	0,82	1,59
0,77	646	GLU	OE2	NH2	ARG	651	0,82	1,59
0,77	646	GLU	OE1	NH2	ARG	651	0,82	1,59
0,77	646	GLU	OE1	NH1	ARG	651	0,82	1,59
0,73	647	GLU	OE1	NH2	ARG	637	0,6	1,33
0,71	650	ASP	OD2	NH2	ARG	672	0,78	1,49
0,71	650	ASP	OD2	NH1	ARG	672	0,78	1,49
0,71	650	ASP	OD1	NH2	ARG	672	0,78	1,49
0,71	650	ASP	OD1	NH1	ARG	672	0,78	1,49
0,48	657	GLU	OE2	NZ	LYS	629	0,64	1,12
0,48	657	GLU	OE1	NZ	LYS	629	0,64	1,12
0,65	659	ASP	OD2	ND1	HIS	683	0,45	1,1
0,65	659	ASP	OD2	NE2	HIS	683	0,45	1,1
0,65	659	ASP	OD1	NE2	HIS	683	0,45	1,1



0,65	659	ASP	OD1	ND1	HIS	683	0,45	1,1
0,65	659	ASP	OD2	NZ	LYS	686	0,67	1,32
0,73	660	ASP	OD2	NH2	ARG	656	0,84	1,57
0,73	660	ASP	OD2	NH1	ARG	656	0,84	1,57
0,73	660	ASP	OD1	NH2	ARG	656	0,84	1,57
0,73	660	ASP	OD1	NH1	ARG	656	0,84	1,57
0,68	668	GLU	OE2	NH2	ARG	671	0,82	1,5
0,68	668	GLU	OE2	NH1	ARG	671	0,82	1,5
0,68	668	GLU	OE1	NH1	ARG	671	0,82	1,5
0,68	668	GLU	OE2	NH1	ARG	672	0,78	1,46
0,68	668	GLU	OE2	NH2	ARG	672	0,78	1,46
0,68	668	GLU	OE1	NH1	ARG	672	0,78	1,46
0,51	680	GLU	OE2	ND1	HIS	683	0,45	0,96
0,51	680	GLU	OE2	NE2	HIS	683	0,45	0,96
0,51	680	GLU	OE1	ND1	HIS	683	0,45	0,96

Para tanto, utilizamos o pacote de programas do STING Millenium no servidor da Empresa Informática Agropecuária ([http://mirrors.rcsb.org/SMS/index\\_m\\_mirror.html](http://mirrors.rcsb.org/SMS/index_m_mirror.html)), onde observa-se que os resíduos compreendidos da Met354 a Ile743 (especialmente os compreendidos na porção C-terminal dos resíduos 600 – 700 da SERCA) não realizam interações atrativas significativas, i.e., com valores maiores que 2,0 entre outros resíduos da própria cadeia (interação intercadeia), assim sendo, poderiam participar da interface proposta para a dimerização, em uma interface de interação entre as cadeias de monômeros da SERCA.

Com base nessas informações, e na proposição de uma interface de associação entre P-ATPases, realizamos posteriormente um mapa de superfície da porção C-terminal do sítio de fosforilação (dados não mostrados), verificamos a existência de resíduos de caráter hidrofóbico voltados para fora da estrutura 3D da SERCA (Toyoshima et al., 2000). Este tem sido o ponto de partida para a proposição de um modelo tridimensional para o dímero de LCD desta P-ATPase, que também nos auxiliaram na escolha dos mutantes e no estabelecimento de um modelo estrutural para oligômeros de P-ATPases.

A partir desse conjunto de dados foi possível a mutagêneses sítio-dirigida. A mutagêneses sítio-dirigida *in vitro* é uma técnica que permite estudar a relação estrutura-função do gene e da proteína pela modificação das seqüências do vetor e que facilita as estratégias de clonagem e expressão.

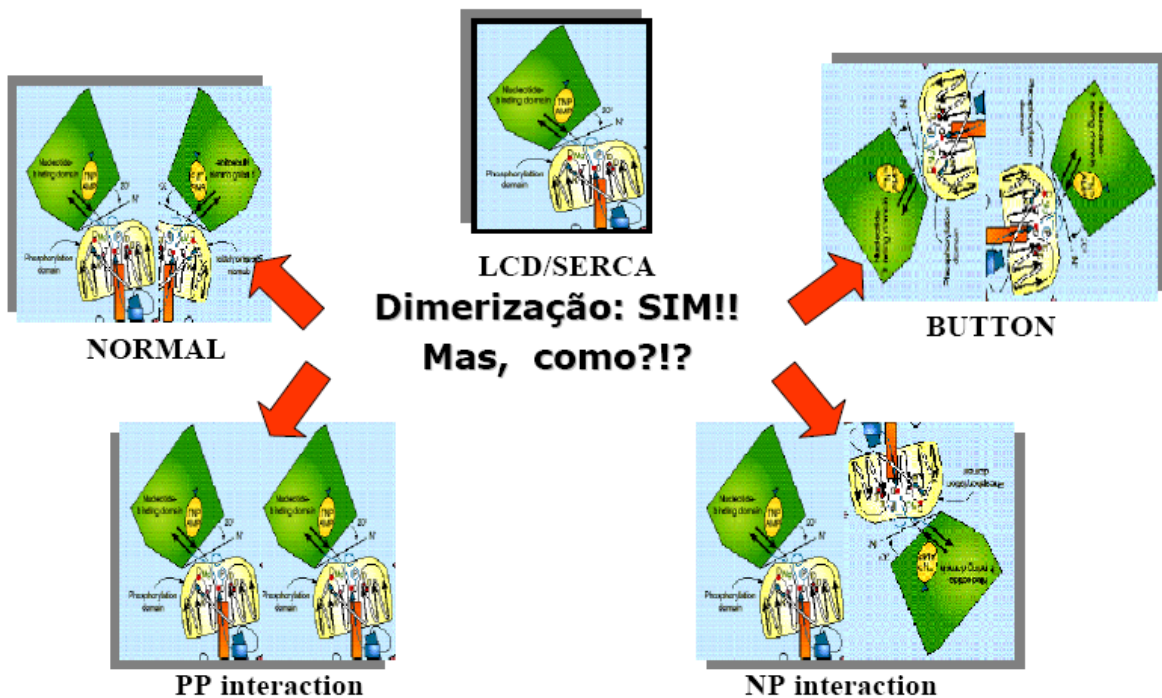
A seqüência da LCD da  $\text{Ca}^{2+}$ -ATPase de retículo sarco(endo)plasmático já havia sido previamente clonada em *Escherichia coli* (Carvalho-Alves et al., 2000), e que serviu para o desenho dos oligonucleotídeos dos diferentes mutantes.

Os resíduos de aminoácidos que foram substituídos pela técnica de mutagêneses sítio-dirigida encontram-se expostos ao solvente e localizados na região C-terminal do domínio de fosforilação. E, adicionalmente, consideramos os papéis desenvolvidos por esses resíduos na interação com outras proteínas regulatórias, tais como o fosfolamban, e na sua atividade enzimática. Visto que alguns dos mutantes propostos neste projeto para os quais desenhamos oligonucleotídeos já haviam sido descritos na literatura, mas nenhum dos candidatos-mutantes alterava as características funcionais da enzima (MacLennan et al., 1997), então somente observaríamos o papel desempenhado no processo de interação proteína-proteína. Ressalta-se que estes resíduos são altamente conservados entre as mais diversas P-ATPases.

Os clones recombinantes provenientes da transformação de *E. coli* M-15[pREP-4] (Qiagen) serão crescidos e processados para purificação dos mutantes da LCD/SERCA, conforme adotado na purificação LCD/SERCA selvagem, em presença de 6M de uréia (Carvalho-Alves et al., 2000). Todos esses mutantes da LCD/SERCA, ainda serão estudados.

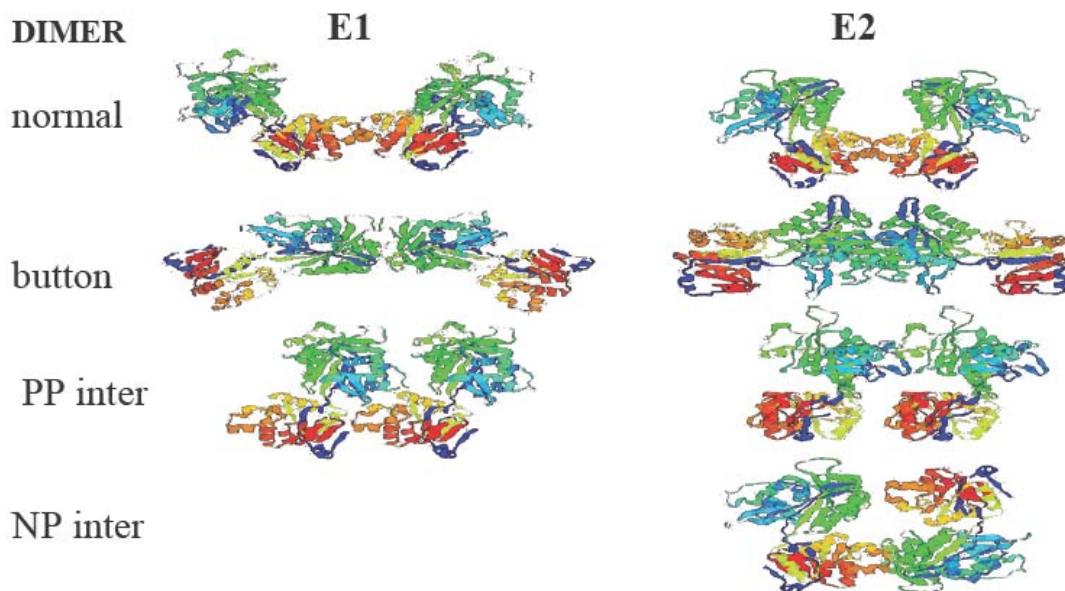
## Arranjo Tridimensional do Complexo dimérico do LCD/SERCA

Apesar desses resultados obtidos, em que demonstramos que o sítio de fosforilação parece ser importante na associação entre P-ATPases, restava-nos saber como se dava essa associação (Figura 15). Uma vez que pouco é sabido sobre o arranjo tridimensional desses complexos e a relação estrutura-função com a proteína nativa em membranas.



**Figura 15. Eventos de Dimerização.** Representação em “cartoon” dos possíveis eventos de dimerização para o LCD/SERCA. No centro, representação do monômero de LCD/SERCA que sofre o fenômeno de auto-associação (Carvalho-Alves et al., 2000), e dos lados, quatro representações de possíveis estágios e/ou estados de associação. O estado normal é condizente com a associação feita pela região C-terminal do sítio de fosforilação tendo cadeias polipeptídicas em sentidos opostos (de acordo com Toyoshima et al., 1993) – modelo sugerido pelo manuscrito desse trabalho (manuscrito 1 da seção P-ATPases); O “button”, corresponde a um modelo de interação, no qual as regiões que estendem para os segmentos transmembranares na enzima nativa participam na associação; O “NP interaction”, as interações entre as cadeias ocorreriam entre o domínio N de um monômero com o domínio P do outro; No “PP interaction” a interação ocorre pela porção C-terminal do sítio de fosforilação, contudo as cadeias estão lateralmente ligadas e não em sentidos opostos. Os modelos “button” e “NP interaction” não seriam passíveis de ocorrência na membrana nativa.

Ainda utilizando modelagem molecular, criamos diferentes modelos hipotéticos/teóricos diméricos para o LCD/SERCA (nos estados conformacionais E1 e E2; Figura 16).



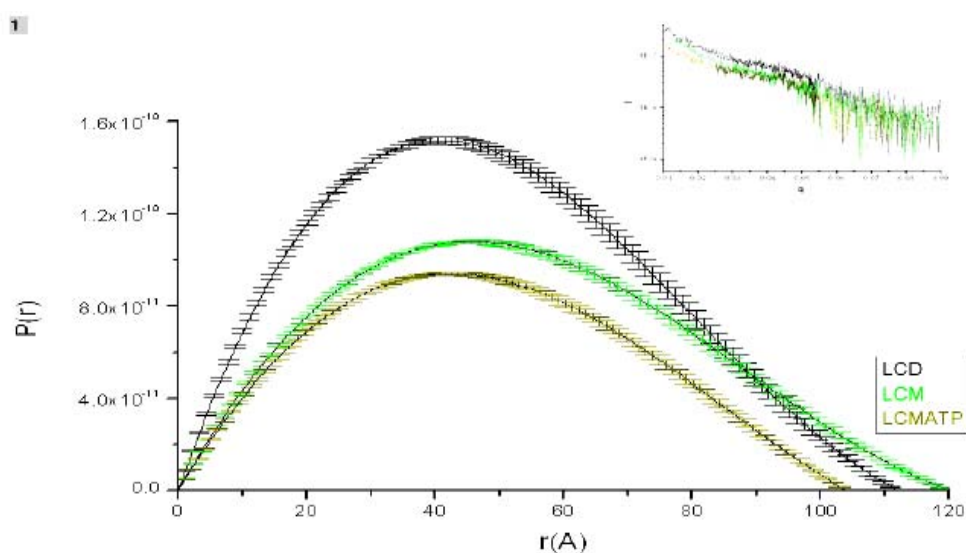
**Figura 16. Modelos Teóricos de LCD Dimerização.** Representação de diferentes modelos teóricos para o dímero de LCD/SERCA, gerados no SPDViewer. Os modelos foram construídos nos estados conformacionais da SERCA e refletindo os possíveis eventos de dimerização para o LCD/SERCA discutido na Figura 12.

Estudos anteriores do nosso grupo já tinham demonstrado por SAXS e raio Stokes que o LCD/SERCA formam dímeros (Carvalho-Alves et al., 2000), então seria interessante investigar a organização estrutural do dímero de LCD/SERCA, já que análises de SAXS de macromoléculas biológicas, em solução, fornecem informações estruturais de baixa resolução ( $> 10 \text{ \AA}$ ) e podem ser realizados em várias condições experimentais, além de permitir a determinação de parâmetros estruturais, tais como raio de giro, diâmetro máximo, peso molecular, grau de compactação/enovelamento (estado de desnaturação) de proteínas.

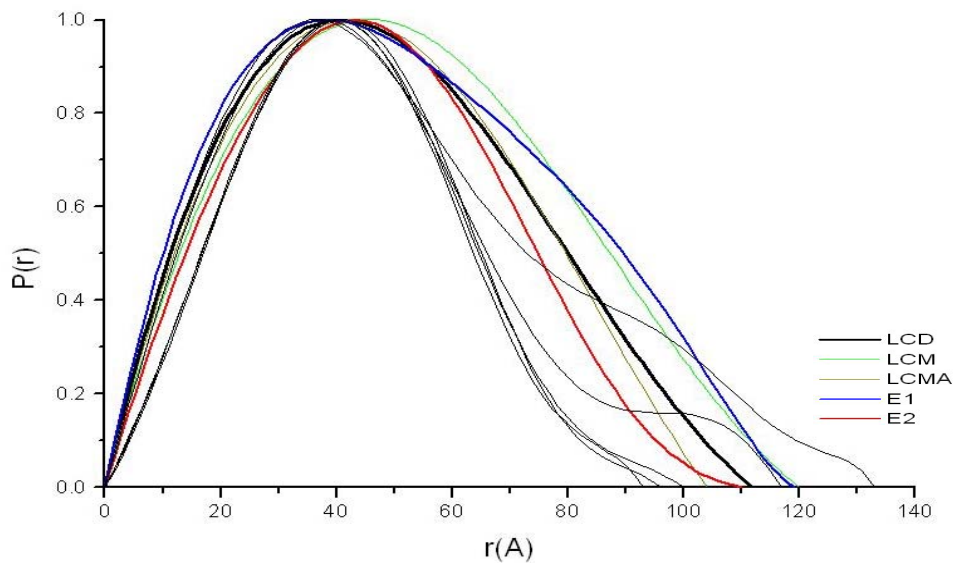
Assim, submetemos proposta de estudar o arranjo 3D do LCD/SERCA ao Laboratório Nacional de Luz Sincontron (LNLS) para a linha de

Espalhamento de Raios-X a Baixo Ângulo (D11A-SAXS), que foi aprovada sob o número D11A-SAXS-3426.

Dados experimentais obtidos na linha de SAXS para o LCD/SERCA na presença ou não de ligantes (i.e., co-fatores, como ATP e Mg-ATPase) (Figura 17) foram comparados com dados obtidos para os diferentes modelos teóricos descritos acima (Figura 18). Então, pela comparação de parâmetros estruturais (raios de giro e diâmetro máximo) obtidos experimentalmente na linha de SAXS com parâmetros teóricos gerados para diferentes modelos diméricos do LCD/SERCA usando Modelagem Molecular e o programa CRY SOL (Tabela II), podemos ter uma compreensão do formato estrutural dos dímeros e que contribua na proposição de um modelo estrutural.



**Figura 17. Função de distribuição (Pr) para o LCD/SERCA.** Dados experimentais de LCD/SERCA com Pi; Mg ou Mg-ATP expostos ao SAXS: LCD, LCM e LCMATP (linhas preta, verde e marrom), respectivamente. Inset: Espalhamento de raios-X para LCD/SERCA sob diferentes condições experimentais.



**Figura 18. Comparação dos dados de SAXS.** Fit da função de distribuição dos dados experimentais (LCD, LCM, LCMA – que representam o LCD/SERCA sob diferentes condições experimentais, contra dados de espalhamento de raios-X dos diferentes modelos de dímeros do LCD/SERCA. Os perfis da função de distribuição ( $P(r)$ ) dos dados experimentais são similares aos obtidos para os modelos teóricos correspondentes ao “Normal” tanto para o estado conformacional E1 quanto no estado E2 (em azul e vermelho, respectivamente). Os demais modelos teóricos gerados mostram perfis de  $P(r)$  diferentes quando comparados com os dados experimentais, e por isso não foram destacados.

Os dados experimentais da função de distribuição  $P(r)$  do LCD/SERCA sob diferentes condições experimentais, i.e., com ou sem ligantes: Mg ou Mg-ATP (Fig. 17), demonstram que a ligação desses co-fatores causa ligeiras modificações conformacionais, o que pode refletir em intermediários do ciclo catalítico; e que, interessantemente, quando comparados com os diferentes modelos de dímero do LCD/SERCA são similares aos obtidos pelos modelos “normal” (E1 e E2, ver Figuras 15, 16 e 18) enquanto os demais modelos (i.e., “Button”, “PP interaction” e “NP interaction”) são nitidamente diferentes (Figuras 15, 16 e 18 e Tabela II).

**TABELA II – Parâmetros Estruturais.** Comparação de parâmetros estruturais ( $D_{max}$  – diâmetro máximo, e Raio de giro; ambos em Angstrom) experimentais e teóricos obtidos para diferentes modelos diméricos de LCD/SERCA.

<b>Dados</b>	<b>Estrutura/Modelo</b>	<b>Parâmetros estrutural</b>	
		<b><math>D_{max}</math></b>	<b>Raio de giro</b>
teórico	Normal (E1)	119	40,79
teórico	Normal (E2)	110	36,01
teórico	PPinteraction (E1)	100	32,62
teórico	PPinteraction (E2)	93	32,12
teórico	NPinteraction (E2)	96	32,31
teórico	Button(E1)	133	41,37
teórico	Button(E2)	117	36,15
experimental	LCD	112	37,87
experimental	LCM	120	41,11
experimental	LCMATP	104	37,00

Desse modo, o conjunto de resultados aqui apresentados corroboram a identificação de resíduos envolvidos na associação do LCD de P-ATPases e a proposição de um modelo para essa associação, sendo um forte indício de que estas enzimas são oligoméricas em suas membranas biológicas, o que acabaria com uma longa polêmica na literatura. Os dados aqui apresentados (ver manuscrito 1 da seção P-ATPase de Resultados e Discussão) sugerem, portanto, que a porção C-terminal do LCD de P-ATPases existem num equilíbrio dinâmico entre monômero e dímero (Carvalho-Alves et al., 2000), característica esta compartilhada entre os subgrupos II (PMCA, SERCA e Na,K-ATPase) e III (H<sup>+</sup>-ATPases) das P-ATPases.

Adicionalmente, observa-se que o formato da partícula dimerica do LCD/SERCA condicionam a participação do sítio de fosforilação em uma conformação relacionada aos observados em cristais tubulares (Toyoshima et al., 1993), corroborando a idéia de que a interação seja através da região C-terminal do domínio P entre duas cadeias de LCD, e em concordância à estruturação (orientação) da P-ATPase na membrana nativa (Figura 18 e Tabela II) e que essa interação seja de relevância funcional para grande parte de P-ATPases.

Por outro lado, pretendemos continuar na caracterização dos resíduos de aminoácidos nesse fenômeno de auto-associação de P-ATPases, e para isso sugerimos o estudo da termodinâmica do processo de associação de dímeros

de LCDs, através de cromatografias de gele filtração em HPLC a diferentes concentrações de polipeptídeos (LCDs), bem como utilizar os mutantes, onde observaríamos qual a(s) consequência(s) da(s) alteração(ões) no(s) resíduo(s) de aminoácido(s) nesse fenômeno. Assim pela proporção entre os picos atribuídos ao monômero e ao dímero, pode-se calcular a constante de dissociação ( $K_d$ ), permitindo uma maior compreensão da dinâmica dos domínios hidrofílicos durante a catálise e explicar o comportamento cinético destas enzimas.

Paralelamente, pode-se marcar os LCDs com isotiocianato de fluoresceína (FITC), uma sonda fluorescente que liga-se covalentemente ao sítio de ligação de nucleotídeos de P-ATPases. Assim as unidades marcadas poderão ter seu estado agregacional estudado através de medidas de polarização de fluorescência do FITC, e que conjuntamente, reforçarão e/ou confirmarão a participação da porção C-terminal do sítio de fosforilação (como também, pelo uso dos mutantes gerados por esse trabalho – identificar quais resíduos de aminoácidos contribuem) na associação entre os monômeros de LCD, e conseqüentemente, comprovar nosso modelo de interação entre as cadeias alfa das P-ATPases.

Desse modo, com base nos conjuntos de dados aqui apresentados concluímos que:

1. O LCDh mostra dois picos principais compatíveis com formas dimérica e monomérica, respectivamente. Já o polipeptídeo LCD $\Delta$ h mostra um único pico e é essencialmente monomérico, assim corrobora que a presença do domínio P é essencial para dimerização;
2. Parâmetros estruturais (raio de giro e diâmetro máximo) do LCD/SERCA obtidos por SAXS sugere que a porção carboxi-terminal do domínio P está envolvida na auto-associação;
3. O modelo dimérico para P-ATPases proposto está de acordo com nossos resultados pelo envolvimento do domínio de fosforilação (especialmente a região C-terminal do LCD) na participação na dimerização tanto para a SERCA1a quanto para a PMA1, como também não interferem nos movimentos propostos dos domínios de ligação de nucleotídeo (N) e de



atuação (compreendida na alça citoplasmática menor - domínio A ) durante o mecanismo de transporte.

## Genoma

Em 2001, o Brasil incorporou-se à era genômica com a publicação da seqüência completa do genoma da bactéria *Xylella fastidiosa*, causadora do amarelinho em culturas de citros (CVC - *Citrus Variegated Chlorosis*) (Simpson et al., 2000). Esse foi o primeiro fitopatógeno no mundo a ser seqüenciado completamente. Depois disso, pelo fato de o Brasil movimentar uma grande parcela de sua economia em produtos agrícolas, foi colocado grande interesse na elucidação de outros genomas de patógenos de culturas de interesse nacional como *Xanthomonas citri*, bactéria causadora do cancro cítrico.

E, seguindo essa tendência nacional, o Estado do Rio de Janeiro lançou o projeto de seqüenciamento do genoma completo da *G. diazotrophicus*, porque ela promove a fixação de nitrogênio diminuindo assim a necessidade do uso de fertilizantes nitrogenados. Este projeto teve a participação de oito grupos de diferentes universidades e instituições de pesquisa. Dentre elas estão a UFRJ, com o Instituto de Bioquímica Médica, o Instituto de Biofísica Carlos Chagas Filho e Instituto de Biologia; a UERJ com o Instituto de Bioquímica e o Instituto de Biologia; A UFRRJ com o Instituto de Biologia; a UENF com o Laboratório de Biotecnologia e a Embrapa Agrobiologia.

O “*data mine*” dos resultados de anotação de todas as ORFs validadas, foi realizado através do acesso direto no banco de dados do sistema Bionotes de anotação, onde todas as informações pertinentes as ORFs validadas pelos membros integrantes do Projeto Riogene, foram acessadas para desenvolvimento de análises estatísticas e observação das principais particularidades do genoma deste importante endófito.

Vale destacar como características gerais do genoma do *G. diazotrophicus* PAL5, consistem na presença de um cromossomo circular com tamanho de 4137,5 Kb, além de 2 plasmídeos circulares designados como A e B, com tamanhos de 38,8 e 16,6 Kb, respectivamente (dados de dezembro de 2006; ver Loureiro, 2007). Este genoma apresenta 3.854 genes (ORFs). Outras considerações sobre genes e grupos funcionais sobre a anotação desse genoma encontram-se descritas em Loureiro, 2007. Assim observa-se 563 ORFs relacionados ao sistema de transporte (Tabela III).

**Tabela III – Descrição das ORFs identificadas com pertencentes ou relacionadas com o sistema de transportes no genoma da *G. diazotrophicus*.**

<b>ORFs</b>	<b>Função da seqüência</b>	<b>TC Number</b>
GD0019	putative sugar transporter protein	2.A.1.1
GD0022	putative cation transporter protein	2.A.4
GD0042	putative sugar transporter protein	2.A.1
GD0046	Outer membrane protein (Omp A domain)	1.B.17
GD0056	putative DNA uptake protein	9.A.21
GD0058	putative ABC transporter	3.A.1.210
GD0096	putative membrane protein	9.B
GD0105	putative membrane protein	9.B
GD0131	traG Conjugal transfer protein TRAG	3.A.7
GD0133	trbB type II secretion system protein	3.A.5
GD0134	trbB conjugal transfer protein trbB	3.A.7
GD0135	trbE conjugal transfer protein trbE	3.A.7
GD0136	trbJ putative type IV secretion system protein	3.A.7
GD0137	trbL putative type IV secretion system protein	3.A.7
GD0138	trbF putative type IV secretion system protein	3.A.7
GD0139	trbG putative type IV secretion system protein	3.A.7
GD0140	trbI putative type IV secretion system protein	3.A.7
GD0146	mdtB Multidrug resistance protein	2.A.6.2
GD0147	mdtA Multidrug resistance protein	2.A.6.2
GD0148	RND efflux system, outer membrane lipoprotein, NodT	2.A.6.2
GD0169	putative membrane protein	9.B
GD0195	putative MFS permease	2.A.1.14
GD0198	TonB-dependent receptor	1.B.14
GD0199	putative sugar transporter	2.A.1.6
GD0210	accB Biotin carboxyl carrier protein of acetyl-CoA	3.B.1
GD0214	putative glucarate transporter	2.A.1.14
GD0215	TonB-dependent receptor	1.B.14
GD0222	NAD(P)H dehydrogenase (quinone)	2.A.37
GD0232	nhaA Na(+)/H(+) antiporter 1	2.A.33
GD0243	Multidrug resistance protein	2.A.1.2
GD0244	secretion protein HlyD	2.A.6
GD0245	outer membrane protein	2.A.6.2
GD0257	putative ABC-type transporter, substrate binding protein	3.A.1.5
GD0259	Inner membrane ABC transporter permease protein	3.A.1.5
GD0260	Dipeptide transport system permease protein	3.A.1.5
GD0261	ABC transporter ATP-binding protein	3.A.1.5
GD0265	TonB-dependent receptor	1.B.14
GD0276	ptsP Phosphoenolpyruvate-protein phosphotransferase ptsP	8.A.7
GD0278	glpF Glycerol uptake facilitator protein	1.A.8
GD0291	Putative Na(+)/H(+) exchanger	2.A.36
GD0316	putative membrane protein	9.B
GD0331	putative transport protein	8.A
GD0332	putative transport protein	9.B
GD0343	corB putative Magnesium-transporting ATPase	3.A.3.4
GD0344	putative transport protein	1.A.30.2
GD0350	major facilitator superfamily transporter	2.A.1
GD0351	arsB Arsenical pump membrane protein	2.A.45
GD0352	major facilitator superfamily transporter	2.A.1.14

<b>GD0357</b>	pntA NAD(P) transhydrogenase subunit alpha part 1	3.D.2
<b>GD0358</b>	pntB NAD(P) transhydrogenase subunit alpha part 2	3.D.2
<b>GD0359</b>	pntB NAD(P) transhydrogenase subunit beta	3.D.2
<b>GD0362</b>	citA Citrate-proton symporter	2.A.1.6
<b>GD0365</b>	putative drug resistance transporter	2.A.1.3
<b>GD0370</b>	putative periplasmic binding proteins	3.A.1
<b>GD0371</b>	putative ABC transporter ATP-binding protein	3.A.1
<b>GD0374</b>	putative ribose transport system permease protein	2.A.7
<b>GD0432</b>	putative TonB-dependent receptor	1.B.14
<b>GD0438</b>	putative molybdenum import ATP-binding protein modC	3.A.1.8.1
<b>GD0439</b>	putative molybdenum transport system permease protein modB	3.A.1.8.1
<b>GD0440</b>	putative molybdate-binding periplasmic protein precursor	3.A.1.8.1
<b>GD0470</b>	Chemoreceptor mcpA	9.B
<b>GD0471</b>	Oxidoreductase (aldo/keto reductase) protein	8.A.5
<b>GD0477</b>	putative extracellular solute-binding protein	3.A.1
<b>GD0479</b>	putative ABC transporter, permease protein	3.A.1
<b>GD0480</b>	putative ABC transporter, permease protein	3.A.1
<b>GD0481</b>	putative ABC transporter, ATP-binding protein	3.A.1
<b>GD0494</b>	pseudopilin G precursor	3.A.15
<b>GD0495</b>	putative prepilin peptidase and N-methyltransferase O	3.A.15
<b>GD0496</b>	putative type II secretion pathway kinase GspE	3.A.15
<b>GD0497</b>	putative general secretion pathway protein F	3.A.15
<b>GD0498</b>	putative pseudopilin H precursor	3.A.15
<b>GD0499</b>	putative pseudopilin I precursor	3.A.15
<b>GD0500</b>	putative pseudopilin J precursor	3.A.15
<b>GD0501</b>	putative general secretion pathway protein L	3.A.15
<b>GD0502</b>	putative general secretion pathway protein M	3.A.15
<b>GD0503</b>	putative general secretion pathway protein N	3.A.15
<b>GD0504</b>	putative general secretion pathway protein D	3.A.15
<b>GD0505</b>	putative general secretion pathway protein K	3.A.15
<b>GD0514</b>	putative membrane protein	9.B
<b>GD0517</b>	putative membrane protein	9.B
<b>GD0518</b>	putative outer membrane protein (OmpA/MotB)	1.B.17
<b>GD0520</b>	putative MotA/TolQ/ExbB proton channel family protein	1.A.30
<b>GD0541</b>	putative voltage-gated ClC-type chloride channel	1.A.11
<b>GD0549</b>	putative MFS transporter	2.A.1.2
<b>GD0568</b>	TonB-dependent receptor	1.B.14.1
<b>GD0569</b>	putative TonB-dependent receptor	1.B.14.1
<b>GD0574</b>	putative porin protein	1.B.5.1
<b>GD0575</b>	putative sulfate-binding protein precursor	3.A.1.6.1
<b>GD0576</b>	putative sulfate transport system permease protein	3.A.1.6.1
<b>GD0577</b>	putative sulfate transport system permease protein cysW	3.A.1.6.1
<b>GD0578</b>	putative sulfate/thiosulfate import ATP-binding protein cysA	3.A.1.6.1
<b>GD0579</b>	putative acriflavin resistance protein	2.A.6.2
<b>GD0580</b>	putative multidrug resistance protein mdtA precursor	2.A.6.2
<b>GD0581</b>	putative toluene efflux pump outer membrane protein ttgC precursor	2.A.6.2
<b>GD0611</b>	putative NADH-quinone oxidoreductase chain M	3.D.1
<b>GD0622</b>	putative lysine exporter protein	2.A.75.1
<b>GD0624</b>	putative sugar transporter	2.A.1.7
<b>GD0627</b>	putative galactose-proton symporter	2.A.1.1
<b>GD0645</b>	putative transporter protein	2.A.1.14
<b>GD0652</b>	putative membrane protein	9.B
<b>GD0653</b>	putative membrane protein	9.B

GD0656	kup1 potassium transport system protein	2.A.72
GD0666	galP Galactose-proton symporter	2.A.1.1
GD0667	conserved hypothetical protein	9.B
GD0670	mdtC Multidrug resistance protein	2.A.6.2
GD0671	hypothetical protein	9.B
GD0674	Iron permease protein	9.A.10
GD0678	mntH manganese transport protein mntH	2.A.55
GD0685	conserved hypothetical protein	9.B.63
GD0689	fecE Iron(III) dicitrate transport ATP-binding protein	3.A.1.14
GD0690	putative ferrichrome ABC transporter permease protein	3.A.1.14
GD0692	TonB-dependent receptor	1.B.14
GD0706	gudP putative glucarate transporter	2.A.1.14
GD0709	major facilitator superfamily (MFS) transporter	2.A.1
GD0714	Preprotein translocase secE subunit	3.A.5
GD0715	putative MFS transporter	2.A.1.3
GD0718	atpH ATP synthase delta chain	3.A.2.1
GD0719	atpA ATP synthase subunit alpha	3.A.2.1
GD0720	atpG ATP synthase gamma chain	3.A.2.1
GD0721	atpD ATP synthase subunit beta	3.A.2.1
GD0722	atpC ATP synthase epsilon chain	3.A.2.1
GD0731	amtB Ammonium transporter	2.A.49
GD0745	hypothetical protein	9.B
GD0761	major facilitator superfamily transporter	2.A.1
GD0790	Secretion protein HlyD	2.A.6
GD0791	putative membrane protein	2.A.6
GD0796	Porin B precursor	1.B.19
GD0803	putative surface antigen	9.B
GD0810	Purine nucleoside permease	2.A.41
GD0813	putative ABC transporter ATP-binding protein	3.A.1
GD0814	putative membrane protein	9.B
GD0821	conserved hypothetical protein	9.B
GD0825	secA Preprotein translocase secA subunit	3.A.5
GD0831	tolQ Protein tolQ	1.A.30
GD0832	tolR Biopolymer transport protein	1.A.30
GD0834	tolB Protein tolB precursor	1.A.30
GD0835	omp Outer membrane lipoprotein omp16	1.B.6.1
GD0838	ftsH Cell division protein ftsH	3.A.16
GD0846	putative membrane protein	9.B
GD0853	putative membrane protein	2.A.7
GD0857	tppB Tripeptide permease tppB	2.A.17
GD0901	putative aquaporin	1.A.8
GD0927	kefB Glutathione-regulated potassium-efflux system protein	2.A.37
GD0934	Porin B precursor	1.B.19
GD0945	hypothetical protein	5.A
GD0950	putative Hemolysin-type calcium-binding	1.C
GD1020	major facilitator transporter superfamily	2.A.1.2
GD1027	Major facilitator superfamily	2.A.1
GD1032	hlyD secretion protein HlyD	2.A.6
GD1034	multidrug efflux transporter protein	2.A.1.3
GD1040	Putative NAD(P)H oxidoreductase	2.A.37
GD1046	trbI conjugation TrbI-like protein	3.A.7
GD1047	trbG conjugal transfer protein trbG	3.A.7
GD1048	trbF conjugal transfer protein; TrbF	3.A.7

GD1049	conjugal transfer protein; TrbL	3.A.7
GD1051	trbJ conjugal transfer protein; TrbJ	3.A.7
GD1052	trbE conjugal transfer protein trbE	3.A.7
GD1053	trbB conjugal transfer protein; TrbB	3.A.7
GD1054	trbC Conjugal transfer protein TrbC	3.A.7
GD1055	Type II/IV secretion system protein	3.A.7
GD1057	putative membrane protein	9.B
GD1059	traG Conjugal transfer protein traG	3.A.7
GD1077	putative membrane protein	9.B
GD1084	pucl Probable allantoin permease	2.A.39
GD1139	putative ABC transport	3.A.1.12
GD1164	mdtC putative Multidrug resistance protein	2.A.6.2
GD1165	putative Purine-cytosine permease	2.A.39
GD1176	putative bicyclomycin resistance protein	2.A.1.2
GD1183	putative mechanosensitive ion channel	1.A.23
GD1193	putative sodium cotransporter	2.A.21.5.10
GD1199	putative MFS transporter	2.A.1
GD1200	probable TonB-dependent receptor	1.B.14.1
GD1213	Hypothetical transmembrane protein	9.B
GD1224	putative phosphoenolpyruvate-protein phosphotransferase	8.A.7.1
GD1225	putative phosphoryl transfer system, HPr	4.A.6
GD1226	Putative phosphotransferase enzyme IIA	4.A.6
GD1231	putative ABC transporter protein	3.A.1.105.2
GD1232	putative ABC transporter protein	3.A.1.105.2
GD1238	putative TonB-dependent Receptor protein	1.B.14.1
GD1238	putative sugar transport protein	2.A.1.6
GD1246	putative ABC transporter protein	3.A.1.15
GD1247	putative ABC transporter protein	3.A.1.15
GD1248	putative ABC transporter protein	3.A.1.15
GD1256	Putative ABC transporter permeases	3.A.1
GD1257	putative Magnesium transport protein corA	1.A.35.1
GD1257	putative ABC transporter, ATP-binding protein	3.A.1.6.1
GD1262	putative TonB-dependent receptor	1.B.14.1
GD1267	putative TonB-dependent Receptor protein	1.B.14
GD1276	putative Outer membrane protein	1.B.39.1
GD1283	putative membrane protein	2.A.66.1
GD1285	putative acriflavine resistance protein	8.A.1.6
GD1286	putative acriflavine resistance protein B	2.A.2.6
GD1326	putative Outer membrane protein oprM	2.A.6
GD1338	putative DNA translocase ftsK	3.A.12.1.2
GD1349	putative major facilitator protein	2.A.1
GD1356	putative sodium symporter protein	2.A.21
GD1380	putative tonB-dependent receptor	1.B.14.1
GD1385	putative antibiotic resistance transport membrane protein	2.A.1.2
GD1390	putative Biotin carboxyl carrier protein of acetyl-CoA carboxylase	3.B.1.1
GD1395	putative Heme exporter protein B	3.A.1.107.1
GD1396	putative cytochrome c biogenesis ATP-binding export protein ccmA	3.A.1.107
GD1398	putative ABC transporter protein	3.A.1
GD1399	putative amino acid ABC transporter, periplasmic amino acid-binding protein	3.A.1
GD1400	putative inner membrane amino-acid ABC transporter permease protein	3.A.1
GD1427	putative Biopolymer transport exbB protein	1.A.30
GD1428	putative Biopolymer transport	1.A.30
GD1429	putative Biopolymer transport exbD2 protein	1.A.30

GD1472	putative permease protein	2.A.40
GD1478	putative TonB-dependent receptor	1.B.14.1
GD1514	putative TonB-dependent Receptor protein	1.B.14.1
GD1529	D-ribose-binding protein	3.A.1.2.1
GD1536	Ribose transport system permease protein rbsC	3.A.1.2.1
GD1537	Ribose transport system permease protein rbsC	3.A.1.2.1
GD1538	Ribose transport ATP-binding protein rbsA	3.A.1.2.9
GD1539	Ribose ABC transporter, substrate-binding protein	3.A.1.2.1
GD1553	Copper resistance protein A	9.A.10
GD1554	Copper resistance protein B	9.A.10
GD1555	putative inner membrane transport protein	2.A.1.2
GD1558	putative heavy-metal ion transporter	2.A.6.1
GD1559	putative resistance protein	2.A.6
GD1560	putative type I secretion outer membrane protein,...	1.B.17
GD1572	putative hexuronate transporter	2.A.1.14
GD1574	Putative ABC transporter permease protein	3.A.1.2
GD1606	Probable potassium transport system protein kup	2.A.72
GD1608	Potassium-transporting ATPase A chain	3.A.3.7.1
GD1609	Potassium-transporting ATPase B chain	3.A.3.7.1
GD1610	Potassium-transporting ATPase C chain	3.A.3.7.1
GD1611	Sensor protein kdpD	2.A.1
GD1643	putative membrane protein	3.A.1
GD1644	putative extracellular solute-binding protein	3.A.1
GD1645	putative ABC-transporter ATP-binding protein	3.A.1
GD1652	Outer membrane protein	2.A.6.2.23
GD1660	outer membrane protein	2.A.6.2
GD1661	putative secretion protein	2.A.6
GD1662	putative transporter protein	2.A.1
GD1664	putative carbohydrate-selective porin	1.B.19.1.1
GD1687	putative oxidoreductase	8.A.5
GD1761	Type II/IV secretion system protein	3.A.7
GD1769	putative membrane protein	2.A
GD1770	putative membrane protein	2.A
GD1783	mscL Large-conductance mechanosensitive channel	1.A.22.1
GD1787	secD Protein-export membrane protein secD	2.A.6
GD1791	uup ABC transporter ATP-binding protein uup	3.A.1.120
GD1798	cydD Transport ATP-binding protein cydD	3.A.1.129
GD1799	cydC Transport ATP-binding protein cydC	3.A.1.129
GD1804	glpF Glycerol uptake facilitator protein	1.A.8
GD1814	tolC Outer membrane protein tolC precursor	1.B.17
GD1817	Putative nitrate transport ATP-binding protein	3.A.1.16
GD1818	putative glycine betaine transport system permease	3.A.1.12
GD1838	putative Na <sup>+</sup> /H <sup>+</sup> antiporter	2.A.36
GD1843	Mechanosensitive ion channel	1.A.23
GD1844	cationic amino acid transporter	2.A.3.3
GD1868	cvrA Cell volume regulation protein A	2.A.36
GD1895	putative transporter protein	2.A.28
GD1896	Magnesium and cobalt efflux protein corC	9.B.37
GD1900	surface antigen	1.B.33
GD2004	modE Molybdenum transport protein modE	3.A.1.8
GD2006	TonB-dependent receptor	1.B.14.1
GD2030	Mg chelatase-related protein	3.A.1
GD2065	Probable allantoin permease	2.A.39.3

GD2073	putative ABC transporter ATP-binding protein	3.A.1.120
GD2074	Outer membrane protein	2.A.6.2
GD2075	mdtO Multidrug resistance protein mdtO	8.A.1
GD2076	mdtN Multidrug resistance protein mdtN	8.A.1
GD2092	cyoD Cytochrome o ubiquinol oxidase protein cyoD	3.D.4
GD2093	cyoC Cytochrome o ubiquinol oxidase subunit 3	3.D.4
GD2094	cyoB Ubiquinol oxidase subunit 1	3.D.4
GD2095	cyoA Ubiquinol oxidase subunit 2 precursor	3.D.3
GD2097	putative membrane protein	9.B
GD2116	petC Cytochrome c	3.D.3
GD2117	petB Cytochrome b	3.D.3
GD2121	btuC Vitamin B12 import system permease protein btuC	3.A.1.14
GD2122	putative periplasmic binding protein	3.A.1
GD2124	ccp Cytochrome c551 peroxidase precursor	5.B.3.1
GD2127	putative transporter protein	9.B
GD2128	secB Protein-export protein secB	9.B
GD2135	Inner membrane transport protein	2.A.41
GD2161	ctaB Protoheme IX farnesyltransferase	3.D.4
GD2162	ctaD Cytochrome c oxidase subunit 1	3.D.4
GD2179	Probable ABC transporter ATP-binding protein	3.A.1.4
GD2196	oxaA Inner membrane protein oxaA	2.A.9
GD2206	yaeT Outer membrane protein assembly factor yaeT precursor	1.B.33.1
GD2225	putative membrane protein	2.A.1
GD2230	galP Galactose-proton symporter	2.A.1.1
GD2248	Outer membrane lipoprotein	1.B.6.1
GD2261	norM Multidrug resistance protein norM (Multidrug-efflux transporter)	2.A.66.1
GD2284	tatC Sec-independent protein translocase protein	2.A.64
GD2294	gabP GABA permease	2.A.3.1
GD2302	Permease protein	2.A.1.40
GD2356	ABC transporter, ATP-binding protein	3.A.1
GD2360	Major facilitator superfamily	2.A.1
GD2374	ftsY Cell division protein ftsY	3.A.5.1
GD2381	ABC-type Fe <sup>3+</sup> transport system ATP-binding component	3.A.1
GD2382	ABC-type Fe <sup>3+</sup> transport system permease component	3.A.1
GD2383	ABC-type Fe <sup>3+</sup> transport system periplasmic component	3.A.1
GD2391	uvrA UvrABC system protein A	3.A.1
GD2411	amiA Probable N-acetylmuramoyl-L-alanine amidase amiA	9.B
GD2415	putative membrane protein	9.B
GD2423	nrgA Ammonia channel precursor (Ammonia transporter)	2.A.49
GD2432	zitB Zinc transporter zitB	2.A.4.1.4
GD2433	fieF Cation-efflux pump fieF	2.A.4.1
GD2434	clcA H <sup>(+)</sup> /Cl <sup>(-)</sup> exchange transporter clcA	1.A.11
GD2457	kup potassium transport system protein	2.A.72.1
GD2463	Major facilitator transporter protein	2.A.1.2
GD2482	ctrD Capsule polysaccharide export ATP-binding protein ctrD	3.A.1
GD2483	ctrC Capsule polysaccharide export inner-membrane protein	8.A.4
GD2484	ctrB Capsule polysaccharide export inner-membrane protein	8.A.4
GD2486	ctrA Putative polysaccharide export protein	1.B.18
GD2487	Inward rectifier potassium channel	1.A.2.1
GD2503	Lysine exporter protein	2.A.75
GD2508	putative inner membrane protein	2.A.1
GD2513	secretion protein HlyD	2.A.6
GD2514	Cobalt-zinc-cadmium resistance protein	2.A.6.1.2



GD2515	efflux pump outer membrane protein	2.A.6.2
GD2517	rbsC Ribose transport system permease	3.A.1.2
GD2518	rbsB D-ribose-binding protein precursor	3.A.1.2
GD2528	fbpC Ferric cations import ATP-binding protein fbpC	3.A.1
GD2536	nuoN NADH-quinone oxidoreductase chain N	3.D.1
GD2537	nuoM NADH-quinone oxidoreductase chain M	3.D.1
GD2538	nuoL NADH-quinone oxidoreductase chain L	3.D.1
GD2539	nuoK NADH-quinone oxidoreductase chain K	3.D.1
GD2540	nuoJ NADH-quinone oxidoreductase chain J	3.D.1
GD2541	nuoI NADH-quinone oxidoreductase subunit I	3.D.1
GD2542	nuoH NADH-quinone oxidoreductase subunit H	3.D.1
GD2543	nuoG NADH-quinone oxidoreductase chain 3	3.D.1
GD2544	nuoF NADH-quinone oxidoreductase chain F	3.D.1
GD2545	nuoE NADH-quinone oxidoreductase chain E	3.D.1
GD2546	nuoCD ADH-quinone oxidoreductase chain C/D	3.D.1
GD2547	nuoB NADH-quinone oxidoreductase chain B	3.D.1
GD2548	nuoA NADH-quinone oxidoreductase chain A	3.D.1
GD2565	putative membrane protein	9.B
GD2571	Succinoglycan biosynthesis transport protein	8.A.3
GD2577	Putative O-antigen transporter	2.A.66
GD2590	putative small multidrug resistance protein	2.A.7
GD2595	Sodium/substrate symport protein	2.A.21
GD2596	putative membrane protein	9.B
GD2601	putative membrane protein	9.B
GD2606	Putative proline/betaine transporter	2.A.1.6
GD2609	major facilitator family transporter	2.A.1
GD2613	xylH Xylose transport system permease protein xylH	3.A.1.2
GD2614	xylG D-xylose transport ATP-binding protein xylG	3.A.1.2
GD2615	xylF D-xylose-binding periplasmic protein precursor	3.A.1.2
GD2616	gumJ Lipopolysaccharide biosynthesis protein	2.A.66
GD2648	major facilitator superfamily (MFS) transporter	2.A.1
GD2653	amino acid permease	2.A.3.1.9
GD2662	kefC Glutathione-regulated potassium-efflux system protein	2.A.37.1
GD2675	potD Spermidine/putrescine-binding periplasmic protein	3.A.1.11
GD2678	putative membrane protein	9.B
GD2679	ctaE Cytochrome c oxidase subunit 3	3.D.4
GD2680	ctaD Cytochrome c oxidase subunit 1	3.D.4
GD2681	ctaC Cytochrome c oxidase subunit 2 precursor	3.D.4
GD2706	4Fe-4S ferredoxin, iron-sulfur binding	5.A.3
GD2710	putative transporter protein	2.A.76
GD2712	rbsB D-ribose-binding protein precursor	3.A.1.2
GD2713	rbsC Ribose transport system permease protein rbsC	3.A.1.2
GD2714	rbsA Ribose transport ATP-binding protein rbsA	3.A.1.2
GD2718	rbsC Ribose transport system permease protein	3.A.1.2
GD2719	ABC transporter, ATP-binding protein	3.A.1.2
GD2721	putative transporter protein	3.A.1.1
GD2722	putative transporter protein	3.A.1.1
GD2723	putative transporter protein	3.A.1.1
GD2724	putative sugar uptake ABC transporter ATP-binding protein	3.A.1.1
GD2727	oprB Porin B precursor	1.B.19.1
GD2728	rbsC Ribose transport system permease protein	3.A.1.2
GD2730	nrtA Nitrate transport protein nrtA	3.A.1.16
GD2731	nrtB Nitrate transport permease protein nrtB	3.A.1.16

GD2732	nrtC Nitrate transport ATP-binding protein nrtC	3.A.1.16
GD2741	metQ D-methionine-binding lipoprotein metQ	3.A.1.24
GD2742	metN D-methionine transport ATP-binding protein metN	3.A.1.24
GD2743	metI D-methionine transport system permease protein	3.A.1.24
GD2744	fucl L-fucose isomerase	3.A
GD2745	rbsB D-ribose-binding periplasmic protein precursor	3.A.1.2.1
GD2746	rbsC Ribose transport system permease protein	3.A.1.2.1
GD2762	putative membrane protein	9.B
GD2780	Membrane lipoprotein	3.A.1.2
GD2781	ABC transporter, ATP-binding protein	3.A.1.2
GD2782	putative amino acid transporter	3.A.1.2
GD2783	putative amino acid transporter	3.A.1.2
GD2789	TonB-dependent receptor	1.B.14.1
GD2792	TonB-dependent receptor	1.B.14.1
GD2797	ABC transporter ATP-binding protein	3.A.1.3
GD2798	putative membrane protein	3.A.1.3
GD2800	Inner membrane amino-acid ABC transporter permease	3.A.1.3
GD2801	artJ Arginine-binding periplasmic protein 2 precursor	3.A.1.3
GD2809	TonB-dependent receptor	1.B.14
GD2828	trbI conjugation TrbI-like protein	3.A.7
GD2830	trbF Conjugal transfer protein trbF	3.A.7
GD2832	trbL conjugal transfer protein; TrbL	3.A.7
GD2834	trbE putative conjugal transfer protein;	3.A.7
GD2836	trbC Conjugal transfer protein trbC	3.A.7
GD2837	Type II/IV secretion system protein	3.A.7
GD2841	traG Conjugal transfer protein traG	3.A.7.1
GD2874	Copper-transporting P-type ATPase	3.A.3.5
GD2875	clcA H(+)/Cl(-) exchange transporter	1.A.11
GD2927	nonribosomal peptide synthetases (NPRS)	1.D.1
GD2929	major facilitator superfamily (MFS) transporter	2.A.1
GD2944	putative multidrug resistance protein	2.A.1.3
GD2945	putative multidrug resistance protein	2.A.1.3
GD2946	multidrug efflux MFS outer membrane protein	2.A.6.2
GD2961	traF putative conjugal transfer protein traF	3.A.7
GD2995	putative multidrug resistance protein	2.A.1
GD3001	trbI putative conjugal transfer protein trbI	3.A.7
GD3003	CagX putative conjugal transfer protein	3.A.7
GD3004	trbF putative conjugal transfer protein	3.A.7
GD3005	trbL TrbL/VirB6 putative conjugal transfer protein	3.A.7
GD3007	trbE putative conjugal transfer protein trbE	3.A.7
GD3008	trbD putative conjugal transfer protein trbD	3.A.7
GD3009	trbC putative conjugal transfer protein	3.A.7
GD3010	trbB conjugal transfer protein trbB	3.A.7
GD3013	traG Conjugal transfer protein traG	3.A.7
GD3028	putative permease	2.A
GD3029	putative permease protein	2.A
GD3053	Putative sulfate transporter	2.A.53
GD3055	ABC transporter substrate-binding protein	3.A.1
GD3057	Putative ABC transport system permease protein	3.A.1
GD3112	putative amino acid permease protein	2.A.3.3
GD3121	lolD Lipoprotein-releasing system ATP-binding protein lolD	3.A.1
GD3122	lolC Lipoprotein-releasing system transmembrane protein	3.A.1
GD3127	nouN NADH-quinone oxidoreductase	3.D.1

GD3128	nouL NADH-quinone oxidoreductase	3.D.1
GD3129	nouK NADH-quinone oxidoreductase	3.D.1
GD3130	nouJ NADH-quinone oxidoreductase	3.D.1
GD3131	nouL NADH-quinone oxidoreductase chain I	3.D.1
GD3132	nouH NADH-quinone oxidoreductase	3.D.1
GD3133	nouG NADH-quinone oxidoreductase	3.D.1
GD3134	nouF NADH-quinone oxidoreductase	3.D.1
GD3135	nouE NADH-quinone oxidoreductase	3.D.1
GD3136	nouD NADH-quinone oxidoreductase	3.D.1
GD3137	nouC NADH-quinone oxidoreductase	3.D.1
GD3138	nuoB NADH-quinone oxidoreductase chain B	3.D.1
GD3139	nouA NADH-quinone oxidoreductase	3.D.1
GD3149	Predicted integral membrane protein	9.B
GD3159	sufD FeS assembly protein SufD	3.A.1
GD3160	sufC FeS assembly protein SufC	3.A.1
GD3161	sufB FeS assembly protein SufB	3.A.1
GD3163	Predicted membrane protein	2.A
GD3170	putative ABC transporter	3.A.1
GD3174	putative membrane protein	8.A.1
GD3181	putative lipoprotein (SmpA/OmlA family)	8.A
GD3195	ffh Signal recognition particle protein	3.A.5
GD3197	phoU Phosphate transport system protein	3.A.1.7
GD3198	pstB Phosphate import ATP-binding protein pstB	3.A.1.7
GD3199	pstA Phosphate transport system permease protein	3.A.1.7
GD3201	pstC Phosphate transport system permease protein	3.A.1.7
GD3205	putative ABC transporter	3.A.1
GD3210	putative oligopeptide transporter	2.A.67
GD3226	putative sugar transporter	2.A.1.1
GD3228	putative thiamine transport ATP-binding protein	3.A.1
GD3235	putative cobalamin binding protein	9.B
GD3236	putative heavy metal transporter	3.A.3.5
GD3237	putative heavy metal transporter	3.A.3
GD3257	putative 4-amino-4-deoxy-L-arabinose transferase	3.A.1
GD3284	Ubiquinol-cytochrome c reductase iron-sulfur subunit	3.D.3
GD3289	Peptidoglycan synthetase ftsI	2.A.7
GD3290	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	2.A.7
GD3291	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	2.A.7
GD3292	Phospho-N-acetylmuramoyl-pentapeptide-transferase	4.A
GD3293	UDP-N-acetylmuramoylalanine--D-glutamate ligase	2.A.7
GD3295	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	2.A.7
GD3296	UDP-N-acetylmuramate--L-alanine ligase	2.A.7
GD3297	UDP-N-acetylenolpyruvoylglucosamine reductase	2.A.7
GD3305	putative membrane protein	9.B
GD3313	putative ABC transporter	3.A.1
GD3315	putative amino acid transporter	3.A.1
GD3316	putative amino acid transporter	3.A.1
GD3321	putative membrane protein	2.A.1
GD3327	putative transporter protein	2.A.1
GD3329	putative ABC transporter	3.A.1
GD3330	putative ABC transporter	3.A.1
GD3355	Outer membrane lipoprotein carrier protein LolA	1.B
GD3361	putative transporter protein	2.A.1.3

GD3363	Galactose-proton symporter	2.A.1.1
GD3365	Gluconate permease	2.A.8
GD3389	putative transporter protein	2.A.1.3
GD3393	two-component sensor histidine kinase/response regulator hybrid protein	2.A.21
GD3394	putative histidine kinase	2.A.21
GD3401	putative membrane protein	9.B
GD3402	putative ABC transporter	3.A.1
GD3417	NADH-quinone oxidoreductase chain B	3.D.1
GD3420	putative membrane protein	2.A
GD3421	putative membrane protein	9.B
GD3422	putative membrane protein	9.B
GD3423	probable Na <sup>+</sup> -translocating NADH-quinone dehydrogenase	3.D.5
GD3424	putative oxidoreductase	8.A.5
GD3431	putative permease	2.A
GD3432	putative permease protein	2.A
GD3433	Organic solvent tolerance protein precursor	1.B.42.1.1
GD3459	Multidrug resistance protein B	2.A.1.3
GD3460	Multidrug resistance protein A	2.A.1.3
GD3461	putative ABC transporter protein, ATP-binding cassette	3.A.1.6.1
GD3475	Multidrug resistance protein mdtA	2.A.6.2
GD3476	Multidrug resistance protein mdtB	2.A.6.2
GD3477	Multidrug resistance protein mdtC	2.A.6.2
GD3478	Putative outer membrane drug efflux protein	2.A.6.2
GD3484	hypothetical membrane protein	9.B
GD3495	Preprotein translocase secY subunit	3.A.5.1.1
GD3512	putative secretion protein	8.A.1
GD3513	putative ABC transporter protein	3.A.1.109
GD3522	putative autotransporter protein	1.B.12
GD3525	conserved hypothetical protein	89.B.20
GD3527	putative outer membrane efflux protein	1.B.17
GD3528	putative carbohydrate-selective porin	1.B.19
GD3528	putative transporter protein	2.A.6.2
GD3529	acrA putative acriflavin resistance protein	2.A.6.2
GD3530	putative two-component sensor histidine kinase	2.A.21
GD3530	putative outer membrane protien	2.A.6
GD3545	putative peptide ABC transport protein	3.A.1.5
GD3546	putative peptide ABC transport protein	3.A.1.5
GD3547	putative transport/binding protein	3.A.1.5
GD3548	cycZ heme exporter protein C	3.A.1.107
GD3548	putative transport/binding protein	3.A.1.5
GD3549	putative peptide bind protein	3.A.1.5
GD3552	cycK Cytochrome c-type biogenesis protein	9.B.14
GD3559	putative membrane protein	9.B
GD3568	putative membrane protein	9.B
GD3569	putative transporter protein	2.A.1
GD3570	putative TonB-dependent receptor	1.B.14
GD3589	putative membrane protein	9.B
GD3602	ABC transporter ATP-binding protein	3.A.1.120
GD3605	putative membrane protein	9.B
GD3753	putative TonB-dependent Receptor protein	1.B.14.1.2
GD3756	putative sulfate trasporter	2.A.53.4.1
GD3765	Carbohydrate-selective porin	1.B.19.1.1
GD3772	Arsenical pump membrane protein	3.A.4.1.1

<b>GD3803</b>	putative amino acid permease	2.A.3.1
<b>GD3812</b>	putative Outer membrane protein	2.A.6.2
<b>GD3823</b>	putative TonB-dependent receptor	1.B.14.1
<b>GD3824</b>	putative tripeptide permease	2.A.17
<b>GD3829</b>	putative phosphate transport system permease protein pstC	3.A.1
<b>GD3830</b>	putative phosphate transport system permease protein pstA	3.A.1
<b>GD3831</b>	Phosphate import ATP-binding protein pstB	3.A.1
<b>GD3835</b>	putative phosphate-binding periplasmic protein precursor	3.A.1.7
<b>GD3836</b>	putative TonB-dependent Receptor protein	1.B.14.1.4
<b>GD3838</b>	putative TonB-dependent Receptor protein	1.B.14.1.3
<b>GD3840</b>	putative phosphate transporter	2.A.20.1
<b>GD3842</b>	putative TonB-dependent Receptor protein	1.B.14.1.3
<b>GD3843</b>	putative Outer membrane protein	2.A.6.2.23
<b>GD3844</b>	putative ABC transporter	3.A.1.3
<b>GD3845</b>	putative transport protein	3.A.1.3
<b>GD3846</b>	putative ABC transport system, ATP-binding protein	3.A.1.2.9
<b>GD3850</b>	putative transport ATP-binding protein	3.A.1.2.1
<b>GD3852</b>	putative transport system permease protein	3.A.1.2.9
<b>GD3853</b>	putative transport system permease protein	3.A.1.2.8
<b>GD3854</b>	conserved hypothetical protein	3.A.1.2.9
<b>GD3886</b>	Secretion protein HlyD	8.A.1.1.1
<b>GD3887</b>	putative Outer membrane protein oprM	2.A.6.2.23
<b>GD3888</b>	putative chloride channel protein	1.A.11.6.1
<b>GD3967</b>	putative outer membrane efflux protein	2.A.6.2.21
<b>GD3969</b>	putative secretion protein HlyD	2.A.6.2.20
<b>GD3972</b>	putative C4-dicarboxylate transport protein	2.A.23.1.2
<b>GD3975</b>	Multidrug resistance protein mdtB	2.A.6.2.12
<b>GD3976</b>	putative secretion protein HlyD	2.A.6.2.20
	putative outer membrane factor	1.B.17
	putative Na(+)/H(+) antiporter	2.A.36.3.2
	putative efflux transporter, RND family. Pseudogene	2.A.6
	putative heavy metal efflux protein. Pseudogene.	2.A.6.1
	putative ABC transporter component	3.A.1
	putative conjugal transfer protein trbG/VirB9	3.A.7.1.1
	putative conjugal transfer protein trbI/VirB10	3.A.7.1.1
	putative conjugal transfer protein	3.A.7.1.1
	putative conjugal transfer protein	3.A.7.1.1
	putative conjugal transfer protein	3.A.7.1.1
	putative conjugal transfer protein	3.A.7.1.1

A quantidade total de proteínas reconhecidas como sistema de transporte e a porcentagem relativa ao número total de frames abertos da leitura (ORFs) foram comparadas a algumas bactérias que tiveram seus genomas divulgados (Tabela IV).

**Tabela IV – Comparação entre o sistema de transportes no genoma da *G. diazotrophicus* com outros genomas publicados.**

Organismos	Proteobacteria	Tamanho do Genoma (Mb)	Total ORFs	Total de Transportadores	No. de Transportadores por Mb genoma	Total Transportadores como Percentagem de ORFs Totais
<i>A. tumefaciens</i>	Alpha	5,3	5419	357	69,25	6,6
<i>B. japonicum</i>	Alpha	9,11	8317	550	60,37	6,6
<i>M. loti</i>	Alpha	7,59	6752	883	57,05	13,1
<i>S. meliloti</i>	Alpha	6,7	6204	826	57,76	13,3
<i>G. oxydans</i>	Alpha	2,92	2664	146	50,00	5,5
<b><i>G. diazotrophicus</i></b>	<b>Alpha</b>	<b>4,3</b>	<b>3854</b>	<b>563</b>	<b>130,93</b>	<b>14,6</b>
<i>R. prowazekii</i>	Alpha	1,1	835	48	43,63	5,7
<i>Azoarcus</i> sp.	Beta	4,74	3992	200	42,1	5
<i>C. violaceum</i>	Beta	4,7	4431	489	104,0	11
<i>X. fastidiosa</i>	Gamma	2,83	2904	78	29,1	2,7
<i>E. coli</i> K-12	Gamma	4,6	4288	354	76,96	8,3

Desse modo, identificamos que os sistemas de transportes representam a maior (14,6%) classe de genes no genoma *G. diazotrophicus*. Vale ressaltar que o grupo que espécies do grupo alfa-Proteobacteria exibem diferenças correspondentes nos índices de transportadores em decorrência a variedade de “estilos de vida”, por exemplos: organismos relacionados a rizosferas, tais como *Sinorhizobium meliloti* e *Mesorhizobium loti* apresentam 13,3% e 13,1%, respectivamente; enquanto patogênicos intracelulares obrigatórios ou simbioses, como *Rickettsia prowazekii* apresenta 5,7% de proteínas transportadores; e, especialmente, que micróbios associados a planta/solo codificam um robusto sistema de transportadores, refletindo uma versatilidade dos organismos a exposição de uma escala larga de substratos diferentes em seu ambiente natural (Ren e Paulsen, 2005).

Um olhar mais apurado no sistema de transporte é vital à compreensão da potencialidade metabólica de organismos completamente seqüenciados. Entretanto, é frequentemente problemático efetuar tais anotações, por causa da ocorrência de famílias grandes e complexas de transportadores, como as transportadores ABC e MFS superfamílias, e a presença de múltiplos parálogos em muitos organismos (Ren e Paulsen, 2005). De fato, identificamos 130 genes pertencentes ao ABC transportadores e 48 da superfamília MFS, além de uma grande quantidade de genes para exportar metais pesados e outros compostos tóxicos, o que é corroborado pelo percentual de transportadores ativos no genoma (Tabela V).

**Tabela V – Principais classes de transportadores anotados no genoma de *G. diazotrophicus*.**

Categoria de Transportador	Números de ORFs anotadas	% em relação a ORFs de transportadores anotadas	% em relação a ORFs totais anotadas*
Transportadores primários ativos	257	45.6	6.4
Transportadores Secundários	167	29.7	4.2
Poros e canais	68	12.1	1.7
Outros	71	12.6	1.8
<b>TOTAL</b>	<b>563</b>	<b>100</b>	<b>14.1</b>

O maior grupo de ORFs (257 ORFs) codificam para transportes ligados a transportadores ativos (45,6%), dos quais 130 são ABC transportadores. Na verdade, ABC transportadores constituem o principal sistema de transporte em *G. diazotrophicus* PAL5 (~23,1% de todos ORFs ligados ao sistema de transporte). Além disso, a maioria (~70%) das ORFs de ABC transportadores em *G. diazotrophicus* PAL5 são dedicadas a aquisição de nutrientes, tais como: fosfatos, fosfonatos, sulfato/molibdato, aminoácidos (glutamato/aspartato, leucina/isoleucina/valine, arginina/ornitina, taurina, histidina), metais (cobre, ferro, magnésio, manganês/zinco, níquel, potássio), nitrato/nitrito, espermidina/putrescina, dipeptídeos/oligopeptídeos, e açúcares (ribose, xilose, arabinose,  $\beta$ -galactose, maltose e glicerol-3, fosfato), ao passo que as restantes estão relacionadas com a multi-resistência. Ressalta-se, ainda, que a grande maioria destes ABC transportadores estão organizados em genes agregados (que provavelmente funcionam como óperons) composto por três componentes: uma componente ATPase e duas proteínas auxiliares.

Dados de proteômica de *G. diazotrophicus* PAL5 em condições de fixação de nitrogênio mostram que os transportadores representam a maioria das proteínas induzidas (30%) durante a fase estacionária (fase S) de crescimento sugerindo uma estratégia de resistência a limitação de nutrientes. Na fase logarítmica (fase L) de crescimento, observa-se a participação de genes envolvidos no metabolismo de DNA e a produção e conversão de energia, que estão de acordo com o alto grau de requerimento energético (Lery LMS, dados não publicados; comunicação pessoal).

Ressalta-se uma “expressão” diferencial de uma proteína com função reguladora, um canal osmosensitivo de  $K^+$  histidina cinase (KdpD). Em *E. coli* e outras bactérias, KdpD/KdpE são, respectivamente, o sensor cinase e o sistema regulador da expressão do operon *kdpFABC*, que codifica um transportador de alta afinidade de  $K^+$ . Em *G. diazotrophicus*, tal sistema poderá ser um estratégia para manter a homeostase citoplasmática e controle do pH (pela extrusão de  $H^+$  durante o influxo de  $K^+$ ), uma vez que as culturas foram cultivadas em pH 5,5 e durante o crescimento pH externo (pHo) caiu para valores muito baixos (~2,7) (Lery LMS, dados não publicados; comunicação pessoal).

Entre as proteínas envolvidas no metabolismo energético uma ATP sintase (*atpD*) foi identificada. A expressão dessa *atpD* por *G. diazotrophicus* sob essas condições crescimento (fase L, e cultivo sem controle de pH) foi interessante e pode fazer parte da mesma estratégia envolvendo o sistema Kdp, acima mencionado, para manter a homeostase citoplasmática e o pH. *G. diazotrophicus* é uma bactéria ácido-tolerante, e sua capacidade de sobreviver em ambientes ácidos podem ser envolver mecanismos de manutenção do pH interno (pHi) com a participação de Kdp e ATP sintase (Lery LMS, dados não publicados; comunicação pessoal).

Além disso, podemos encontrar 40 genes da família RND e 41 genes de secreção tipo IV que estão relacionados com a exportação lipooligosacarídeos e em sistemas de conjugação (transporte de proteínas, DNA e outras macromoléculas); o que talvez tenha permitido a aquisição de uma grande quantidade de genes de transferência laterais, respectivamente; como também a presença de 38 genes responsáveis pela produção de glicosiltransferases, que são genes requeridos para biossíntese de polissacarídeos de superfície celular (principalmente succinoglicanos). Acredita-se que estes apresentem importante papel de reconhecimento celular, no decurso de mecanismos de interação planta-bactéria os quais correspondem a requisitos necessários ao reconhecimento planta-bactéria, e conseqüente ocorrência do processo de fixação biológica de nitrogênio (Krause et al., 2006)

Ressalta-se ainda a ocorrência de 23 genes relacionados com o transporte de ferro (TonB) no genoma de *G. diazotrophicus*. Ferro é um



elemento essencial para algumas proteínas, entre eles o complexo nitrogenase, citocromos e oxidoreductases, relacionadas ou não com cadeia transportadores de elétrons. A presença destas proteínas, associada a genes responsáveis pela síntese de bacteriocinas e antibióticos pode conferir à *G. diazotrophicus* uma vantagem competitiva em relação a outros endosimbiontes vegetais. Se isso for verdade, poderia se pensar em mutantes superexpressando genes RND e/ou outros genes do ABC clusters que condicionariam uma melhor colonização de *G. diazotrophicus* e otimizando a fixação biológica de nitrogênio, sendo assim estudos futuros necessitam ser realizados para confirmar essa proposição.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

## V. REFERÊNCIAS BIBLIOGRÁFICAS

---

- Altschul, S.F. Fundamentals of database searching. In: Trends guide to bioinformatics. Elsevier Science, 1988. p. 7-9.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search program. *Nucleic Acids Res.* 25, 3389-3402.
- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.*, 215(3): 403-410, 1990.
- Amzel, L.M., and Pedersen, P.L. Proton ATPases: structure and mechanism. *Ann. Rev. Biochem.* (1983), 52, 801 – 824.
- Andersen, J.P. (1989). Monomer-oligomer equilibrium of sarcoplasmic Ca<sup>2+</sup>-ATPase and the role of the subunit interaction in the Ca<sup>2+</sup> pump mechanism. *Bioch Biophys Acta.* 988, 47-72.
- Andersen, J.P. and Jorgensen, P.L. (1985). Conformational states of sarcoplasmic reticulum ATPase as studied by proteolytic cleavage. *J Membr Biol.* 88, 187-198.
- Andersen, J.P. and Vilsen, B. (1995). Structure-function relationships of cation translocation by Ca<sup>2+</sup>- and Na<sup>+</sup>, K<sup>+</sup>-ATPases studies by site-directed mutagenesis. *FEBS Letters* 359, 101-106.
- Andersen, J.P. Monomer-oligomer equilibrium of sarcoplasmic reticulum Ca-ATPase and the role of subunit interaction in the Ca<sup>2+</sup> pump mechanism, *Bioch. Biophys. Acta* 988 (1989) 47-72.
- Arai, R., Wriggers, W., Nishikawa, Y., Nagamune, T. e Fujisawa, T. (2004) Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. *Prot. Struct. Funct. Bioinform.*, 57, 829- 838.
- Arata, Y., Nishi, T., Kawasaki-Nishi, S., Shao, E., Wilkens, S. and Forgac, M. (2002). Structure, subunit function and regulation of the coated vesicle and yeast vacuolar(H<sup>+</sup>)-ATPases. *Bioch Biophys Acta.* 155, 71-74.
- Aravind, L., Galperin, M.Y. and Koonin, E.V. (1998). The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *TIBS.* 23, 127-9.
- Auer, M.; Scarborough, G.A. & Kühlbrandt W. Three-dimensional map of the plasma membrane H<sup>+</sup>-ATPase in the open conformation, *Nature* 392 (1998) 840-843.
- Avery, P.T.; MacLeod, C.M. & McCarthy, M. Studies on the nature of the substance inducing transformation of pneumococcus types. *J. Exp. Med.* (1944), 79: 137-158.
- Axelsen, K.B. and Palmgren, M. G. (1998). Evolution of substrate specificities in the P-type ATPase superfamily. *J. Mol Evol.* 46, 64-101.
- Baldani, J.I. & Baldani, V.L.D. History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *An. Acad. Bras. Ciênc.*, 77(3): 549-579, 2005.
- Bateman, A .; . Coin, L. .; . Durbin, R. .; . Finn, R.D. .; . Hollich, V. .; . Griffiths-Jones, S. .; . Khanna, A. .; . Marshall, M .; . Moxon, S .; . Sonnhammer, E.L. .; . Studholme, D.J .; . Yeats, C . & Eddy, S.R. . The Pfam protein families database. *Nucleic Acids Res.*, 32(1): D138-D141, 2004.
- Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J. & Wheeler, D.L. GenBank. *Nucleic Acids Res.*, 31(1): 23-7, 2005.
- Berendsen, H. J. C., D. Van Der Spoel, et al. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, v.91, p.43-56. 1995.
- Bhandari P. & Gowrishankar, J. An Escherichia coli host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer, *J. Bact.* 164 (1997) 4403-4406.

- Bigelow, D.J. & Inesi, G. Contributions of chemical derivatization and spectroscopic studies to the characterization of the Ca<sup>2+</sup> transport ATPase of sarcoplasmic reticulum, *Bioch. Biophys. Acta* 1113 (1992) 323-338.
- Blanco, G. and Mercer, R.W. (1998). Isozymes of the Na-K-ATPase: heterogeneity in structure, diversity in function. *Am J Physiol.* 275, F633-F650.
- Blanco, G.; Koster, J.C. & Mercer R.W. The alpha subunit of the Na,K-ATPase specifically and stably associates into oligomers, *Proc Natl Acad Sci U S A* 91 (1994) 8542-8546.
- Blanco, Y.; Arroyo, M.; Legaz, M.E. & Vicente, C. Isolation from *Gluconacetobacter diazotrophicus* cell walls of specific receptors for sugarcane glycoproteins, which act as recognition factors. *J. Chromatog.*, 1093: 204-211, 2005.
- Blaustein, M.P. and Lederer, W. J. (1999). Sodium/calcium exchange: its physiological implications. *Physiological Reviews.* 79(3),763-854.
- Boeckmann, B. ; Bairoch, A. ; Apweiler, R. ; Blatter, M.C. ; Estreicher, A. ; Gasteiger, E. ; Martin, M.J. ; Michoud, K. ; O'Donovan, C. ; Phan, I. ; Pilbout, S. & Schneider, M. . The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31(1): 365-370, 2003.
- Bowman, B.J and Bowman, E.J. (1986). H<sup>+</sup>-ATPases forma mitochondria, plasma membrane and vacuoles of fungal cells. *J Membran Biol.* 94,83-97.
- Bowman, B.J. Berenski, C.J. & Jung, C.Y. Size of the plasma membrane H<sup>+</sup>-ATPase from *Neurospora crassa* determined by radiation inactivation and comparison with the sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase from skeletal muscle, *J. Biol. Chem.* 15 (1985) 8726-8730.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analyt Bichem.* 72, 248-554.
- Briskin, D. P. & Reynolds-Niesman, I. Change in target molecular size of the red beet plasma membrane ATPase during solubilization and reconstitution, *Plant Physiol.* 90 (1989) 394-397.
- Bull, P.C., Thomas, G.R., Rommens, J.M. and Cox, D.W. (1993) The wilson disease gene is a putative copper transporting p-type ATPase similar to the Menkes gene. *Nat Genet.* 5, 327-37.
- Bult, C.J. et al. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science.* 273, 1058-1073.
- Capieaux, E., Rapin, C., Thinés, D., Dupont, Y. and Goffeau, A. (1993). Overexpression in *Escherichia coli* and purification of an ATP-binding peptide from yeast plasma membrane H<sup>+</sup>-ATPase. *J Biol Chem.* 268, 21895-21900.
- Carlin, A., Shi, W., Dey, S. and Rosen, B.P. (1995)The ars operon of *Escherichia coli* confers arsenical and antimonial resistance. *J. Bacteriol* 177, 981-986.
- Carneiro, A.C.M. Acesso a recursos genéticos, conhecimentos tradicionais associados e repartição de benefícios. *Revista da ABPI* (2007), 88: 3-16.
- Carraro, DM & Kitajima, J.P. Seqüenciamento e bioinformática de genomas bacterianos. *Biotecnologia Ciência & Desenvolvimento - nº 28, 16 -20.* 2002
- Carvalho-Alves, P.C.; Hering, V.R.; Oliveira, J.M.S.; Salinas, R.K. & Verjovski-Almeida, S. Requirement of the hinge domain for dimerization of Ca<sup>2+</sup>-ATPase large cytoplasmic portion expressed in bacteria, *Bioch. Biophys. Acta* 1467 (2000) 73-84.
- Castellani L., Hardwicke, P.M. & Vilbert, P. Dimer ribbons in the three-dimensional structure of sarcoplasmic reticulum, *J. Mol. Biol.* 185 (1985) 579-594.
- Cavalcante, V.A. & Döbereiner, J. A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. *Plant Soil.*, 108: 23-31, 1988.
- Cavières, J.D. Calmodulin and the target size of the (Ca<sup>2+</sup> + Mg<sup>2+</sup>)-ATPase of human red-cell ghosts *Bioch. Biophys. Acta* 771 (1984) 241-244.
- Champeil, P. et al, (1998). Characterization of a protease-resistant domain of the cytosolic portion of sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase. *J Biol Chem.* 273, 6619-6631.

- Chargaff, E.; Vischer E.; Doniger, R.; Green, C. & Misani, F. The composition of the desoxyribose nucleic acid of thymus and spleen. *J. Biol. Chem.* (1949), 177: 405-416.
- Chen, Z., Jones, L.R. O'Brian, J.J., Moorman, J.R. and Cala, S.E. (1998) Structural domain in phospholemman. *Am Heart Assoc.* 367-78.
- Clark, M., R. D. Cramer, et al. Validation of the General-Purpose Tripos 5.2 Force-Field. *Journal of Computational Chemistry*, v.10, n.8, DEC, p.982-1012. 1989.
- Clarke, D.M., Loo, T.W., Inesi, G. and MacLennan, D.H. (1989) Location of high affinity Ca<sup>2+</sup>-binding sites within the predicted transmembrane domain of the Ca<sup>2+</sup>-ATPase. *Nature.* 339, 476-78.
- Coelho-Sampaio, T.; Ferreira, S.F; Benaim, G. & Vyeira, A. Dissociation of purified erythrocyte Ca(2+)-ATPase by hydrostatic pressure, *J. Biol. Chem.* 266 (1991) 22266-22272.
- Costa, C.J.; Gatto, C. & Kaplan, J.H. Interactions between Na,K-ATPase alpha-subunit ATP-binding domains, *J. Biol. Chem.* 278 (2003) 9176-9184.
- Crick, F.H.C. Central dogma of molecular biology. *Nature* (1970), 227: 561-563.
- Cryklaff M., Goormaghtigh, E. & Scarborough, G.A. 2-D structure of the *Neurospora crassa* plasma membrane ATPase as determined by electron cryomicroscopy, *EMBO J.* 14 (1995) 1854-1857.
- Dalton, R. Microfossils: Squaring up over ancient life. 2002 *Nature* 417, 782-784.
- Danko, S., Dahio, T., Yamasaki, K., Kamidochi, M., Suzuki, H. and Toyoshima, C. (2001). ADP-insensitive phosphoenzyme intermediate of sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase has a compact conformation resistant to proteinase K, V8 protease and trypsin. *Febs Letters.* 489, 277-282.
- Deamer, D. W. (1997) The first living Systems: a bioenergetic perspective. *Microbiology Molecular Biology Reviews.* 61, 239-61.
- Delavoie, F.; McIntosh, D.; Henao, F.; Pernazi, G.; Champeil, P.; Stokes, D.L.; Lacapère, J.-J. Projection map of covalently phosphorylated Ca-ATPase from tubular crystals, *Ann. N. Y. Acad. Sci.* 986 (2003) 17-19.
- Delcher, A.L. ; Harmon, D. ; Kasif, S. ; White, O. & Salzberg, S.L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.*, 27(23): 4636-4641, 1999.
- DePristo, M.A., de Bakker, P.I. e Blundell, T.L. (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure*, 12, 831-838.
- Dimroth, P. (1997) primary sodium ion translocating enzymes. *Bioch Biophys Acta.* 1318, 11-51.
- Dohaue, R.A.Jr. and Bebee, R.L. (1999) BL21-SI competent cells for protein expression in *E. coli*. *Focus.* 21, 49-51.
- Donnet, C.; Arystarkhova, E. & Sweadner, K. Thermal denaturation of the Na,K-ATPase provides evidence for alpha-alpha oligomeric interaction and gamma subunit association with the C-terminal domain, *J. Biol. Chem.* 276 (2001) 7357-7365.
- Downs, D.M. & Escalante-Semerena, J.C. Impact of Genomics and Genetics on the Elucidation of Bacterial Metabolism. *Methods*, 20: 47-54, 2000.
- Dux, L. and Martonosi, A.N. (1983). Two-dimensional arrays of proteins in sarcoplasmic reticulum and purified Ca<sup>2+</sup>-ATPase vesicles treated with vanadate. *J Biol Chem.* 258, 2599-2603.
- Dux, L., Pikula, S., Mullner, N. and Martonosi, A.N. (1987). Crystallization of the Ca<sup>2+</sup>-ATPase in detergent-solubilized sarcoplasmic reticulum. *J Biol Chem.* 262, 6439-6342.
- Ewing, B.; Hillier, L.; Wendl, M. & Green, P. Base-calling of automated sequencer traces using Phred I accuracy assessment. *Genome Res.*, 8(3): 175-185, 1998.
- Fagan, M.J. and Saier, Jr. M.H. (1994). P-type ATPases of eukaryotes and bacteria: sequence analyses and construction of phylogenetic trees. *J Mol Evol.* 38, 57-99.
- Falquet, L. ; Pagni, M. ; Bucher, P. ; Hulo, N. ; Sigrist, C.J. ; Hofmann, K. & Bairoch, A. The PROSITE database, its status in 2002. *Nucleic Acids Res.*, 30: 235-238, 2002.
- Falson, P. et al. (1997). The cytoplasmic loop between putative transmembrane segments 6 and 7 in sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase binds Ca<sup>2+</sup> and is functionally important. *J*

- Biol Chem. 272, 17258-17262.
- Fedorov, A.N. & Baldwin, T.O. (1999) Process of biosynthetic protein folding determines the rapid formation of native structure. *J. Mol. Biol.*, 294, 579-586.
- Feezor, R.J.; Cheng, A.; Paddock, H.N.; Baker, H.V. & Moldawer, L.L. Functional genomics and gene expression profiling in sepsis: beyond class prediction. *Clin Infect Dis.*, 41(S7): S 427-435, 2005.
- Felsenstein, J. Department of Genetics, University of Washington, PHYLIP: the PHYLogeny Inference Package. Disponível em: <http://evolution.genetics.washington.edu/phylip.html>. Acesso em: 10 de maio de 2006.
- Férraille, E. and Doucet, A. (1999). Sodium-Potassium-adenositriphosphate-depedent sodium transport in the kidney: hormonal control. *Physiological Reviews.* 81(1),345-418.
- Fichant G, Basse MJ, & Quentin Y. ABCdb: an online resource for ABC transporter repertories from sequenced archaeal and bacterial genomes. *FEMS Microbiol Lett.* 2006; 256(2):333-339.
- Fisher, K. & Newton, W.E. Nitrogenase proteins from *Gluconacetobacter diazotrophicus*, a sugarcane-colonizing bacterium. *Biochim. Bioph. ACTA*, 1750: 154-165, 2005.
- Fleischmann, R.D. et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science.* 269, 496-512.
- Fraser, C.M et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science.* 270, 397-403.
- Frauenfelder, H., Sligar, S.G. e Wolynes, P.G. (1991) The energy landscape and motions of proteins. *Science*, 254, 1598-1603.
- Fuentes-Ramírez, L.E.; Caballero-Mellado, J.; Sepúlveda, J. & Martínez-Romero, E. Colonization of sugarcane by *Acetobacter diazotrophicus* is inhibited by high N-fertilization. *FEMS Microbiol. Ecol.*, 29: 117-128, 1999.
- Gabel, M., Siebers, A., Epstein, W. and Alterdorf, K. (1998). Assembly of the Kdp complex, the multi-subunit K<sup>+</sup>-transport ATPase of *E. coli*. *Bioch Biophys Acta.* 1415, 77-84.
- Gaber, R.F. (1992). Molecular genetics of yeast ion transport. *Int Rev. Cytol.* 137A: 299-353.
- Garcia, E.S. Biodiversidade, Biotecnologia e Saúde. *Cad. Saúde Públ.*, Rio de Janeiro, 11 (3): 491-494, 1995.
- Gatto, C., Wang, A.X. and Kaplan, J.H. (1998). The M4M5 cytoplasmic loop of the Na,K-ATPase, overexpressed in *Escherichia coli*, binds nucleoside triphosphates with the same selectivity as the intact native protein. *J Biol Chem.* 273, 10578-10585.
- Gillis, M.; Kersters, K.; Hoste, B.; Janssens, D.; Kroppenstedt, R.M.; Stephan, M.P.; Teixeira, K.R.S.; Döbereiner, J. & Deley, J. *Acetobacter diazotrophicus* sp. nov., a nitrogen-fixing acetic acid bacterium associated with sugarcane. *Int. J. Syst. Bacteriol.* 39: 361-364, 1989.
- Goffeau, A. et al. (1997). The yeast genome directory. *Nature (suppl)*387, 1-105.
- Goormaghtigh, E. C. Chadwick, G.A. Scarborough, Monomers of the *Neurospora* plasma membrane H<sup>+</sup>-ATPase catalyze efficient proton translocation, *J. Biol. Chem.* 261 (1986) 7466-7471.
- Gordon, D.; Desmarais, C. & Green, P. Automated finishing with autofinish. *Genome Res.* 11(4): 614-625, 2001.
- Gordon, D; Abajian, C . & Green, P . Consed: a graphical tool for sequence finishing. *Genome Res.*, 8(3): 195-202, 1998.
- Hanel, A.M. and Jencks, W.P. (1991). *Biochemistry.* 30, 11320-30.
- Harger, C.; Skupski, M.; Bingham, J.; Farmer, A.; Hoisie, S.; Hraber, P.; Kiphart, D.; Krakowski, L.; McLeod, M.; Schwertfeger, J.; Seluja, G.; Siepel, A.; Singh, G.; Stamper, D.; Steadman, P.; Thayer, N.; Thompson, R.; Wargo, P.; Waugh, M.; Zhuang, J.J . & Schad, P.A.. The Genome Sequence DataBase (GSDB): improving data quality and data access. *Nucleic Acids Res.*, 26: 21-26, 1998.
- Hartley, J.L. Temple, G.F. and Brasch, M.A. (2000). DNA cloning using in vitro site-specific

- recombination. *Gen Res.* 10, 1788-1795.
- Hayashi, Y. Kameyama, K. Kobayashi, T. Hagiwara, E.; Shinji, N. & Takagi, T. Oligomeric structure of solubilized Na<sup>+</sup>/K<sup>+</sup>-ATPase linked to E1/E2 conformation, *Ann. N. Y. Acad. Sci.* 834 (1997) 19-29.
- Hershey, A.D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* (1952), 36: 39-56.
- Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.*, 9(9): 868-877, 1999.
- Inesi, G. & Watanabe, S. Temperature dependence of ATP hydrolysis and calcium uptake by fragmented sarcoplasmic membranes, *Arch. Biochem. Biophys.* 121 (1967) 665-671.
- Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* (1961), 3: 318-356.
- Jahn, T.; Dietrich, J.; B. Andersen, Leidvik, B. Otter, C.; Briving, C.; Kühlbrandt, W. & Palmgren, M.G. Large scale expression, purification and 2D crystallization of recombinant plant plasma membrane H<sup>+</sup>-ATPase, *J. Mol. Biol.* 309 (2001) 465-476.
- Jones, D.T. Threader – Protein fold recognition by optimal protein sequence threading. Disponível em <http://bioinf.cs.ucl.ac.uk/threader/>. Acesso em 10 de maio de 2006.
- Jorgensen, P.L., Nielsen, J.M., Rasmussen, J.K. and Pedersen, P.A. (1998). Structure-function relationships of E1-E2 transitions and cation binding in Na,K-pump protein. *Bioch Biophys Acta.* 1365, 65-70.
- Kafatos, F.C. A Revolutionary Landscape: the Restructuring of Biology and its Convergence with Medicine. 2002, 319(4): 861-867.
- Kalabokis, V.N.; Bozzola, J.J.; Castelli, L. & Hardwicke, P.M.D. A possible role for the dimer ribbon state of scallop sarcoplasmic reticulum. Dimmer ribbons are associated with stabilization of the Ca(2<sup>+</sup>)-free Ca-ATPase *J. Biol. Chem.* 266 (1991) 22044-22050.
- Kanazawa, T. and Boyer, P.D. (1973). *J Biol Chem.* 248, 3163-3172.
- Karner, M. B.; DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409, 507-510 (2001).
- Kaya, S. Abe, K. Taniguchi, K. Yazawa, M. Katoh, T. Kikumoto, M. Oiwa, K. & Hayashi, Y. Oligomeric structure of P-type ATPases observed by single molecule detection technique, *Ann. N. Y. Acad. Sci.* 968 (2003) 278-280.
- Koonin, E.V.; Mushegian, A. R. & Rudd, K.E. Sequencing and analysis of bacterial genomes. *Curr. Biology.*, 6(4): 404-416, 1996.
- Koster, J.C. Blanco, G. & Mercer, R.W. A cytoplasmic region of the Na,K-ATPase alpha-subunit is necessary for specific alpha/alpha association, *J. Biol. Chem.* 270 (1995) 14332-14339.
- Krause, A; Ramakumar, A.; Bartels, D; Battistoni, F.; Bekel, T.; Boch, J.; Böhm, M.; Friedrich, F.; Hurek, T.; Krause, L.; Linke, B.; McHardy, A.C.; Sarkar, A.; Schneiker, S.; Syed, A.A.; Thauer, R.; Vorhölter, F.J.; Weidner, S.; Purhler, A.; Reinhold-Hurek, B.; Kaiser, O. & Goesmann, A. Complete genome of the mutualistic N<sub>2</sub>-fixing grass endophyte *Azoarcus* sp. strain BH72. *Nat. Biotech.*, 24(11): 1385-1391, 2006.
- Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics.* 17(12):1244-5. (2001).
- Kyrpides, N. Genomes OnLine Database (GOLD): a monitor of complete and ongoing genome projects world wide. *Bioinformatics.*, 15: 773-774, 1999.
- Lacapere, J.-J. and Guillian, F. (1993). *Eur. J Biochem.* 211, 117-126.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage t4. *Nature.* 193, 680-85.
- Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination, *Annu. Rev. Biochem.* 58 (1989) 913-949.
- Le Maire, M., Viel, A. and Moller, J.V. (1989) Size exclusion Chromatography and Universal calibration of gel columns. *Anal Biochem.* 177, 50-56.
- Lee, A.G. A calcium pump made visible, *Curr. Opin Struct. Biol.* 12 (2002) 547-554.

- Lee, A.G. and East, J.M. (2001) *Biochem J.* 356, 665-683.
- Lee, S.; Reth, A.; Meletzus, D.; Sevilla, M. & Kennedy, C. Characterization of a Major Cluster of nif, fix, and Associated Genes in a Sugarcane Endophyte, *Acetobacter diazotrophicus*. *J. Bacteriol.* 182(24): 7088-7091, 2000.
- Lemos, M. Workflow para bioinformática. 2004. 224p. Tese de Doutorado em Ciências da Computação. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- Lemos, M.; Seibel, L.F.B. & Casanova, M.A. 2003. Sistema de Anotações em Biosseqüências. Monografia da ciência da computação número 04/03, Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- Levi, V, Rossi, J.P.F.C. Castello, P.R. & Flecha, L.G. Structural significance of the plasma membrane calcium pump oligomerization *Biophys. J.* 82 (2002) 437-446.
- Levi, V. Rossi, J.P.F.C. Castello, P.R. & Flecha, L.G. Oligomerization of the plasma membrane calcium pump involves two regions with different thermal stability, *FEBS Letters* 483 (2000) 99-103.
- Levi, V., Rossi, J.P.F.C., Castello, P.R. and Flecha, L.G. (2000) Oligomerization of the plasma membrane calcium pump involves two regions with different thermal stability. *FEBS Letters*. 483, 99-103.
- Lewis, S.M and Thomas, D. D. (1992). *Biochemistry.* 31, 7381-89.
- Lindorff-Larsen, K., Best, R.B., DePristo, M.A., Dobson, C.M. e Vendruscolo, M. (2005) Simultaneous determination of protein structure and dynamics. *Nature*, 433, 128-132.
- Loureiro, M. M. Montagem e Finalização do genoma de *Gluconacetobacter diazotrophicus* PAL5. 2007. Tese de Doutorado em Química Biológica. Rio de Janeiro. Universidade Federal do Rio de Janeiro-CCS-Instituto de Bioquímica Médica. Rio de Janeiro.
- Lowe, T.M. & Eddy, S.R. . tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5): 955-964, 1997. Luciński, R.; Polcyn, W. & Ratajczak, L. Nitrate reduction and nitrogen fixation in symbiotic association *Rhizobium* - legumes. *Acta Bioch. Pol.*, 49(2): 537-546, 2002.
- Lutsenko, S. & Kaplan, J.H. Organization of P-type ATPases: significance of structural diversity, *Biochemistry* 34 (1995) 15607-15612.
- MacLennan, D.H and Green, N.M. (2000). Pumping ions. *Nature.* 405, 633-34.
- MacLennan, D.H., Rice, W.J. & Green, N.M. The mechanism of Ca<sup>2+</sup> transport by sarco(endo)plasmic reticulum Ca<sup>2+</sup>-ATPases, *J. Biol. Chem.* 272 (1997) 28815-28818.
- Maguire, P. & Ohlendieck, K. Oligomerization of sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase from rabbit skeletal muscle, *FEBS Letters* 396 (1996) 115-118.
- Mahaney, J.E. Froehlich, J.P. & Thomas, D.D. Conformational transitions of the sarcoplasmic reticulum Ca-ATPase studied by time-resolved EPR and quenched-flow kinetics *Biochemistry* 34 (1995) 4864-4879.
- Martonosi, A.N. The structure and interactions of Ca(2+)-ATPase, *Biosci. Resp.* 15 (1995) 263-281.
- Martonosi, A.N. (1996). Structure-functions relationships in the Ca<sup>2+</sup>-ATPase of sarcoplasmic reticulum: facts, speculations and questions for the future. *Biochem Biophys Acta.* 1275, 111-117.
- Masuda, H. and de Meis, L. Phosphorylation of sarcoplasmic reticulum membrane by orthophosphate (1973). *Biochemistry.* 12, 4581-85.
- McIntosh, D.B. (1992). Glutaraldehyde cross-links Lys-492 and Arg-678 at the active site of sarcoplasmic reticulum Ca-ATPase. *J Biol Chem.* 267, 22328-22335.
- McIntosh, D.B., Wooley, D.G., MacLennan, D.H., Vilsen, B. and Andersen J.P. (1999). Interaction of nucleosides with Asp351 and the conserved phosphorylation od sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase. *J Biol Chem.* 274, 25227-25236.
- Merino, J.M. Gutiérrez-Merino, C. & Henao, F. Plausible stoichiometry of the interacting nucleotide-binding sites in the Ca(2+)-ATPase from sarcoplasmic reticulum membranes, *Arch. Biochem. Biophys.* 368 (1999) 298-302.



- Minitz, E. and Guillian, F. (1997). Ca<sup>2+</sup> transport by the sarcoplasmic reticulum ATPase. *Bioch Biophys Acta* 1318, 52-70.
- MMA-Ministério do Meio Ambiente. Primeiro relatório nacional para a convenção sobre diversidade biológica. Brasília: MMA, 1998, 283 p.
- Moller, J.V. Lind, K.E. & Andersen, J.P. Enzyme kinetics and substrate stabilization of detergent-solubilized and membraneous (Ca<sup>2+</sup> + Mg<sup>2+</sup>)-activated ATPase from sarcoplasmic reticulum. Effect of protein-protein interactions, *J. Biol. Chem.* 255 (1980) 1912-1920.
- Moller, J.V.; Jull, B.; le Maire, M. Structural organization, ion transport, and energy transduction of P-type ATPases, *Bioch. Biophys. Acta* 1286 (1996) 1-51.
- Morii, M. Hayata, Y. Mizoguchi, K. & Takeguchi, N. Oligomeric regulation of gastric H<sup>+</sup>,K<sup>+</sup>-ATPase *J. Biol. Chem.* 271 (1996) 4068-4072.
- Morsomme, P. Chami, M. Marco, S. Nader, J. Jetchum, K.A. Goffeau, A. & Rigaud, J.-L. Characterization of a hyperthermophilic P-type ATPase from *Methanococcus jannaschii* expressed in yeast, *J. Biol. Chem.* 277 (2002) 29608-29616.
- Morsomme, P. and Boutry, M. (2000). The plant plasma membrane H<sup>+</sup>-ATPase: structure, function and regulation. *Bioch Biophys Acta.* 1465, 1-16.
- Moszer, I. The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. *FEBS Lett.*, 430: 28-36, 1998.
- Moutin, M.-J., Cuillel, M., Rapin, C., Miras, R., Anger, M., Lompré, A.-M. and Dupont, Y. (1994). Measurements of ATP binding on the large cytoplasmic loop of the sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase overexpressed in *Escherichia coli*. *J Biol Chem.* 269, 11147-54.
- Moutin, M.-J., Rapin, C., Miras, R., Vinçon, M., Dupont, Y. and McIntosh, D.B. (1998). Autonomous folding of the recombinant large cytoplasmic loop of sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase probed by affinity labelling and trypsin digestion. *Eur J Biochem.* 251, 682-690.
- Munõz-Rojas, J. & Caballero-Mellado, J. Population Dynamics of *Gluconacetobacter diazotrophicus* in Sugarcane Cultivars and Its Effect on Plant Growth. *Microb. Ecol.*, 46: 454-464, 2003.
- Muthukumarasamy, R.; Govindarajan, M.; Vadivelu, M. & Revathi, G. N-fertilizer saving by the inoculation of *Gluconacetobacter diazotrophicus* and *Herbaspirillum* sp. in micropropagated sugarcane plants. *Microbiol. Res.*, 161: 238-245, 2006.
- Muthukumarasamy, R.; Rebatí, G.; Seshadri, S. & Lakshminarasimhan, C. *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*), a promising diazotrophic endophyte in tropics. *Curr. Science.*, 83(2): 137-145, 2002.
- NCBI – National Center for Biotechnology Information. Open Reading Frame Finder. Disponível em: <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>. Acesso em 10 de maio de 2006a.
- NCBI – National Center for Biotechnology Information. The Blast Databases. Disponível em: <ftp://ftp.ncbi.nih.gov/blast/db/>. Acesso em 10 de maio de 2006b.
- Nelson, N. (1988) Structure, Function and Evolution of Proton-ATPases. *Plant Physiol.* 86, 1-3.
- Nelson, N. (1994). Energizing porters by proton-motive force. *J. Exp. Biol.* 196, 7-14.
- Nirenberg, M.W & Mattaei, J.H. The dependence of cell-free protein in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci.* (1961), 47: 1588-1602.
- Njus, D. (2000a). Biological Membranes (<http://sun.science.wayne.edu/~bio669/>)
- Njus, D. (2000b). Active transport (<http://sun.science.wayne.edu/~bio669/>)
- Nogueira, E.M.; Vinagre, F.; Masuda, H.P.; Vargas, C.; Pádua, V.L.M.; Silva, F.R.; Santos, R.V.; Baldani, J.I.; Ferreira, P.C.G. & Hemery, A.S. Expression of sugar cane genes induced by inoculation with *Gluconacetobacter diazotrophicus* and *Herbaspirillum rubrisubalbicans*. *Genet. Mol. Biol.*, 24: 199-206, 2001.
- Noji, H. and Yoshida, M. (2000). The rotary machinery of the cell, ATPase synthase. *J. Biol Chem.* 276, 1665-68.
- Orlowski, S. and Champeil, P. The two calcium ions initially bound to nonphosphorylated

- sarcoplasmic reticulum Ca-ATPase can no longer be kinetically distinguished when they dissociate from phosphorylated ATPase toward the lumen (1991). *Biochemistry*. 11331-42.
- Palmgren, M.G. (1998). Proton gradients and plant growth: role of the plasma membrane H<sup>+</sup>-ATPase. *Adv. Bot. Res.* 28, 1-70.
- Palmgren, M.G. and Axelsen, K.B. (1998). Evolution of P-type ATPases. *Bioch Biophys Acta*. 1365, 37-54.
- Paulsen IT, Sliwinski MK, Saier MH Jr. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *J Mol Biol.* 1998 277(3):573-92.
- Pedersen, P.L. and Carafoli, E. (1987) Ion motive ATPases. I. ubiquity, properties, and significance to cell function. *Trends Biochem Sci.* 4, 146-150.
- Pich, U. and Karlisch, S.J.D. (1980). Indications for an oligomeric structure and conformation changes in sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase labelled selectively with fluorescein. *Bioch Biophys Acta*. 626, 255-261.
- Piñon, D.; Casas, M.; Blanch, M.; Fontaniella, B.; Blanco, Y.; Vicente, C.; Solas, M.T. & Legaz, M.E. *Gluconacetobacter diazotrophicus*, a sugar cane endosymbiont, produces a bacteriocin against *Xanthomonas albilineans*, a sugar cane pathogen. *Res. Microbiol.* 53: 345-351, 2002.
- Portillo, F. (2000). Regulation of plasma membrane H<sup>+</sup>-ATPase in fungi and plants. *Bioch Biophys Acta*. 1469, 31-42.
- Prosdocimi, F. et al., Manual do usuário. *Biotecnologia Ciência e Desenvolvimento*, 29 (2002): 12-25.
- Prust, C.; Hoffmeister, M.; Liesegang, H.; Wiezer, A.; Fricke, W.F.; Ehrenreich, A.; Gottschalk, G. & Deppenmeier, U. Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. *Nat Biotechnol.*, 23(2): 195-200, 2005.
- Ptashne, M. (1992). *A Genetic Switch: Phage (Lambda) and higher Organisms* (Cambridge, MA: Cell Press).
- Ranganathan, R. (1994) Evolutionary origins of ion channels. *Proc Natl Acad Sci.* 91, 3484-86.
- Rees, D.C. and Howard, J.B. (1999). Structural bioenergetics and energy transduction mechanism. *J. Mol Biol.* 293, 343-350.
- Ren Q, Paulsen IT. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Comput Biol.* 2005 (3):e27.
- Rensing, C., Ghosh, M., and Rosen, B. P. (1999) families of soft-metal-ion-transporting ATPases. *J. Bacteriology.* 181(19), 5891-5897.
- Rice, W.J. Young, H.S. Martin, D.W. Sachs, J.R. & Stokes, D.L. Structure of Na<sup>+</sup>,K<sup>+</sup>-ATPase at 11-A resolution: comparison with Ca<sup>2+</sup>-ATPase in E1 and E2 states, *Biophys. J.* 80 (2001) 2187-2197.
- Rider, I.S. and Dijkstra, B. (1999). Identification of the Mg<sup>2+</sup>-binding site in the P-type ATPase and phosphatase members of the HAD (haloacid dehalogenase) superfamily by structural similarity to response regulator protein CheY. *Biochem J.* 339, 223-226.
- Rose, S.T.R. The Biology of the future and the future of biology. *J. Mol. Biol.* (2002), 319: 877-884.
- Rosen, B.P., Bhattacharjee, H. and Shi, W. (1995). Mechanism of metalloregulation of anion-translocation ATPase. *J. Bioenerg Biomembr* 27, 85-91.
- Saier, Jr. M. H. Vectorial metabolism and Evolution of transport systems. *J. Bacteriology.* 182 (18), 5029-5035.
- Saier, M.H., Jr., Tran, C.V. and Barabote, R.D. (2006). TCDB: The transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res.* 34: D181–D186.
- Sali, A. MODELLER – A Program for protein structure modeling. Disponível em <http://salilab.org/modeller/manual/manual.html>. Acesso em 10 de maio de 2006.

- Sambrook, J. and Russell, D.W. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 2001.
- Sanchez, R & Sali, A. 1999 ModBase: a database of comparative protein structure models. *Bioinformatics*. 1999; 15(12):1060-1.
- Sanger Institute. Informatics Analysis Software: Alfresco, November 2003. Disponível em <http://www.sanger.ac.uk/Software/Alfresco>. Acesso em 10 de maio de 2006.
- Sanger, F. & Coulson, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* (1975), 94: 441-448.
- Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.*, 74: 5463-5467, 1977.
- Santos, M.B.Q. Montagem e finalização do genoma de *Gluconacetobacter diazotrophicus* PAL5. 2005, 105 p. Dissertação de Mestrado em Química Biológica. CCS-Instituto de Bioquímica Médica. Rio de Janeiro.
- Saraste, M. and Walker, J. E. Membranes proteins Channels, pump and change separators. *Curr Opin Struct Biol* 8, 477-479.
- Scarborough, G.A. (1999) Structure and function of the P-type ATPases. *Curr Opin Cell Biol.* 11, 517-522.
- Schwartz, I. Microbial Genomics: From Sequence to Function. *Genomics*. 6(5): 493-495, 2000.
- Serrano R., Kielland-Brandt, M.C. & Fink, G.R. Yeast plasma membrane ATPase is essential for growth and has homology with (Na<sup>+</sup> + K<sup>+</sup>), K<sup>+</sup>- and Ca<sup>2+</sup>-ATPases, *Nature* 319 (1986) 689-693.
- Serrano, R. (1988). Structure and function of proton translocating ATPase in plasma membrane of plants and fungi. *Biochem Biophys Acta.* 947, 1-28.
- Shortle, D., Simons, K.T. & Baker, D. (1998) Clustering of low-energy conformations near the native structures of small proteins. *Proc. Natl. Acad. Sci. USA*, 95, 11158-11162.
- Simkiss, K. (1998) Cell membranes; barriers, regulators and transducer? *Comp. Bioch Physiol Part A.* 120, 17-22.
- Simmerman, H.K.B. and Jones, L.R. (1998). Phospholamban: protein structure, mechanism of action and role in cardiac function. *Phys Rev.* 78, 921-47.
- Simpson, A.J. et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* (2000), 406: 151-157.
- Sistema Bionotes. *Gluconacetobacter diazotrophicus* community. Disponível em <http://www.bionotes.com.br>. Acessado em 10 de maio de 2006a.
- Sistema Bionotes. Riogene: Anotação do genoma do *Gluconacetobacter diazotrophicus*. Disponível em: <http://www.bionotes.com.br>. Acessado em 10 de maio de 2006b.
- Skou, J.C. (1957) The influence of some cations on an adenosine triphosphate from peripheral nerves. *Bioch Biophys Acta.* 23, 394-401.
- Skou, J.C. (1989) The identification of the sodium-pump as the membrane-bound Na<sup>+</sup>/K<sup>+</sup>-ATPase: a commentary on "the influence of some cations on adenosine triphosphate from peripheral nerves. *Bioch Biophys Acta.* 1000, 435-446.
- Smirnova, I.N., Kasho, V.N. and Faller, L.D. (1998). Inferences about the catalytic domain of P-type ATPases from the tertiary structures of enzymes that catalyze the same elementary reaction. *FEBS Letters.* 431, 309-314.
- Soloz, M. and Vulpe, C. (1996). CPx-type ATPases: a class of P-type ATPases that pump heavy metals. *TIBS* 21, 237-41.
- Stokes, D.L and Green, N.M. (2000) Modeling a dehalogenase fold into the 8A density map for Ca<sup>2+</sup>-ATPase defines a new domain structure. *Biophys J.* 78, 2000.
- Stokes, D.L and Wagenknecht, T. (2000). Calcium transport the sarcoplasmic reticulum. *Eur J Biochem.* 267, 5274-79.
- Stokes, D.L. and Green, N.M. (1990). Three-dimensional crystal of Ca-ATPase from sarcoplasmic reticulum. *Biophys J.* 57, 1-14.
- Stokes, D.L., Auer, M., Zhang, P. and Kühlbrandt, W. (1999). Comparison of H<sup>+</sup>-ATPase and

- Ca<sup>2+</sup>-ATPase suggests that a large conformation change initiates P-type ion pump reaction cycles. *Curr Biol.* 9, 672-679.
- Stryer, L. (1994). *Bioquímica*. 4ª edição. Editora Guanabara Koogan S.A., Rio de Janeiro.
- Sze, H., Li, X. and Palmgren, M.G. (1999). Energization of plant cell membranes by H<sup>+</sup>-Pumping ATPases; regulation and biosynthesis. *The Plant Cell.* 11, 677-689.
- Szepanski, S., Veit, M., Pleschka, S., Klenk, H.D., Schmidt, M.F.G. & Herrier, G. (1994) Posttranslational folding of the influenza C virus glycoprotein HEF - defective processing in cells expressing the cloned gene. *J. Gen. Virol.*, 75, 1023-1030.
- Tang, X., Halleck, M.S., Schlegel, R.A and Williamson, P. (1996). A subfamily of P-type ATPases with aminophospholipid transporting activity. *Science.* 272, 1495-97.
- Tatusova, T.A. and Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Letters.* 174, 247-250.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876-4882. (1997).
- TIGR – The Institute for Genomic Research. RBSFinder. Disponível em <http://www.tigr.org/software/genefinding.shtml>. Acesso em 10 de maio de 2006.
- Toyoshima, C. & Nomura, H. Structural changes in the calcium pump accompanying the dissociation of calcium, *Nature* 418 (2002) 605-611.
- Toyoshima, C. Nakasto, M. Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution, *Nature* 405 (2000) 647-655.
- Toyoshima, C.; Sasabe, H. & Stokes, D.L. Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane, *Nature* 362 (1993) 469-471.
- Tran, C.M and Farley. R.A. (1999). Catalytic activity of an isolated domain of Na,K-ATPase expressed in *Escherichia coli*. *Biophys J.* 77, 258-266.
- Vargas, C.; Pádua V.L.M.; Nogueira, E.M; Vinagre, F.; Masuda, H.P.; Silva, F.R.; Baldani, J.I.; Ferreira, P.C.G. & Hemerly, A.S. Signaling pathways mediating the association between sugarcane and endophytic diazotrophic bacteria: a genomic approach. *Symbiosis*, 35: 159-180, 2003.
- Vasilets, L.A. and Schwarz, W. (1993). Structure-function relationships of cation binding in the Na<sup>+</sup>/K<sup>+</sup>-ATPase. *Bioch Biophys Acta.* 1154, 201-222.
- Vulpe, C., Levinson, B., Whitney, S., Packman, S. and Gitschier (1993). Isolation of a candidate gene for Menkes disease and evidence that it codes a copper-transporting ATPase. *Nat. Genet.* 3, 7-13.
- Watson, J.D. & Crick, F.H.C. Genetical implications of the structure for desoxyribose nucleic acid. *Nature* (1953b), 171: 964-967.
- Watson, J.D. & Crick, F.H.C. Molecular structure of nucleic acids: a structure for desoxyribose nucleic acid. *Nature* (1953a), 171: 737-738.
- Wertén PJ, Remigy HW, de Groot BL, Fotiadis D, Philippsen A, Stahlberg H, Grubmüller H, Engel A. Progress in the analysis of membrane protein structure and function. *FEBS Lett.* 2002; 529(1):65-72.
- Woese, CR. A new biology for a new century. *Microbiol Mol Biol Rev.* 2004; 68(2):173-186.
- Wolfersberger, M. G. (1994) Uniporters, symporters and antiporters. *J. Exp. Biol.* 196, 5-6.
- Wu, C.H. ; . Huang, H. ; . Arminski, L. ; . Castro-Alvear, J. ; . Chen, Y. ; . Hu, Z.Z. ; . Ledley, R.S. ; . Lewis, K.C. ; . Mewes, H.W. ; . Orcutt, B.C. ; . Suzek, B.E. ; . Tsugita, A. ; . Vinayaka, C.R. ; . Yeh, L.S. ; . Zhang, J. . & . Barker, W.C. . The protein information resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res.*, 30(1): 35-37, 2002.
- WU-BLAST – Washington University BLAST Archives-BLAST Manual. Disponível em <http://blast.wustl.edu/doc/blast1.pdf>. Acesso em 10 de maio de 2006.
- Xu, C., Rice, W.J., He, W. and Stokes, D.L. (2002). A structural model for the catalytic cycle of Ca<sup>2+</sup>-ATPase. *J Mol Biol.* 316, 201-211.

- Yamamoto, H., Imamura, Y., Tagaya, T., Fukui, T. and Kawakita, M (1989). Ca<sup>2+</sup>-dependent conformational change of the ATP-binding site of Ca<sup>2+</sup>-transporting ATPase of sarcoplasmic reticulum as revealed by an alteration of the target-site specificity of adenosine triphosphopyridoxal. *J Biochem.* 106, 1221-25.
- Young, H.S. Jones, L.R. & Stokes, D.L. Locating phospholamban in co-crystals with Ca(2+)-ATPase by cryoelectron microscopy, *Biophys. J.* 81 (2001) 884-894.
- Zanin GM, Santana CC, Bon EP, Giordano RC, de Moraes FF, Andrietta SR, de Carvalho Neto CC, Macedo IC, Fo DL, Ramos LP, Fontana JD. Brazilian bioethanol program. *Appl Biochem Biotechnol.* 2000 84-86:1147-1161.
- Zhang P., Toyoshima, C. Yonekura, K. Green, N.M. & Stokes, D.L Structure of the calcium pump from sarcoplasmic reticulum at 8-A resolution, *Nature* 392 (1998) 835-839.

**APÊNDICES:**  
**OUTROS PROJETOS E PUBLICAÇÕES**

## OUTROS PROJETOS E PUBLICAÇÕES

---

Nesta seção serão apresentadas outras publicações que realizadas durante o doutoramento.

Os resultados serão divididos em três seções distintas. A primeira, denominada “*Prospecção Microbiana*” descreve o projeto de prospecção da microbiota em um ambiente impactado por metais pesados.

A segunda parte, denominada: “*Modelagem Comparativa*”, englobará três manuscritos, onde descrevemos a utilização da metodologia de modelagem comparativa para caracterizar estruturalmente enzimas com potencial biotecnológico.

Na terceira parte, intitulada: “*Archaea e Biotecnologia*” englobará duas revisões sobre o domínio Archaea, abrangendo aspectos gerais desde a filogenia e fisiologia, as estratégias de adaptações, quanto aplicações biotecnológicas e a genômica desses microorganismos.

## PARTE 1: PROSPECÇÃO MICROBIANA

### Prospecção Microbiana no Estado do Rio de Janeiro

A biodiversidade hoje existente é o resultado da evolução durante 3,9 bilhões de anos, sobretudo decorrente do surgimento das primeiras formas procarióticas<sup>5</sup> de vida. E, atualmente, as formas procarióticas formam a maior parte da biomassa total do planeta, apresentando uma vasta variedade de espécies, uma enorme diversidade genética, metabólica e fisiológica, e desempenham funções cruciais nos processos biogeoquímicos do planeta, tais como a degradação da matéria orgânica, provisão de componentes atmosféricos, fixação de nitrogênio e fotossíntese.

Apesar de sua grande importância na manutenção da biosfera, estima-se que menos de 1% dos microrganismos existentes no planeta tenham sido caracterizados e descritos (Amann *et al.*, 1990), devido principalmente às limitações dos métodos tradicionais de isolamento e cultivo de microrganismos (Palleroni, 1997).

Todavia, no início da década de 1970, com as metodologias de mapeamento gênico e seqüenciamento de genes e de genomas completos, deu-se um grande impulso a nossa compreensão de como as informações gênicas determinam o funcionamento dos organismos e de sua capacidade adaptativa ao meio ambiente. Ao mesmo tempo, as caracterizações moleculares (metodologias independente de cultivo) (Giovannoni *et al.*, 1990; Olsen *et al.*, 1986; Pace, 1997) permitiram ao estudo da diversidade microbiológica uma melhor compreensão das funções exercidas pelos organismos nos ambientes, bem como de suas interações com outros componentes da diversidade (DeLong, 2005).

No Brasil, apesar de uma crescente contribuição e destaque no cenário científico mundial, há ainda a necessidade de potencializar sua competência científica-tecnológica para que venham a ser melhor conhecidas e aproveitadas

---

<sup>5</sup> Primeiras formas procarióticas – A evidência que temos do início da vida na Terra são restos de cianobactérias que viveram há 3,45 bilhões de anos em rochas sedimentares na Austrália (Rex Dalton, *Nature* 417 (2002): 782 – 784).



todas as potencialidades econômicas de sua biodiversidade. Segundo uma estimativa média, o Brasil soma 20% das espécies da biota mundial, assim, espera-se que uma parcela substancial das novas espécies do mundo serão encontradas no Brasil, especialmente de táxons muito diversos e ainda pouco estudados (por exemplo, bactérias, ácaros e nematódeos de vida livre, himenópteros parasitas) (Tabela I; Lewinsohn & Prado, 2005).

**TABELA A.I – Biodiversidade.** Totais, em milhares, de espécies conhecidas e não conhecidas ou registradas, para grandes grupos taxonômicos no mundo e no Brasil. Reproduzido de Lewinsohn & Prado, 2005.

Grupo Principal (x1000)	MUNDO		BRASIL	
	Valor Preferencial	Média	Limite inferior	Limite Superior
Vírus	400,0	52,6	40,1	70,4
Monera	1000,0	131,4	100,2	175,9
Fungos	1500,0	197,1	150,3	263,9
Protozoários	200,0	26,3	20,0	35,2
Algas	400,0	52,6	40,1	70,4
Plantas	320,0	51,5	48,5	54,5
Nematoda	400,0	52,6	40,1	70,4
Crustacea	150,0	19,7	15,0	26,4
Arachnida	750,0	98,5	75,2	132,0
Insetos	8000,0	1051,0	801,8	1407,6
Mollusca	200,0	26,3	20,0	35,2
Cordata	50,0	7,9	7,2	8,8
Outros	250,0	32,8	25,1	44,0
<b>TOTAL</b>	<b>13620,0</b>	<b>1800,3</b>	<b>1383,6</b>	<b>2394,7</b>

Neste sentido, a diversidade microbiológica brasileira reveste-se de uma importância estratégica ímpar considerando o fato de apresentar-se ainda praticamente inexplorada. Portanto, torna-se necessário a exploração e caracterização dessa diversidade, que pode permitir o isolamento de microrganismos e/ou bioprodutos para implementação de novas tecnologias, como a diminuição da utilização de insumos agrícolas como os fertilizantes e agroquímicos, a redução da poluição ambiental, além de muitos outros produtos derivados de microrganismos. É importante ressaltar que grande parte dos avanços da biotecnologia moderna e da agricultura são derivados das descobertas recentes nas áreas de genética, fisiologia e metabolismo de microrganismos (Cardoso *et al.*, 2003).

Entretanto, as técnicas de microbiologia e microscopia convencionais são insuficientes para responder a estas questões. A maioria dos organismos em amostras ambientais não pode ser detectada através da microscopia convencional, pois eles aderem às partículas de solo tornando-se imperceptíveis. Atualmente, os especialistas reconhecem que apenas uma pequena fração dos microrganismos que ocorre naturalmente foi até agora isolada e caracterizada, e meios de cultura seletivos não são capazes de mimetizar as condições que microrganismos particulares requerem para sua proliferação em seu hábitat natural (Torsvik *et al.*, 1990).

Porém, os recentes avanços nas técnicas de biologia molecular, principalmente pela comparação análise filogenética<sup>6</sup> da subunidade menor (SSU ou 16S) do RNA ribossomal (rRNA) - um marcador filogenético universal - previamente estabelecida por Carl Woese em 1987 (Woese, 1987), produziu uma renovação na microbiologia ambiental. Uma nova subárea, definida como Ecologia Microbiana Molecular, surgiu da integração entre estas diferentes áreas representando um avanço no nosso conhecimento de abundância, diversidade, e função dos procariontes (Rosado *et al.*, 1997).

Em muitos casos, entretanto, o simples conhecimento da diversidade filogenética em um ambiente não é muito útil para o entendimento dos processos de interação metabólica e os fatores que os controlam. Mas, por outro lado, a abordagem molecular pode ajudar na identificação de microrganismos que são ecologicamente relevantes, devido suas abundâncias ou atividades. Esses microrganismos podem ser o tema de estudos detalhados ou alvos para experimentos de cultivo diretos (Clementino, 2006).

Porém são escassos os estudos que utilizam metodologias de Ecologia Molecular no territorial nacional (Peixoto *et al.*, 2002; Piza *et al.*, 2004). A partir desse cenário de imensidão territorial do Brasil e limitado conhecimento da diversidade microbiana (bactérias e arqueas) o grupo de Biologia Molecular do Instituto de Bioquímica, desde 2003, vem investindo esforços na identificação e caracterização de comunidades microbianas no Estado do Rio de Janeiro (Clementino *et al.*, 2007; Vieira *et al.*, 2007a,b), e dentre os ambientes

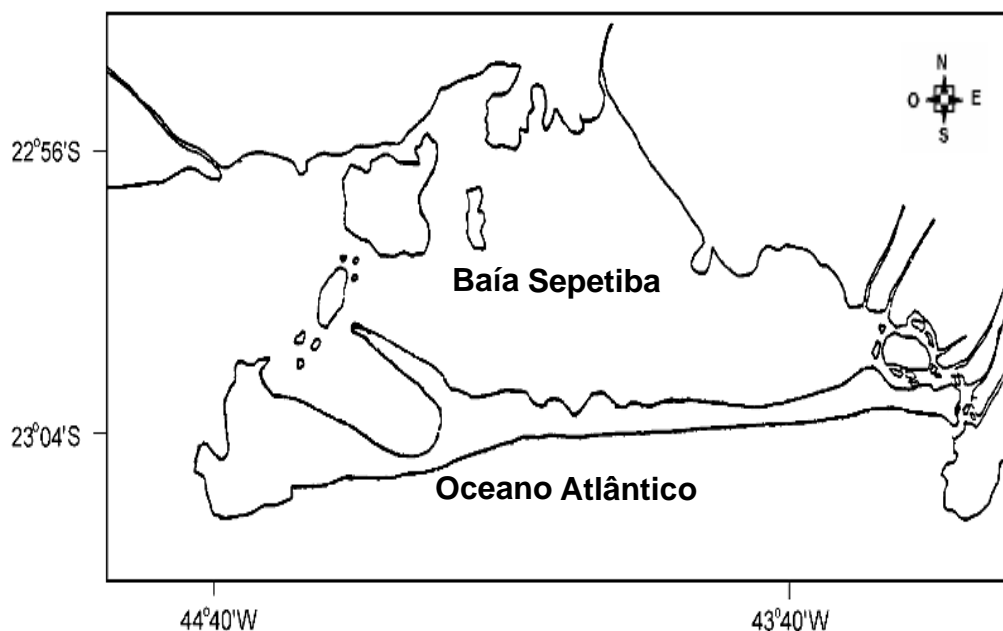
---

<sup>6</sup> Análise filogenética molecular – É um método de reconstrução filogenética que possui como base a comparação entre espécies de seqüências de DNA ou proteínas homólogas.

estudados encontra-se uma área fortemente impactada por metais pesados na Baía de Sepetiba, que descrevemos a seguir.

### **A Baía de Sepetiba**

A Baía de Sepetiba, localizada a oeste da cidade do Rio de Janeiro, tem uma área de 305km<sup>2</sup> e uma bacia de drenagem de 2.617km<sup>2</sup> (Figura A1; Mollisani *et al.*, 2003), que até a década de 60 caracterizava-se pela pesca e pelo turismo. Nos últimos trinta anos sofreu as conseqüências da expansão da região metropolitana do Rio de Janeiro, que resultou em grandes modificações nas estruturas espaciais, socio-econômicas e ecológicas da área em questão.



**Figura A1. Localização da Baía de Sepetiba.**

Com o desenvolvimento industrial e adensamento dos núcleos urbanos, a Baía de Sepetiba tornou-se o segundo principal corpo receptor de efluentes industriais do Estado, já que nela encontramos as mais importantes fontes poluidoras: uma indústria beneficiadora de zinco, uma siderúrgica de grande porte, dezenas de pequenas metalúrgicas, uma usina termoelétrica, um grande aterro sanitário, esgoto urbano não tratado, além de atividades portuárias (Barcelos, 1995; Forte, 1996; Amado Filho *et al.*, 1999).

Dentre os poluentes destacam-se os metais pesados. Altos níveis de zinco (Zn) e cádmio (Cd) são encontrados em organismos bentônicos (algas, moluscos, crustáceos e ostras) da Baía de Sepetiba, sendo esses níveis bem maiores que em outras áreas costeiras do Estado do Rio de Janeiro (Lima et al. 1986; Pfeiffer et al. 1985; Carvalho et al. 1991; Karez et al. 1994; Kurita & Pfeiffer 1999). Também apresenta valores de cromo (Cr) em moluscos e peixes acima da concentração máxima permitida (MPC) fixada pela legislação brasileira, e coincide com o início das operações de beneficiamento de minérios pela Companhia Mercantil e Industrial Ingá no início da década de 60 (Molisani *et al.*, 2003).

Dentre as indústrias poluidoras, destaca-se a Cia Mercantil e Industrial Ingá, empresa instalada em 1963 que processava minério para a produção de zinco de alta pureza, obtido através dos minérios calamínicos ( $2\text{ZnO}\cdot\text{SiO}_2\cdot\text{H}_2\text{O}$ ) e willemíticos ( $\text{Zn}_2\text{SiO}_4$ ) que contém impurezas de outros metais, gerando grandes quantidades de resíduos, como o cádmio e chumbo (Figura A2).



**Figura A2. Cia. Ingá Mercantil.** Visão área dos rejeitos contendo metais pesados e dos tanques de contenção (ao centro; indicado pela seta) e na parte inferior a direita, a Baía de Sepetiba (Porto de Itaguai). Imagem extraída do Google Earth.

Esses resíduos foram acumulados no pátio da empresa, em verdadeira montanha (aproximadamente 1 a 3 milhões de toneladas) cercada por dique de contenção, ao longo de pelo menos 30 anos. Para reduzir a contaminação da Baía pelos resíduos da indústria, foram instalados tanques de acumulação de águas, estação de tratamento de resíduos e águas residuárias e outros

sistemas de controle, apresentando duas principais formas de poluição, a saber: i. a rotineira, que está associada aos efluentes líquidos represados e a disposição inadequada dos rejeitos que acabam sendo carregados para a Baía de Sepetiba; e ii. a acidental, que é aquela decorrente do rompimento dos sistemas de contenção e disposição de resíduos com conseqüências calamitosas que leva a situações extremas de poluição.

Com a falência da empresa no ano de 1998, a indústria foi abandonada e os sistemas de controles desativados, o que acarretou a continuidade do processo de contaminação, que se estende desde o início de suas atividades, aumentando os riscos de ocorrência de acidente ambiental de grandes proporções através do rompimento do dique de contenção dos resíduos.

Tais condições representam barreiras naturais à presença de diversas formas de vida, principalmente macroorganismos (plantas, moluscos e peixes), dado os desafios da seleção natural impostos. No entanto, contribuem para a adaptação de microrganismos a esses ambientes, de tal modo que, o estudo da diversidade de populações microbianas em ambientes extremos tem recebido especial atenção, pois podem permitir o isolamento de microrganismos e/ou bioprodutos para implementação de tecnologias para biorremediação e/ou para a prática industrial (Cardoso *et al.*, 2003).

**Manuscrito 1:** “*Prokaryotic Community of Acidic and Heavy Metal Contaminated Drainage Waters of Zinc Mine Residues in Sepetiba Bay, Brazil*”.

Em 2003, em conjunto com o grupo coordenado pelo Prof. Orlando escrevemos um artigo de revisão intitulado: “*Archaea: Potencial Biotecnológico – Utilização e Aplicação de Arqueas na Biotecnologia*” para a Revista Biotecnologia, Ciência e Desenvolvimento. A partir desse trabalho, interessávamos fazer uma prospecção da ocorrência de representantes de Domínio Archaea no Estado do Rio de Janeiro, mediante a criação de bibliotecas para o gene 16S rDNA, estimulados pela falta de dados na literatura como também pela potencialidade da utilização desses microorganismos em processos biotecnológicos. Um dos ambientes inseridos nessa prospecção foi o pátio da massa falida da Companhia Mercantil e Industrial Ingá.

A Cia. Ingá beneficiava minérios de zinco, próximo ao Porto de Sepetiba, para obtenção de lingotes de zinco. Ao longo de aproximadamente 30 anos de atividade na região, a empresa acumulou no seu pátio subprodutos ricos em metais pesados (como Zn, Pb e Cd) desse processo de beneficiamento mineral. No final da década de 90, a Cia. Ingá decretou falência, relegando a população do Estado do Rio de Janeiro (especialmente aos moradores do município de Itaguaí) o maior passivo ambiental do Estado.

Contudo, para iniciarmos o projeto de bioprospecção na Cia. Ingá necessitávamos estabelecer uma colaboração com o grupo do Prof. Dr. João Alfredo Medeiros (Instituto de Química/UFRJ), que é o coordenador do monitoramento ambiental da Cia. Ingá (indicado pelo Ministério Público Estadual).

Estudos preliminares foram realizados pelo grupo do Professor Alexandre Rosado utilizando métodos convencionais de cultivo e isolamento (Martins et al., 2003), contudo não abordando os sistemas de contenção de águas residuárias da montanha de rejeitos do pátio da empresa. Esses autores identificaram somente membros do domínio Bacteria, não vislumbrando a presença de representantes do domínio Archaea.

**PROKARYOTIC COMMUNITIES OF ACIDIC AND HEAVY METAL  
CONTAMINATED DRAINAGE WATERS OF ZINC MINE RESIDUES IN  
SEPETIBA BAY, BRAZIL.**

Almeida, W.I.<sup>1\*</sup>, Cardoso, A.M.<sup>1</sup>, Silveira, C.B.<sup>1</sup>, Vieira, R.P.<sup>1</sup>, Costa, R.G.<sup>1</sup>, Gonzalez, A.S.M.<sup>2</sup>, Paranhos, R.<sup>2</sup>, Medeiros, J.A.<sup>3</sup>, F.A.D. Freitas<sup>4</sup>, Albano, R.M.<sup>4</sup> and Martins, O.B.<sup>1</sup>

<sup>1</sup>Instituto de Bioquímica Médica, <sup>2</sup>Biologia, <sup>3</sup>Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

<sup>4</sup>Departamento de Bioquímica, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil.

**\*Correspondence to:**

Wellington Inácio de Almeida  
Instituto de Bioquímica Médica  
Universidade Federal do Rio de Janeiro  
Centro de Ciências da Saúde - Bloco D, subsolo, sala 5.  
21941-590 - Rio de Janeiro – Brazil.  
Tel.: +5521 25626751  
Fax: +5521 22708647  
E-mail: [welington@bioqmed.ufrj.br](mailto:welington@bioqmed.ufrj.br)

**Running title:** Archaea and Bacteria in Extreme Environments

## Abstract

Molecular approaches were used to characterize the phylogenetic composition and structure of microbial communities of acid mine drainage (AMD) waters associated with Zn mine tailings in Itaguai city (near Sepetiba Bay, Rio de Janeiro, Brazil). In addition, we analyzed the physical, chemical and microbiological parameters linked to natural and anthropogenic polluted environments in Sepetiba Bay. Sequencing and phylogenetic analysis of 16S rRNA genes revealed that approximately half of the phylotypes fell into lineages that are poorly represented by cultured organisms or have thus far been represented by only a few environmental sequences. LIBSHUFF statistics shows that bacteria libraries are significantly different while the archaeal libraries are more similar between metal impacted vs. pristine habitats. Thus, the present study extends our knowledge of the microbial diversity populating AMD environments and the detection of a substantial number of uncultured Archaea and Bacteria suggest that heavy metal contaminated waters may contain novel prokaryote phylotypes.



## Introduction

Prokaryotes can be found in highly acidic waters which originate naturally by volcanic activity or artificially by mining. Acid mine drainage (AMD) water is a growing worldwide environmental problem due to active and abandoned mines (Johnson, 2003). In particular, acidic drainage is originated by the exposure of sulphidic minerals to  $O_2$ , resulting in the formation of soluble sulphates. These minerals with high ferrous iron content become oxidized in contact with water producing ferric ions and  $H_2$ . This process is also often enhanced by microbial catalysis (Tan et al., 2007). These ions when leached into streams acidify the water which can normally reach pH values below 3. Additionally, other metal ions such as Zn, Hg, Ni, Cr, Cd, Cu, Mn, Al, As and Pb occur in AMD waters at final concentrations far above the permissible levels.

In Rio de Janeiro state, the Cia. Mercantil e Industrial Ingá, a zinc smelter plant located in Madeira Island (Sepetiba Bay), has been disused and causes environmental problems. The plant went bankrupt and was shutdown in 1998. When the mine was active, acidic water, after being partially treated, was actively pumped and discharged in the bay. When government funding was withdrawn, all treatment ceased and the rising mine water, with a pH of 2.8 and a high zinc content, was passively discharged. The abandoned industrial area accumulated three million tons of solid residues. During a rainfall in 1998, about 50 million liters of acidic metal laden water were accidentally released in the bay, and after this, water containing one hundred tons of Zn per year has been discharged (Medeiros, unpublished data). Sepetiba bay harbors about 400 industries, including metallurgical and petrochemicals and fishing and tourist activities. Studies in the bay have identified significant changes in sedimentation

rates, concentration of inorganic pollutants and most recently, eutrophication (ref). Of great concern is the impact from increasingly large heavy metal loads being discharged from industrial emissions into the bay, to the atmosphere and to nearby rivers. Consequently, high concentrations of heavy metals, especially Zn, can be observed in the bay's compartments (animal and plant tissues, water and sediments).

Concentrations of Zn in natural seawater range from 3.9 to 4.9  $\mu\text{g. L}^{-1}$ . About 37% of the Zn is in organic form and the main inorganic ion species are  $\text{ZnOH}^+$ ,  $\text{Zn}^{2+}$ , and  $\text{ZnCO}_3$ . The low pH is probably responsible for metal mobilization and enhanced toxicity in water runoff from mines. Acidic, metal contaminated water generated by the microbial accelerated dissolution of sulfide minerals, are a source of Sepetiba Bay pollution with Zn concentrations ranging from 4.4 to 1800  $\mu\text{g. L}^{-1}$  (Molisani et al., 2003), and the effect of this pollution can be lethal or sub lethal to components of the biota, such as phytoplankton, zooplankton, benthos, fish, birds, mammals and finally to humans (Coelho and Fonseca, 1986).

However, the diversity of microorganisms and the roles they play in situ remain largely unknown as normally only a small proportion of the bacteria present in these extreme environments are readily culturable. Thus, the purpose of the research described here was to investigate the bacterial and archaeal diversity in water samples from the contaminated settlement tanks of the bankrupt Cia. Ingá and compare them to natural bay environments through sequencing of 16S rDNA genes in clone libraries. We also discuss the microbiological relationship between abiotic parameters found in these environments. Prokaryotic diversity was studied by phylogenetic analysis and

rarefaction curves. Coverage of libraries was accessed by LIBSHUFF statistics. Knowledge of these alloctonous microbes will help in understanding the role of microorganisms in AMD generation and ultimately, how these may be exploited for environmental remediation.

## MATERIALS AND METHODS

**Site Description.** The zinc smelter plant (22°54' S and 43°54' W), is located at Itaguaí City, Rio de Janeiro state, southeastern Brazil (Figure 1). It covers an area of approximately 120 km<sup>2</sup>. The climate in this region is typically hot-humid tropical, with an annual mean precipitation of 1,400 mm and evaporation of 960 mm, which contribute to the total fluvial load of heavy metals in Sepetiba Bay. Emission rate estimates suggest that the main Zn and Cd source to Sepetiba Bay are effluents from the Zn-Cd ore processing plant, which added 24 t y<sup>-1</sup> of Cd and 3,660 t y<sup>-1</sup> of Zn into the bay's basin until 1998, when the plant closed. How much of this contribution is still reaching the bay after the closure of the plant is unknown (Molisani *et al.*, 2003).

Superficial water samples (1m deep) were collected in different sites at contaminated tanks (A1 and A2) and at an underground leakage point from the main tank (CM) at Cia. Mercantil e Industrial Ingá on March 16<sup>th</sup>, 2006. In addition, two control habitats were sampled; one at a relatively pristine freshwater environment, Marrecas Lagoon (LM) and one marine sample in Sepetiba Bay (BS). Samples were immediately kept on ice for transport to the laboratory and processed for analysis.

**Site characterization of physical, chemical and microbiological parameters.** Microbial abundance was determined by flow cytometry after nucleic acid staining with syto13 fluorochrome at 2.5  $\mu\text{M}$  in samples fixed with 2% paraformaldehyde (del Giorgio et al., 1996; Gasol and del Giorgio, 2000). Microbial activity was analyzed by a  $^3\text{H}$ -leucine method (Kirchman et al., 1985; Smith and Azam, 1992). Physical and chemical parameters were determined according to standard oceanographic methods (Grasshoff et al., 1983; Parsons et al., 1984). Electric conductivity, pH and Eh were measured *in situ* using specific electrodes. Heavy metal measurements were carried out by atomic absorption spectrophotometry (Atomic Absorption Spectrometer, Perkin Elmer 3100).

**DNA extraction.** Water was filtered and DNA was subsequently prepared by a standard method (Somerville et al., 1989). Briefly, water filtration in 3.0  $\mu\text{m}$  ester-cellulose filter, was performed to capture symbionts of microalgae, zooplankton or prokaryotes attached to particulate materials. The remaining free-living planktonic microbes were concentrated on a Sterivex-filter (0.22  $\mu\text{m}$ ) (Millipore). About 50  $\mu\text{l}$  of freshly prepared lysozyme (1 mg/ml) was added to filter units containing 1.8 ml of lysis buffer, and incubated at 37°C for 45 min. Then 50  $\mu\text{l}$  of freshly prepared proteinase K (0.2 mg/ml) and 200  $\mu\text{l}$  of 10% sodium dodecyl sulfate (SDS) were added to the filter, and incubated at 55 °C for 1 h. Lysates were removed from the filter units with sterile 3 ml syringes, each rinsed with 1 ml of lysis buffer, and incubated for 15 min. The rinse buffer and lysates were pooled. Crude lysates were extracted twice with phenol-chloroform-isoamyl alcohol (25:24:1, pH 8.0) and once with chloroform-isoamyl

alcohol (24:1). The nucleic acids in the aqueous phase were precipitated with 2 volumes of ethanol at  $-20^{\circ}\text{C}$  overnight, centrifuged at 7,000 rpm for 15 min, washed with 70% ethanol, dried and then dissolved with 100  $\mu\text{l}$  TE (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). The quality of DNAs was checked in agarose gels stained with Syber Green after electrophoresis and the image digitalized with Storm Image Scanner (GE Health Care).

**PCR amplification of 16S rDNA gene.** Four archaeal and three bacterial 16S rDNA gene libraries were constructed from free-living planktonic prokaryotes collected at distinct environments (A1, A2, CM, LM and BS). PCR was performed in 50  $\mu\text{L}$  reaction mixtures (2.5 mM MgCl<sub>2</sub>, 0.2 mM dNTPs, 50 pmol of each primer, 2.5 U of Platinum *Taq* DNA polymerase, and PCR buffer (Invitrogen) for approximately 100 ng of sample DNA. Three oligonucleotides were used: universal prokaryote reverse primer 907ABR (5'-TTTGAGTTTMTTAATGCC-3') (Weisbeurg et al., 1991), and either universal Bacteria forward primer 27BF (5'-AGAGTTTGATCATGGCTCAG-3') (Lane, 1991) or universal Archaea forward primer 21AF (5'-TTCCGGTTGATCCTGCCGGA-3') (DeLong, 1992). PCR amplification began with a 5-min denaturing step at  $94^{\circ}\text{C}$ ; followed by 30 cycles of  $94^{\circ}\text{C}$  for 1.30 min,  $50^{\circ}\text{C}$  for 1.30 min, and  $72^{\circ}\text{C}$  for 2 min. The final cycle was an extension at  $72^{\circ}\text{C}$  for 10 min.

**Cloning and sequencing of PCR products.** PCR bands with the expected molecular weight (about 880 bp) were excised from the 1% agarose gel, and eluted using the GFX-DNA and gel band purification kit (GE Health Care).

Amplicons were cloned into the plasmid vector pGEM-T (Promega), according to the manufacturer's recommendations, and transformed into DH-10b *Escherichia coli* competent cells. Clones from each library were submitted to sequence analysis. DNA from each clone was prepared (400 ng) and sequences were obtained after sequencing reactions by capillary electrophoresis on a MegaBace1000 using the DYENamic dye terminator cycle sequencing kit (GE Healthcare). Chromatograms were transformed into Fasta format sequences with Phred software (Edwing et al., 1998). A total of 140 valid sequences with Phred score  $\geq 20$  were compared with sequences in the Ribosomal Database Project II (RDP II). Chimeric sequences were identified and removed using CHECK-CHIMERA (Cole et al., 2003). Alignments with representative archaeal sequences obtained at GenBank databases were carried out using ClustalX program (Thompson et al., 1997) to compare sequences.

**Sequence trimming and analysis.** Plasmid vector sequences and regions with quality below 20, according to Phred scores at the 3'- and 5'-ends of the 16S rRNA gene inserts were removed. Thus, only high quality sequences were used in the analyses. Sequences with less than 300 bp were excluded from subsequent analyses. The sequences obtained were compared with those in the GenBank by the BLAST search tool. Phylogenetic trees were calculated by the Kimura two-parameter model (Kimura, 1980) and the neighbor-joining algorithm (Saitou & Nei, 1987) using the MEGA software. One thousand bootstraps were performed to assign confidence levels. The diversity of the phylotypes was further examined using the DOTUR software to obtain

LIBSHUFF statistics (Schloss et al., 2004; Schloss and Handelsman, 2005) and rarefaction analysis (Hurlbert, 1971; Heck et al., 1975). The partial 16S rRNA gene sequences reported in this study was submitted to GenBank/NCBI database under accession numbers (EF463932 to EF464065 and EF517812 to EF517814).

## RESULTS AND DISCUSSION

Sepetiba Bay (Fig.1B) is a semi-closed water body connected to the sea in the east and supports 40 km<sup>2</sup> of mangrove forests, which are best developed at the inner eastern end and play role in providing nursery and feeding for the bay's fisheries. Although the industrialization process in the area is relatively recent, occurring mostly in the past 40 years, pollution levels have already matched levels of polluted industrial areas of Europe and North America (Lacerda and Pfeiffer, 1987).

In previous reports, our group has been investigating the microbial diversity in natural and impacted habitats from Rio de Janeiro state, Brazil, a tropical region of South America (Clementino et al., 2007; Vieira et al., 2007a, b), and thus we decided to extend our study to this heavy metal impacted area. For this purpose, we have constructed the 16S rDNA libraries and analyzed physicochemical parameters of different tank waters from Cia. Ingá and from the surrounding environment in Sepatiba bay.

**Chemical and microbiological parameters.** Temperature values ranged from 27 to 30 °C, and were typical of summer superficial waters in this region (Table I). All heavy-metal polluted waters showed low pH values, varying from 2.0 to

4.5. Nutrient analyses of nitrogen compounds showed tenfold higher concentrations of ammonia in tanks of Cia. Ingá, when compared with LM freshwater. Phosphorus was present at high concentrations at the underground leakage point CM, and decreased to nM levels in natural freshwater in LM (Table I). The acidic solution contained high levels of zinc and other heavy metals like cadmium and lead were detected at low concentrations (Table II).

Prokaryotic counts varied from  $0.5 \times 10^6$  cells.mL<sup>-1</sup> in pristine LM freshwater to  $7.3 \times 10^6$  cells.mL<sup>-1</sup> in the heavily polluted tanks A1, A2 and CM. The main tank A1 presented low prokaryotic activity, around 10 ngC.L<sup>-1</sup>.h<sup>-1</sup> in opposition to the high value of 270 ngC.L<sup>-1</sup>.h<sup>-1</sup> found in LM (Figure 2). Higher SPM and chlorophyll values observed in CM station support the higher specific metabolic activity in the moderately oxygenated and nutrient rich waters of the underground leakage point from the main tank (Table I) when compared to A1 and A2 tanks. Accordingly, the two control habitats (BS and LM) presented the highest metabolic activity.

The food chain in the tanks is very constrained and it is probably exclusively microbial. Primary productivity in the tanks is the sum of photosynthetic and chemolithotropic activity, especially because no sources of organic matter and nitrate seem to be available in the tanks. Primary producers also need some inorganic compounds and elements (P, Ca, K, Fe, Cu, Zn or Mg). These can be found in the tanks in concentrations that exceed these requirements (Tables I and II), as has been described for other heavy metal polluted acidic waters, such as Tinto River (Almis et al., 2001).

These metal impacted Ingá tank waters did not display macroscopic forms of life, but prokaryotic cells were quite abundant, probably due to high



nutrient availability and low grazing. However, specific production shows that microbial metabolic rates are higher in natural communities and lower in metal impacted assemblages. Furthermore, tolerance to high zinc levels is linked to low metabolic rates in these microbial assemblages (Fig. 2C).

Additionally, there are large passive residuals (Fig. 1C), which are AMD producers in its basin. These are persistent in the environment unless natural or man-made processes are initiated to counteract the acidity. Another concern is the presence of heavy metals in the sediment which are distributed by dragging operations during maintenance and enhancement of Sepetiba Harbor decreasing the environmental quality. Both situations create a serious threat to the large biodiversity and the potential economic uses of the bay.

AMD remediation is generally performed through chemical treatment of the source, thereby preventing its formation. The most common methods used for AMD remediation are passive (wetland) methods, conventional active methods of adding limestone, quicklime or soda ash to promote metal precipitation and other biological routes such as biosorbents and rotating biological contactors (Wildeman 1993, Groudev et al. 1999, Shutes 2001, Brown et al. 2002). Recently, the AMD bioremediation approach which incorporates both biological and chemical methods is an alternative process to conventional environmental cleanup technologies (Macaskie et al. 1995, Boswell et al. 1998, Bonthron et al. 2000). However, if the source is an abandoned area, it is sometimes impossible to gain access to the areas that are creating and contributing to acid mine drainage problems.

In this sense, remediation of tailing ecosystem, either by phytoextraction or by direct microbial bioremediation *in situ* greatly depends on microorganisms

which are well known to change metal speciation and mobility and, consequently, affect the performance of plants in such environments (Gadd, 2000; Lasat, 2002). Therefore, adjustment of the physical and chemical parameters into a range that is more conducive to microbial activity should be an effective support for biological recovery (Anderson et al., 2003).

However, until now, most microbial information in tailing ecosystems comes from acid mine drainage (Southman and Beveridge, 1992). Microbial community structure has been described only in a few categories of tailings, such as sulfidric mine waste and uranium tailings, by culture-based methods (Southman and Beveridge, 1992; Schippers et al., 1995). Therefore, a comprehensive study of the microbial composition in mine tailings still remains to be uncovered due to the great chemical and physical heterogeneity of tailings around the world, with little or no information about mining or AMD areas in tropical regions (Dudka and Adriano, 1997).

**Clone libraries, sequencing and diversity.** A total of 300 bacterial clones and 100 archaeal clones were randomly selected and sequenced. Chimeric sequences were identified and removed using the CHECK-CHIMERA software at RDP II (<http://rdp.cme.msu.edu>). A total of 118 good sequences, 88 bacterial and 30 archaeal sequences were analyzed. Valid sequences with phred score  $\geq 20$  were used for database query with online BLAST search in GenBank (<http://www.ncbi.nlm.nih.gov>). Phylogenetic analysis indicated that the 88 bacterial clones were related to uncultivated environmental microorganisms (Figure 3). About 90% of the clones found in the contaminated waters were related to freshwater microorganisms. Others were associated with acidophilic

species related to *Acidocella*, *Acidosphaera* and *Acidiphilium* (Figure 3). We observed several bacterial clones in the impacted tanks that were phylogenetically affiliated to *Acidithiobacillus* or *Leptospirillum* genera, which are well-known contributors to acid generation in AMD environments (data not shown) and Proteobacteria, belonging to species capable of ferric iron and sulfate reduction (Bond et al., 2000; Norris et al., 2002; Brofft et., 2002; Brunnel et al., 2005). However, most of the sequences that corresponded to acidophilic organisms were closely related to the genera *Acidocella*, *Acidosphaera* and *Acidiphilium*, which are capable of reducing ferric to ferrous iron by respiring reduced carbon compounds (Johnson and Bridge, 2002). Interestingly, the main photo-pigment of *Acidiphilium* species is bacteriochlorophyll **a** chelated with zinc (Zn-BChl **a**) in place of magnesium as the central metal (Wakao et al., 1996; Hiraishi et al., 1998) because Zn-BChl **a**, is more stable than Mg-BChl **a** under acidic conditions (Wakao et al., 1996; Kobayashi et al., 1998).

All archaeal clones were closely related to uncultivated species (Figure 4). Members of the Crenarchaeota phyla dominated zinc and acidic impacted site A1 and natural habitats LM and BS. We found four Euryarchaeota phylotypes affiliated to *Methanoregula boonei* (Braeur et al., 2006) only in LM fresh water, which is suggestive of eutrophication and methanogenesis processes happening in Marrecas Lagoon. We identified three marine archaea in BS that are affiliated to a crenarchaeon phylotype retrieved in a previous study in Guanabara Bay (Vieira et al., 2007b), suggesting that this uncultured planktonic crenarchaeon is ubiquitous in the coastal zone of Rio de Janeiro state.

The LM environment showed the CFB group (Cytophaga, Flavobacterium and Bacteriodetes) (Group PFW1 in Fig. 3), *Legionella* genera and *Enterobacteriaceae* which are not found in the tanks. These are widely distributed in different habitats and show close association with human and animal hosts. So, their presence suggests domestic sewage input pollution. Interestingly, the Zn contaminated tanks showed distinct bacterial communities when compared to LM, while the archaeal community shared members between impacted tanks and LM pristine freshwater (Figures 3 and 4).

Rarefaction analysis is a procedure to compare biodiversity from different habitats with distinct sample or clone numbers. Rarefaction analysis at 97% cutoff stringency indicates that the number of sequenced clones from each Archaea and Bacteria library were insufficient to cover the total microbial diversity (Figure 5A). Our clone libraries demand higher clone numbers to make a visible representation of the total environmental microbial diversity. On the other hand, at 80% cutoff, rarefaction curves suggest a reasonable coverage of the main representative groups of each habitat investigated (Figure 5B).

We also performed a quantitative comparison between Archaea (Figure 6A) and Bacteria (Figure 6B) libraries of these impacted and natural habitats by LIBSHUFF statistic procedure. Our analysis shows that bacterial diversity is a bit higher than archaeal. The diversity of Bacteria in acidic, heavy-metal A1, A2 and CM impacted waters is lower than in LM, a natural environment. LIBSHUFF quantitative statistics shows that bacteria libraries are significantly different ( $p=0.001$ ); on the other hand, archaeal libraries are more similar ( $p=0.046$ ) between metal impacted and pristine habitats. Taken together, our data suggest

that bacterial phylotypes were specific from each environment and that they are in accordance with our phylogenetic analyses.

This study showed that the high level of biodiversity found in the impacted tanks underlines an interesting level of adaptation of different prokaryotic microorganisms to low pH and high concentrations of heavy metals. Further characterization of this AMD system should give us insight not only into the role of the different microorganisms living there, but also into their unusual physiology, their possible biotechnological applications, and their contribution to environmental pollution.

### **Final considerations**

Perhaps one of the most intriguing aspects of mining problems is that the majority of AMD formed is the direct result of microbial activity. We are studying the linkage between the geochemistry of AMD and its prokaryotic community structure in order to understand the diversity of organisms that tolerate the heavy metal and acidic extreme environments. The presence of a high number of sequences related to uncultivable prokaryotes suggests the possibility to isolate new species. We can speculate that the alloctonous metal resistant microbiota could be applied to AMD remediation process. This study has shown an intriguing biodiversity of indigenous chemolithoautotrophic and heterotrophic Archaea and Bacteria living in an abandoned mining area.

### **Acknowledgements**

We acknowledge the Genome Sequencing Johanna Döbereiner Facilities IBqM/UFRJ, the Limnology Laboratory of UFRJ for the access to liquid cintilator.

We are grateful to Dr. J. P. Torres (IBCCF/UFRJ) for manuscript review. This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ).

## TABLES

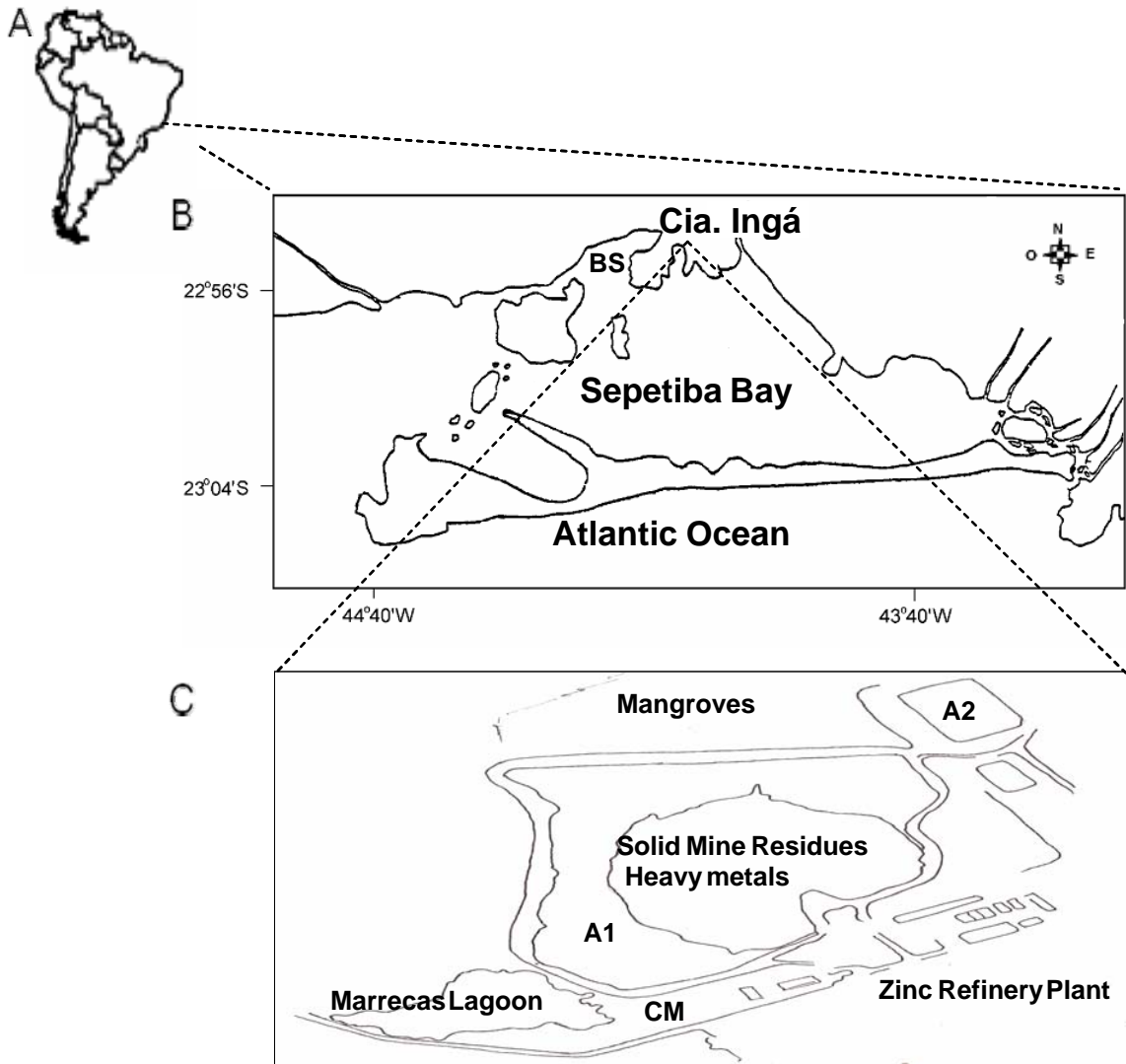
Table I: Physicochemical data of Cia. Ingá and Sepetiba Bay sampling sites.

Parameters	Sites				
	A1	A2	CM	LM	BS
Temperature (°C)	29	30	30	28	27
pH	3.5	3.5	1.5	5.5	7.5
Salinity (S)	0.17	0.16	0.55	0.20	2.37
Chlorophyll (µg/L)	1.02	1.26	7.38	2.78	4.01
Feofitin (µM)	0.74	1.74	16.7	2.16	3.44
Orthophosphate (µM)	0.89	0.09	0.92	0.11	0.78
Phosphorous (µM)	2.61	0.27	7.48	0.92	1.28
Ammonia (µM)	1.62	3.70	2.38	0.25	1.11
Nitrite (µM)	0.58	0.18	3.58	0.16	0.45
Silicate (µM)	630	387	30.6	51.2	51.7
SPM (mg/L)	18.8	12.2	32.4	4.00	49.8

Table II: Heavy metal and mineral data of Cia. Ingá sampling sites.

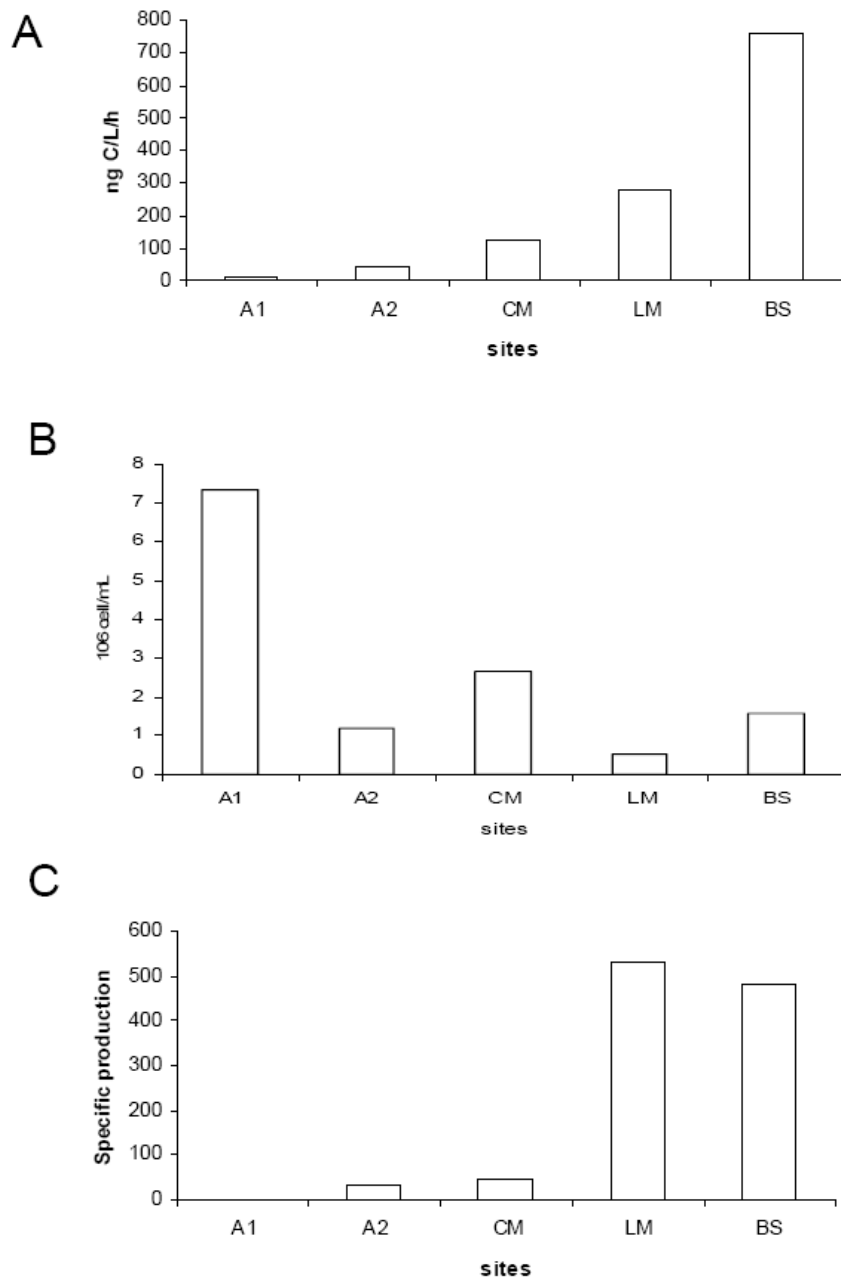
Parameters	Sites			
	A1	A2	LM	CM
Turbidez NTU	7	2.9	1.5	4.85
Conductivity (µS)	6,58	9,450	3,500	13,700
Eh (mV)	703	765	770	790
Zn (mg/L)	1,950	2,975	1,200	3,950
Cd (mg/L)	10	17.5	4.5	27.5
Pb (mg/L)	2	3	1	4
Mg (mg/L)	700	800	425	1,100
Ca (mg/L)	400	500	325	775

**FIGURES**



**Figure 1. Map of Sepetiba Bay, and schematic view of the zinc smelter plant.**





**Figure 2. Microbiologic parameters**

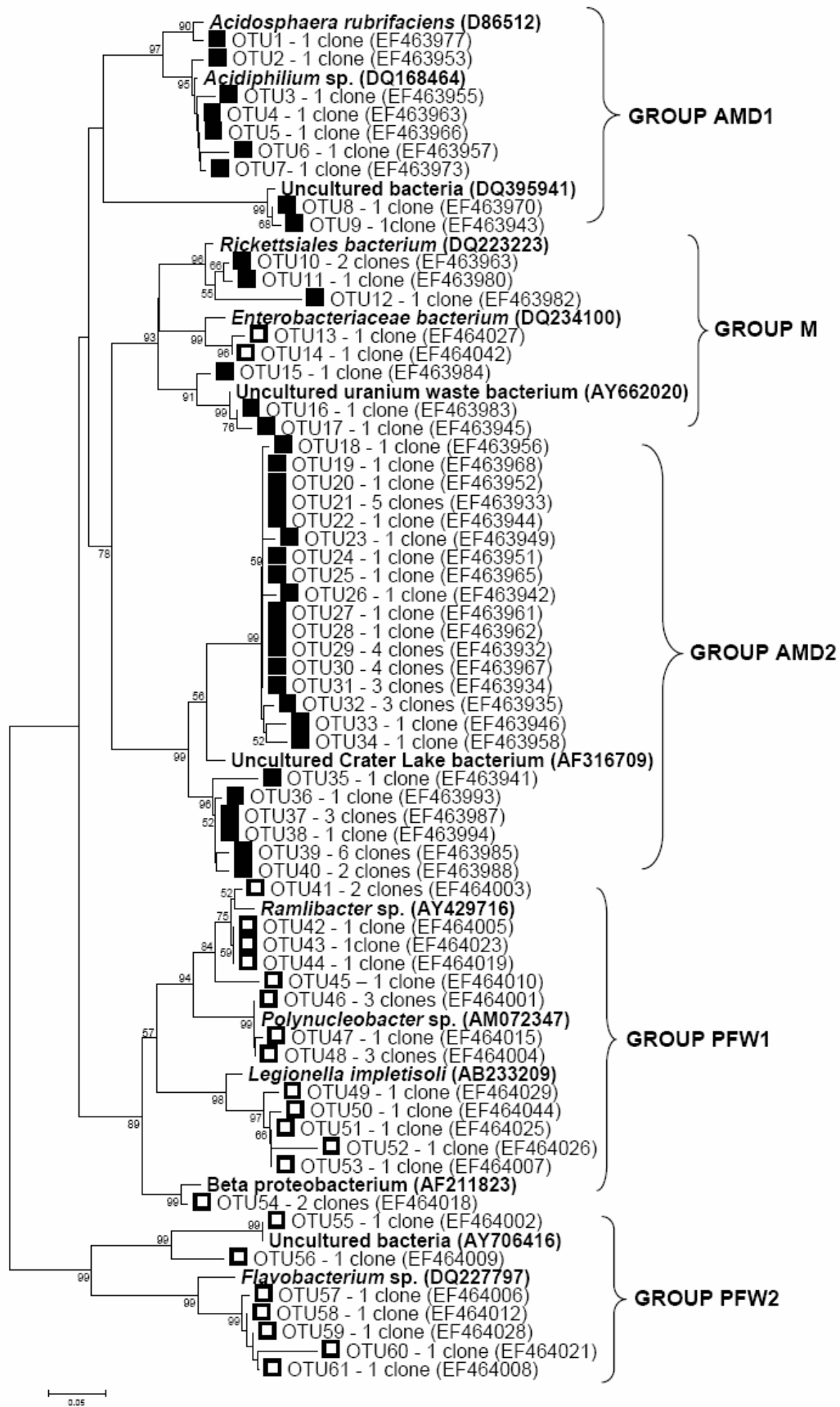


Figure 3. Bacterial phylogenetic tree

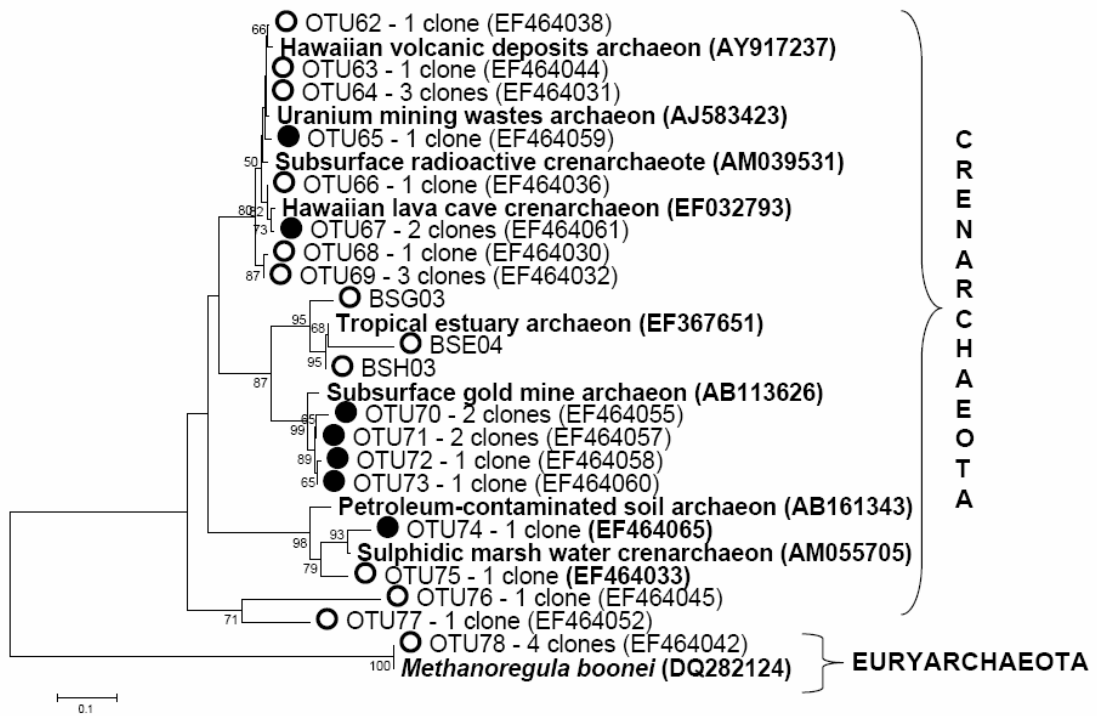


Figure 4. Archaeal phylogenetic tree

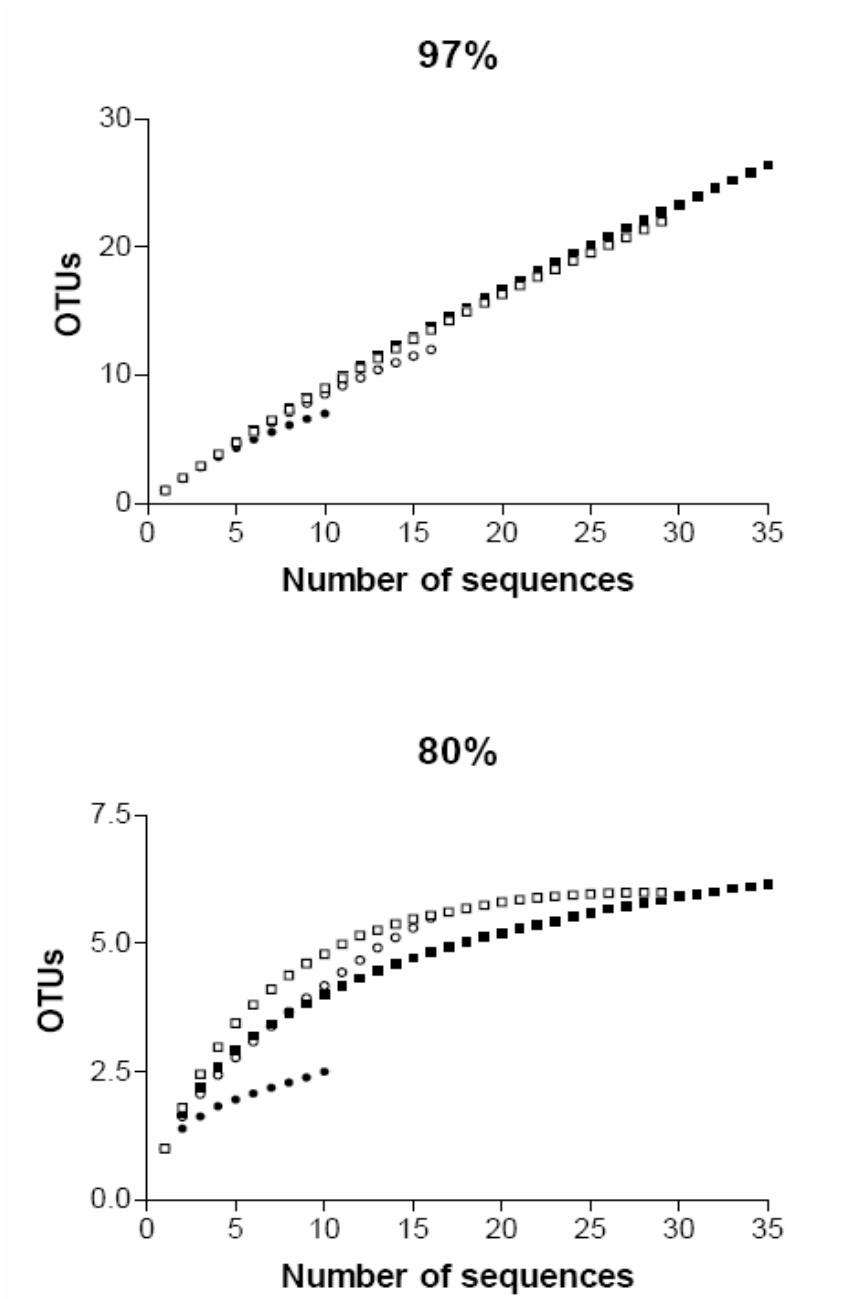


Figure 5. Rarefaction

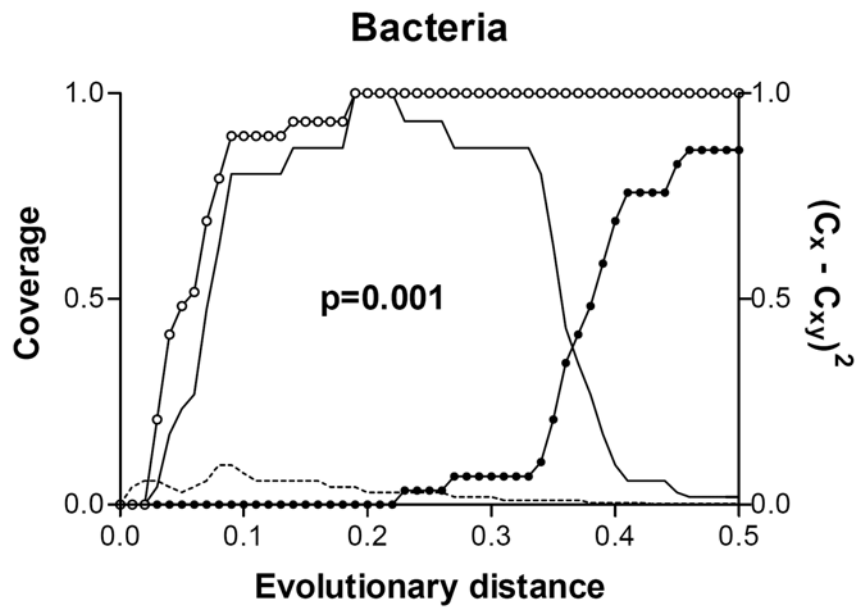
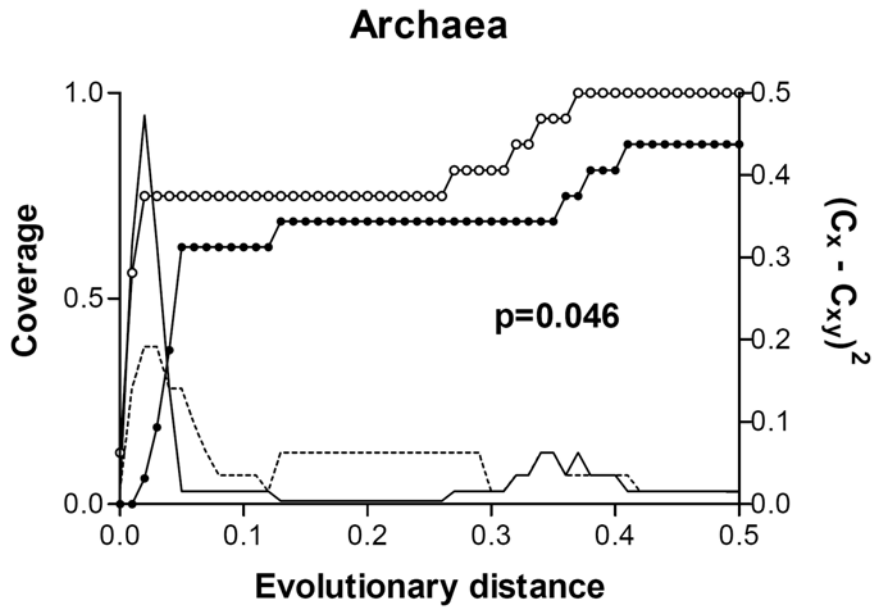


Figure 6. LIBSHUFF statistic

## LEGENDS

**Figure 1:** Sampling sites location: (A) South America; Brazil (B) Cia. Inga position near Sepetiba Bay (BS); (C). Zinc smelter plant area with sampling sites A1, A2, CM, and Marrecas Lagoon (LM).

**Figure 2:** Microbiological parameters of prokaryotic communities obtained at A1, A2, CM, LM and BS stations. (A) Heterotrophic uptake of  $^3\text{H}$ -leucine (B) Abundance of free-living prokaryotes (C) Specific production; prokaryotic metabolic rates

**Figure 3:** Neighbor-joining Bacteria phylogenetic tree construction from partial 16S rDNA sequences. One access number for each OTU is shown on the tree. Bootstrap values (1000 replicates) higher than 50% are shown. Scale bar represents the 5% substitution percentage.

**Figure 4:** Neighbor-joining Archaea phylogenetic tree construction from partial 16S rDNA sequences. One access number for each OTU is shown on the tree. Bootstrap values (1000 replicates) higher than 50% are shown. Scale bar represents the 5% substitution percentage. (A) Crenarchaeota and (B) Euryarchaeota.

**Figure 5:** Rarefaction curves comparative analysis. Archaea ( $\circ, \bullet$ ) and Bacteria ( $\square, \blacksquare$ ) from natural, LM and BS ( $\square, \circ$ ) vs. acidic and metal impacted environments A1, A2 and CM ( $\blacksquare, \bullet$ ). Clusterization stringency at (A) 97% or (B) 80% identity.

**Figure 6:** Results of LIBSHUFF comparisons. Homologous (○) and heterologous (●) coverage curves for 16S rRNA gene libraries from (A) Archaea and (B) Bacteria.

## REFERENCES

- . Amado Filho, GM, Karez, CS, Andrade, LR, Yoneshigue-Valentin, Y, Pfeiffer, WC (1997) Effects on growth and accumulation of zinc I six seaweed species. *Toxicol and Environ* 37:223-228
- . Amann, RI, Stromley, J, Devereux, R, Key, R, Stahl, DA (1992) Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms. *Appl Environ Microbiol* 58: 614-623
- . Andrade, L, Gonzalez, AM, Araújo, FV, Paranhos, R (2003) Flow cytometry assessment of bacterioplankton in tropical marine environments. *J Microbiol Methods* 55: 841-850
- . Clementino, MM, Fernandes, CC, Vieira, RP, Cardoso, AM, Polycarpo, CR, Martins, OB (in press) Archaeal diversity in naturally occurring and impacted environments from a tropical region. *J Appl Microbiol* doi: 10.1111/j.1365-2672.2006.03230.x
- . Cole, JR, Chai, B, Marsh, TL, Farris, RJ, Wang, Q, Kulam, SA et al (2003) The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* 31: 442-443
- . Correa Junior JD, Allodi S, Amado-Filho GM, Farina M. (2000) Zinc accumulation in phosphate granules of *Ucides cordatus* hepatopancreas. *Braz J Med Biol Res.* 33:217-21.
- . DeLong, EF, Karl, DM, (2005) and Genomic perspectives in microbial oceanography. *Nature* 15: 336-342
- . Edwing, B, Hillier, L, Wendl, M, Green P (1998) Basecalling of automated sequencer traces using phred I Accuracy assessment. *Gen Res* 8:175-185
- . Giovannoni, SJ, Stingl, U (2005) Molecular diversity and ecology of microbial plankton. *Nature* 15: 343-348
- . Grasshoff, K, Kremling, K, Erhardt, M (1999) *Methods of seawater analysis.* Wiley-VCH Verlag Germany 3rd Ed 600 pp
- . Hallberg KB, Coupland K, Kimura S, Johnson DB (2006) Macroscopic streamer growths in acidic, metal-rich mine waters in north wales consist of novel and remarkably simple bacterial communities. *Appl Environ Microbiol.* 72:2022-30.
- . Heck, Jr KL, van Belle, G, Simberloff, D (1975) Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56: 1459-1461
- . Hiraishi, A, Matsuzawa, Y, Kanbe T, Wakao N (2000) *Acidosfera rubrifaciens* gen. nov. sp. Nov., an aerobic bacteriochlorophyll-containing bacterium isolated from acidic environments. *Intern J Syst and Evol Microbiol* 50:1539- 1546
- . Huang, X, Madan, A (1999) CAP3: A DNA sequence assembly program. *Gen Res* 9: 868-877
- . Hurlbert, SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52: 577-586
- . Johnson DB, Hallberg KB. (2003) The microbiology of acidic mine waters. *Res Microbiol.* 154:466-73.
- . Johnson, DB, Hallberg KB (2005) Acid mine drainage remediation options: a review *Sci Tot Environ* 338: 3-14



- . Junior, RGSL, Araújo, FG, Maia, MF, Pinto, ASSB (2002) Evaluation of heavy metals in fish of the Sepetiba and Ilha Grande Bays, Rio de Janeiro, Brazil. *Environ Res* 89: 171-179
- . Kamjunke N, Tittel J, Krumbeck H, Beulker C, Poerschmann J. (2005) High heterotrophic bacterial production in acidic, iron-rich mining lakes. *Microb Ecol.* 49:425-33.
- . Kimura, M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111-20
- . Kirchman, DL, K'nees, E, Hodson, R (1985) Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. *Appl Environ Microbiol* 49: 599-607
- . Kudo, Y, Nakajima, T, Miyaki, T, Oyaizu, H (1997) Methanogen flora of paddy soils in Japan. *FEMS Microbiol Ecol* 22: 39-48
- . Kumar, S, Tamura, K, Jakobsen, IB, Nei, M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245
- . Lopez-Archilla AI, Marin I, Amils R. (2001) Microbial Community Composition and Ecology of an Acidic Aquatic Environment: The Tinto River, Spain. *Microb Ecol.* 41:20-35.
- . Lohr AJ, Laverman AM, Braster M, van Straalen NM, Roling WF. (2006) Microbial communities in the world's largest acidic volcanic lake, Kawah Ijen in Indonesia, and in the Banyupahit river originating from it. *Microb Ecol.* 2:609-18.
- . Molisani, MM, Martins, RV, Machado HHM, Bidone ED, Lacerda LD (2003) Environmental changes in Sepetiba Bay, SE Brazil. *Reg Environ Change* 4:17-27
- . Murray, AE, Preston, CM, Massana, R, Taylor, LT, Blakis, A, Wu, K, DeLong, EF (1998) Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island Antarctica. *Appl Environ Microbiol* 64: 2585-2595
- . Nicomrat D, Dick WA, Tuovinen OH. (2006) Assessment of the microbial community in a constructed wetland that receives acid coal mine drainage. *Microb Ecol.* 51:83-9.
- . Parsons, TR, Maita, Y, Lalli, CM (1984) A manual of chemical and biological methods for seawater analysis. Oxford Pergamon Press 173 pp
- . Rappe, MS, Giovannoni, SJ (2003) The uncultured microbial majority. *Ann Rev Microbiol* 57: 369-394
- . Robertson, CE, Harris, JK, Spear, JR, Pace, NR (2005) Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol* 8: 638-642
- . Saitou, N, Nei, M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425
- . Schleper, C, Jurgens, G, Jonuscheit, M (2005) Genomic studies of uncultivated Archaea. *Nat Rev Microbiol* 3: 479-488
- . Singleton, DR, Furlong, MA, Ratbhun, SL, Whitman WB (2001) Quantitative comparisons of 16S rRNA Gene Sequence Libraries from Environmental Samples. *Appl Environ Microbiol* 67: 4374-4376
- . Somerville, CC, Knight, IT, Straube, WL, Colwell, RR (1989) Simple rapid method for direct isolation of nucleic acids from aquatic environments. *Appl Environ Microbiol* 55: 548-554

- . Takai K, Moser DP, DeFlaun M, Onstott TC, Fredrickson JK. (2001) Archaeal diversity in waters from deep South African gold mines. *Appl Environ Microbiol.* 67:5750-60.
- . Tan GL, Shu WS, Hallberg KB, Li F, Lan CY, Huang LN. (2006) Cultivation-dependent and cultivation-independent characterization of the microbial community in acid mine drainage associated with acidic Pb/Zn mine tailings at Lechang, Guangdong, China. *FEMS Microbiol Ecol.* 59:118-26.
- . Thompson, JD, Gibson, TJ, Plewniak, F, Jeanmougin, F, Higgins, D G (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24: 4876-4882
- . Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature.* 4:37-43.
- . Wendt-Potthoff K, Koschorreck M. (2002) Functional groups and activities of bacteria in a highly acidic volcanic mountain stream and lake in Patagonia, Argentina. *Microb Ecol.* 43:92-106.
- . Zang, HB, Yang MX, Shi W, Zheng Y, Sha T, Zhao ZW (2007) Bacterial diversity in mine tailings compared by cultivation and cultivation-independent methods and their resistance to lead and cadmium. *Microb Ecol.* 10.1007/s00248-007-9229-y
- . Zeng, Y, Li, H, Jiao, N (in press) Phylogenetic diversity of planktonic archaea in the estuarine region of East China Sea. *Microbiol Res*

## PARTE 2: MODELAGEM COMPARATIVA

**Manuscrito 2:** “*Homology Modeling of the Oxygenase Component of Carbazole 1,9a-Dioxygenase (CarAa) Involved in Petroleum Denitrogenation Pathway of Pseudomonas sp.*”.

Importantes colaborações foram feitas neste período (2003-2006) E, deram-se com um conjunto de trabalhos envolvendo uma variedade de enzimas dos grupos do Prof. Tito Alves (PEQ/COPPE - UFRJ) e do Prof. Orlando Martins (IBqM - UFRJ) envolvidas nos projetos de pesquisa originais de alguns de seus alunos.

Dentre estas colaborações devo citar primeiramente, a que pude estabelecer com a aluna de Doutorado do PEQ, Ariane Leites Larentis, cuja tese envolvia a clonagem e expressão de enzimas de *Pseudomonas sp.* envolvidas na rota de desnitração de carbazol. O carbazol é um composto heterocíclico presente no petróleo, cuja destilação provoca o envenenamento de catalisadores da indústria petroquímica, e a queima de combustíveis provenientes do petróleo provoca a emissão de NOx (gases causadores de chuva ácida e da destruição da camada de ozônio).

A rota de desnitração de carbazol de *Pseudomonas sp.* especificamente as reações que levam este composto até ácido antranílico, envolve o emprego de três complexos enzimáticos: a carbazol 1,9a-dioxygenase, a dioxygenase de rmetaclivagem e uma hidrolase. O primeiro complexo enzimático é formado por três subunidades CarAa, CarAc e CarAd.

Nosso primeiro trabalho envolveu a modelagem da subunidade CarAa deste complexo, a qual é responsável pela última etapa do transporte de elétrons necessário para a redução do carbazol à 2'-aminobifenil-2,3-diol. Neste trabalho propomos, utilizando como base as cadeias alfa da naftaleno 1,2 dioxygenase de *Pseudomonas putida*, um modelo estrutural formado por três subunidades idênticas formando um homotrímero triangular. Além disso, os resíduos Glu176-Asp180-His183-His187 e His71-His93-Cys69-Cys90, foram identificados como formadores dos domínios de ferro mononuclear e de Rieske

[2Fe-2S] respectivamente, que são tidos como responsáveis pela catálise da reação de dioxigenação.

## Homology Modeling of the Oxygenase Component of Carbazole 1,9a-Dioxygenase (CarAa) Involved in Petroleum Denitrogenation Pathway of *Pseudomonas* sp.

Ariane Leites Larentis<sup>1,2\*</sup>, Rodrigo Volcan Almeida<sup>1,2</sup>, Alexander Machado Cardoso<sup>2</sup>, Wellington Inácio Almeida<sup>2</sup>, Shaila Cíntia Rössle<sup>3</sup>, Paulo Mascarello Bisch<sup>3</sup>, Orlando Bonifácio Martins<sup>2</sup> and Tito Lívio Moitinho Alves<sup>1</sup>

<sup>1</sup>Laboratório de Bioprocessos; Programa de Engenharia Química; COPPE; <sup>2</sup>Laboratório de Biologia Molecular; Instituto de Bioquímica Médica; <sup>3</sup>Laboratório de Física Biológica; Instituto de Biofísica Carlos Chagas Filho; UFRJ; Av. Brigadeiro Trompowski s/nº; 21941-590; ariane@peq.coppe.ufrj.br; Rio de Janeiro - RJ - Brasil

### ABSTRACT

The understanding of the enzymatic mechanism, selectivity and specificity of the carbazole 1,9a-dioxygenase, that catalyzes the first reaction in the carbazole degradation pathway in *Pseudomonas* sp., is an important step in the development of a biocatalyst for denitrogenation and remediation of these recalcitrant compounds present in petroleum. For this purpose, the  $\alpha_3$  trimeric structure of the CarAa component from carbazole 1,9a-dioxygenase was obtained through comparative molecular modeling techniques, using the homologue enzyme naphthalene 1,2-dioxygenase (INDO) as a reference structure. A possible structure for the catalytic site was identified, obtaining the mononuclear iron and Rieske [2Fe-2S] cluster domains, responsible for electron transfer to dioxygenation reaction. Asp180 was suggested to be the residue attaching these two domains, between two  $\alpha$  subunits in the trimer, as occurs in the template protein. The model also indicates that the residue that completes the mononuclear iron domain is Asp333.

**Key words:** Comparative modeling, structure, carbazole, dioxygenase, petroleum, denitrogenation

### INTRODUCTION

Carbazole and its dibenzopyrrole derivatives are recalcitrant aromatic, heterocyclic and nitrogenated compounds, commonly present in petroleum. These nitrogenated molecules, found in oils and other fossil fuels are potential pollutants, because they are related to environmental problems such as nitrogen oxides (NO<sub>x</sub>) generation and emission, responsible for acid rain formation and destruction of the ozone layer. Also, these nitrogenated compounds present an economic impact in the petroleum refining process

because they act as catalyst “poisons” in fuel cracking (FCC), as inhibitors in hydrodesulfurization (HDS), and they can modify the quality of oil products (Benedik et al., 1998). The coupling of biodegradation metabolic pathways of these nitrogenated compounds to the expensive and hazardous hydrotreating processes used for N and S removal (HDN and HDS, respectively) is currently a strategy of great biotechnological interest. Due to the high selectivity and mild conditions of most bioprocesses, they could be envisaged as alternative to hydrotreating processes that utilize high temperature and pressure, as well

\* Author for correspondence

as affect other petroleum constituents (Benedik et al., 1998).

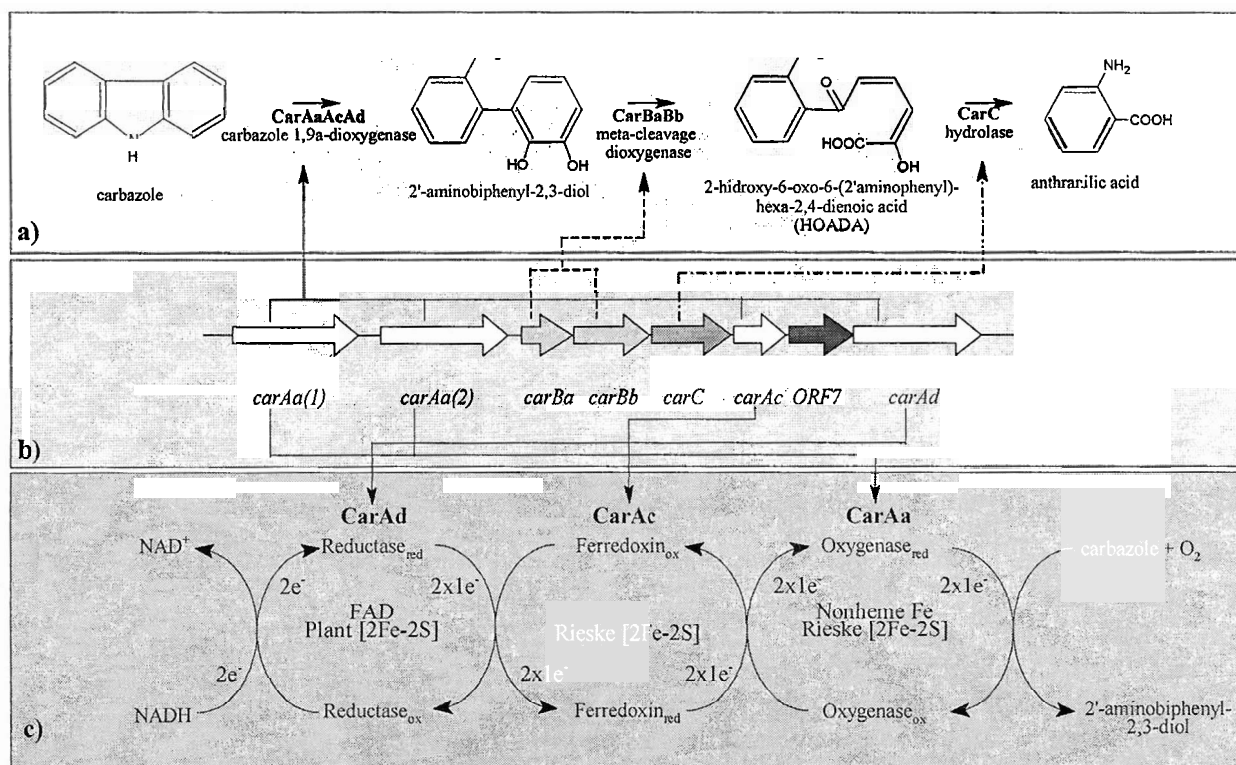
Some genes and enzymes involved in microbial pathways of carbazole-degrading bacteria are described in the literature. The molecular basis of the *Pseudomonas* sp. carbazole denitrogenation route is already known (Sato et al., 1997<sup>a,b</sup>). The map of the eight genes encoding the three enzymatic complexes involved in this pathway (*car* operon) is shown in Fig. 1 (a and b). The first enzyme, carbazole 1,9a-dioxygenase, converts carbazole into 2'-aminobiphenyl-2,3-diol by an angular dioxygenation and depends on the coexistence of three components, encoded by *carAa*, *carAc* and *carAd* genes, all of them with [2Fe-2S] clusters. The component encoded by *carAa* also presents a mononuclear iron domain that possibly acts as the catalytic site for oxygen activation for dioxygenation.

The protein encoded by *carAd* presents, besides the iron-sulfur cluster, FAD and NAD-binding domains. The second enzyme, a *meta*-cleavage dioxygenase,

is encoded by both *carBa* and *carBb* genes, while the third one, encoded by *carC* gene, presents an *a/b* hydrolase characteristic domain, and catalyzes the hydrolysis of HOADA to produce anthranilic acid. Finally, *ORF7* encodes a protein that is not essential for carbazole dioxygenase activity and has no homology with other degradative proteins (Sato et al., 1997<sup>a,b</sup>). The *car* operon is located in a megaplasmid in these bacteria.

The reaction step that removes the nitrogen from anthranilic acid, producing catechol, occurs by another dioxygenation reaction catalyzed by an enzymatic complex encoded by three genes (*antABC*). These are present in an operon named *ant*, with a copy in the same megaplasmid and another in bacterial chromosomes (Nojiri et al., 2001<sup>a</sup>).

Carbazole 1,9a-dioxygenase is known to have a broad substrate range and can catalyze diverse oxygenations: angular and lateral dioxygenations (also termed *cis*-dihydroxylation) and monooxygenations (Nojiri et al., 1999; Nam et al., 2001; Nam et al.,



**Figure 1** - (a) Carbazole degradation pathway to anthranilic acid from *Pseudomonas* sp. (b) Map of genes from operon *car* required for carbazole degradation. (c) Components and functions of the electron transfer mechanism of carbazole 1,9a dioxygenase: CarAd (containing FAD and plant-type [2Fe-2S] cluster binding sites) acts as a reductase component and transfers electrons from NADH to a ferredoxin component that presents a Rieske [2Fe-2S] cluster (CarAc). These electrons flow to the catalytic a3 oxygenase component that contains a Rieske [2Fe-2S] cluster and a mononuclear iron center (modified from Sato et al., 1997<sup>a,b</sup>; Benedik et al., 1998, Nam et al., 2002<sup>a</sup>).

2002<sup>b</sup>). It belongs to a bacterial multicomponent enzyme group involved in biodegradation pathways of aromatic compounds that presents Rieske iron-sulfur cluster non-heme iron dioxygenases mechanisms (Wackett, 2002). These enzymes use reduced pyridine nucleotide as the initial two-electron source for dioxygen activation. These electrons are passed to a protein component that contains a flavin cofactor acting as a two to one electron redox switch. Each one of these electrons is donated to a [2Fe-2S] center, also presented in the protein containing FAD domain or in a separate ferredoxin component. In the last step, electrons are transferred to the oxygenase component that contains a Rieske [2Fe-2S] cluster and a mononuclear iron center (Wackett, 2002). Although the carbazole dioxygenase component structures have not been yet determined, CarAa, CarAc and CarAd carbazole dioxygenase components from *Pseudomonas resinovorans* CA10 have previously been purified and enzymatically and spectroscopically characterized by Nam et al. (2002<sup>a,b</sup>). According to these studies, the 13 kDa monomer of CarAc and the 37 kDa of CarAd act, respectively, as the ferredoxin and reductase components in the electron transport from NADH to the terminal oxygenase component CarAa, whose monomer has 44 kDa. This protein was purified and identified as an  $\alpha_3$  homotrimeric structure with 132 kDa (Nam et al., 2002<sup>a,b</sup>). The proposed electron-transfer reactions, as well as the characteristic domains and functions of carbazole dioxygenase components, are presented in Figure 1c (Nam et al., 2002<sup>a,b</sup>).

A knowledge of these enzymes, their function and selectivities to different nitrogenated compounds, is essential for developing biocatalysts for petroleum biorefining and bioremediation. Hence, structural characterization of subunits and catalytic sites of the enzymatic complexes involved in carbazole degradation is a useful approach to understand the reaction mechanisms, as the structure of most carbazole enzymes are not yet published, except for HOADA hydrolase (Habe et al., 2003).

Based on this, the homology modeling of the catalytic component of carbazole 1,9a-dioxygenase (CarAa) was carried out employing naphthalene 1,2-dioxygenase (Kauppi et al., 1998) as a template in order to predict target structure. The comparative structural modeling of CarAa will help to gain an understanding of enzymatic mechanism throughout the protein complex formation involved in carbazole degradation.

## MATERIALS AND METHODS

Structural comparative modeling allows the prediction of the structure of a target protein sequence (the catalytic oxygenase component of carbazole 1,9a-dioxygenase, CarAa) based on its alignment to one or more homologue proteins of known molecular coordinates, available in the *Protein Data Bank* (PDB). The comparative modeling is performed in four iterative sequential steps (Peitsch, 1995 and 1996; Martí-Renom et al., 2000; Santos Filho and Alencastro, 2003): template selection, target-template alignment, model building of the target and model evaluation.

### Template selection and alignment

In order to identify protein templates with a high degree of similarity, BLASTP (*Basic Local Alignment Search Tool-Protein-Protein*) (Altschul et al., 1990) against PDB was performed. This program is available at <http://www.ncbi.nlm.nih.gov/blast/>. The alignment between the 384-amino acid sequence of CarAa and 449-amino acid template sequence (a chains of naphthalene 1,2-dioxygenase, deposited in PDB with the access code 1NDO) was made by T-COFFEE (Notredame et al., 2000) and improved manually based on template structure information.

### Comparative modeling

Sequence alignment was used in the model construction of the CarAa monomer by the spatial restrictions technique with the software MODELLER 6v2 (Šali and Blundell, 1993). Spatial coordinates of each chain of the CarAa trimeric complex ( $\alpha_3$  structure), were obtained by SPDBViewer (Guex and Peitsch, 1997) by superposition of the monomer model with a chains of the  $\alpha_3\beta_3$  original structure of 1NDO.

### Evaluation of the model

Model evaluation was performed using Ramachandran Plots (Ramachandran et al., 1963) and Residue Energy Plots carried out by the programs PROCHECK (Laskowsky et al., 1993) and PROSAR (Sippl, 1993), respectively.

The residues and domains characterization, as well as the computation of bond distances, were performed through visual inspection with the help of RASMOL (Sayle and Milner-White, 1995).

## RESULTS AND DISCUSSION

The search for a template for CarAa through BLASTP against all protein structures deposited in the PDB indicated naphthalene 1,2-dioxygenase (1NDO) from *Pseudomonas putida* (Kauppi et al., 1998) as a possible homologue structure for comparative modeling of the target protein.

The naphthalene dioxygenase is a multimeric mononuclear non-heme iron oxygenase enzyme composed of non-identical subunits (a and b) in a mushroom-shaped a3b3 hexamer (Kauppi et al., 1998), that degrades naphthalene by a lateral dioxygenation, resulting in the formation of a *cis*-dihydrodiol compound (Nojiri et al., 2001<sup>b</sup>). The structure of 1NDO resolved by Kauppi et al. (1998) presents in the Rieske [2Fe-2S] center, an iron atom coordinated by Cys81 and Cys101 and another one coordinated by His83 and His104. The catalytic mononuclear iron domain is coordinated by His208, His213 and Asp362. His208 is hydrogen bonded to Asp205, which binds His104, attaching the mononuclear center to the Rieske cluster.

According to the Batie classification system for multicomponent dioxygenases (Nam et al., 2001), carbazole 1,9a-dioxygenase and naphthalene 1,2-dioxygenase from *Pseudomonas* sp. belong to the same group based on the composition of the electron transport chain (number of constituent components and the nature of the redox centers). An important difference between carbazole and naphthalene dioxygenases is that the oxygenase component of the enzyme that degrades carbazole does not present the b chain, that probably has only a structural role in a3b3-type dioxygenases (Kauppi et al., 1998; Nam et al., 2001), but forms an a chain trimer similar to naphthalene dioxygenase, strengthening its utilization as a template structure (Nam et al., 2002<sup>a,b</sup>). According to Sato et al. (1997<sup>a</sup>), the enzyme carbazole 1,9-a dioxygenase is able to degrade both substrates, but with different conversion results. While 99% of carbazole is degraded by carbazole 1,9-a dioxygenase, this conversion is reduced to 77% for naphthalene. Therefore, this experimental data gives further support to the use of naphthalene dioxygenase as a template for obtaining the carbazole dioxygenase structure by homology.

The CarAa-a1NDO alignment results (Fig. 2) show that the entire CarAa sequence presents 20% of

identity and 30% of similarity with a chains of naphthalene 1,2-dioxygenase. Nevertheless, correct alignments for the catalytic site domains (Rieske [2Fe-2S] cluster and mononuclear iron center) were obtained, with only a few small gaps. The Rieske iron-sulfur cluster is coordinated by two histidine and two cysteine ligands with a signature Cys-X-His-X<sub>16-17</sub>-Cys-X<sub>2</sub>-His, while the mononuclear iron center presents a motif like Glu-X<sub>3,4</sub>-Asp-X<sub>2</sub>-His-X<sub>4,5</sub>-His (Sato et al., 1997<sup>b</sup>; Jiang et al., 1996).

In the CarAa amino acid sequence, these domains can be identified as His71-His93-Cys69-Cys90 and Glu176-Asp180-His183-His187, respectively. Three aspartate residues could complete the mononuclear iron domain in CarAa: Asp332, Asp333 or Asp335. In this model, Asp333 presented the best fit to the residue Asp362 in 1NDO (Figure 2).

The trimeric structure obtained by homology, presented in Fig. 3, is consistent with available experimental results and the proposed mechanism for carbazole dioxygenase (Nam et al., 2002<sup>a,b</sup>). Rieske and mononuclear iron sites are located at the interface between a subunits, assembling a trigonal geometry and allowing the existence of three active sites for each holoenzyme. This is consistent with electron transfer mechanisms described in the literature for dioxygenases. The active sites are formed by a Rieske iron-sulfur cluster belonging to one a subunit juxtaposed within 12-14 Å from the mononuclear iron center found in the adjacent a subunit. The distance between the iron centers belonging to the same a subunit is too large to perform an efficient electron transfer. This is similar to naphthalene dioxygenase (Kauppi et al., 1998; Wackett, 2002).

In addition, the existence of an access channel to the catalytic site for substrate binding in naphthalene dioxygenase can also be identified in the proposed carbazole dioxygenase structure (Fig. 3b). The residues that possibly compose this channel are hydrophobic, as in the template. These domains, however, differently from the Rieske cluster and mononuclear iron center, are expected to vary for different groups of dioxygenases (Kauppi et al., 1998).

The modeling of the Rieske site, which acts in electron transfer to the catalytic site, presents consistent results in comparison to the reference structure (Fig. 4). The modeling results for the EX<sub>3,4</sub>DX<sub>2</sub>HX<sub>4,5</sub>H motif, which is considered to act



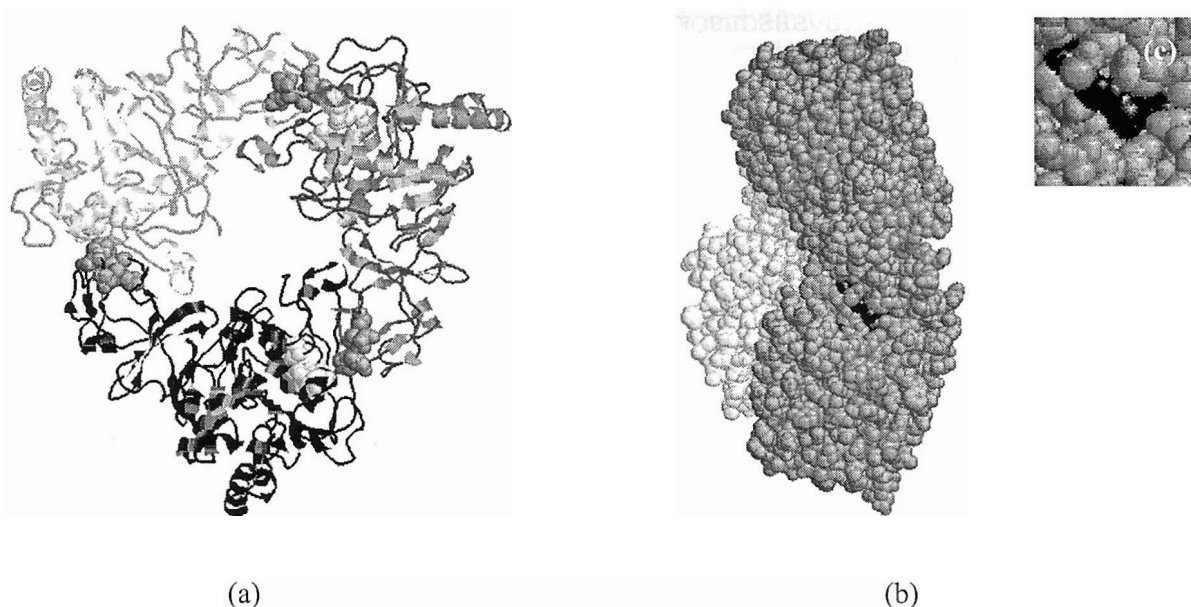
1NDO	1	-MNYNKKILVSESGLSQKHLIHGDEELFQHEL- KTIFARNWLFLLTHDSLIPAPGDYVTAK
CarAa	1	MANVDEAILKRKVGWA-----PYVDAKLGFRNHWPVMSKEIDE-GEPKTLK
		* :: ** . * : : * * . : * : . . * * : * *
1NDO	59	MGIDEVIVSRQNDGSIRAFNLVCRHRGKTL-VSVEAGNAKGFVCSYHGWF----GSNGE
CarAa	48	LLGENLLVNRI-DGKLYCLKDRCLHRGVQLSVKVECKTKSTITCWYHAWTYRWEDGVLCD
		: ::::*.* **.: : : * *** * *.***. . . :.* **.* : * :
1NDO	114	LQSVPFKDYGESLNKKCLGLKEVARVESFHGFYIGCF-DQEAPPLMDYLGDAAWYLEP
CarAa	107	ILTNPTSAQIGRQKL-----KTPVQEAQKCVFIYLGDDPPPLARD-----TP
		: : * . : : :.* :. *.: * : : : * :.*** *
1NDO	173	MFKHSGGLELVGPPGKVVIKANWKAPAENFVGDAYHVGWTHASSLRSGESIFSSLAGNAA
CarAa	151	PNFLDDMEILGK--NQIIKSNWRLAVENGFDPSHI-YIHKDSILVKD---NDLALPLG
		. . . : * : * : : * * : * . * : : * . * : : * . * : : * . * .
1NDO	233	LPPEGAGLQMTSKYSGMGVLWDGYS--GVHSADLPELMAFGGAKQERLNKEIGDVRAR
CarAa	204	FAPGGDRKQQTRVDDDDVGRKGVYDLIGEHLGVPVFEGTIGGEVVREGAYGEKIVANDIS
		:.* * * * . . . : . * . * * . . . : . . : : : * *
1NDO	291	IYRSH-LNCTVFPNNSMLTCSGVFKVWNPIDANTTEVWPTYAIVEKDMPEDLKRRLADSVQ
CarAa	264	IWLPGVLKVNPFNPDMQ----FEWYVPIDENTH--YYFQTLGKPCANDEERKNYEQEF
		*: . * : . *** .* : * : * * * * : : : * . : * : * : . .
1NDO	350	RTFGPAGFWES---DDNDNMETASQNGKKYQSRSDLLSNLGFGEVDVYGDVYPGVVGKS
CarAa	318	ESKWKPMALEGFNDDI----WAREAMVDFYADKGVNEILFEVD-----EAIVAWR
		: . . * . ** * : . : : * . . : : : * * * . : * .
1NDO	410	AIGETSYRGFYRAYQAHVSSSNWAEFEHASSTWHTELTKTT
CarAa	370	KLASEHNQGIQT--QAHVSG-----
		. . . : * : * * * * .

**Figure 2** - Alignment between CarAa and 1NDO a chain sequences. Identical amino acids are indicated by (,); similar amino acids by (: or (.) or (.)); different amino acids ( ), while (-) indicates gaps added to the sequences. The Rieske (CXHX<sub>16-17</sub>CX<sub>2</sub>H) motif (gray), the mononuclear iron (EX<sub>3,4</sub>DX<sub>2</sub>HX<sub>4,5</sub>H) motif and the aspartate (Asp333) that completes the mononuclear catalytic center (light gray) are highlighted.

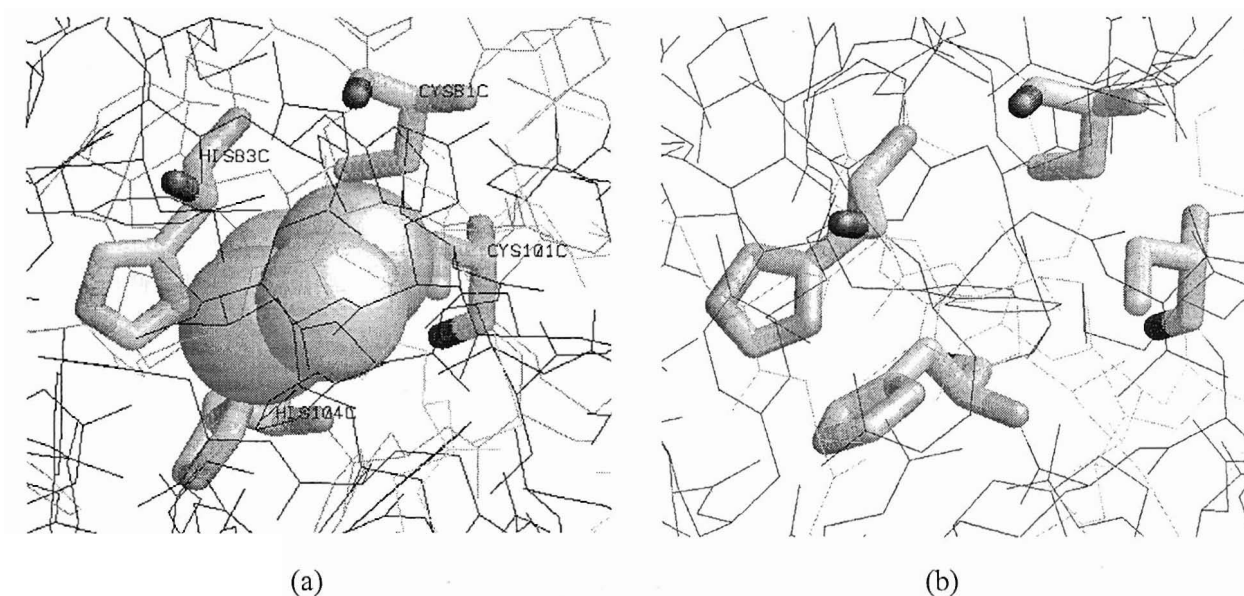
as a mononuclear iron ligand at the site of oxygen activation in the catalytic site of dioxygenases (Jiang et al., 1996; Sato et al., 1997<sup>b</sup>), are shown in Fig. 5, compared to the template structure.

The naphthalene dioxygenase mononuclear iron, the site of aromatic substrate cleavage with oxygen, is coordinated by three residues (His 208, His213, Asp362), as shown in Figure 5, and a water molecule (Kauppi et al., 1998; Wackett, 2002). Both histidine residues are part of the mononuclear domain signature Glu200-Asp205-His208-His213, conserved in many dioxygenases. Asp205 is responsible for connecting the mononuclear iron domain to the Rieske cluster at the interface with the subjacent a subunit in the trimeric structure, which permits the electron transfer needed by the chemical reaction. In an analogous way, in the CarAa model structure, it is possible to identify the mononuclear iron motif as Glu176-Asp180-His183-His187. The Asp180

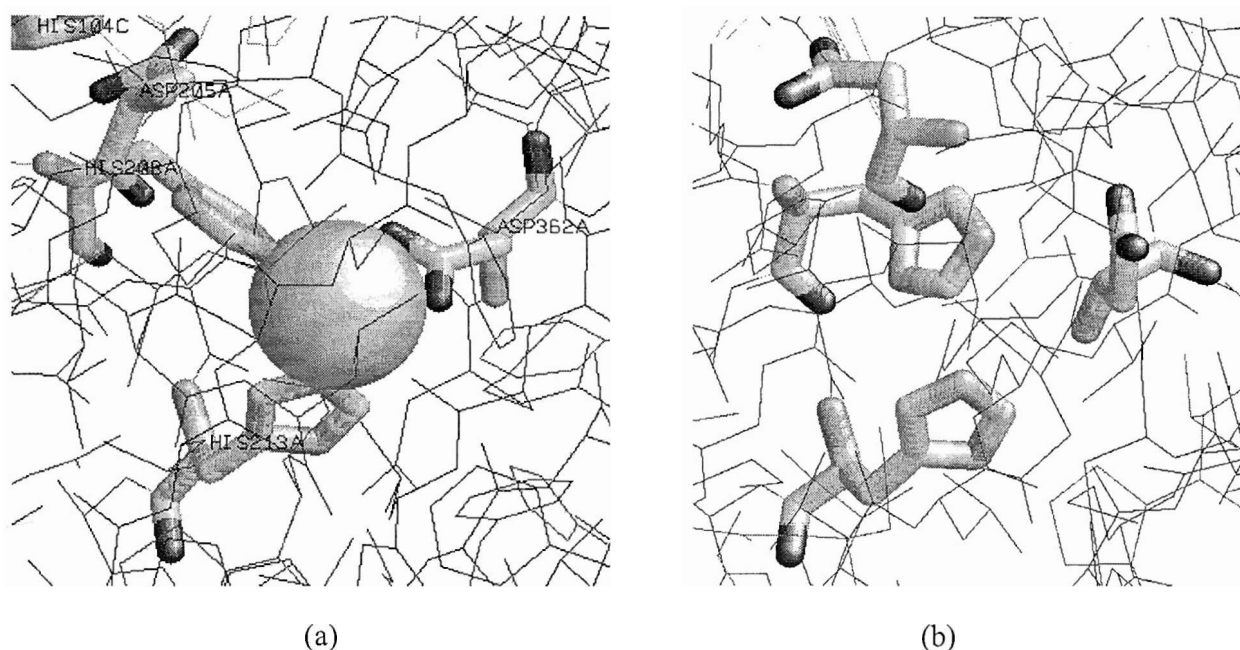
residue, presented in this signature, may be responsible for attaching the mononuclear iron motif to His93 in the Rieske domain, through attractive forces, probably a hydrogen bond, at the interface of two a subunits of CarAa trimer, equivalent to Asp205 from naphthalene dioxygenase. The Asp362 residue completes the catalytic site and the coordination to the iron atom in naphthalene dioxygenase. In accordance to this reference structure, it is proposed that Asp333 completes the mononuclear domain in carbazole dioxygenase, as discussed in the CarAa-a1NDO alignment description.



**Figure 3** - Model of CarAa trimer obtained by homology using 1NDO a chains. (a) Loops,  $\alpha$ -helices and  $\beta$ -sheets structures for the three chains of the CarAa structure (black, gray and light gray). The Rieske (gray) and the mononuclear iron (light gray) centers are also indicated in spacefill representation; (b) lateral visualization of compact trimeric structure; (c) in detail, the possible access channel for substrate binding in the active site (the Rieske and the mononuclear centers are represented by the black residues).



**Figure 4** - Comparison between Rieske sites from 1NDO (formed by His83, His104, Cys81 and Cys101 residues) (a) and the modeled structure for CarAa (formed by residues His71, His93, Cys69 and Cys90) (b). The atoms represented in the amino acids are: C (light gray), O (gray) and N (dark gray). The light gray extremities of the Cys residues represent S atoms. The spheres in the reference structure represent the [2Fe-2S] cluster.



**Figure 5** - Mononuclear iron domain from 1NDO, formed by residues Asp205 (responsible for connection of mononuclear site to adjacent a subunit interface), His208, His213 and Asp362 (a) and from the proposed model for the oxygenase component of carbazole 1,9a-dioxygenase, formed by amino acids Asp180, His183, His187 and Asp333 (Asp180 attaches the Rieske and mononuclear iron domains at two a subunit interfaces in the trimeric structure of CarAa) (b). The atoms represented in the amino acids are: C (light gray), O (gray) and N (dark gray). The sphere in the reference structure represents the iron atom from catalytic site.

## CONCLUSIONS

With the homology modeling of the CarAa component from carbazole 1,9a-dioxygenase, performed using the a3 1NDO structure as a template, a trimeric enzyme structure was successfully obtained with acceptable residue energies and 95% of the residues located at spatially possible angles, according to the Ramachandran plots (data not shown).

A possible structure for the catalytic site was identified, obtaining the mononuclear iron domain, with the consensus sequence Glu176-Asp180-His183-His187, and the Rieske [2Fe-2S] cluster (His71-His93-Cys69-Cys90), described as being responsible for electron transfer in the dioxygenation reaction. It is proposed that the residue Asp180 participates in the attachment between these two domains, at the interface of two a subunits in the trimeric model, as in the template protein. The model also indicates that the residue that completes the catalytic domain is Asp333.

However, in order to understand how the electron flow occurs, allowing the dioxygenation cleavage of carbazole, the structures of the other two

components, CarAc and CarAd, and the enzymatic complex formation have to be constructed. This complex (oxygenase, ferredoxin and reductase components) is not yet described in the literature for any dioxygenase. Molecular modeling can also be used for other enzymes involved in the carbazole degradation pathway. In conclusion, the model obtained is a suitable first approach for structural characterization, considering the lack of experimental three-dimensional data. This kind of theoretical structural approach can be used to gain a better understanding of the enzymatic mechanism.

## ACKNOWLEDGEMENTS

CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior), FAPERJ (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro) and CTPetro/FINEP-Nº21.01.0278.00 supported this work.

## RESUMO

O entendimento do mecanismo da enzima carbazol 1,9a-dioxigenase, que catalisa a reação inicial da rota de degradação de carbazol em *Pseudomonas* sp., assim como a compreensão de sua seletividade e especificidade, são etapas importantes no desenvolvimento de um biocatalisador para desnitração e remediação destes compostos presentes em petróleo. Para isto, a estrutura a3 do trímero da componente CarAa da carbazol 1,9a-dioxigenase foi obtida através das técnicas de modelagem molecular por homologia, utilizando a enzima 1,2-naftaleno dioxigenase (1NDO) como estrutura de referência. A possível estrutura do sítio catalítico foi identificada, obtendo-se os domínios de ferro mononuclear e o de Rieske [2Fe-2S], descrito como responsável pela transferência de elétrons para a reação de dioxigenação. A modelagem comparativa indica que o resíduo Asp180 conecta estes domínios, na interface de duas subunidades a do trímero, como observado na proteína molde. O modelo sugere também que o resíduo que completa o domínio de ferro mononuclear seja Asp333.

## 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**(3), 403-410.
- Benedik, M. J.; Gibbs, P. R.; Riddle, R. R. and Willson, R. C. (1998), Microbial denitrogenation of fossil fuels. *T. Biotechnol.*, **16** (9), 390-395.
- Guex, N. and Peitsch, M. C. (1997), SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, **18**, 2714-2723.
- Habe, H.; Morii, K.; Fushinobu, S.; Nam, J.-W.; Ayabe, Y.; Yoshida, T.; Wakagi, T.; Yamane, H.; Nojiri, H. and Omori, T. (2003), Crystal structure of a histidine-tagged serine hydrolase involved in the carbazole degradation (CarC enzyme). *Biochem. Biophys. Res. Commun.* **303** (2), 631-639.
- Jiang, H.; Parales, R. E.; Lynch, N. A. and Gibson, D. T. (1996), Site-Directed Mutagenesis of Conserved Amino Acids in the Alpha Subunit of Toluene Dioxigenase: Potential Mononuclear Non-Heme Iron Coordination Sites. *J. Bacteriol.*, **178** (11), 3133-3139.
- Kauppi, B.; Lee, K.; Carredano, E.; Parales, R. E.; Gibson, D. T.; Eklund, H. and Ramaswamy, S. (1998), Structure of an aromatic-ring-hydroxylating dioxigenase - naphthalene 1,2-dioxigenase. *Structure*, **6**(5), 571-586.
- Laskowski, R. A.; MacArthur, M. W.; Moss, D. S. and Thornton, J. M. (1993), Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.
- Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F. and Šali, A. (2000), Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291-325.
- Nam, J.-W.; Nojiri, H.; Yoshida, T.; Habe, H.; Yamane, H. and Omori, T. (2001), New classification system for oxygenase components involved in ring-hydroxylating oxygenations. *Biosc. Biotechnol. Biochem.*, **65** (2), 254-263.
- Nam, J.-W.; Fujimoto, Z.; Mizuno, H.; Yamane, H.; Yoshida, T.; Habe, H.; Nojiri, H. and Omori, T. (2002<sup>a</sup>), Crystallization and preliminary crystallographic analysis of the terminal oxygenase component of carbazole 1,9a-dioxigenase of *Pseudomonas resinovorans* strain CA10. *Acta Cryst.*, **D58**, 1350-1352.
- Nam, J.-W.; Nojiri, H.; Noguchi, H.; Uchimura, H.; Yoshida, T.; Habe, H.; Yamane, H. and Omori, T. (2002<sup>b</sup>), Purification and Characterization of Carbazole 1,9a-Dioxigenase, a Three-Component Dioxigenase System of *Pseudomonas resinovorans* Strain CA10. *Appl. Environ. Microbiol.*, **68** (12), 5882-5890.
- Notredame, C.; Higgins, D. G. and Heringa, J. (2000), T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302** (1), 205-217.
- Nojiri, H.; Nam, J.-W.; Kosaka, M.; Morii, K.-I.; Takemura, T.; Furihata, K.; Yamane, H. and Omori, T. (1999), Diverse Oxygenations Catalyzed by Carbazole 1,9a-Dioxigenase from *Pseudomonas* sp. Strain CA10. *J. Bacteriol.*, **181** (10), 3105-3113.
- Nojiri, H.; Sekiguchi, H.; Maeda, K.; Urata, M.; Nakai, S.I.; Yoshida, T.; Habe, H. and Omori, T. (2001<sup>a</sup>), Genetic Characterization and Evolutionary Implications of a *car* Gene Cluster in the Carbazole Degradator *Pseudomonas* sp. Strain CA10. *J. Bacteriol.*, **183** (12), 3663-3679.
- Nojiri, H.; Habe, H. and Omori, T. (2001<sup>b</sup>), Bacterial degradation of aromatic compounds via angular dioxigenation. *J. Gen. Appl. Microbiol.* **47** (6), 279-305.
- Peitsch, M. C. (1995), Protein modeling by E-mail. *Bio/Technology*, **13**, 658-660.
- Peitsch, M. C. (1996), ProMod and Swiss-Model: Internet-based tools for automated comparative protein modeling. *Biochem. Soc. Trans.*, **24**, 274-279.

- Ramachandran, G. N., Ramakrishnan C. and Sasisekharan, V. (1963), Stereochemistry of Polypeptide Chain Conformations. *J. Mol. Biol.*, **7**, 95-99.
- Šali, A. and Blundell, T. L. (1993), Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779-815.
- Santos Filho, O. A. and Alencastro, R. B. (2003), Modelagem de Proteínas por Homologia. *Quim. Nova* **26**(2), 253-259.
- Sato, S. I.; Ouchiyama, N.; Kimura, T.; Nojiri, H.; Yamane, H. and Omori, T. (1997<sup>a</sup>), Cloning of Genes Involved in Carbazole Degradation of *Pseudomonas* sp. Strain CA10: Nucleotide Sequences of Genes and Characterization of *meta*-Cleavage Enzymes and Hydrolase. *J. Bacteriol.*, **179** (15), 4841-4849.
- Sato, S.I.; Nam, J.W.; Kasuga, K.; Nojiri, H.; Yamane, H. and Omori, T. (1997<sup>b</sup>), Identification and Characterization of Genes Encoding Carbazole 1,9a-Dioxygenase in *Pseudomonas* sp. Strain CA10. *J. Bacteriol.*, **179** (15), 4850-4858.
- Sayle, R. A. and Milner-White, E. J. (1995), RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20** (9), 374-376.
- Sippl, M. J. (1993), Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**, 355-352.
- Wackett, L. P. (2002), Review - Mechanism and applications of Rieske non-heme iron dioxygenases. *Enz. Microbial Technol.*, **31**, 577-587.

Received: July 18, 2003;

Revised: February 09, 2004;

Accepted: October 15, 2004.

**Manuscrito 3:** “*Expression and Homology Modeling of 2'-Aminobiphenyl-2,3-diol-1,2-Dioxygenase from Pseudomonas stutzeri Carbazole Degradation Pathway*”.

Outro trabalho publicado em colaboração com a Ariane, diz respeito a segunda enzima da rota, a 2' aminobifenil-2,3-diol-1,2-dioxigenase (CarB). Neste trabalho foram apresentados os resultados da clonagem, expressão, caracterização prévia e modelagem comparativa da estrutura desta enzima. A CarB é a enzima responsável pela reação intermediária da rota que leva o carbazol até ácido antranílico. Estes dois genes foram clonados em sua forma nativa e com uma cauda de histidina, utilizando-se respectivamente os plasmídeos pDEST-14 e pDEST-17 (Gateway - Invitrogen). A expressão e a atividade para ambas construções gênicas foram similares, demonstrando que a cauda de histidina não influencia de forma significativa o comportamento da enzima. Os modelos foram construídos para cada monômero da enzima, utilizando-se como molde a estrutura da protocatecol 4,5-dioxigenase (LigAB) de *Sphingomonas paucimobilis* SYK6. Com estes modelos foi possível sugerir: a estrutura do heterotetrâmero; os resíduos His12-His53-Glu230 que juntos coordenam um átomo de Fe(II) necessário para o mecanismo catalítico, como formadores do sítio ativo; a importância dos resíduos His117 e His182 que se ligam através de ligações de hidrogênio às hidroxilas do anel catecol do substrato. Além disso, foi possível observar a substituição dos resíduos Ser269 e Asn270 da LigAB por Trp257 e Phe258 na CarB, sugerindo uma possível razão para esta última enzima não ser ativa contra protocatecol. Estes resultados foram publicados na revista *Cell Biochemistry and Biophysics* sob o título: “*Expression and Homology Modeling of 2'-Aminobiphenyl-2,3-diol-1,2-Dioxygenase from Pseudomonas stutzeri Carbazole Degradation Pathway*”.

ORIGINAL ARTICLE

## Expression and Homology Modeling of 2'-Aminobiphenyl-2,3-Diol-1,2-Dioxygenase From *Pseudomonas stutzeri* Carbazole Degradation Pathway

Ariane Leites Larentis,<sup>1,2,\*</sup> Rodrigo Volcan Almeida,<sup>1,2</sup> Shaila Cintia Rössle,<sup>3</sup>  
Alexander Machado Cardoso,<sup>2</sup> Welington Inácio Almeida,<sup>2</sup>  
Paulo Mascarello Bisch,<sup>3</sup> Tito Livio Moitinho Alves,<sup>1</sup>  
and Orlando Bonifácio Martins<sup>2</sup>

<sup>1</sup>Laboratório de Bioprocessos, Programa de Engenharia Química, COPPE, <sup>2</sup>Laboratório de Biologia Molecular, Instituto de Bioquímica Médica, and <sup>3</sup>Laboratório de Física Biológica, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

### Abstract

The enzyme 2'-aminobiphenyl-2,3-diol-1,2-dioxygenase (CarB), encoded by two genes (*carBa* and *carBb*), is an  $\alpha_2\beta_2$  heterotetramer that presents *meta*-cleavage activity toward the hydroxylated aromatic ring in the carbazole degradation pathway from petroleum-degrader bacteria *Pseudomonas* spp. The 1082-base pair polymerase chain reaction product corresponding to *carBaBb* genes from *Pseudomonas stutzeri* ATCC 31258 was cloned by site-specific recombination and expressed in high levels in *Escherichia coli* BL21-SI with a histidine-tag and in native form. The CarB activity toward 2,3-dihydroxybiphenyl was similar for these two constructions. The  $\alpha_2\beta_2$ -heterotetrameric 3D model of CarB dioxygenase was proposed by homology modeling using the protocatechuate 4,5-dioxygenase (LigAB) structure as template. Accordingly, His12, His53, and Glu230 coordinate the Fe(II) in the catalytic site at the subunit CarBb. The model also indicates that His182 is the catalytic base responsible for deprotonating one of the hydroxyl group of the substrate by a hydrogen bond. The hydrophobic residues Trp257 and Phe258 in the CarB structure substituted the LigAB amino acid residues Ser269 and Asn270. These data could explain why the CarB was active for 2,3-dihydroxybiphenyl and not for protocatechuate.

**Index Entries:** *meta*-Cleavage enzyme; extradiol; dioxygenase; heterologous gene expression; comparative modeling; petroleum denitrogenation; biodegradation; 2,3-dihydroxybiphenyl; 2-hydroxy-6-oxo-6-(2'-aminophenyl)-hexa-2,4-dienoic acid; HOPDA.

### INTRODUCTION

To have insight into the carbazole degradation pathway for developing biocatalysts for petroleum biorefining and bioremediation, the 2'-aminobiphenyl-2,3-diol-1,2-dioxygenase (CarB) from *Pseudomonas stutzeri* ATCC 31258 was cloned using site-specific recombination system, expressed in high levels and in active form toward

2,3-dihydroxybiphenyl (2,3-DHBP) (analogous to 2'-aminobiphenyl-2,3-diol) in *Escherichia coli*. The Fe(II) dependence of this enzyme can be maintained in vitro by using a recombinant resting cell system, providing conditions for the production of high levels of active biocatalysts. To gain a better understanding of the enzymatic mechanism, the  $\alpha_2\beta_2$ -heterotetrameric CarB was structurally characterized by comparative modeling. The structural analysis of the model was able to explain why the enzyme is active toward the substrate 2,3-DHBP but not toward protocatechuate, although the

\* Author to whom all correspondence and reprint requests should be addressed. E-mail: ariane@peq.coppe.ufrj.br

protocatechuate 4,5-dioxygenase (LigAB) structure was used as template for the modeling procedure.

The enzyme CarB is an extradiol dioxygenase that catalyzes the *meta*-cleavage of the catechol ring of the compound 2'-aminobiphenyl-2,3-diol by adding a dioxygen molecule to produce 2-hydroxy-6-oxo-6-(2'-aminophenyl)-hexa-2,4-dienoic acid. That compound bears the aminophenyl and one hydroxyl group and incorporates a COOH group and a double bond with oxygen atom in the opened ring. The aforementioned enzyme is encoded by both *carBa* and *carBb* genes and is an  $\alpha_2\beta_2$  heterotetramer (1). It is the second enzyme that takes part in the carbazole degradation pathway to anthranilic acid from *Pseudomonas* spp. and some other carbazole-degrader bacteria that are able to grow in carbazole as sole carbon and nitrogen source. This route involves two different dioxygenases (carbazole and 2'-aminobiphenyl-2,3-diol) and a hydrolase, whose genes are present in a 6-kb operon in a megaplasmid found in these bacteria (2,3). Because carbazole and its dibenzopyrrole derivatives are heterocyclic nitrogenated aromatic compounds commonly found in fossil fuels, they are potential pollutants (nitrogen oxide emissions, acid rain formation, and destruction of the ozone layer), and knowledge of those degrader bacteria and their biodegradation pathways is currently of great biotechnological interest for developing biocatalysts for petroleum biorefining and bioremediation (4).

The extradiol dioxygenases, such as CarB, differ from the intradiol-type by the distinct oxidation state of the iron in the active site and lead to distinct products when they catalyze reactions of catechol aromatic ring fission on the same substrate. The active site of the extradiol-type presents nonheme ferrous iron [Fe(II)] and catalyzes the ring-cleavage at C–C bonds adjacent to the vicinal hydroxyl groups (*meta*-cleavage), whereas the intradiol dioxygenases are able to cleave C–C bond between the vicinal hydroxyl groups (*ortho*-cleavage) with a nonheme ferric iron in the active site (1,5). Three classes (I, II, and III) of the extradiol dioxygenases have already been identified: classes I and II present similar primary structures and may have a common ancestor (class II enzymes seem to have evolved from class I by gene duplication and present two domains with nearly the same pattern), whereas class III and II do not share similarity in the amino acid and structural levels (1). However, common features around the active sites of class II and III extradiol dioxygenases suggest that they share the same catalytic mechanism despite their completely different 3D structures (5). In the proposed mechanism, the two hydroxyl groups of the catechol directly coordinate to the Fe(II) active site in a trigonal bipyramidal geometry. Then the O<sub>2</sub> molecule binds to the ferrous ion and the iron-mediated charge transfers from the catechol to O<sub>2</sub>, resulting in the O<sub>2</sub> activation,

which attacks the C–C bond of the substrate catechol ring (5–7).

In the subunit CarBb of the 2'-aminobiphenyl-2,3-diol-1,2-dioxygenase, the catalytic site is composed by a ferrous ion present in a conserved LigB-domain (1), whereas CarBa does not present significant homologies to other nucleotide or amino acid sequences (2). According to the literature results (1,2,8), the presence of both subunits is essential for CarB *meta*-cleavage activity. Besides 2'-aminobiphenyl-2,3-diol, CarB presents activity toward its analogous 2,3-dihydroxybiphenyl and 2,2',3-trihydroxybiphenyl but not toward catechol and its methyl derivatives, as well as protocatechuate, which is the substrate for protocatechuate LigAB (5), in spite of the homology of both enzymes.

## MATERIALS AND METHODS

### *Strains and Growth Conditions*

*E. coli* strain DH10B was used as host for plasmid DNA cloning and sequencing. *E. coli* strain BL21-SI was used for salt-inducible protein expression under the control of T7 promoter and protease deficient for minimizing heterologous protein degradation (9). *E. coli* strain DH10B was cultivated in LB medium (1% NaCl, 1% bactotryptone, and 0.5% yeast extract), and BL21-SI was grown in the same medium lacking NaCl (LBON) with 100  $\mu$ g of ampicillin per milliliter, at 37°C with vigorous shaking.

### *Cloning of Genes Encoding CarB*

Genes encoding the two subunits of 2'-aminobiphenyl-2,3-diol-1,2-dioxygenase (272 bp of *carBa* and 809 bp of *carBb* genes) were amplified by polymerase chain reaction (PCR) from *P. stutzeri* ATCC 31258 (10) total DNA by using the proofreading Platinum® Pfx DNA Polymerase (Invitrogen, Carlsbad, CA) and specific primers. The 1082-bp PCR product of *carBaBb* genes was cloned by site-specific recombination (11) using two commercial expression vectors (Invitrogen): pDEST™17 (histidine [His]-tagged) and pDEST™14 (native form). Plasmid DNA was prepared from *E. coli* cells by alkaline lysis (12). Recombinant clones were confirmed by PCR, restriction enzyme digestion, and by sequencing. The cloning was described in detail in Larentis et al. (13).

### *CarB Expression, Preparation of Cell Extract, Purification, and Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis*

*E. coli* BL21-SI harboring recombinant expression vectors pDEST14 and pDEST17 were cultivated in 10 mL of LBON medium at 37°C until absorbance 0.8 at 600 nm was reached. Then, CarB expression was



induced with 0.3 M NaCl, and cells were incubated for 24 h. Samples (1 mL, before salt induction and for 24 h) were harvested, and the pellets were stored at  $-20^{\circ}\text{C}$ . The pellets from 1-mL samples were resuspended in GET (50 mM glucose, 10 mM EDTA, pH 8.0, and 25 mM Tris-HCl, pH 7.4), sonicated, and submitted to total protein concentration measurement by Bradford method using bovine serum albumin as the standard (14). CarB cloned with His-tag was purified by affinity chromatography on a nickel nitrilotriacetate resin (QIAGEN, Valencia, CA) following standard procedures. Then, 18% SDS-PAGE was performed with 20  $\mu\text{g}$  of cell extract in a Bio-Rad (Hercules, CA) apparatus essentially as described by Laemmli (15). Gel was stained with Coomassie brilliant blue R-250.

### Determination of meta-Cleavage Activity

Recombinant *E. coli* BL21-SI cells were cultivated overnight in 10 mL of LB medium at  $37^{\circ}\text{C}$ . A volume (200–250  $\mu\text{L}$ ) of this saturated culture was harvested, and the pellets were resuspended in Tris-HCl, pH 7.5, to a final buffer concentration of 50 mM in 1 mL of resting cell reaction. The same volume of culture was sonicated (five cycles of 10 s with 1-min intervals) to test cell extract activity. *meta*-Cleavage of the cell extract samples (sonicated) was also measured after incubation for 10 min on ice with 50  $\mu\text{M}$  ascorbic acid and 35  $\mu\text{M}$  ammonium iron(II) sulfate (1,16).

CarB activity was assayed by monitoring the increase in absorbance at 434 nm for the yellow metabolite 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid (HOPDA) produced by the degradation of 100  $\mu\text{M}$  2,3-DHBP (analogous to 2'-aminobiphenyl-2,3-diol), at  $35^{\circ}\text{C}$ . The molar extinction coefficient of HOPDA is  $13,200\text{ cm}^{-1}\text{M}^{-1}$  (1,2).

### Structural Characterization by Homology Modeling

The comparative modeling is performed in four iterative sequential steps (17): template selection, target-template alignment, model building of the target protein, and model evaluation.

The template selection was performed with Basic Local Alignment Search Tool-Protein-Protein (BLASTP) (18) against Protein Data Bank (PDB) for subunit CarBb. For CarBa, the template identification was performed by the threading methodology, i.e., by adjusting the amino acid sequence of CarBa into protein three-dimensional structures with program GenTHREADER (19). The templates for CarBa and CarBb were identified, respectively, as the 139-amino acid  $\alpha$  (LigA)- and 302-amino acid  $\beta$  (LigB)-chains of protocatechuate 4,5-dioxygenase (LigAB) from *Sphingomonas paucimobilis* SYK6 (5). The protocatechuate-LigAB complex (access

code 1B4U on PDB) was used as a template to analyze the substrate-CarB interactions.

The alignments between target and template sequences were made by T-COFFEE (20) and improved manually. These sequence alignments were used in the model construction of the CarBb and CarBa monomers by the spatial restrictions technique with the software MODELLER 6v2 (21). After these procedures, the final model was refined by a relaxation of the structure using Amber force field (22), through 100 steps of Steepest Descent and 200 steps of Gradient Conjugated algorithms using the program HyperChem (23). Spatial coordinates of the dimers CarBa and CarBb as well as the  $\alpha_2\beta_2$  heterotetramer were obtained using SPDBViewer (24) by superposition of the monomer models with each  $\alpha$ - and  $\beta$ -chain of the  $\alpha_2\beta_2$  template structure of 1B4U, respectively.

Model evaluation was performed by stereochemical quality analysis using the program PROCHECK (25); the compatibility of the atomic model (3D) with its own amino acid sequence (one-dimensional, 1D) was made by VERIFY3D (26), and errors in the structures were analyzed by WHAT-CHECK (27); the software PROVE (28) was used to assess the quality of the model based on the calculations of atomic volumes; nonbonded atomic interactions in the structures were analyzed by ERRAT (29). The residue and domain characterizations and the computation of bond distances were performed through visual inspection with the help of SPDBViewer (24).

## RESULTS

### Cloning, Expression, and Activity Measurements

The 1082-bp fragment corresponding to *carBaBb* genes from *P. stutzeri* ATCC 31258 were amplified by PCR and cloned by the site-specific recombination system into two expression vectors: pDEST17 (His-tagged) and pDEST14 (native form). Sequencing indicates that the nucleotide sequence presents at least 99% of identity with *Pseudomonas* sp. CA10 (2) and *P. stutzeri* OM1 (8).

The proteins were successfully expressed by salt induction in *E. coli* BL21-SI, as confirmed by the appearance of two bands corresponding to CarBb and CarBa expected protein sizes in 18% SDS-PAGE (29 and 10 kDa, respectively), with higher expression levels for CarBb (Fig. 1).

The *meta*-cleavage activity for both His-tagged and native-form recombinant resting cells were measured at  $35^{\circ}\text{C}$  and pH 7.5 by monitoring the absorbance at 434 nm of the appearance of HOPDA (1,2) by the degradation of 2,3-DHBP. The behavior observed for CarB cloned with and without the His-tag was similar (Table 1). The activity of CarB for the degradation of 2,3-DHBP was measured in three different conditions: resting

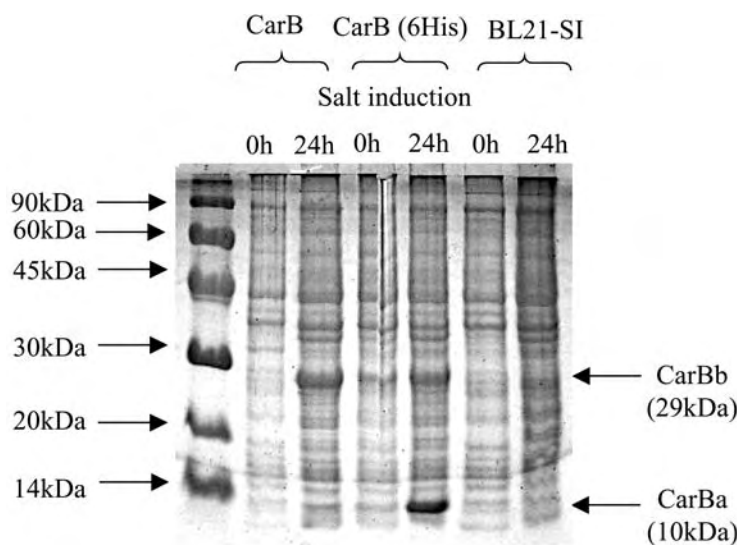


Fig. 1. SDS-PAGE of CarBa and CarBb obtained by BL21-SI cell extract samples without (0 h) and with salt induction (24 h). The samples are the recombinant protein expressed in native form (CarB), with His-tag (CarB-6His) and the strain BL21-SI without the cloning construction as negative control. Proteins were stained with Coomassie brilliant blue R-250. The protein molecular weight standard is LMW (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

Table 1

The *meta*-Cleavage Activity of CarB Cloned in His-Tagged and Native Form for 100  $\mu$ M 2,3-DHBP at 35°C and 50 mM Tris-HCl Buffer, pH 7.5, by Using Resting Cells, Cell Extract, and Reactivated With Ascorbic Acid and Ammonium Iron(II) Sulfate

Cell	Resting cells	Cell extract	Cell extract + ascorbate + Fe(II)
His-tagged	0.82	0.02	0.42
Native form	0.53	0.02	0.55

Activity was measured in micromoles of HOPDA per milligram of protein per minute. Activity was not detected in the negative control (BL21-SI strain). Errors are less than 10%.

cells, cell extract, and activated cell extract. No activity was detected with cell extract, but after the incubation with ascorbic acid and ammonium iron(II) sulfate on ice, the enzyme recovered the *meta*-cleavage activity (Table 1). CarB activity results from recombinant cells were comparable with those reported previously for the *meta*-cleavage enzyme from *Pseudomonas* sp. CA10 (1,2), *P. stutzeri* OM1 (8), *Pseudomonas* sp. LD2 (30), *Sphingomonas* CB3 (31), and enzymatic activity was detected in cells stored at 4°C for 30 d. However, the His-tagged protein eluted near homogeneity from nickel affinity column was unstable. The instability of the dioxygenases during purification was reported in

the literature. The instability of the dioxygenases during purification was reported in the literature (30).

### Construction of the 3D Structure of CarB

Each monomer model (CarBa and CarBb) was constructed using the homology modeling technique based on the crystallographic structure of the protocatechuate LigAB from *S. paucimobilis* (5).

The global alignment between the sequence of 29-kDa CarBb and the  $\beta$ -chain of the template (LigB) performed with T-COFFEE showed 31% of identity and more than 67% of similarity (Fig. 2). The model stereochemistry and geometry were considered good by PROCHECK analyses (25), with 97% of the residues in the allowed regions of the Ramachandran plot (32), including those used for structural analysis. VERIFY3D results indicated that almost 92% of the residues in the model are compatible with CarBb amino acid sequence (26). WHAT-CHECK, PROVE, and ERRAT analyses confirmed that the model is satisfactory (27–29).

The 10-kDa CarBa subunit 3D model was constructed with residues 30 to 139 of the  $\alpha$ -chain of the template (LigA) and presented 26% of identity and 61% of similarity in the alignment used in the modeling (Fig. 2). This manually improved alignment was selected by analysis of structural similarity to the template performed in SPDBViewer, after building several models. Because the BLAST search for a template did not find any significant similarity for homology inference between CarBa and PDB-available protein structures,



Fig. 2. (A) Alignment between CarBa and LigA sequences and (B) between CarBb and LigB. Identical amino acids are indicated by stars (\*); similar amino acids by (: or (;) different amino acids by (,) whereas [minus] indicates gaps added to the sequences. The catalytic residues of CarB (His12, His53, and Glu230) presented in CarBb and of LigAB (His12, His61, and Glu242) are highlighted (light gray).

the threading approach was used for template identification. This “profile-based threading” methodology has been developed so far for the identification of a protein structure from sequence information alone in the absence of apparent sequence identity to any other protein (33). Thus, the template search for CarBa was performed by GenTHREADER (19), indicating the LigA subunit as a protein with a high enough score for CarBa 3D model construction. The model presented no residues in disallowed regions in the Ramachandran plot, with 92% in the most favored regions (25,32). VERIFY3D analysis was similar to CarBb (26). The evaluation made by WHAT-CHECK, PROVE, and ERRAT also indicated that the model is satisfactory (27–29).

Evaluation of the monomer models indicated that the stereochemistry and energetic parameters are acceptable. The  $\alpha_2\beta_2$ -heterotetrameric structure of CarB was obtained using the molecular coordinates of CarBa and CarBb monomer 3D models (Fig. 3).

### Structural Analysis of CarB 3D Model

The  $\alpha_2\beta_2$ -heterotetrameric model is composed by two bound CarBb subunits ( $\beta$ -chain), and each subunit contacts one small CarBa unit ( $\alpha$ -chain); the two  $\alpha$ -subunits do not make direct contact. CarBa is formed by  $\alpha$ -helices and coils, whereas CarBb presents an  $\alpha/\beta$ -structure very similar to the template (5). The enzymatic catalytic site was characterized, indicating that the coordination of



Fig. 3. Loops,  $\alpha$ -helices, and  $\beta$ -sheets of the CarBa subunit (black chains) and CarBb (gray and light gray chains) of the structural model obtained for  $\alpha_2\beta_2$ -heterotetrameric CarB by homology. The catalytic residues (His12, His53, and Glu230) present in both CarBb chains at the interface with CarBa are also indicated in space-fill representation (gray and light gray).

Fe(II) in the catalytic site of the enzymatic complex of CarB occurs by three amino acid residues (His12, His53, and Glu230), completely conserved in the class III extradiol dioxygenases, as indicated by Sato et al. (2) and Sugimoto et al. (5). These residues in the active site are correctly positioned when compared to LigAB crystallographic structure, as shown in Fig. 4A. His12 and His53 are located on loops and seem to be relatively flexible, whereas Glu230 is very inflexible because it is set into an  $\alpha$ -helix. They are located in a deep cleft of the subunit CarBb, similar to the template (5). According to the template structure, the active site is probably completed by one water molecule, and it is covered by the CarBa subunit, which acts as a lid. The residues Ser54 to Thr75 from CarBa cover the active-site pocket, similar to those in the template (residues 85 to 110 from LigA). All of these residues in the LigAB are located in an  $\alpha$ -helix that closes the open end of the pocket, and, for this reason, the substrate is completely inaccessible by the solvent. Thus, in the absence of the substrate, five water molecules occupy the cavity, and this seems to be essential for the activity of this enzyme class (5).

The model also indicates that residues His117 and His182 in CarBb sequence are important for hydrogen bonding with both hydroxyl groups of the substrate catechol ring, and for that reason, for catalytic activity. Ile13, Leu14, Met15, and His117 of CarBb and some residues of the CarBa (Gln62 and Tyr65) subunit are the hydrophobic amino acid suggested to form the catalytic pocket that presents a complementary surface shape to the substrate (Fig. 4B). Apparently, the substitution of Ser269 and Asn270 residues present in the LigB

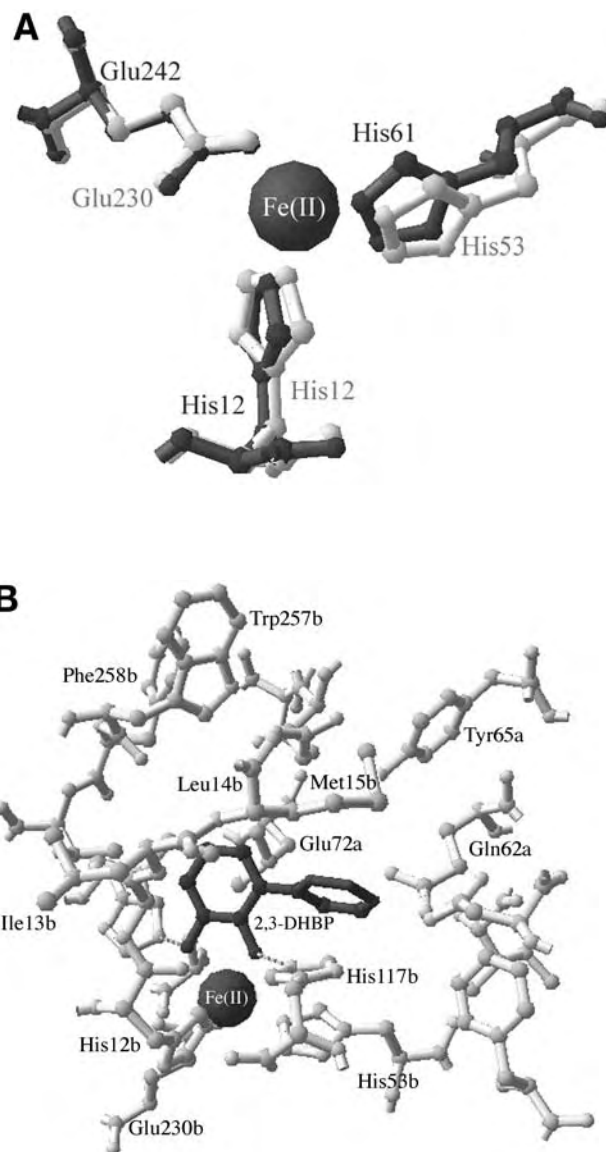


Fig. 4. (A) Representation of the catalytic site in the subunit CarBb in light gray [Fe(II) coordinated by residues His12, His53, and Glu230] superimposed to the catalytic site of the template LigB in dark gray (residues His12, His61, and Glu242). (B) The complex structure of CarB, 2,3-DHBP, and hydrogen bonds (dots) between amino acid in the catalytic site and the hydroxyl groups of the substrate, and hydrophobic residues that could form the substrate-binding pocket. The suffixes "a" and "b" attached to a residue refer to CarBa and CarBb subunits, respectively. (C) Superposition of the catalytic site (ferrous iron) of the complex LigAB-protocatechuate (PCA) in dark gray to the catalytic site of CarB (light gray). The dots represent the hydrogen bonds formed in the surrounding amino acid residues that are important for ring fission catalysis of PCA. Notice that, different from the LigAB-PCA, no hydrogen bonds are formed in the complex CarB-2,3-DHBP except between the OH groups with hydrophobic residues. (Figure continues)

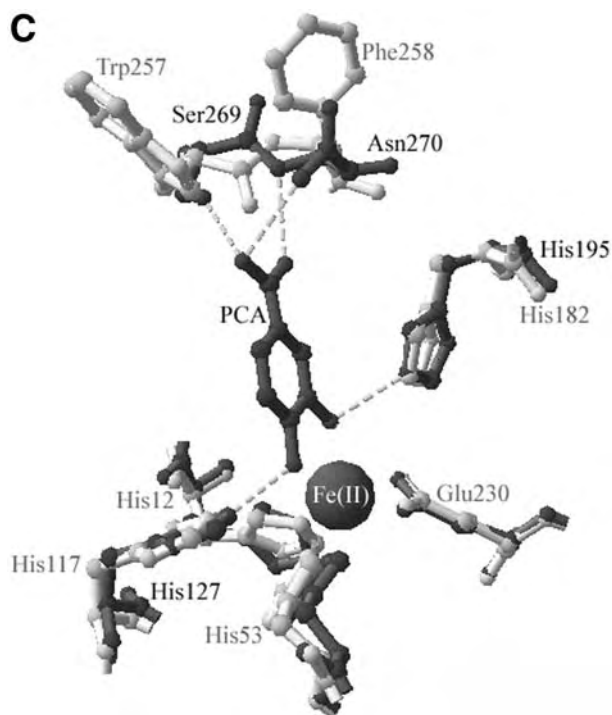


Fig. 4. (continued)

sequence by hydrophobic residues (Trp257 and Phe258, respectively) in CarB sequence do not allow the hydrogen bond formation between the protocatechuate and these residues (Fig. 4C).

## DISCUSSION

Information about the enzymes involved in carbazole degradation pathways, their function, and their selectivities for different nitrogenated compounds is essential for developing biocatalysts for petroleum biorefining and bioremediation. Hence, structural characterization of subunits and catalytic sites of the enzymatic complexes involved in carbazole degradation is a useful approach to understand the reaction mechanisms, considering the lack of experimental 3D data. Although the crystallization of CarBaBb has already been described by Iwata et al. (34), the enzyme structure has not been reported yet. Thus, the theoretical structural approach described here can be used to gain a better understanding of the enzymatic mechanism. With the combination of cloning and modeling analysis, we can characterize the enzymes and study protein interactions and reaction mechanisms to evaluate and to enhance this system for carbazole degradation.

The measurements of *meta*-cleavage activity of recombinant cells (Table 1) indicated the CarB require-

ment of the ferrous ion, which is supplied by the cell environment in the resting cell reaction. CarB dependence on Fe(II) also was shown by sequence homology by Sato et al. (2). The instability of CarB eluted in pure form from affinity chromatography may be caused by loss of structural Fe(II), low ionic strength (30), or other factors.

Structural *in silico* characterization of the enzyme confirmed that the amino acid residues His12, His53, and Glu230 (Fig. 4A) are coordinated to Fe(II) in the catalytic site of CarB. Because the comparative modeling was able to produce consistent models for CarBb and CarBa and also for the  $\alpha_2\beta_2$ -heterotetrameric structure of CarB (Fig. 3), the model was useful to explain some experimental behavior of the *meta*-cleavage enzyme and to give further insight on the catalytic mechanism.

Structural analysis of the model active site were made in comparison with protocatechuate LigAB (5) and with 2,3-dihydroxybiphenyl-1,2-dioxygenase (BphC) (6,16), by superposition of 2,3-DHBP and protocatechuate. The model indicates that residues His117 and His182 in the CarBb sequence are important for hydrogen bonding and for catalytic activity. His182 is the corresponding amino acid to His195 for LigAB and His194 for BphC, which are suggested to be the catalytic base responsible for deprotonating the hydroxyl group, through the formation of a hydrogen bond (5,16). A hydrogen bond between His182 and the OH from 2,3-DHBP was verified in the CarB model (Fig. 4B), supporting the hypothesis that this residue could be the base for the catalysis in CarB. The other OH group of the substrate catechol ring is hydrogen bonded to His117, corresponding to His127 of LigAB, which was confirmed with computation of H bonds in the model.

The residues Ile13, Leu14, Met15, His117, Leu184, and Thr259 of CarBb could form the catalytic binding pocket. In the LigAB structure, Ile13, Pro14, Leu197, His195, and Thr271 (LigB) and also the LigA Phe103 contact the catechol ring of the substrate (protocatechuate), and there is a correlation between the van der Waals surface of bound substrate and of those residues (5,16). The binding pocket also may comprise some residues from CarBa (indicated in Fig. 4B). The Phe103 residue of LigA corresponds to the negative residue Glu72 in the alignment of CarBa, which may interact by hydrogen bonding with the NH<sub>2</sub> group of the CarB substrate 2'-aminobiphenyl-2,3-diol (Fig. 4B). Uragami et al. (16) described the shape of the substrate binding pocket and proposed that its hydrophobic character is important to substrate binding. Also, Sugimoto et al. (5) showed that the precise complementary surface shape between the substrate and its binding pocket seems to be the dominant factor in stabilizing substrate binding, rather than the iron and hydroxyl group coordination. The model analysis indicated that the hydrophobic residues near the catalytic

site, as described above, probably permit the binding of the aromatic ring opening of the substrates as well as forming a pocket that fits the compound 2,3-DHBP, which may explain the high affinity between CarB and 2,3-DHPB. This result is interesting because the model was constructed using the enzyme that degrades protocatechuate and not the substrate for CarB specifically. Even though the sequence and structures of CarB and BphC are different, both are able to degrade 2,3-DHBP (1,5,16), which is probably related to the nature of the binding pocket. This is in accordance to literature reports (5), i.e., those enzymes present similar active sites and reaction mechanisms. Although the models were constructed using alignments that do not present very high percentages of sequential identity, the modeling, as well as the molecular structures already reported in the literature support, points to the hydrophobic character of the residues that form the binding pocket in the ring-opening catecholic dioxygenases being maintained. Furthermore, the points raised in the Discussion are not invalidated, mainly because the most important structural conclusions are taken from the active site, whose residues and geometries are completely conserved in these enzymes (5).

As indicated in Fig. 4C, Ser269 and Asn270 in LigB were substituted by hydrophobic residues Trp257 and Phe258 in CarBb structure, which do not form hydrogen bonds with the protocatechuate. Sugimoto et al. (5) describe that these bonds are essential for protocatechuate stabilization in the active site of LigAB for ligand degradation, once LigAB shows little activity for catechol and its derivatives lacking the carboxylate group. This feature can explain why CarB, even with a similar structure to LigAB, is not able to degrade protocatechuate.

Although the structures of class I, II, and III of extradiol dioxygenases were completely unrelated and do not exhibit any similarity, their active sites share common characteristics (nonheme ferrous coordination in a similar geometrical rearrangement), strongly suggesting the same reaction mechanism (5,7). Thus, for CarB and LigAB (both class III), the same ring fission reaction mechanism can be inferred. This assumption is reinforced by the complete conservation of the amino acid in the active site, the coherent structural characterization, and the similar geometry shared by the enzymes.

## ACKNOWLEDGMENTS

We thank Drs. Helen C. Ferraz and Marcelo H. Herbst for fruitful discussions and comments on the manuscript. This work was supported by grants from Conselho Nacional de Desenvolvimento Científico e

Tecnológico (CNPq), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), and CTPetro/FINEP-no.21.01.0278.00.

## REFERENCES

1. Iwata, K., Nojiri, H., Shimizu, K., Yoshida, T., Habe, H., and Omori, T. (2003) Expression, purification, and characterization of 2'-aminobiphenyl-2,3-diol 1,2-dioxygenase from carbazole-degrader *Pseudomonas resinovorans* strain CA10. *Biosci. Biotechnol. Biochem.* **67**, 300–307.
2. Sato, S. I., Ouchiyama, N., Kimura, T., Nojiri, H., Yamane, H., and Omori, T. (1997) Cloning of genes involved in carbazole degradation of *Pseudomonas* sp. strain CA10: nucleotide sequences of genes and characterization of *meta*-cleavage enzymes and hydrolase. *J. Bacteriol.* **179**, 4841–4849.
3. Sato, S. I., Nam, J.-W., Kasuga, K., Nojiri, H., Yamane, H., and Omori, T. (1997) Identification and characterization of genes encoding carbazole 1,9a-dioxygenase in *Pseudomonas* sp. strain CA10. *J. Bacteriol.* **179**, 4850–4858.
4. Benedik, M. J., Gibbs, P. R., Riddle, R. R., and Willson, R. C. (1998) Microbial denitrogenation of fossil fuels. *Trends Biotechnol.* **16**, 390–395.
5. Sugimoto, K., Senda, T., Aoshima, H., Masai, E., Fukuda, M., and Mitsui, Y. (1999) Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. *Structure* **7**, 953–965.
6. Senda, T., Sugiyama, K., Narita, H., et al. (1996) Three-dimensional structures of free form and two substrate complexes of an extradiol ring-cleavage type dioxygenase, the BphC enzyme from *Pseudomonas* sp. strain KKS102. *J. Mol. Biol.* **255**, 735–752.
7. Vaillancourt, F. H., Labbé, G., Drouin, N. M., Fortin, P. D., and Eltis, L. D. (2002) The mechanism-based inactivation of 2,3-dihydroxybiphenyl 1,2-dioxygenase by catecholic substrates. *J. Biol. Chem.* **277**, 2019–2027.
8. Ouchiyama, N., Miyachi, S., and Omori, T. (1998) Cloning and nucleotide sequence of carbazole catabolic genes from *Pseudomonas stutzeri* strain OM1, isolated from activated sludge. *J. Gen. Appl. Microbiol.* **44**, 57–63.
9. Bhandari, P. and Gowrishankar, J. (1997) An *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. *J. Bacteriol.* **179**, 4403–4406.
10. Hisatsuka, K. and Sato, M. (1994) Microbial transformation of carbazole to anthranilic acid by *Pseudomonas stutzeri*. *Biosci. Biotech. Biochem.* **58**, 213–214.
11. Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795.
12. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
13. Larentis, A. L., Alves, T. L. M., and Martins, O. B. (2005) Cloning and expression of *meta*-cleavage enzyme (CarB) of carbazole degradation pathway from *Pseudomonas stutzeri*. *Braz. Arch. Biol. Technol.* **48**, 127–134.

14. Bradford, M. M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
15. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
16. Uragami, Y., Senda, T., Sugimoto, K., et al. (2001) Crystal structures of substrate free and complex forms of reactivated BphC, an extradiol type ring-cleavage dioxygenase. *J. Inorg. Biochem.* **83**, 269–279.
17. Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325.
18. 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.
19. Jones, D. T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287**, 797–815.
20. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
21. Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
22. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986) An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230–252.
23. Froimowitz, M. (1993) HyperChem: a software package for computational chemistry and molecular modeling. *Biotechniques* **14**, 1010–1013.
24. Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723.
25. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
26. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85.
27. Hooft, R. W. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures. *Nature* **381**, 272.
28. Pontius, J., Richelle, J., and Wodak, S. J. (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* **264**, 121–136.
29. Colovos, C. and Yeates, T. O. (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519.
30. Gibbs, P. R., Riddle, R. R., Marchal, L., Benedik, M. J., and Willson, R. C. (2003) Purification and characterization of 2'-aminobiphenyl-2,3-diol 1,2-dioxygenase from *Pseudomonas* sp. LD2. *Protein Expr. Purif.* **32**, 35–43.
31. Shepherd, J. M. and Lloyd-Jones, G. (1998) Novel carbazole degradation genes of *Sphingomonas* CB3: sequence analysis, transcription, and molecular ecology. *Biochem. Biophys. Res. Commun.* **247**, 129–135.
32. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain conformations. *J. Mol. Biol.* **7**, 95–99.
33. Rost, B., Schneider, R., and Sander, C. (1997) Protein fold recognition by prediction-based threading. *J. Mol. Biol.* **270**, 471–480.
34. Iwata, K., Noguchi, H., Usami, Y., et al. (2004) Crystallization and preliminary crystallographic analysis of the 2'-aminobiphenyl-2,3-diol 1,2-dioxygenase from the carbazole-degrader *Pseudomonas resinovorans* strain CA10. *Acta Crystallogr.* **D60**, 2340–2342.

**Manuscrito 4:** “Cloning, expression, partial characterization and structural modeling of a novel esterase from *Pyrococcus furiosus*”.

Em colaboração com o aluno de doutorado do Prof. Tito Alves e Prof. Orlando Martins, Rodrigo Volcan Almeida, objetivávamos obter um modelo estrutural para a lipase de *P. furiosus*, através do qual pudéssemos extrair um conjunto de informações a respeito de seu sítio ativo, sobre as razões estruturais de sua termoestabilidade e também de aminoácidos responsáveis pelo seu pH ótimo de atuação. Desta forma lançou-se mão da metodologia de “threading”, visto que lípases não muito conservadas em sua seqüência primária; a qual busca conservações a partir da estrutura terciária. Com esta metodologia encontramos um possível molde: a estrutura da prolil oligopeptidase de porco (*Sus scrofa*).

Este resultado, além de indicar a possibilidade de seguir adiante nos experimentos de modelagem, também nos abriu outras possibilidades de investigação como, por exemplo, as relações estruturais entre estas duas famílias de enzimas e suas relações evolutivas, visto que estamos tratando de duas enzimas com atividades diferentes, de organismos bastante diferentes, mas com estruturas terciárias bastante similares.

De posse de um molde que nos possibilitou a construção do modelo, foi possível através de sua análise, identificar os resíduos Ser149 - Asp233 - His264 como integrantes de seu sítio ativo e, além disso, propor, através do conhecimento das razões que governam a dependência da atividade de lipases com o pH do meio, mutações (Lys186Glu e Glu19SLys) que alterassem o pH ótimo de atuação da lipase de *P. furiosus*, gerando assim dois protótipos enzimáticos, um com atuação em pH mais ácido e outro em pH mais alcalino que a enzima selvagem.



## Cloning, expression, partial characterization and structural modeling of a novel esterase from *Pyrococcus furiosus*

Rodrigo Volcan Almeida<sup>a,b,\*</sup>, Sylvia Maria Campbell Alquéres<sup>b</sup>, Ariane Leites Larentis<sup>a,b</sup>, Shaila Cíntia Rössle<sup>c</sup>, Alexander Machado Cardoso<sup>b</sup>, Welington Inácio Almeida<sup>b</sup>, Paulo Mascarello Bisch<sup>c</sup>, Tito Lívio Moitinho Alves<sup>a</sup>, Orlando Bonifácio Martins<sup>b</sup>

<sup>a</sup> Laboratório de Bioprocessos, Programa de Engenharia Química, COPPE, UFRJ, Brazil

<sup>b</sup> Laboratório de Biologia Molecular, Programa de Biotecnologia e Biologia Molecular, IBqM, UFRJ, Brazil

<sup>c</sup> Laboratório de Física Biológica, IBCCF, UFRJ, Brazil

Received 13 October 2005; received in revised form 16 February 2006; accepted 17 February 2006

### Abstract

In this report the ORF *PF2001* from the hyperthermophilic archaeon *Pyrococcus furiosus* was identified and its protein sequence was characterized in silico, revealing characteristics of the conserved domains of the dipeptidyl aminopeptidases, hydrolases of the  $\alpha/\beta$  superfamily, esterases and lipases. In order to understand its function, the ORF *PF2001* without the 60 bp of the 5'-terminus, responsible for encoding a signal peptide (*PF2001* $\Delta$ 60), was cloned and expressed in *Escherichia coli* BL21(DE3) pLysS, and the recombinant enzyme was characterized for esterase, lipase and protease activities. The total protein extract from *E. coli* harboring the plasmid containing the ORF *PF2001* $\Delta$ 60 exhibited its highest activity towards the substrate 4-methylumbelliferyl-heptanoate (C7) and lower activities towards 4-methylumbelliferyl-acetate (C2) and 4-methylumbelliferyl-palmitate (C16). The enzyme was thermostable for 120 min at 75 °C and was completely inhibited by 1 mM PMSF. A theoretical structural model was constructed by comparative modeling using as template a prolyl oligopeptidase from *Sus scrofa*. Although no protease activity was detected a putative catalytic triad (Ser149, Asp233 and His264) with high similarity to the template was identified. The structural characteristics that confer enzymatic specificity to the *P. furiosus* enzyme are discussed. Taken together the data strongly suggest that ORF *PF2001* from *P. furiosus* is responsible for encoding a novel esterase.

© 2006 Elsevier Inc. All rights reserved.

**Keywords:** Esterases; Lipases; Proteases; Comparative modeling; Homology modeling; Extremozymes; Thermophilic enzymes; Archaea

### 1. Introduction

The search for new enzymes to catalyze a diversity of reactions with high efficiency and stability has been a constant pursuit of enzymologists, biochemists and chemical engineers. This search has been directed primarily toward exploration of microbial diversity [1–6], and sometimes this biodiversity is accessed through genomic or cDNA libraries [7–11]. This approach takes advantage of the great biodiversity of the Earth in order to find enzymes that catalyze the reaction of interest. More recently, protein engineering techniques such as site-directed mutagenesis and error-prone PCR have been introduced making it

possible to modify known enzymes to meet specific requirements [12–15].

In the last decade, the increase in number of sequencing projects has yielded gene and protein databases with very useful information for screening of new potential biocatalysts [16,17]. However, this information has not been fully exploited, since it is estimated that 30% of all deposited sequences in a completely sequenced genome are hypothetical proteins.

Esterases, lipases and proteases are very important biocatalysts for industrial purposes. They can be utilized in laundry, food (lipases and proteases), oil chemistry (lipases), fine chemistry, pharmaceutical (esterases, lipases and proteases) and paper industries (lipases), as well as in biodiesel production (lipases), waste treatment (lipases and proteases) and other biotechnological applications [18–22].

The hyperthermophilic archaea have received increasing attention because of the high thermostability of their enzymes

\* Corresponding author at: Universidade Federal do Rio de Janeiro, Av. Brigadeiro Trompowski s/no, Cidade Universitária, Centro de Tecnologia, bloco G, sala G115, Rio de Janeiro, RJ 21941-590, Brazil. Fax: +55 21 2562 8300.

E-mail address: [volcan@peq.coppe.ufrj.br](mailto:volcan@peq.coppe.ufrj.br) (R.V. Almeida).

[23]. *Pyrococcus furiosus* is a hyperthermophilic Archaeon that grows optimally at 100 °C [24], has its genome completely sequenced [25] and is a source of several important enzymes: the well-known commercially available DNA polymerase [26] and other potential biocatalysts, including  $\alpha$ -amylases [27–29],  $\beta$ -glucosidase [30,31], esterase [7], carboxyl peptidase [32] and proteases [33,34].

In this work the ORF *PF2001* from *P. furiosus* and its protein sequence were characterized in silico, revealing conserved domains for peptidase, lipase and esterase. This gene was cloned, expressed, characterized and structurally modeled and the results show that the enzyme is a novel esterase.

## 2. Materials and methods

### 2.1. Enzymes, strains, and chemicals

Restriction enzymes and the vector pGEM-T were purchased from Promega. Platinum Taq DNA polymerase high fidelity and *Escherichia coli* DH10B were purchased from Invitrogen. *E. coli* BL21(DE3) pLysS and the vector pET32a were obtained from Novagen. Ultra-pure deoxynucleotide solution (dNTPs), isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG), antibody anti-His tag and the ECL Western Blotting System were purchased from Amersham Pharmacia/GE Healthcare; 4-methylumbelliferyl-acetate (MUF-Ace), 4-methylumbelliferyl-heptanoate (MUF-Hep) and 4-methylumbelliferyl-palmitate (MUF-Pal) were purchased from Sigma. All other chemicals were of the highest reagent grade commercially available.

### 2.2. Data mining

Gene data mining was performed using the free softwares BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) [35] and CD-SEARCH (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [36]. N-terminal sequence analysis of the translational product of *PF2001* gene was performed by SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) [37].

### 2.3. Cloning

The mature gene (*PF2001* $\Delta$ 60) was amplified by PCR using the primers: 5'-AGATCTGGGTACCGACGACGACGACAAGGGGTATAAAATGGTAAATCCAC CT-3' and 5'-GTCGACTCATCCCATCCACCTTTAAGGAACTCTC-3'. The Bgl II and Sal I restriction sites are underlined in the forward and reverse primers, respectively. The enterokinase site is indicated in bold on the forward primer. The PCR product was inserted into pGEM-T Easy Vector and transformed in *E. coli* DH10B. DNA was dideoxy-cycle-sequenced with fluorescent terminators (Big Dye, Applied Biosystems, Foster City, CA) in an Applied Biosystems ABI Prism 377 automated DNA sequencer. The confirmed recombinant plasmid was digested by Bgl II and Sal I endonucleases and the product was purified and inserted into pET-32a digested with the same enzymes. This vector fuses thioredoxin to the amino-terminal region of the expressed protein. The recombinant product was primarily transformed in maintenance host *E. coli* DH10B and further in expression host *E. coli* BL21(DE3) pLysS.

### 2.4. Expression

The transformed cells were grown in LB medium (0.5% yeast extract, 1% tryptone and 0.5% NaCl) containing ampicillin (100 ng/mL) and chloramphenicol (12.5 ng/mL). After incubation with shaking at 37 °C until the  $A_{600\text{nm}}$  reached 0.6, the induction was carried out by adding IPTG at a final concentration of 0.5 mM and incubating for 3 h at 37 °C. The induced cells were collected by centrifugation and stored at –20 °C.

The expression level was analyzed by Western blot using the antibody anti-His tag and the ECL kit following the instructions of the manufacturers.

### 2.5. Enzymatic activity assays

The stored cells were resuspended in phosphate buffer (50 mM pH 7.0). After ultrasonic cell disruption, the total protein extract was centrifuged at 4000  $\times$  g for 15 min at 4 °C. The supernatant was incubated at 90 °C for 10 min and centrifuged at 10000  $\times$  g for 20 min at 4 °C. This supernatant was used for esterase/lipase and protease activity assays.

Esterase/lipase assays were carried out using 4-methylumbelliferone (MUF)-derivatives as substrates (4-methylumbelliferyl acetate, heptanoate and palmitate) in a Varian Cary Eclipse spectrofluorimeter, as described elsewhere [38]. The reaction mixture contained 0.6 mL of 50 mM phosphate buffer pH 7.0 containing 0.1% arabic gum and 0.4% Triton X-100<sup>®</sup>, and 2.4  $\mu$ L of MUF-derivative substrate stock solution (25 mM in ethyleneglycol monomethylether). Sixty microliters of protein extract (35–45  $\mu$ g of total protein) were added to start the reaction. Enzymatic activity was determined at 37 °C by measuring the increase of fluorescence emission ( $\lambda_{\text{ex}} = 323$  nm and  $\lambda_{\text{em}} = 448$  nm) due to the release of MUF caused by hydrolysis for the first 10 min. The reaction was carried out at 37 °C due to unspecific hydrolysis of MUF-acetate at higher temperatures. A calibration curve was constructed with 4-methylumbelliferone (MUF). One unit of activity was defined as the amount of enzyme required to release 1  $\mu$ mol of MUF per minute under conditions described above. For optimum temperature determination, the reaction was assayed using MUF-heptanoate as substrate at 40, 50, 60, 70 and 80 °C with 50 mM phosphate buffer pH 7.0. For optimum pH determination, the reactions were performed with 100 mM bis-tris-propane buffer at pH 6, 7, 8 and 9. Protease assays were carried out according to Kunitz [39] at 37 °C, using casein as substrate. Protein concentration was determined by the Bradford method [40].

### 2.6. Thermostability analysis

The enzyme samples in phosphate buffer (50 mM pH 7.0) containing 0.1% arabic gum and 0.4% Triton X-100<sup>®</sup> were incubated in a heating block at 75 and 100 °C. Tubes were removed periodically, chilled on ice, centrifuged briefly and assayed for esterase activity using MUF-heptanoate at 37 °C.

### 2.7. Comparative modeling

The template identification was performed by threading methodology, i.e., by fitting the amino-acid sequence of *P. furiosus* enzyme into protein three-dimensional structures available in the Protein Data Bank [41]. This procedure was carried out with the program THREADER 3 [42]. The alignments between the amino-acid sequences of the protein template and *P. furiosus* enzyme were made by T-COFFEE [43] and manually adjusted. Six different sequence alignments were tested in the model construction of *P. furiosus* enzyme by the spatial restrictions technique using the *model* routine from software MODELLER 6v2 [44]. These preliminary models were evaluated through measurements of spatial distances between hydrogen-bonded residues in the proposed active site (O $\gamma$ -Ser149/N $\epsilon$ -His264 and N $\delta$ -His264/O $\delta$ -Asp233) with the help of SPDBViewer v3.7(SP5) [45]. The best model was refined by a relaxation of the structure using Amber force field [46], through 100 steps of Steepest Descent and 200 steps of Gradient Conjugated algorithms using the program HyperChem [47]. Final model evaluation was performed by several methods: stereochemical quality was analyzed using the program PROCHECK [48], errors in the structures were analyzed by WHAT-CHECK [49], the software PROVE [50] was used to assess the quality of the model based on the calculations of atomic volumes, and non-bonded atomic interactions in the structures were analyzed by ERRAT [51].

## 3. Results

### 3.1. Cloning, expression and characterization of *PF2001* $\Delta$ 60

*P. furiosus* and *P. abyssi* are producers of esterases [2,7]. However, the genes responsible for esterase activities in these organisms have not yet been identified. A

gene from *P. abyssi*, described as a hypothetical 2-acetyl-1-alkylglycerophosphocholine esterase (E.C. 3.1.1.47), was found in the Genbank database. This gene exhibits a high similarity (60% identity) with the ORF *PF2001* from *P. furiosus*

(data not shown). Although the alignment showed high similarity, the 5' and 3' regions presented low similarity with a target hypothetical 2-acetyl-1-alkylglycerophosphocholine esterase. For this reason, these regions were analyzed and the

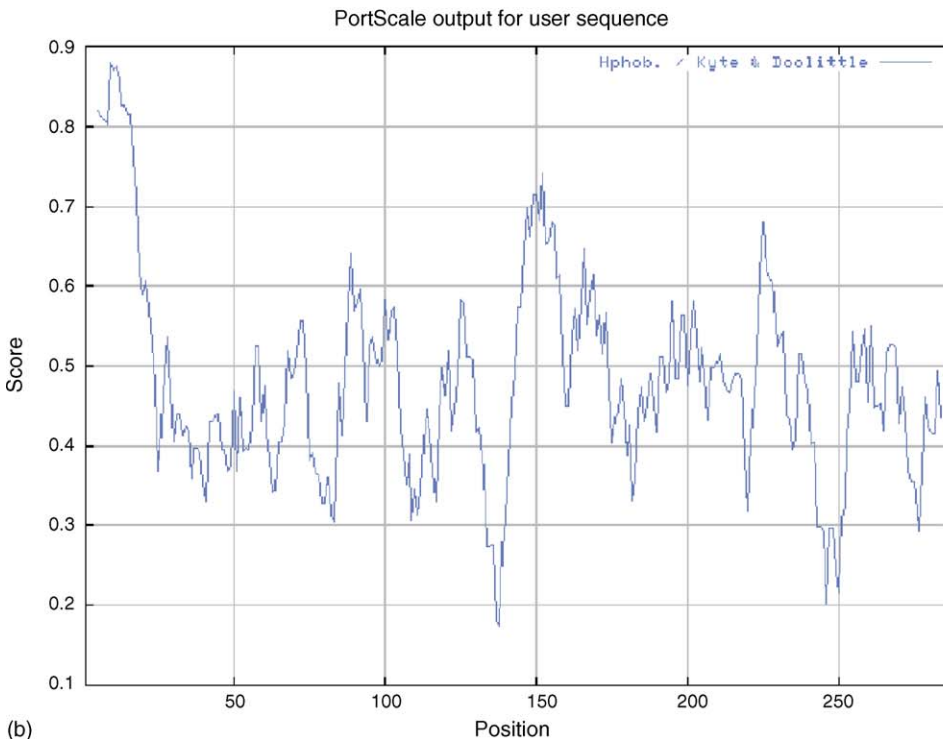
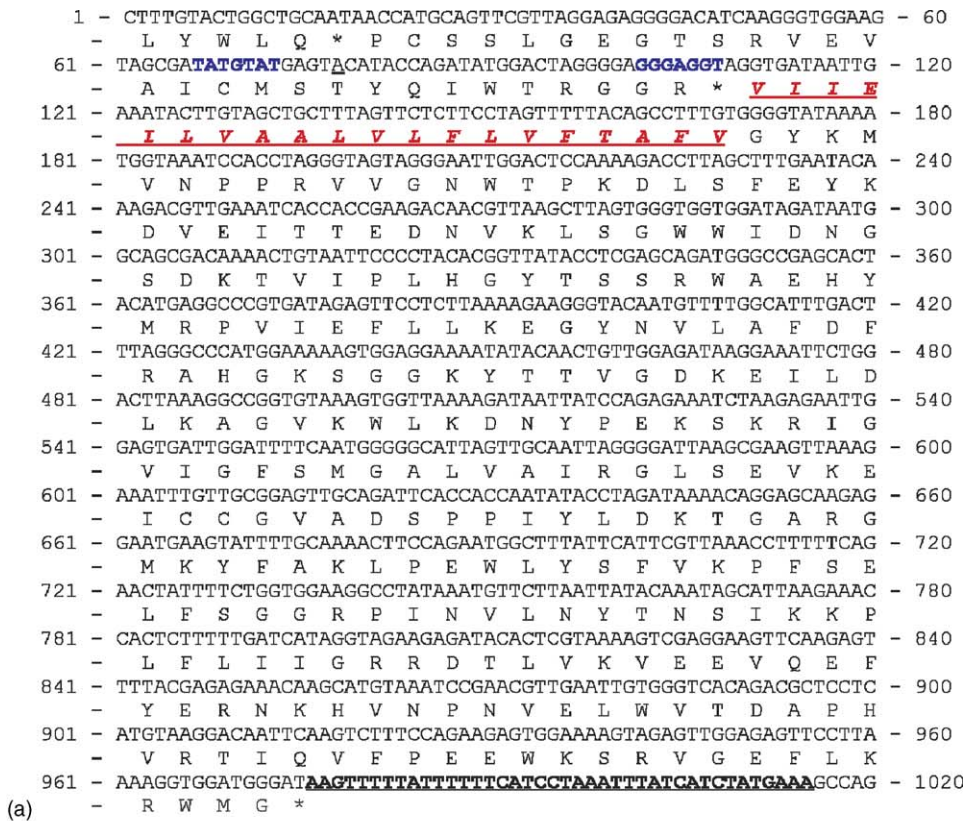


Fig. 1. Nucleotide sequence of *PF2001* gene from *Pyrococcus furiosus* and its deduced amino acid sequence (a); hydropathicity scale (b). Possible Pribnow box and Shine-Dalgarno sites (bold); signal peptide (italic underlined); stop codon (\*); pyridine-rich regions (bold underlined).

probable promoter and terminator sequences were identified (Fig. 1a).

The sequence GTG was proposed as the start codon, and the GTGAGGT sequence was found nine base pairs upstream to the GTG start codon. This sequence is similar to the putative Shine-Dalgarno GGTGGA and GAGGT of the  $\alpha$ -amylase gene from *P. furiosus* [22] and the glyceraldehyde 3-phosphate dehydrogenase gene from *P. woesei* [52]. Furthermore, a probable TATGTAT Pribnow-box sequence was identified and a region with a high content of pyrimidines, as in the  $\alpha$ -amylase gene from *P. furiosus* [28], was found after the TAA stop codon. The proposed ORF contains 864 bp, which yields a protein with a molecular weight of 32.9 kDa. The content of G + C is 47.8%, which is greater than the 40.8% presented by *P. furiosus* genome [53]. Adenine and thymine are the nucleotides preferred in the third position of the codons, as observed throughout its genome (<http://www.kazusa.or.jp/codon/>). N-terminal analysis of the translational product from ORF *PF2001*, using SignalP [37] showed that the first 20 amino acids comprise a probable signal peptide with high hydrophobic content (Fig. 1b).

Next, the conserved domains present in the encoded protein were analyzed using CD-SEARCH [36]. Three related domains were identified: dipeptidyl aminopeptidases (COG1506), hydrolases of the  $\alpha/\beta$  superfamily (COG1073), esterases and lipases (COG1647). The dipeptidyl peptidase domain is found in several proteins which cleave peptide bonds, including prolyl amino peptidases. These enzymes are serine proteases with different primary structure orientation of the catalytic site when compared to chymotrypsin (His-Asp-Ser) and subtilisin families (Asp-His-Ser). Their catalytic sites (Ser-Asp-His) are similar to those of esterases and lipases. Furthermore, the protein members

of the  $\alpha/\beta$  hydrolase family include dipeptidyl aminopeptidase, esterase and lipase domains. These data suggest that the hypothetical protein from *P. furiosus* could be a member of the protease and esterase/lipase families. In order to test this hypothesis, the mature gene (*PF2001* $\Delta$ 60) was cloned. The Western blot of the protein extract from *E. coli* BL21(DE3) pLysS harboring the plasmid containing the ORF *PF2001* $\Delta$ 60 as revealed with His tag antibody shows the expression of a 47.8 kDa protein (Fig. 2a), which is consistent with the molecular weight of expressed fusion protein. Another band with lower molecular weight (37 kDa) also can be visualized (Fig. 2a), probably due to proteolysis.

Fig. 2b illustrates the specificity of the enzyme from *P. furiosus* for substrates with different chain lengths. As can be observed, the enzyme hydrolyzed MUF-Ace and MUF-Pal at similar rates, approximately 13% of the rate of MUF-Hep hydrolysis. Taking into account that esterases are, by definition, enzymes that have the ability to hydrolyze ester substrates with short-chain fatty acids ( $\leq$ C10), whereas lipases hydrolyze long-chain acylglycerols ( $\geq$ C10) [54], the ORF *PF2001* encodes an esterase. Furthermore, the enzyme from *P. furiosus* was assayed for protease activity and no hydrolysis of casein was detected.

A typical kinetic experiment of MUF-Hep hydrolysis using protein extract from *E. coli* BL21(DE3) pLysS harboring the plasmid containing the ORF *PF2001* $\Delta$ 60, illustrates the inhibition of *P. furiosus* enzyme by PMSF (Fig. 2c). This result suggests that a serine is present in the active site.

The temperature dependence for esterase activity, measured at pH 7.0 using protein extract from the recombinant *E. coli*, showed an optimum at 60 °C, and retained 65% of the maximal activity at 80 °C (Fig. 3a). The enzyme was stable for 2 h at 75 °C,

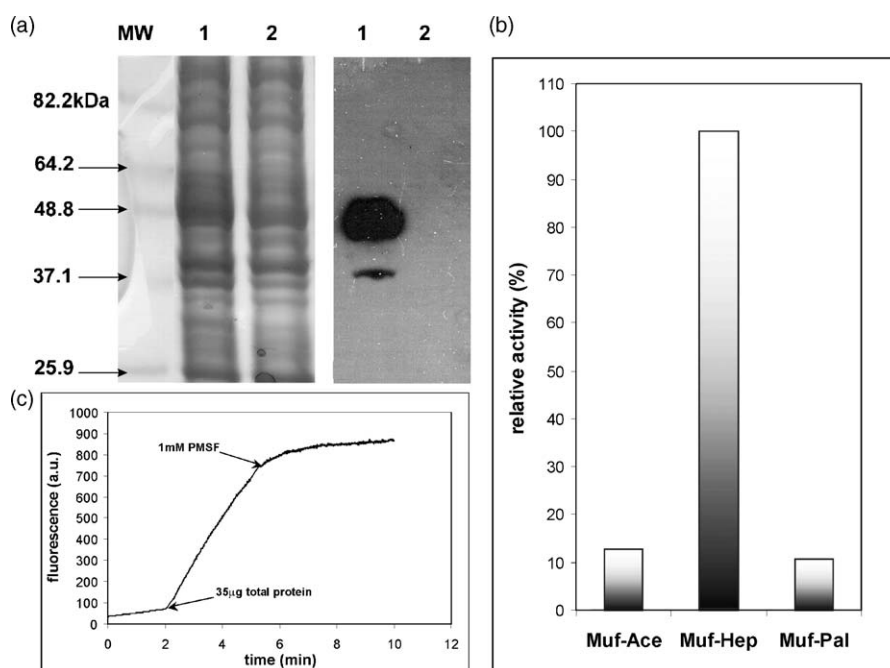


Fig. 2. Expression of the enzyme from *P. furiosus*. SDS-PAGE and Western blot (a); specificity of *P. furiosus* enzyme towards MUF-derivatives (b); hydrolysis of MUF-Hep and inhibition by PMSF (c). Panel (a) shows molecular weight (MW), total protein extract of recombinant *E. coli* BL21(DE3) pLysS (1) and total protein extract of *E. coli* BL21(DE3) pLysS (2).

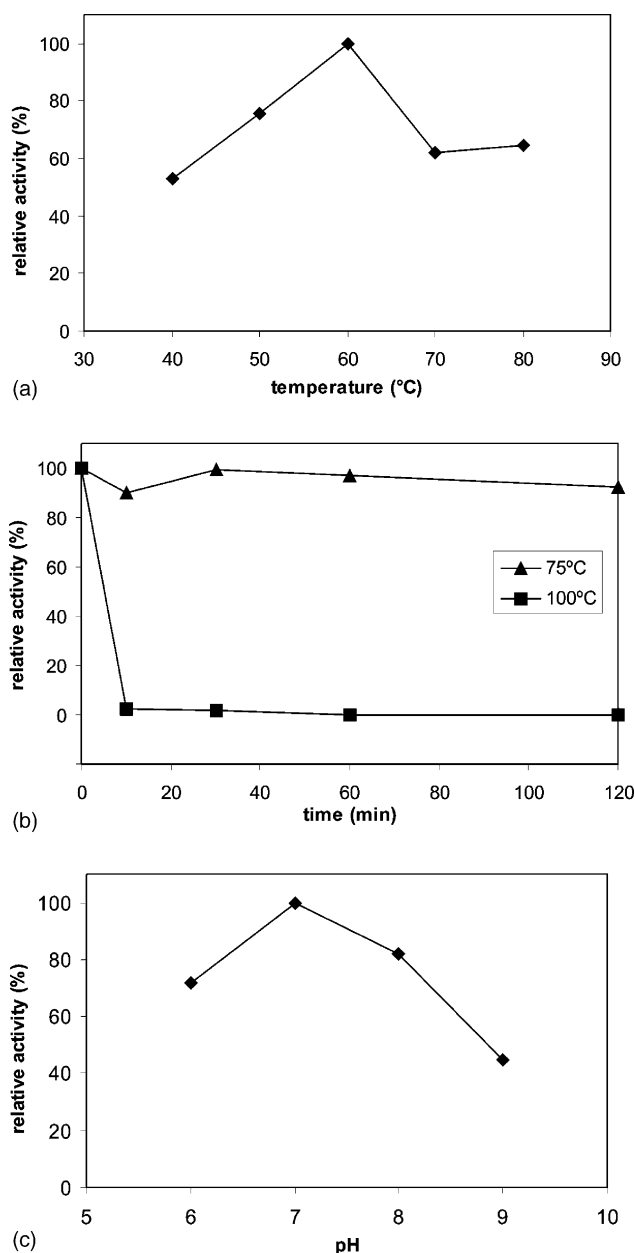


Fig. 3. Some characteristics of recombinant enzyme after thermal treatment (90 °C for 10 min). Effect of temperature on esterase activity (a); thermostability (b); effect of pH on esterase activity (c). The temperature effect and thermostability were measured in 50 mM phosphate buffer pH 7.0. The pH effect was measured out at 37 °C in 100 mM bis-tris-propane.

but lost its activity completely when pre-incubated at 100 °C for 10 min (Fig. 3b).

Fig. 3c illustrates the activity assays carried out at different pH values. As can be seen, the optimal activity was found at pH 7.0, with 71% and 82% of the maximal activity at pH 6.0 and 8.0, respectively.

### 3.2. Structural modeling

The search for a template for *P. furiosus* enzyme using THREADER against all protein structures deposited in the PDB indicated the prolyl oligopeptidase from muscles of *Sus scrofa*

(PDB code 1QFM), involved in the hydrolysis of small peptides (up to 30 amino acids) [55], as the best template. This enzyme has 710 amino-acid residues and two characteristic domains (Fig. 4a): the  $\beta$ -propeller (residues 73–427) and the catalytic domain (residues 1–72 and 428–710). The  $\beta$ -propeller domain (Fig. 4b) in *Sus scrofa* prolyl oligopeptidase has no structural similarity with any part of the esterase from *P. furiosus*. Nevertheless, the catalytic domain (Fig. 4c) presents the characteristic  $\alpha/\beta$  hydrolase fold and the catalytic triad (Ser554-Asp641-His680), which is located in a large cavity between the two enzyme domains. The residues 428–710 of this domain are similar to the sequence of the enzyme from *P. furiosus* and they were used to build the alignments for structural modeling.

The first exploratory alignments indicated the residues Ser149, Asp233 and His264 as the probable catalytic triad, but in models obtained with this alignment the catalytic triad was not well structured. For this reason several maneuvers to improve the structural model obtained from the alignment shown in Fig. 5 were carried out. The distances O $\gamma$ -Ser149/N $\epsilon$ -His264 and N $\delta$ -His264/O $\delta$ -Asp233 were improved by adjusting Ser149, Asp233 and His264 rotamers and the His264 loop (by fixing Asn254 and Pro272 positions) and by the optimization procedure described in a previous section. In the optimized model, distances were 3.42 Å for O $\gamma$ -Ser149/N $\epsilon$ -His264 and 2.60 Å for N $\delta$ -His264/O $\delta$ -Asp233, which are compatible with catalytic triads for esterases, lipases and proteases (Fig. 6).

The quality of the model was analyzed as described in a previous section and the results of stereochemistry and geometry were considered good by PROCHECK analyses [48], with 98.6% of the residues in the allowed regions of the Ramachandran plot [56], remarkably those used for structural analysis. It is interesting to note that Ser149, probably the catalytic serine in the triad, is positioned in a less favorable region, but the same occurs in the template 1QFM and in other  $\alpha/\beta$  hydrolases. This conformation is the result of a rigid loop formed by the pentapeptide (Gly-X-Ser-X-Gly) in a  $\beta$ -sheet-loop- $\alpha$ -helix motif, forcing the serine residue to adopt a conformation with unfavorable main-chain angles ( $\phi$  and  $\psi$  [57]). Analysis using WHAT-CHECK, PROVE and ERRAT [49–51] also indicated that the model is satisfactory.

## 4. Discussion

The protein encoded by ORF *PF2001* from *P. furiosus* presents a conserved domain characteristic of proteases, lipases and esterases. Since these enzymes are very important for industrial applications [18–22], ORF *PF2001* has the potential for success in biotechnological applications, which is improved when the thermoactivity and thermostability characteristic [7,26–34] of enzymes from *P. furiosus* are considered. In order to assess this hypothesis the ORF *PF2001* without the 60 bp of the 5'-terminus (*PF2001* $\Delta$ 60) was cloned and expressed in *E. coli* BL21(DE3) pLysS, and the recombinant enzyme was characterized with respect esterase, lipase and protease activities.

When the recombinant enzyme was assayed using 4-methylumbelliferyl acetate, heptanoate and palmitate, it exhibited its highest activity towards 4-methylumbelliferyl



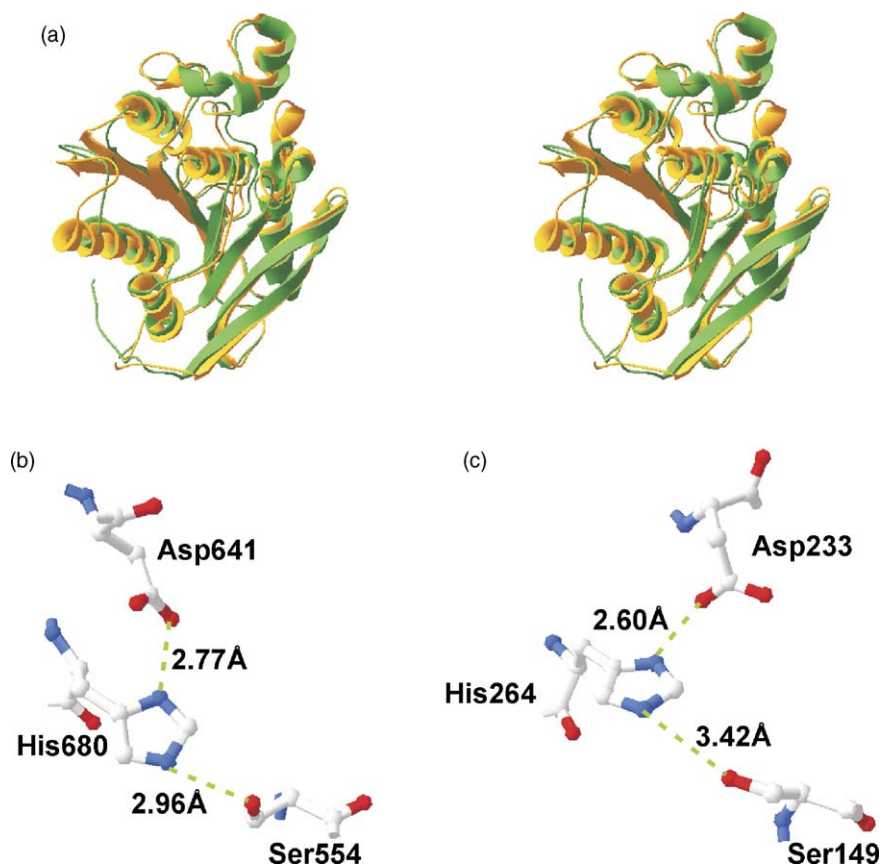


Fig. 6. Superposition of modified prolyl oligopeptidase (dark gray) and *P. furiosus* enzyme model (light gray) (a); and their respective active sites (b and c). Colored by atoms, carbon (light gray), nitrogen (gray) and oxygen (dark gray). The lines and numbers represents the distances between O $\gamma$ -Ser149/N $\epsilon$ -His264 and N $\delta$ -His264/O $\delta$ -Asp233 in angstroms. The picture was drawn with SPDBViewer v3.7(SP5).

respectively. In addition, they are active toward peptide substrates, and the *A. pernix* enzyme also has esterase activity [59]. The *P. furiosus* esterase described in this work presents similar folding of the catalytic domain as the enzymes described above, although it does not have the  $\beta$ -propeller domain. It has been suggested [55] that the  $\beta$ -propeller domain regulates the enzyme activity via substrate selection by size exclusion, allowing only the small peptides access to the active site and protecting proteins from proteolysis in the cytosol. Therefore, the *P. furiosus* esterase should hydrolyze proteins, but no protease activity was detected for *P. furiosus* esterase. We speculate that the absence of a  $\beta$ -propeller does not directly imply that an enzyme can hydrolyze peptides with larger molecular weights. In this regard, we can speculate that another probable function for the  $\beta$ -propeller is to favor a conformation that confers on the enzyme the capacity to hydrolyze peptide linkages.

#### Acknowledgements

CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) and FAPERJ (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro) supported this work. We thank Prof. Maysa Clementino (Instituto Nacional de Controle da Qualidade em Saúde, FIOCRUZ) for donation of *P. furiosus* genomic DNA, Núcleo de Estudos

de Genoma Johanna Döbereiner (Universidade Federal do Rio de Janeiro), Dr. Rodolfo M. Albano (Universidade do Estado do Rio de Janeiro) for DNA sequencing and Dr. Martha Sorenson (Universidade Federal do Rio de Janeiro) for discussion and comments on the manuscript.

#### References

- [1] Schmidt-Dannert C, Sztajer H, Stöcklein W, Menge U, Schmid RD. Screening, purification and properties of a thermophilic lipase from *Bacillus thermocatenulatus*. *Biochim Biophys Acta* 1994;1214:43–53.
- [2] Cornec L, Robineau J, Rolland JL, Dietrich J, Barbier G. Thermostable esterases screened on hyperthermophilic archaeal and bacterial strains isolated from deep-sea hydrothermal vents: characterization of esterase activity of hyperthermophilic archaeum, *Pyrococcus abyssi*. *J Mar Biotechnol* 1998;6:104–10.
- [3] Kademi A, Ait-Abdelkader N, Fakhreddine L, Baratti JC. A thermostable esterase activity from newly isolated moderate thermophilic bacterial strains. *Enzyme Microb Technol* 1999;24:332–8.
- [4] Martín S, Márquez MC, Sánchez-Porro C, Mellado E, Arahall DR, Ventosa A. *Marinobacter lipolyticus* sp. nov., a novel moderate halophile with lipolytic activity. *Int J Syst Evol Microbiol* 2003;53:1383–7.
- [5] Sánchez-Porro C, Mellado E, Bertoldo C, Antranikian G, Ventosa A. Screening and characterization of the protease CP1 produced by the moderately halophilic bacterium *Pseudoalteromonas* sp. strain CP76. *Extremophiles* 2003;7:221–8.
- [6] Sánchez-Porro C, Martín S, Mellado E, Ventosa A. Diversity of moderately halophilic bacteria producing extracellular hydrolytic enzymes. *J Appl Microbiol* 2003;94:295–300.

- [7] Ikeda M, Clark D. Molecular cloning of extremely thermostable esterase gene from hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli*. *Biotechnol Bioeng* 1998;57:624–9.
- [8] Morana A, Di Prizito N, Aurilia V, Rossi M, Cannio R. A carboxylesterase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: cloning of the gene, characterization of the protein. *Gene* 2002;283:107–15.
- [9] Tripathi MK, Roy U, Jinwal UK, Jain SK, Roy PK. Cloning, sequencing and structural features of a novel *Streptococcus* lipase. *Enzyme Microb Technol* 2004;34:437–45.
- [10] Valenzuela JG, Francischetti IMB, Pham VM, Garfield MK, Ribeiro JMC. Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. *Insect Biochem Mol Biol* 2003;33:717–32.
- [11] Suzuki T, Srivastava AS, Kurokawa T. cDNA cloning and phylogenetic analysis of pancreatic serine proteases from Japanese flounder, *Paralichthys olivaceus*. *Comp Biochem Physiol B: Biochem Mol Biol* 2002;131:63–70.
- [12] Arnold FH. Directed evolution: creating biocatalysts for the future. *Chem Eng Sci* 1996;51:5091–102.
- [13] Rubingh DN. Protein engineering from a bioindustrial point of view. *Curr Opin Biotechnol* 1997;8:417–22.
- [14] Svendsen A. Lipase protein engineering. *Biochim Biophys Acta* 2000;1543:223–38.
- [15] Bornscheuer UT, Pohl M. Improved biocatalysts by directed evolution and rational protein design. *Curr Opin Chem Biol* 2001;5:137–43.
- [16] Marrs B, Delagrave S, Murphy D. Novel approaches for discovering industrial enzymes. *Curr Opin Microbiol* 1999;2:241–5.
- [17] Sehgal AC, Kelly RM. Strategic selection of hyperthermophilic esterases for resolution of 2-arylpropionic esters. *Biotechnol Prog* 2003;19:1410–6.
- [18] Rao MB, Tanksale AM, Ghatge MS, Deshpande VV. Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* 1998;62:597–635.
- [19] Sharma R, Chisti Y, Banerjee UC. Production, purification, characterization and applications of lipases. *Biotechnol Adv* 2001;19:627–62.
- [20] Jaeger KE, Eggert T. Lipases for biotechnology. *Curr Opin Biotechnol* 2002;13:390–7.
- [21] Reetz MT. Lipases as practical biocatalysts. *Curr Opin Chem Biol* 2002;6:145–50.
- [22] Maurer K. Detergent proteases. *Curr Opin Biotechnol* 2004;15:330–4.
- [23] Eichler J. Biotechnological uses of archaeal extremozymes. *Biotechnol Adv* 2001;19:261–78.
- [24] Fiala G, Stetter KO. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100 °C. *Arch Microbiol* 1986;145:56–61.
- [25] Maeder DL, Weiss RB, Dunn DM, Cherry JL, González JM, DiRuggiero J, et al. Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* 1999;152:1299–305.
- [26] Lundberg KS, Shoemaker DD, Adams MWW, Short JM, Sorge JA, Mathur EJ. High fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* 1991;108:1–6.
- [27] Koch R, Zabłowski P, Spreinat A, Antranikian G. Extremely thermostable amylolytic enzyme from archaeobacterium *Pyrococcus furiosus*. *FEMS Microbiol Lett* 1990;71:21–6.
- [28] Laderman KA, Asada K, Uemori T, Mukai H, Taguchi Y, Kato I, et al.  $\alpha$ -amylase from *Pyrococcus furiosus*: cloning and sequencing of the gene and expression in *Escherichia coli*. *J Biol Chem* 1993;268:24402–7.
- [29] Jorgensen S, Vorgias EC, Antranikian G. Cloning, sequencing, characterization, and expression of an extracellular  $\alpha$ -amylase from hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis*. *J Biol Chem* 1997;272:16335–42.
- [30] Voorhost WGB, Eggen RIL, Luesink EJ, De Vos WM. Characterization of celB gene coding for  $\beta$ -glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site directed mutation in *Escherichia coli*. *J Bacteriol* 1995;177:7105–11.
- [31] de Roode BM, Van Der Meer TD, Kaper T, Franssen MCR, Van Der Padt A, Van Der Oost J, et al. The catalytic potency of  $\beta$ -glucosidase from *Pyrococcus furiosus* in the direct glucosylation reaction. *Enzyme Microb Technol* 2001;29:621–4.
- [32] Tsunawasa S, Nakura S, Tanigawa T, Kato I. Pyrrolidone carboxyl peptidase from hyperthermophilic archaeon *Pyrococcus furiosus*: cloning and overexpression in *Escherichia coli* of the gene, and its application to protein sequence analysis. *J Biochem* 1998;124:778–83.
- [33] Blumentals II, Robinson AS, Kely RM. Characterization of sodium dodecyl sulfate resistant proteolytic activity in the hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *Appl Environ Microbiol* 1990;56:1992–8.
- [34] Eggen R, Geerling A, Watts J, de Vos WM. Characterization of pyrolysin, a hyperthermoactive serine protease from the archaeobacterium *Pyrococcus furiosus*. *FEMS Microbiol Lett* 1990;71:17–20.
- [35] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.
- [36] Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, et al. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* 2003;31:383–7.
- [37] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–95.
- [38] Prim N, Sánchez M, Ruiz C, Pastor FIJ, Díaz P. Use of methylumbelliferyl-derivative substrates for lipase activity characterization. *J Mol Catal B: Enzym* 2003;22:339–46.
- [39] Kunitz M. Crystalline soybean trypsin inhibitor. II. General properties. *J Gen Physiol* 1947;30:291–310.
- [40] Bradford MM. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of the protein dye binding. *Anal Biochem* 1976;72:248–54.
- [41] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* 2000;28:235–42.
- [42] Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* 1999;287:797–815.
- [43] Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205–17.
- [44] Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
- [45] Guex N, Peitsch MC. Swiss-Model and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;18:2714–23.
- [46] Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom force field for simulations of proteins and nucleic acids. *J Comp Chem* 1986;7:230–52.
- [47] Froimowitz M. HyperChem: a software package for computational chemistry and molecular modeling. *Biotechniques* 1993;14:1010–3.
- [48] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. Procheck: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26:283–91.
- [49] Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature* 1996;381:272.
- [50] Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol* 1996;264:121–36.
- [51] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 1993;2:1511–9.
- [52] Zwickl P, Fabry S, Bogedain C, Haas A, Hensel R. Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeobacterium *Pyrococcus woesei*: characterization of the enzyme, cloning and sequencing of the gene, and expression in *Escherichia coli*. *J Bacteriol* 1990;172:4329–38.
- [53] Lecompte O, Ripp R, Puzos-Barbe V, Duprat S, Heilig R, Dietrich J, et al. Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. *Genome Res* 2001;11:981–93.
- [54] Jaeger KE, Dijkstra BW, Reetz MT. Bacterial biocatalysts: molecular biology, three-dimensional structures and biotechnological applications of lipases. *Annu Rev Microbiol* 1999;53:315–51.



- [55] Fülöp V, Bocskei S, Polgár L. Prolyl oligopeptidase: an unusual  $\beta$ -propeller domain regulates proteolysis. *Cell* 1998;94:161–70.
- [56] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain conformations. *J Mol Biol* 1963;7:95–9.
- [57] Schrag JD, Cygler M. Lipases and  $\alpha/\beta$  hydrolase fold. *Meth Enzymol* 1997;284:85–107.
- [58] Manco G, Giosuè E, D'Auria S, Herman P, Carrea G, Rossi M. Cloning, overexpression, and properties of a new thermophilic and thermostable esterase with sequence similarity to hormone-sensitive lipase subfamily from the archaeon *Archaeoglobus fulgidus*. *Arch Biochem Biophys* 2000;373:182–92.
- [59] Gao R, Feng Y, Ishikawa K, Ishida H, Ando S, Kosugi Y, et al. Cloning, purification and properties of a hyperthermophilic esterase from archaeon *Aeropyrum pernix* K1. *J Mol Catal B: Enzym* 2003;24–25:1–8.
- [60] Kim S, Lee SB. Thermostable esterase from a thermoacidophilic archaeon: purification and characterization for enzymatic resolution of a chiral compound. *Biosci Biotech Biochem* 2004;68:2289–98.
- [61] Wood ANP, Fernandez-Lafuente R, Cowan Da. Purification and partial characterization of a novel thermophilic carboxylesterase with high mesophilic specific activity. *Enzyme Microb Technol* 1995;17: 816–25.
- [62] Manco G, Adinolfi E, Pisani FM, Ottolina G, Carrea G, Rossi M. Overexpression and properties of a new thermophilic and thermostable esterase from *Bacillus acidocaldarius* with sequence similarity to hormone-sensitive lipase subfamily. *Biochem J* 1998;15(332):203–12.
- [63] Zhang J, Liu J, Zhou J, Ren Y, Dai X, Xiang H. Thermostable esterase from *Thermoanaerobacter tengcongensis*: high-level expression, purification and characterization. *Biotechnol Lett* 2003;25:1463–7.
- [64] Ateslier ZBB, Metin, K. Production and partial characterization of a novel thermostable esterase from a thermophilic *Bacillus* sp., *Enzyme Microb. Technol.* 2006;38:628–35.
- [65] Fojan P, Jonson PH, Petersen MTN, Petersen SB. What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie* 2000;82:1033–41.
- [66] Das AK, Bellizzi III JJ, Tandel S, Biehl E, Clardy J, Hofmann SL. Structural basis for insensitivity of a serine enzyme (palmitoyl-protein thioesterase) to phenylmethylsulfonyl fluoride. *J Biol Chem* 2000;275:23847–51.
- [67] Polgár L. Structural relationship between lipases and peptidases of the prolyl oligopeptidase family. *FEBS Lett* 1992;311:281–4.
- [68] Bartlam M, Wang G, Yang H, Gao R, Zhao X, Xie G, et al. Crystal structure of an acylpeptide hydrolase/esterase from *Aeropyrum pernix* K1. *Structure* 2004;12:1481–8.
- [69] Harris MN, Madura JD, Ming L, Harwood VJ. Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile *Pyrococcus furiosus*. *J Biol Chem* 2001;276:19310–7.

### **PARTE 3: ARCHAEAS E BIOTECNOLOGIA**

Em 1990, Woese, Kandler e Wheelies (WOESE et al., 1990) propuseram a substituição da visão bipartida da vida por um esquema tripartido, baseado em três subreinos ou domínios: Bacteria, Archaea e Eucarya; com base em uma revisão do aumento nos conhecimentos de bioquímica, da genética e de evidências filogenéticas, que quando vistas coletivamente, sugeriram que as Archaeas eram merecedoras de uma situação taxonômica igual àquela de eucariotos e eubacterias. Os organismos procarióticos estão atualmente distribuídos pelos domínios, Archaea e Bacteria.

O Domínio Archaea é caracterizado por microrganismos procarióticos evolutivamente distintos dos organismos alocados no Domínio Bacteria (ADAMS, 1995). Uma grande variedade de Archaea possui metabolismo anaeróbico obrigatório, enquanto que outras espécies são encontradas em ambientes extremos, tais como fontes geotermiais, habitats com elevada salinidade, solos e sistemas aquáticos altamente ácidos ou alcalinos. Pode-se dizer que certas espécies de Archaea definem claramente os limites de tolerância biológica nos extremos físicos e químicos da vida na Terra.

A adaptação dos organismos a esses ambientes obrigou-os a desenvolver componentes celulares e estratégias bioquímicas para a sobrevivência da espécie. Devido às características “excêntricas”, bem como às suas propriedades únicas, esses microrganismos geram bioprodutos que podem ser empregados em condições drásticas, que freqüentemente ocorrem em processos industriais. Os componentes moleculares deles retirados possuem muitas vezes propriedades que os tornam especialmente adequados para utilização nesses processos. Neste contexto, é hoje geralmente aceito que estes microrganismos constituem um precioso repositório de moléculas de interesse industrial e um excelente recurso para o desenvolvimento de novas aplicações biotecnológicas.

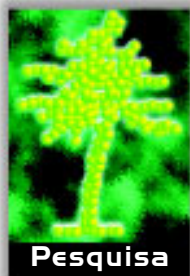
No entanto, o Domínio Archaea é pouco conhecido e explorado em todos os ramos da Microbiologia. No Brasil, algumas iniciativas atuais ainda são incipientes e necessitam de apoio de agências de fomento e desenvolvimento científico para a sua consolidação, apesar de reconhecida

internacionalmente como prioritária para desenvolvimento, particularmente em ecologia microbiana e bioprospecção para biotecnologia, como também não se conhecem registros de isolados do Domínio Archaea caracterizados a nível de espécie de linhagens em coleções de culturas brasileiras (CANHOS & VAZOLLER, 1999).

**Manuscrito 5:** "*Archaea: Potencial Biotecnológico – Utilização e Aplicação de Arqueas na Biotecnologia*".

Com a minha transferência física para o Laboratório de Biologia Molecular, para o desenvolvimento do meu projeto de doutoramento, acabei me envolvendo naturalmente em outros projetos também ali desenvolvidos.

Em 2003, em conjunto com o grupo coordenado pelo Prof. Orlando que reúne especialistas em diferentes áreas da biologia de Archaeas (Diversidade, Biossíntese de Proteínas e Enzimas Hipertermofílicas) escrevi o artigo de revisão sob o título: "*Archaea: Potencial Biotecnológico – Utilização e Aplicação de Arqueas na Biotecnologia*" para a revista Biotecnologia, Ciência e Desenvolvimento. Este artigo, em língua portuguesa, reúne um resumo dos principais avanços da biologia, bioquímica e da tecnologia, com o descobrimento das Archaeas como o terceiro domínio da vida, principalmente com relação a filogenia e fisiologia, os ambientes extremos, suas utilizações e potencialidades em biotecnologia e sua genômica.



# Archaea: Potencial Biotecnológico

Utilização e aplicação de arqueas na biotecnologia

## Alexander Machado Cardoso

Departamento de Bioquímica Médica,  
Instituto de Ciências Biomédicas, UFRJ.  
amcardoso@bioqmed.ufrj.br

## Maysa B. Mandetta Clementino

Departamento de Microbiologia, Instituto Nacional  
de Controle de Qualidade em Saúde, FIOCRUZ.  
maysa@incqs.fiocruz.br

## Orlando Bonifácio Martins

Departamento de Bioquímica Médica,  
Instituto de Ciências Biomédicas, UFRJ.  
omartins@bioqmed.ufrj.br

## Ricardo Pilz Vieira

Departamento de Biologia Marinha,  
Instituto de Biologia, UFRJ.  
pilz@bioqmed.ufrj.br

## Rodrigo Volcan Almeida

Programa de Engenharia Química,  
COPPE, UFRJ.  
volcan@peq.coppe.ufrj.br

## Sylvia M. Campbell Alqueres

Departamento de Bioquímica Médica, Instituto de  
Ciências Biomédicas, UFRJ.  
alqueres@bioqmed.ufrj.br

## Welington Inácio de Almeida

Departamento de Bioquímica Médica,  
Instituto de Ciências Biomédicas, UFRJ.  
welington@bioqmed.ufrj.br

Ilustrações cedidas pelos autores

## 1. Introdução

O domínio *Archaea* é formado principalmente por organismos extremofílicos, isto é, microrganismos que não apenas toleram, mas crescem otimamente em ambientes normalmente considerados inóspitos para a vida, como fontes termais, águas extremamente salgadas, temperaturas baixas e condições extremas de pH. Pode-se dizer que certas espécies de arqueas definem claramente os limites de tolerância biológica nos extremos físicos e químicos da vida. O estudo dos microrganismos provenientes desses ambientes extremos pode nos fornecer informações valiosas acerca da origem da vida na Terra, bem como das estratégias adaptativas aos ambientes onde esta prosperou (Woese, 1998).

A adaptação de organismos a esses ambientes obrigou-os a desenvolver componentes celulares e estratégias bioquímicas para sua sobrevivência. Por outro lado, devido às características "exóticas" que têm, e às suas propriedades únicas, esses microrganismos geram bioprodutos que podem ser empregados em condições drásticas, que freqüentemente ocorrem em processos industriais. Os componentes moleculares deles retirados possuem muitas vezes propriedades que os tornam especialmente adequados para serem utilizados nesses processos. Nesse contexto, é hoje geralmente aceito que esses microrganismos constituem um precioso repositório de moléculas de interesse industrial e um excelente recurso para o desenvolvimento de novas aplicações biotecnológicas.

202

Os benefícios científicos esperados de um conhecimento maior da biologia das arqueas incluem, entre outros, a compreensão das funções exercidas por esses organismos nos ambientes aquáticos e terrestres, bem como suas interações com outros componentes da biodiversidade. Os benefícios econômicos e estratégicos estão relacionados com a descoberta de microrganismos potencialmente exploráveis nos processos biotecnológicos para obtenção de agentes terapêuticos, probióticos, produtos químicos, enzimas e polímeros para aplicações industriais e tecnológicas, biorremediação e biolixiviação de poluentes e recuperação de minérios. Outros benefícios incluem a otimização da capacidade microbiana para processamento de alimentos, tratamento e/ou remediação de resíduos (esgoto doméstico e lixo). Embora ainda não sejam totalmente conhecidas as estratégias moleculares para sua sobrevivência em tais ambientes inóspitos, sabe-se que esses organismos possuem enzimas adaptadas a tais ambientes, e isso desperta o interesse tanto acadêmico quanto industrial.

## 2. Filogenia e Fisiologia

Há cerca de vinte anos, Carl Woese e colaboradores sugeriram que os organismos vivos fossem classificados em três grupos principais: *Archaea*, *Bacteria* e *Eukarya*, com base no estudo das seqüências das moléculas do gene 16S do RNA ribossomal (16S rRNA). Esses grupos são chamados de domínios e acredita-se que surgiram através de

**Tabela 1: Diferenças fundamentais entre os três domínios.**

Características	Bactéria	Archaea	Eucarya
Membrana nuclear	Ausente	Ausente	Presente
Número de cromossomos	1	1	>1
Parede Celular	Peptideoglicano	Pseudo-peptideoglicano, glicoproteínas e outros	Celulose em plantas, quitina em fungos e nenhuma em animais
Mureína na parede celular	Sim	Não	Não
Lipídeos da membrana celular	Glicerídeos ligados a éster, não ramificado; saturado ou mono-insaturado	Isoprenóide; glicerol diéter ou di-glicerol tetraéter	Glicerídeos ligados a éster; não ramificado; poli-insaturado
Organelas (mitocôndria e cloroplastos)	Ausente	Ausente	Presente
Ribossomo	70S	70S	80S
Síntese de proteínas inibida por cloranfenicol e estreptomicina	Sim	Não	Não
Síntese de proteínas inibida pela toxina da difteria	Não	Sim	Sim

vias evolutivas distintas a partir de um ancestral comum. A noção de dicotomia da vida entre eucariontes e procariontes, que ainda domina a biologia e influencia, em particular, a percepção sobre o domínio *Archaea*, está sendo lentamente revista por grupos atuantes em microbiologia. A diversidade e a biologia das arqueas representam uma enorme contribuição à compreensão da Ecologia Microbiana (Woese *et al.*, 1990).

O domínio *Archaea* consiste de três divisões: *Crenarchaeota*, que contém as arqueas hipertermofílicas redutoras de enxofre; *Euryarchaeota*, que compreende uma grande diversidade de organismos, incluídas as espécies metanogênicas, as halofílicas extremas e algumas espécies hipertermofílicas; e *Korarchaeota*, uma divisão descrita mais recentemente, que engloba organismos hipertermofílicos pouco conhecidos, identificados a partir de seqüências do gene 16S do rRNA isolados de fontes termais terrestres, porém ainda não cultivados em laboratório.

Após serem divididos os três grandes domínios a partir do seqüenciamento do 16S rRNA, estudos subsequentes mostraram que cada domínio está associado a uma série de fenótipos. Alguns desses fenótipos

são únicos de cada domínio, enquanto outros são compartilhados entre dois ou até entre todos os três domínios, como pode ser observado na Tabela 1.

### 3. Ambientes extremos

Os primeiros organismos identificados pertencentes ao domínio *Archaea* viviam em ambientes extremos de temperatura, salinidade ou acidez, sugerindo que a preferência por tais habitats, era um traço característico do grupo. Estudos mais recentes mostraram várias eubactérias e organismos eucarióticos que sobrevivem também em ambientes extremos, como observaram igualmente a presença de arqueas em ambientes mais amenos, demonstrando, dessa forma, a contribuição desse grupo na biomassa global (Forterre, 1997). Entretanto, as arqueas parecem ser os únicos organismos descobertos até o presente momento que podem sobreviver a temperaturas acima de 95°C, e o fenótipo hipertermofílico só é encontrado nesse domínio da vida. Uma outra característica exclusiva de *Archaea* é o metabolismo metanogênico: não se conhecem eubactérias nem eucariotos capazes de produzir metano como resíduo de seu metabolismo.

O habitat das arqueas halofílicas extremas é hipersalino e as espécies em cultivo laboratorial requerem para o crescimento, entre 1,5 a 4 M de NaCl, o que significa um ambiente com cerca de 10 vezes a salinidade encontrada na água do mar. As metanogênicas são organismos obrigatoriamente anaeróbios e liberam gás metano (CH<sub>4</sub>) como resíduo metabólico. São encontradas em ambientes com ausência de oxigênio e abundância de matéria orgânica, como pântanos, açudes, lagos, sedimentos marinhos e rúmen de bovinos. Elas retiram hidrogênio e gás carbônico desses ambientes e os utilizam em seu metabolismo. Vivem como simbioses de uma grande variedade de protozoários também anaeróbios, convertendo produtos finais de fermentação em gás metano. São de grande importância ao ambiente em que vivem pela alta eficiência de sua enzima hidrogenase que, mantendo uma baixa pressão parcial de H<sub>2</sub> para que a metanogênese ocorra, permite que os demais organismos fermentadores façam reoxidação do NADH, o que corresponde a um maior rendimento de ATP e um aumento da biomassa (Brock *et al.*, 1994).

As arqueas termoacidófilas compõem um grupo heterogêneo, defini-

**Tabela 2: Estratégias de adaptação das arqueas aos ambientes extremos.**

Ambiente	Problemas	Mecanismos de adaptação
<b>Hipersalino</b>	Lise celular por diferença de pressão osmótica	Mantém altas concentrações de K <sup>+</sup> intracelularmente A parede celular é composta por glicoproteínas que tem uma maior porcentagem de aminoácidos ácidos que atraem os íons Na <sup>+</sup> para o redor da célula. A ligação dos íons com a parede celular estabiliza esta, impedindo a lise
	Desnaturação das proteínas	Essas proteínas expõem aminoácidos de carga negativa, de maneira que, quando os íons positivos entram em contato com a proteína, esta é estabilizada
<b>Altas temperaturas</b>	Lise celular	Maior rigidez da parede celular
	Desnaturação do DNA	Maior quantidade de histonas Maior empacotamento do DNA
	Desnaturação das proteínas	Aumento do número de pontes de hidrogênio Aumento das interações hidrofóbicas Menor porcentagem de aminoácidos termolábeis
<b>Baixas temperaturas</b>	Inativação das proteínas	Menor quantidade de pontes de hidrogênio, dissulfeto e interações hidrofóbicas, o que aumenta a flexibilidade da proteína

do pela capacidade dos organismos de crescerem em altas temperaturas que vão de 55°C a 85°C, com pH que varia de 1,0 até 6,0. O gênero *Sulfolobus* apresenta parede celular composta principalmente por lipoproteína e carboidratos; oxidam H<sub>2</sub>S, mas o princípio para essa oxidação ainda não foi esclarecido. Os principais substratos de crescimento parecem ser fontes quentes e também solos quentes que contenham enxofre, e que, então, oxidam tal elemento em ácido sulfúrico, responsável pela acidez desses habitats. O grupo dos termoplasmas é caracterizado pela ausência de parede celular e é encontrado em minas de carvão, com quantidades substanciais de sulfeto ferroso.

Uma das perguntas que mais intrigam os pesquisadores é como esses incríveis organismos conseguem viver em tais ambientes? Deixam constantemente uma interrogação sobre qual será o limite para o desenvolvimento da vida. Embora muitas dúvidas ainda parem sobre os mecanismos de adaptação a tais ambientes hostis, muitos fatores já são apontados como os responsáveis pela resistência desses organismos. A Tabela 2 reúne os principais mecanismos de adaptação aos problemas causados pelos ambientes extremos.



Meios de Cultivo para Arqueas Termofílicas

A pesquisa envolvendo esses organismos tem sido intensificada nas duas últimas décadas por duas razões principais: pelo conhecimento das condições sob as quais a vida pode existir através do estudo de muitos habitats nunca antes explorados e pelo atual reconhecimento do potencial biotecnológico desses organismos, bem como de seus produtos.

#### 4. Aplicações Biotecnológicas

A tecnologia enzimática experimentou um grande avanço quando as enzimas microbianas passaram a ser utilizadas, principalmente por causa da grande variedade de reações que essas enzimas são capazes de catalisar. Com o descobrimento dos microrganismos extremofílicos (em sua maioria arqueas), o que ampliou ainda mais a

diversidade microbiana, a potencial faixa de processos para utilização de enzimas também se ampliou, principalmente porque as extremozimas (enzimas provenientes de microrganismos extremófilos) sendo naturalmente estáveis em ambientes extremos, vieram suprir a demanda industrial, para a qual, de certa forma, sempre estiveram em desvantagem as enzimas tradicionais. Um problema inerente às extremozimas é a dificuldade de produzi-las utilizando microrganismos selvagens, já que estes, em geral, necessitam de condições especiais para se reproduzirem, como ambientes anaeróbios estritos, altas temperaturas, meios definidos, etc. Contudo esse problema pode ser contornado expressando essas enzimas em outros organismos de mais fácil manipulação, já que existem vários exemplos em que essas enzimas, quando expressas em microrganismos heterólogos, mantêm suas características originais (Eichler, 2001).

Dentre as enzimas de arqueas de grande potencial para a aplicação biotecnológica, destacam-se as hipertermofílicas, psicrófilas, alcalofílicas, halofílicas e barofílicas.

Entre as enzimas de arqueas que têm recebido maior atenção, estão as termozimas, sendo que os principais processos de potencial utilização

dessas enzimas são o beneficiamento do amido, a manufatura e o branqueamento da polpa para produção de papel e a bem estabelecida prática laboratorial da reação em cadeia da polimerase (PCR), entre outras.

#### 4.1. Processamento do Amido

O amido é um dos polímeros mais abundantes na natureza, estando presente principalmente nos vegetais, onde ele é utilizado para o armazenamento de energia na forma de grânulos insolúveis, compostos basicamente de amilose e amilopectina, que são diferentes polímeros de glicose. Além de servir diretamente como alimento na dieta animal, o amido presente nos vegetais pode ser hidrolisado, gerando glicose, maltose e xarope de oligossacarídeos, que, por sua vez, são utilizados para produção de outros químicos ou como substratos em fermentações (Bentley e Willians, 1996; Vieille e Zeikus, 2001).

De uma forma geral, o processo de hidrólise do amido envolve a liquefação e a sacarificação, as quais ocorrem em altas temperaturas. Durante a liquefação, os grãos de amido são gelatinizados em soluções aquosas entre 105°C e 110°C, num pH entre 5,8 a 6,5, quando então, são utilizadas  $\alpha$ -amilases termoestáveis a 95°C, que hidrolisam parcialmente as ligações  $\alpha$ -1,4. Nesse processo, o controle de temperatura e pH é muito importante pois, se a temperatura estiver abaixo de 105°C, a gelatinização ocorre parcialmente, e se aumentar muito, há a inativação das  $\alpha$ -amilases; se o pH estiver mais ácido que 5,5 também ocorre a inativação dessas enzimas, mas, se vai acima de 6,5 são gerados muitos subprodutos. Após a liquefação, o produto é con-

vertido a sacarídeos de baixo peso molecular e, posteriormente, em glicose, utilizando-se pululanase e glicoamilase, e em maltose, utilizando-se pululanase e  $\beta$ -amilase.

Nesse processo, potencialmente poderiam ser feitas duas melhorias com a utilização de extremozimas: a primeira seria a utilização de  $\alpha$ -amilases hipertermofílicas com maior tempo de resistência, de forma que não se necessitasse gastar energia com o resfriamento de 105°C para 95°C, da gelatinização para a liquefação. A segunda seria a em que o pH natural de soluções de amidos gelatinizados, que é de, aproximadamente, 4,5 e que força o ajuste para 5,8 na liquefação, e, posteriormente, redução para 4,2; na sacarificação, de maneira que uma  $\alpha$ -amilase apta a trabalhar em pHs mais baixos reduziria, em muito, os custos do processo.

A Tabela 3 mostra algumas enzimas isoladas de arqueas com grande potencial de aplicação na indústria de processamento de amido. Embora nenhuma dessas reúna todas as características necessárias apontadas acima, elas indicam, no entanto, que com técnicas de engenharia de proteínas, esses catalisadores poderiam ser aperfeiçoados.

#### 4.2. Manufatura do Papel

O processo de manufatura do papel envolve, de uma maneira geral, dois estágios: o polpeamento e o branqueamento. O polpeamento é o estágio no qual a estrutura macroscópica da fibra da madeira é quebrada, removendo-se dela a lignina, gerando daí uma fibra mais maleável e com características próprias para a produção do papel (Tolan, 1996; Vieille e Zeikus, 2001). Esse processo é conduzido mecânica ou quimicamente por adição de

ácidos (processo Sulfite) ou bases (processo Kraft) em altas temperaturas. Os processos de polpeamento, por envolverem condições muito drásticas (160°C-190°C em concentrações elevadas de álcalis ou ácidos), são potencialmente próprios para utilização de enzimas extremofílicas. Embora as celulases sejam bem distribuídas entre os domínios *Eukarya* e *Bacteria*, somente uma celulase de *Archaea* é relatada, a endoglicanase de *Pyrococcus furiosus*, que é capaz de hidrolisar ligações  $\beta$ ,1-4 com uma atividade ótima ocorrendo em 100°C e pH 6,0 (Bauer e Kelly, 1998). Contudo, segundo Eichler (2001), várias arqueas termo e hipertermofílicas possuem enzimas  $\beta$ -glicosídicas.

A polpa resultante de um dos processos acima é levada ao branqueamento, que ocorrerá em menor ou maior grau dependendo da utilização do papel. Nesse processo, a quantidade remanescente de lignina do polpeamento é retirada com a utilização, dependendo do processo, de: cloro, dióxido de cloro, ozônio, peróxido de hidrogênio e altas temperaturas. A utilização desses agentes oxidantes acaba gerando uma quantidade muito grande de poluentes, e, por isso, vem sofrendo sanções das agências ambientais. Uma das alternativas pesquisadas para reduzir esses compostos, principalmente os derivados de cloro, é a utilização de enzimas. Entre as enzimas pesquisadas, as xilanasas, segundo Tolan (1996), são as que possuem uma boa aceitação industrial e têm sido utilizadas em indústrias onde está havendo um decréscimo na utilização dos oxidantes, da ordem de 10% a 15%. Ainda segundo esse autor, as características ótimas para uma xilanase seriam sua ação em temperaturas por volta de 70°C e pH entre 10-12, o que abre uma excelente

Tabela 3. Enzimas com potencial de aplicação na indústria de processamento do amido.

Enzima	Origem	Temperatura	pH	Referência
$\alpha$ -amilase	<i>Pyrococcus furiosus</i>	100°C	5,5-6,0	Dong <i>et al.</i> , 1997
	<i>Pyrococcus woesei</i>	100°C	5,5	Koch <i>et al.</i> , 1997
	<i>Pyrodictium abyssii</i>	100°C	5,0	Niehaus <i>et al.</i> , 1999
Pululanase	<i>Thermus caldophilus</i>	75°C	5,5	Kim <i>et al.</i> , 1996
Glicoamilase	<i>Thermoanaerobacterium thermosaccharolyticum</i>	50 a 60°C	4,0-5,5	Ganghofner <i>et al.</i> , 1998
Amilopululanases	<i>Pyrococcus furiosus</i>	105°C	6,0	Dong <i>et al.</i> , 1997

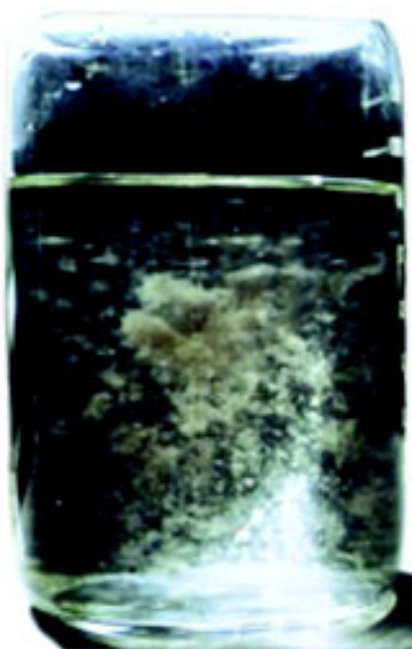
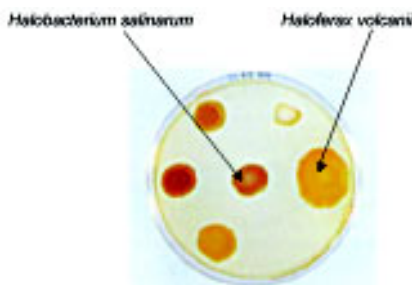


oportunidade para a exploração de xilanasas de arqueas extremofílicas como as de *Pyrococcus furiosus*, *Thermococcus* sp. e *Thermococcus zilligii* (Eichler, 2001).

Um outro problema na fabricação de papel, principalmente pelo polpeamento mecânico, é a deposição do “pitch”, denominação atribuída ao conjunto de substâncias hidrofóbicas da madeira, principalmente triglicerídeos e graxas, que causam muitos problemas na manufatura da polpa e do papel (Jaeger e Reetz, 1998). As lipases já vêm sendo utilizadas na remoção dessas substâncias. Gutierrez *et al.* (2001) comentam que muitos estudos estão sendo realizados, por meio de técnicas de engenharia de proteínas, no intuito de aumentar a amplitude de substratos hidrolisáveis, atividade e estabilidade em pH e temperatura elevados, sendo importante que a enzima esteja ativa a altas temperaturas, uma vez que para um desempenho ótimo, a lipase deveria ser adicionada à polpa a uma temperatura de, aproximadamente, 85°C. Nosso laboratório vem trabalhando no isolamento, clonagem, expressão e caracterização de uma enzima lipolítica de arquea cuja atividade já foi testada a 80°C (manuscrito em preparação).

#### 4.3. Utilização de Polimerases

A reação em cadeia da polimerase (PCR) revolucionou a prática da biologia molecular, ou seja, a surpreendente replicação *in vitro* de seqüências específicas de DNA possibilitou que o isolamento de genes, seu seqüenciamento e mutações específicas, antes uma prática laboriosa, se tornasse uma atividade cotidiana de qualquer laboratório de engenharia genética. O método da reação em cadeia da polimerase, por sua vez, foi extremamente simplificado com a utilização de enzimas hipertermofílicas. Embora a principal DNA polimerase termofílica utilizada seja a de uma bactéria (*Thermus aquaticus*), as DNA polimerases de arqueas, como as de *Pyrococcus furiosus*, *Pyrococcus woesei* e outras, apresentam uma vantagem sobre a *Taq* polimerase, pois elas possuem



*Pyrodictium abyssi*

uma maior capacidade de correção de pareamentos errôneos, diminuindo a frequência de erros na replicação *in vitro* (Lundberg *et al.*, 1991).

Não somente as enzimas adaptadas às altas temperaturas têm potencialidades biotecnológicas. As enzimas de organismos que crescem entre 5°C e 25°C (Psicrofílicos) podem ser empregadas em vários processos e produtos, como proteases, lipases, amilases, b-glicanases em detergentes, pectinases em sucos de frutas, b-galactosidases para produção de leite deslactosado, lipases para maturação de queijos (Herbert, 1992; Eichler, 2001).

#### 4.4. Arqueossomos

As aplicações biotecnológicas das arqueas não se restringem a produção, expressão heteróloga e purificação de extremozimas. Pelo menos uma outra potencialidade biotecnológica deve ser ressaltada. A utilização dos arqueossomos (preparação de lipídeos de membrana arqueana), como coadjuvantes em formulações

de vacinas, proporciona captação 3 a 50 vezes maior pelas células fagocíticas do sistema imune quando comparadas com formulações de lipossomos convencionais, além de geração de resposta imune prolongada (Tolson *et al.*, 1996; Sprott *et al.*, 1997; Krishnan *et al.*, 2000). Todos os estudos realizados *in vitro* e *in vivo* indicam que os arqueossomos são moléculas seguras e não invocam nenhuma toxicidade em ratos. Em geral, os arqueossomos demonstram alta estabilidade ao estresse oxidativo, à alta temperatura, ao pH alcalino, à ação de fosfolipases, de sais biliares e de proteínas do soro. Algumas formulações podem ser esterilizadas em autoclave, sem problemas como fusão ou agregação das vesículas (Patel e Sprott, 1999).

### 5. Genômica

Uma nova era sobre o conhecimento das arqueas começou em 1996 em decorrência do seqüenciamento completo do primeiro genoma de arquea (Bult *et al.*, 1996). Com o subsequente desenvolvimento de novos projetos de seqüenciamento de outros organismos pertencentes ao domínio *Archaea*, foi produzida uma rica amostragem de genomas desse grupo taxonomicamente bastante diversos. Esse repertório de genomas completamente seqüenciados inclui múltiplos representantes das duas maiores divisões de arqueas estabelecidas pela análise filogenética dos genes de rRNA: a *Euryarchaeotae Crenarchaeota*, bem como as principais variantes ecológicas de arquea: os hipertermofílicos, termofílicos moderados e mesofílicos, assim como halofílicos e metanogênicos, formas autotróficas e heterotróficas, e múltiplas espécies que representam organismos anaeróbios e aeróbios. Infelizmente os bancos de dados ainda não contam com seqüências genômicas de alguns organismos pertencentes a ramos de arquea potencialmente importantes, como as misteriosas *Korarchaeota*, que devem ter divergido das demais arqueas nos primórdios da evolução, e a igualmente intrigante nanoarchaea, que parece ter o menor genoma entre todos os organismos de vida livre já descritos (Huber *et al.*, 2002)



Métodos de Preservação de Arqueas

A comparação dos genomas já seqüenciados de arqueas e bactérias permite concluir que a diferença mais marcante entre os domínios *Archaea* e *Bacteria* está na organização de seus sistemas de processamento de informações. A estrutura dos ribossomos e da cromatina, a presença de histonas, assim como a similaridade entre seqüências das proteínas envolvidas na tradução, transcrição, replicação e reparo do DNA; todos esses pontos apontam para uma maior proximidade entre arqueas e eucariotos. Por outro lado, alguns componentes chaves da maquinaria de replicação de DNA não são homólogos nos eucariotos nem nas bactérias. Essa observação permite sugerir a hipótese de que a replicação da dupla fita de DNA como principal forma de replicação do material genético dos seres vivos pode ter, na sua evolução, surgido, independentemente, duas vezes: uma nas bactérias e uma outra no ancestral de arqueas e eucariotos. No entanto, muitas, mas não todas as rotas metabólicas de *Archaea*, são mais parecidas com as de bactérias que as de eucariotos. Esses estudos são concordes quanto ao posicionamento das arqueas como um domínio distinto da vida, com conexões específicas com os eucariotos, e enfatizam a natureza misteriosa e única dos genomas das arqueas (Gaasterland, 1999).

Quando analisamos os 18 genomas de arqueas seqüenciados totalmente até hoje, podemos concluir que 16 proteínas são exclusivas de arqueas, enquanto 61 são exclusivas de arquea-eucariotos. Interessantemente, desses 61 genes, apenas 2 não pertencem à maquinaria do processamento de informação. Portanto, a análise genômica das arqueas já

seqüenciadas corrobora a identificação destas como um grupo de organismos que têm uma base sólida e estável de genes, os quais, primariamente, codificam proteínas envolvidas na replicação e expressão do genoma. Além desses, existe um segundo grupo de genes que é compartilhado pelas arqueas e eucariotos, genes que são claramente associados ao processamento da informação. O fato da afinidade arquea-eucariótica ser quantitativamente pequena, demonstra, no entanto, que o processo de evolução tem sido mais complexo que a simples herança vertical, e tem envolvido uma extensiva transferência lateral de genes entre *Archaea* e *Bacteria*. Após a divergência evolutiva entre as linhagens de arqueas e bactérias, vem ocorrendo uma grande mistura de genes codificantes para enzimas metabólicas, componentes estruturais da célula e outras proteínas que não participam da maquinaria central de processamento da informação (Nelson *et al.*, 1999).

Além dos estudos funcionais, a genômica de *Archaea* é fundamental para o conhecimento que temos de duas transições cruciais na evolução da vida: a primeira é a divergência entre as linhagens de bactérias e as de arquea-eucarióticas, que pode ter envolvido a origem da maquinaria de replicação de DNA. A segunda é a origem dos eucariotos. Em relação a esse ponto, a arquea é uma fonte fantástica de informação, particularmente porque, em muitas situações, ela tem retido as características primitivas, enquanto os eucariotos têm sofrido modificações muito maiores. Um exemplo característico é a subunidade menor da DNA polimerase, que possui todas as marcas de uma fosfatase ativa em arqueas, mas não em eucariotos, onde a atividade fosfatásica está provavelmente inativada. Sem sombra de dúvidas, arquea representa um ancestral comum das linhagens arquea-eucarióticas descendentes. Portanto, a genômica de arquea é a nossa melhor oportunidade de reconstruir essa fase intermediária crítica da evolução da vida (Makarova e Koonin, 2003).

## 6. Considerações Finais

Estudos que envolvem o domínio *Archaea* vêm confirmando as duas hipóteses iniciais de Woese e Fox (1977), isto é, que as arqueas exibem uma diversidade fenotípica no mínimo comparável àquela apresentada pelo domínio *Bacteria* e que os organismos do domínio *Archaea* serão caracterizados por aspectos únicos em âmbito molecular. Outrossim, o fato de *Archaea* exibir um mosaico contendo características dos dois outros domínios continua a estimular discussões entre os evolucionistas (Forterre *et al.*, 2002).

No contexto de extremofilia, a descoberta contemporânea mais surpreendente foi, sem dúvida, a dos organismos hipertermófilos, que estendeu a sobrevivência desse organismo para cerca de 121°C de temperatura em que células vivas proliferam eficientemente. Essa característica notável implica na estabilização de todos os componentes celulares, de modo que a sua funcionalidade seja mantida em condições de temperatura que seriam danosas para a maioria das biomoléculas dos organismos mesófilos. A elucidação das estratégias usadas na estabilização de componentes celulares e, em especial, de proteínas, representa um desafio fascinante para a biologia atual (Kashe e Lovley, 2003).

Os microrganismos apresentam uma imensa diversidade genética e desempenham funções únicas e decisivas na manutenção de ecossistemas, como componentes fundamentais de cadeias alimentares e ciclos biogeoquímicos. É importante ressaltar que grande parte dos avanços da biotecnologia moderna e da agricultura é derivada das descobertas recentes nas áreas de genética, fisiologia e metabolismo de microrganismos.

Considerando que o Brasil possui uma grande extensão territorial com inúmeros e variados ambientes extremos, como: águas termais, salinas, inúmeras estações de tratamento de esgoto, rejeitos industriais, entre outros. A biodiversidade microbiana brasileira, ainda não explorada, pode-se tornar uma fonte para o desenvolvimento biotecnológico do país.

Estamos em plena era biotecnológica, quando os processos bioquímicos são cada vez mais utilizados para a produção de agentes terapêuticos, produtos químicos e biocatalisadores. O grande desafio será incorporar a informação decorrente do estudo desses organismos extremofílicos em novas tecnologias, utilizando o enorme potencial de suas enzimas e biomoléculas.

## 7. Referências Bibliográficas

- Bauer, M. W., Kelly, R. M. (1998). The family b-glucosidases from *Pyrococcus furiosus* and *Agrobacterium faecalis* share a common catalytic mechanism. *Biochem.* **37**:17170-17178.
- Bentley, I. S., Williams, E. C. (1996). Starch conversion. *Ind. Enzymol.* **2**:339-357.
- Björkling, F., Godtfredsen, S. E., Kirk, O. (1991). The Future Impact of Industrial Lipases. *Trends Biotechnol.* **9**:360-363.
- Brock, T. D., Madigan, M. T., Martinko, J. M., Parker, J. (1994). *Biology of Microorganisms*. 7ed. Prentice Hall: New Jersey, 909p.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D. *et al.* (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science.* **273**:1058-1073.
- Dong, G., Vieille, C., Savchenko, A., Zeikus, J. G. (1997). Cloning, sequencing, and expression of the gene encoding extracellular  $\alpha$ -amylase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. *Appl. Environ. Microbiol.* **63**:3569-3576.
- Eichler, J. (2001). Biotechnological uses of archaeal extremozymes. *Biotechnol. Adv.* **19**:261-278.
- Forterre, P. (1997). Archaea: what can we learn from their sequences? *Curr. Opin. Gen. Devel.* **7**:764-770.
- Forterre, P., Brochier C., Philippe, H. (2002). Evolution of the Archaea. *Theor. Popul. Biol.* **4**:409-422.
- Gaasterland, T. (1999). Archaeal genomics. *Curr. Opin. Microbiol.* **2**:542-547.
- Ganghofner, D., Kellermann, J., Staudenbauer, W.L., Bronnenmeier, K. (1998). Purification and properties of an amylopullulanase, a glucoamylase, and an  $\alpha$ -glucosidase in the amylolytic enzyme system of *Thermoanaerobacterium thermosaccharolyticum*. *Biosci. Biotechnol. Biochem.* **62**:302-308.
- Gutierrez, A., Del Rio, J., Martinez, M. J. (2001). The Biotechnology Control of Pitch in Paper Pulp Manufacturing. *Trends Biotechnol.* **19**:340-348.
- Herbert, R. A. (1992). A perspective on the biotechnological potential of extremophiles. *Trends Biotechnol.* **10**:395-402.
- Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., Stetter, K. O. (2002). A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. *Nature.* **417**:63-67.
- Jaeger, K. E., Reetz, M. T. (1998). Microbial Lipases from Vestibule Tools for Biotechnology. *Trends Biotechnol.* **16**:396-402.
- Kashe, K., Lovley, D. R. (2003). Extending the Upper Temperature Limit for Life. *Science.* **301**:934
- Kim, C.-H., Nashiru, O., Ko, J. H. (1996). Purification and biochemical characterization of pullulanase type I from *Thermus caldophilus*. *FEMS Microbiol. Lett.* **138**:147-152.
- Koch, R., Canganella, F., Hippe, H., Jahnke, K. D., Antranikian, G. (1997). Purification and properties of a thermostable pullulanase from a newly isolated thermophilic anaerobic bacterium *Fervidobacterium pennavorans* ven5. *Appl. Microbiol.* **63**:1088-1094.
- Krishnan, L., Dicaire, C. J., Patel, G. B., Sprott, G. D. (2000). Archaeosome vaccine adjuvants induce strong humoral, cell-mediated, and memory responses: comparison to conventional liposomes and alum. *Infect Immun.* **68**: 54-63.
- Lundberg, K. S., Shoemaker, D. D., Adams, M. W., Short, J. M., Sorge, J. A., Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene.* **108**:1-6.
- Makarova, K. S., Koonin, E. V. (2003). Comparative genomics of archaea: how much have we learned in six years, and what's next? *Gen. Biol.* **4**:115-128.
- Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J. *et al.* (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature.* **399**:323-329.
- Niehaus, F., Bertoldo, C., Kahler, M., Antranikian, G. (1999). Extremophiles as a source of novel enzymes for industrial application. *Appl. Microbiol. Biotechnol.* **51**:711-729.
- Patel, G. B., Sprott, G. D. (1999). Archaeobacterial ether lipid liposomes (archaeosomes) as novel vaccine and drug delivery systems. *Crit. Rev. Biotechnol.* **19**:317-57.
- Santos, H., Lamosa, P., da Costa, M. S. (2001). Extremófilos: microrganismos à prova de agressões ambientais extremas. *Bol. Biotec.* **69**:2-10.
- Schiraldi, C., De Rosa, M. (2002). The production of biocatalysts and biomolecules from extremophiles. *Trends Biotechnol.* **20**:515-521.
- Sprott, G. D., Tolson, D. L., Patel, G. B. (1997). Archaeosomes as novel antigen delivery systems. *FEMS Microb. Lett.* **154**: 17-22.
- Tolan, J. S. (1996). Pulp and Paper. *Ind. Enzymol.* **2**:327-338.
- Tolson, D. L., Latta, R. K., Patel, G. B. and Sprott, G. D. (1996) Uptake of archaeobacterial and conventional liposomes by phagocytic cells. *J. Liposome Res.* **6**, 755-776.
- Vieille, C., Zeikus, G. J. (2001). Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol. Mol. Biol. Rev.* **65**:1-43.
- Woese, C. R. (1998). The universal ancestor. *Proc. Natl. Acad. Sci. USA.* **95**:6854-6859.
- Woese, C. R., Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA.* **74**:5088-5090.
- Woese, C. R., Kandler, O., Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc. Natl. Acad. Sci. USA.* **87**:4576-4579.

**Manuscrito 6:** "*Exploring the biotechnological applications in the Archaeal Domain*".

Recentemente, submetemos mais uma revisão sobre o Domínio Archaea para o período *Brazilian Journal of Microbiology* (*in press*).

Neste artigo, além de destacarmos as potencialidades biotecnológicas levantamos aspectos de prospecção e isolamento de organismos desse domínio da Vida no Brasil.

## EXPLORING THE BIOTECHNOLOGICAL APPLICATIONS IN THE ARCHAEOAL DOMAIN

Alquères, S.M.C.<sup>1</sup>; Almeida, R.V.<sup>2</sup>; Clementino, M.M.<sup>3</sup>; Vieira, R.P.<sup>1</sup>; Almeida, W.I.<sup>1</sup>; Cardoso, A.M.<sup>1\*</sup>; Martins, O.B.<sup>1</sup>

<sup>1</sup>Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil; <sup>2</sup>Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil; <sup>3</sup>Instituto Nacional de Controle da Qualidade em Saúde, Fundação Instituto Oswaldo Cruz, Rio de Janeiro, RJ, Brasil.

Submitted: March 25, 2007; Returned to authors for corrections: July 16, 2007; Approved: July 29, 2007.

---

### MINI-REVIEW

---

#### ABSTRACT

Archaea represent a considerable fraction of the prokaryotic world in marine and terrestrial ecosystems, indicating that organisms from this domain might have a large impact on global energy cycles. The extremophilic nature of many archaea has stimulated intense efforts to understand the physiological adaptations for living in extreme environments. Their unusual properties make them a potentially valuable resource in the development of novel biotechnological processes and industrial applications as new pharmaceuticals, cosmetics, nutritional supplements, molecular probes, enzymes, and fine chemicals. In the present mini-review, we show and discuss some exclusive characteristics of Archaea domain and the current knowledge about the biotechnological uses of the archaeal enzymes. The topics are: archaeal characteristics, phylogenetic division, biotechnological applications, isolation and cultivation of new microbes, achievements in genomics, and metagenomic.

**Key words:** Archaea, Biotechnology, Extremozymes, Genomic, Molecular Phylogeny.

---

#### Archaea domain

In the late 1970s, Carl Woese and his colleagues at the University of Illinois studied relationships among prokaryotes and proposed that life should be divided into three domains: Archaea, Bacteria, and Eucarya. They found that there were two distinctly different prokaryotic groups. Those "bacteria" that lived at high temperatures or produced methane were clustered together as a group far away from the usual bacteria (Fig. 1) and the eukaryotes (51).

Woese argued that Archaea, Bacteria, and Eucarya each represent a primary line of descent that diverged early from an ancestral *progenote* with poorly developed genetic machinery. This hypothesis is reflected in the name Archaea, from the Greek *archae*, meaning ancient. Later he treated these groups formally as domains, each consisting of several kingdoms. This separation and organization of tree of life have become very popular, although the idea of the *progenote* itself is not generally supported (50).

For 15 years after their recognition, the Archaea were generally known only as inhabiting hostile environments (27). Within the past two decades, the use of molecular techniques, including PCR-based amplification of 16S rRNA genes, has allowed a culture-independent assessment of microbial diversity, indicating a wide distribution of mostly uncultured archaea in normal habitats, such as ocean and lake waters and soil (10). Nowadays the use of 16S rDNA clone libraries to map the diversity of uncultivated prokaryotes from natural populations has provided a revolutionary advance for interpreting microbial evolutionary relationships. These molecular surveys have produced more than 20,000 archaeal 16S rRNA gene sequences from environmental studies, extending the known groups and increasing the number of novel lineages. In Brazil, our group study the archaeal diversity in tropical environments and show that the detection of a substantial number of uncultured phylotypes suggests that this region harbors a pool of novel archaeoplankton taxa (11,47).

---

\*Corresponding Author. Mailing address: Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Centro de Ciências da Saúde - Bloco D, subsolo, sala 5. 21941-590 - Rio de Janeiro - Brasil. Tel.: +55 21 25626751 - Fax: +55 21 22708647. E-mail: amcardoso@bioqmed.ufrj.br

### Archaeal characteristics

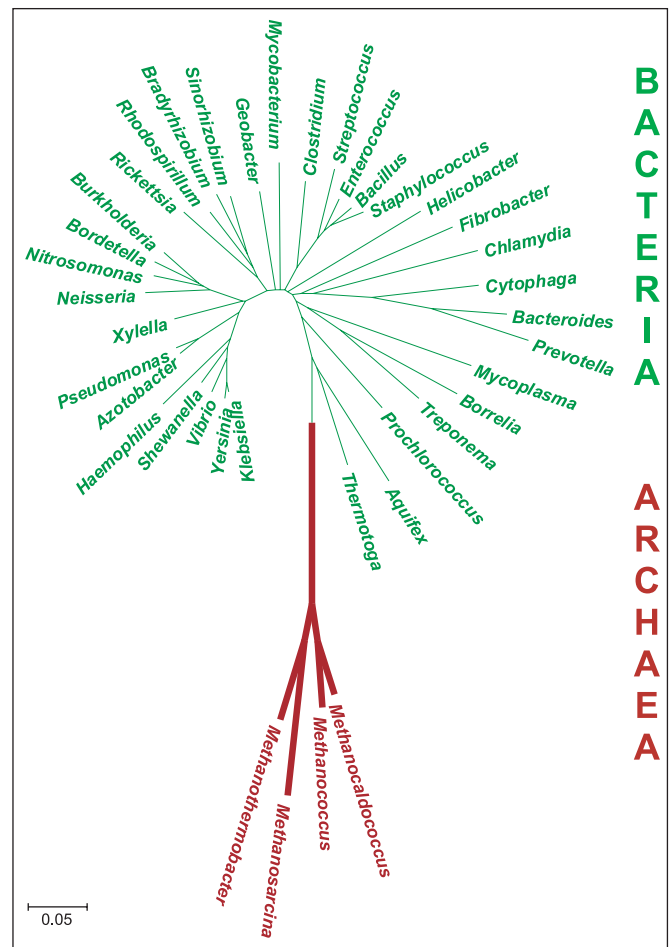
Several features set the Archaea apart, for example, archaea have a single cell membrane containing a peptidoglycan-like wall which is different in bacteria. Furthermore, both eubacteria and eukaryotes have membranes composed mainly of glycerol-ester lipids, whereas archaea have membranes composed of glycerol-ether lipids. These differences may be an adaptation to extreme environments (7). Archaeal organisms also have flagella that are notably different in composition and development from the flagella of bacteria. Individual archaea cells range from 0.1 to over 15  $\mu\text{m}$  in diameter, and some form aggregates or filaments up to 200  $\mu\text{m}$  in length. They occur in various shapes, such as spherical, rod-shaped, spiral, lobed, or rectangular, and they also exhibit a variety of different types of metabolism. Archaeal and bacterial metabolic genes share common evolutionary aspects (6). However, the transcriptional and translational machinery of Archaea is much more similar to Eucarya than Bacteria (28). For instance, archaean translation uses eukaryotic initiation and elongation factors, and their transcription involves TATA-binding proteins and TFIIB (14). The study of protein synthesis in Archaea led to a much deeper understanding of this process not only in the Archaea, but in all domains of life (8,33,52).

Many archaeans are extremophiles, some live at very high temperatures, often above 100°C, as found in geysers and submarine black smokers. Others are found in very cold habitats, highly saline, acidic, or alkaline water. They are able to live in the anoxic mud of marshes and at the bottom of the ocean, even thrive in petroleum deposits deep underground. However, many archaeans are mesophiles, and have been found in environments like marine plankton (5), sewage, and soil (36). Furthermore, many methanogenic archaea are found in the digestive tracts of animals such as ruminants, termites, and humans (15,42).

### Phylogenetic division

The Archaea domain is divided on the basis of 16S rRNA gene sequences into four phylogenetically distinct phylum: Euryarchaeota, Crenarchaeota, Korarchaeota and Nanoarchaeota (Fig. 2).

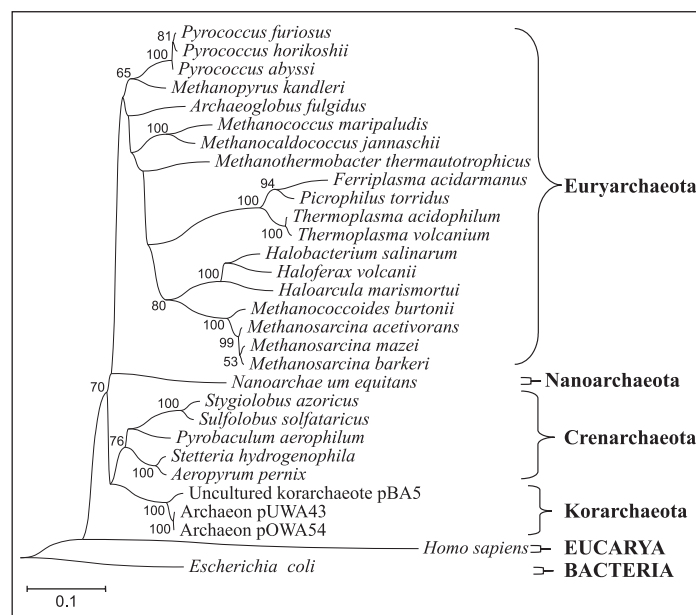
Cultivated crenarchaeotes presents a thermoacidophilic phenotype. The term “cren” means spring or fount and expresses the resemblance of this phenotype to the ancestor of the domain Archaea. Most cultured representatives of the Crenarchaeota are extreme thermophiles or hyperthermophiles. A broad variety of metabolic pathways is evident. Aerobically growing chemolithotrophs gain energy by the oxidation of various sulfur compounds, molecular hydrogen or ferrous iron. Anaerobic chemolithotrophs reduce sulfur, thiosulfate or produce nitrate, hydrogen sulfide or ammonia. Organotrophic growth occurs on complex organic substrates, sugars, amino acids or polymers such as starch and cellulose. However, cultivated species appear to represent a minority in



**Figure 1.** Carl Woese view at the prokaryotic tree of life. We reproduced *in silico* Woese’s experiment by phylogenetic relationships of 16S rDNA sequences from Bacteria and Archaea. We used thirty-five eubacterial and four methanogenic archaeal sequences from GenBank database. Phylogenetic and molecular evolutionary analyses were conducted using MEGA (Kumar *et al.* 2001).

Crenarchaeota, which comprise an important number of uncultured groups from marine plankton, freshwater, and soil samples (13,33,39)

The Euryarchaeota phylum, the term “eury” means broad or wide, contains organisms that are highly diverse in their physiology, morphology and natural habitats. During the last decade, numerous reclassifications within the Euryarchaeota have been carried out, mainly based on the results of 16S rRNA sequence comparisons. Five major groups are known within this kingdom: the obligate anaerobic methanogens, the extreme halophiles, the hyperthermophilic sulfate reducers, the *Thermoplasma* group, and finally, the *Thermococcus*-*Pyrococcus* group (28, 35).



**Figure 2.** Phylogenetic tree of 16S rRNA genes from archaea with genomes representing the four archaeal groups (Nanoarchaeota, Korarchaeota, Crenarchaeota, Euryarchaeota). The numbers at the roots indicate bootstrap support based on a neighbor-joining analysis of 1,000 resampled data sets and are given as percentages with only values of > 50% show. *Escherichia coli* was used as the outgroup.

Interesting, new microbes are still being discovered and investigated. Most recently, Karl Stetter and his colleagues discovered the world's smallest microorganism, *Nanoarchaeum equitans*. Its name roughly translates to "ancient dwarf who rides the fire ball" because of its tendency to latch on to or "ride" the surface of the archaeal microbe *Ignicoccus* ("fireball"). The discovery of this nanosized hyperthermophilic archaeon has led to the creation of a new phylum, Nanoarchaeota (19). This extremophile was found in a 120 m depth submarine hydrothermal vent and thrives in temperatures close to 100°C. With less than 500 kb in its genome, *N. equitans* represents the smallest archaeal genome sequenced until now.

Recently, Korarchaeota-specific primers were designed based on previously published sequences and used to screen a variety of environments. Korarchaeota 16S rRNA genes were amplified from high temperature like Yellowstone National Park hot springs, East Pacific Rise deep-sea hydrothermal vent and also found in other environments (3).

Although this is a well established division, some authors have suggested a revision in this classification as sequences from environmental samples are emerging, showing that the archaeal tree is more complex than expected (35).

## Biotechnological Applications

The extremophilic nature of many Archaea has stimulated intense efforts to understand the physiological adaptations for living in extremes environments and to probe the potential biotechnological applications of their stable cellular components. This is particularly true of their enzymes (called extremozymes), which are able to remain catalytically active under extremes of temperature, salinity, pH and pressure. Many interesting enzymes have been isolated from extremophilic microbes (41). Specific archaeal metabolites have also been purified and characterized and some of them have potential industrial uses (Table 1). However, several technical difficulties have prevented the large-scale industrial application of enzymes and special metabolites from extremophilic sources, the most important being the availability of these compounds (40). Genes encoding several enzymes from extremophiles have been cloned in mesophilic hosts, with the objective of overproducing the enzyme and altering its properties to suit commercial applications. *Escherichia coli*, *Bacillus subtilis* and yeasts have been used successfully as mesophilic hosts for several archaeal genes. Genetic engineering techniques are valuable tools for creating novel biocatalysts that can improve bioprocesses and facilitate the realization of innovative biotransformations. Molecular biology has the potential both to overcome the limitations on enzyme availability and to design new biocatalysts specific to a particular industrial purposes.

## Glucosylating enzymes

Due to the complex structure of starch, a number of enzymes are needed for its degradation. Endo-acting enzymes, such as  $\alpha$ -amylase, hydrolyse linkages in the interior of the starch polymer in a random fashion, which leads to the formation of linear and branched oligosaccharides. Exo-acting enzymes ( $\alpha$  and  $\beta$ -glucoamylases) attack the substrate from the non-reducing end, producing oligo and/or monosaccharides. The finding of new extremely thermostable starch-hydrolysing enzymes, such as amylases and pullulanases, will significantly improve the industrial starch bioconversion process. Investigations concerning recombinant  $\alpha$ -amylases from *Pyrococcus woesei* indicate their suitability for starch processing. Furthermore, the study of recombinant  $\beta$ -galactosidase from *P. woesei* suitable for purpose of low lactose milk and whey production is also presented. The activity of this enzyme in a wide pH range of 4.3-6.6 and high thermostability suggests that it can be used for processing of dairy products at temperatures which restrict microbial growth during a long operating time of continuous-flow reactor with an immobilized enzyme system. Preparation of recombinant  $\alpha$ -amylase and  $\beta$ -galactosidase was facilitated by cloning and expression of genes from *P. woesei* in *E. coli* host. Satisfactory level of recombinant enzymes purification was achieved by thermal denaturation and

**Table 1.** Industrial applications of archaeal products.

Phenotype	Condition	Product	Application
Thermophile	High temperature (45 - 110°C)	Amylases Xylanases Proteases DNA polymerases	Glucose, fructose for sweeteners Paper bleaching Baking, brewing, detergents Genetic engineering
Psychrophile	Low temperature (>15°C)	Proteases Dehydrogenases Amylases	Cheese maturation, dairy production Biosensors Polymer degradation in detergents
Acidophile	Low pH (0 - 4)	Sulfur oxidation	Desulfurization of coal
Alkalophile	High pH (8 - 11)	Cellulases	Polymer degradation in detergents
Halophile	High salt concentration	Whole microorganism	Biopolymers
Piezophile	High pressure	Whole microorganism	Formation of gels and starch granules
Metalophile	High metal concentration	Whole microorganism	Bioremediation, biomineralization
Radiophile	High radiation levels	Whole microorganism	Bioremediation of radionuclide contaminated sites

precipitation of native proteins originated from *E. coli*. The obtained  $\alpha$ -amylase has maximal activity at pH 5.6 and 93°C. The half-life of this preparation (pH 5.6) at 90°C and 110°C was 11 h and 3.5 h, respectively, and retained 24 % of residual activity following incubation for 2 h at 120°C (46).

A hyperthermostable glycosidase enzyme with pullulanase activity at 90°C from *Thermococcus aggregans* was cloned and expressed in *E. coli*. Unlike all other pullulan-hydrolyzing enzymes described, the enzyme is able to attack  $\alpha$ -1,6- as well as  $\alpha$ -1,4-glycosidic linkages, affording a mixture of maltotriose, maltose and glucose. The enzyme is also able to degrade starch, amylose and amylopectin, forming maltotriose and maltose as main products (26).

Halophilic glycosidases have also been identified, including a halophilic  $\beta$  galactosidase from *Haloferax alicantei*. The extremely halophilic galactosidase from *Haloferax alicantei* is optimally active at 4 M NaCl. Purification of the enzyme was facilitated by the ability of sorbitol to stabilize enzyme activity in the absence of salt, which allowed conventional ion-exchange chromatography (18).

The continuous production of halophilic  $\alpha$ -amylase can be performed via whole-cell immobilization of *Halobacterium salinarum* in alginate beads and a polyvinyl alcohol film. The cells were osmotically stable and showed continuous enzyme production for 45 days. The stabilized cells could be permeabilized by chloroform treatment without leakage of the intracellular components. Using this procedure, cells can be reused under improved stabilized conditions for biotechnological applications (4).

### Proteases and Peptidases

The amount of proteolytic enzymes produced world-wide on a commercial scale is larger than that of any of the other

biotechnologically used enzymes. Serine alkaline proteases are used as additives to household detergents for laundering where they must resist denaturation by detergents and alkaline conditions. The leather industry uses proteinases with high keratinolytic and elastolytic activities for soaking. The discovery of proteases that can catalyze reactions under extreme conditions (high temperatures and extremes of pH) will be valuable for industrial applications (23,25).

An extracellular serine protease from the extreme halophile *Halobacterium halobium* may be an excellent catalyst for peptide synthesis, exploiting their reverse reaction, particularly for glycine-containing peptides. The enzyme requires 4 M NaCl for optimal catalytic activity and stability in aqueous solutions (38). The stabilization of halophilic enzymes by organic solvents while lowering the required salt concentration is of practical importance because high NaCl concentrations can be corrosive to metals. Higher substrate solubility in the presence of an organic solvent can be useful in synthetic reactions catalyzed by enzymes from extreme halophiles (21). Many biotechnologically interesting serine proteases have been identified and characterized. Among them is a cell associated serine protease, characterized from *Desulfurococcus* strain SY, which showed a half-life of 4.3 h at 95°C (17). A globular serine protease from *Staphylothermus marinus* was found to be extremely thermostable. This enzyme has residual activity even after 10 min of incubation at 135°C (24). Another gene encoding a subtilisin-like serine protease, named aereolysin, has been cloned from *Pyrobaculum aerophilum* and the protein was modelled on the basis of structures of subtilisin-type proteases (48). Multiple proteolytic activities have been observed in *Pyrococcus furiosus*. The cell envelope associated serine protease of *P. furiosus*, called pyrolysin, was found to be highly stable with a half-life of 20 min at 105°C. The pyrolysin gene



was cloned and sequenced and it was shown that this enzyme is a subtilisin-like serine protease (25,49).

Proteases have also been characterized from the thermoacidophilic archaeon *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius*. In addition to the serine proteases, other types of enzymes have been identified in extremophiles. Three good examples are a thiol protease (propylpeptidase) from *Pyrococcus* sp. and a new type of protease from *P. furiosus* (25).

Psychrophilic enzymes have potential applications in a broad range of industrial, agricultural and medical processes. The recent interest in this field is due to the challenge of finding stable proteases that function in cold water. In addition, peptidyl synthesis studies with mesophilic enzymes have shown that low temperature favors high yields because of reduced hydrolysis of the acyl-enzyme intermediate. In most other processes, reduced energy consumption due to low temperature operation will be a significant advantage (9).

### DNA processing enzymes

One of the most important advances in molecular biology was the development of the polymerase chain reaction (PCR). Thermostable DNA polymerases play a major role in PCR and in a variety of molecular biological applications, e.g. DNA amplification, cloning, sequencing or labeling. Several native and recombinant polymerases have been purified and characterized (32).

DNA polymerase I from the bacterium *Thermus aquaticus*, called *Taq* polymerase, was the first thermostable DNA polymerase utilized in PCR. Archaeal proof-reading polymerases, such as *Pwo* from *Pyrococcus woesei*, *Pfu* from *P. furiosus*, Deep Vent polymerase from the *Pyrococcus* strain GB-D and Vent polymerase from *Thermococcus litoralis*, have an error rate that is up to tenfold lower than that of *Taq* polymerase (44).

### Esterases and lipases

In the field of biotechnology, esterases are receiving increasing attention because of their application in organic biosynthesis. In aqueous solution, esterases catalyze the hydrolytic cleavage of esters to form the constituent acid and alcohol whereas, in organic solutions, the transesterification reaction is promoted. Both the reactants and the products of transesterification are usually highly soluble in the organic phase and the reactants may even form the organic phase themselves. The *P. furiosus* esterase and lipase (1) genes have been cloned in *E. coli* and the functional properties have been determined. The archaeal enzyme is the most thermostable (a half-life of 50 min at 126°C) and thermoactive (optimum temperature of 100°C) esterase known to date (20).

### Archaeosomes

Liposomes are lipid-bilayer bounded vesicles that can be used as a delivery vehicle for certain vaccines, enzymes, and

drugs. The term “archaeosome” was introduced by Sprott and co-workers to describe liposomes made with ether lipids that are unique to the Archaea domain conferring considerable stability on liposomal vesicles (45). Extensive mouse model studies involving intravenous, oral and subcutaneous administration of archaeosomes demonstrated that archaeosomes are safety molecules and they are not toxic (31).

### Archaeal S-layer

Like other elements of the cell envelope, S-layers play an important part in interactions of the microbial cell with the environment (43). It was demonstrated that archaeal S-layers have been shown to be excellent patterning structures in molecular nanotechnology and biomineralization due to their high molecular order, high binding capacity and ability to recrystallize with perfect uniformity on solid surfaces, at the air:water interface and on lipid films (43).

### Bacteriorhodopsin

The halophilic archaeon makes use of light for both energy and sensory transduction by exploiting a family of light-sensitive proteins. The archaeal rhodopsin, a 26.5 kDa protein, has a transmembrane domain of seven helical protein segments which photons are captured via a retinal chromophore (29). The excellent thermodynamic and photochemical stability of bacteriorhodopsin has led to many uses in technical applications like holography, spatial light modulators, artificial retina, neural network optical computing, and volumetric and associative optical memories (22).

### Isolation and cultivation of new microbes

Several studies specifically aimed at the isolation of archaeal DNA rather than at the isolation of the microorganisms themselves are known (13,30,39). However, the isolation of microorganisms from complex mixed cultures and their cultivation in a pure culture is an essential prerequisite for their complete characterization. Different processes are used for this purpose, including the classic plate casting process, dilution methods, and selective enrichment. The majority of cells obtained from nature do not generally form visible colonies on plates (16). It can be very difficult to isolate new microbes when the conditions for their growth are not known or when the microbes grow in a heterogeneous mixed culture in very small numbers (12). Evidently, the biosphere remains replete with unexplored niches filled with extraordinary organisms awaiting the design of new cultivation and isolation strategies (34).

### Genomics

Many microorganisms have been isolated from the Archaea domain, and among the archaeal enzymes that have been purified and characterized, several show potential for industrial use. Archaeal metabolites suitable for industrial applications have

also been found (Table 1). In addition to the development of industrial production processes, basic research is also needed to determine how enzymes have adapted to extreme conditions at the molecular level. It has been shown that modifications in enzymes of psychrophiles to permit conformational flexibility at low temperatures are opposite to those that confer thermostability to proteins of thermophiles (37). The slow incorporation of archaeal tools by industry can be ascribed mainly to the difficulties associated with the development of large-scale processes for the production of archaeal biomasses.

Genome sequencing should help to overcome these difficulties. Cloning into an easy-to-grow mesophilic host may be the preferred method for archaeal enzyme production because it simplifies downstream processing. Strategies that combine molecular biology techniques and customized bioprocess design could greatly improve the yields of archaeal bioproducts. This would result in major cost savings and therefore, enable industrial applications of these unique biomaterials.

#### **Metagenomics: searching for new genes**

Molecular studies of microbial diversity have shown that not cultivated microorganisms can correspond to >99% of the total microbial population (2). Current knowledge of prokaryotic diversity is so poor, in fact, that a significant number of archaea isolated randomly from natural environments will turn out to be a new organism. Furthermore, access to this reservoir of genetic and metabolic diversity is of great interest in biotechnology. The main requirement is the availability of pure, intact, high molecular weight genomic DNA. High-molecular-weight DNA extracted from environmental samples can be cloned directly into large insert cloning vectors, such as cosmids, fosmids, or bacterial artificial chromosomes (BACs) (36). The genome fragments of uncultivated microbes can be useful for expressing protein-encoding genes, determining enzyme structure and function, or dissecting metabolic pathways. Examples of such an approach resulting in the genomic analysis are the construction of large-insert bacterial artificial chromosome (BAC) libraries from the genomic DNA of planktonic marine microbial assemblages (5) and from genomic DNA isolated directly from soil (36).

#### **Perspectives**

Many microorganisms have been isolated from the domain Archaea, and among the archaeal enzymes that have been purified and characterized, several show potential for industrial uses. However, biotechnological processes based on these archaeal tools remain limited, despite the extensive recent literature suggesting possible novel bioprocesses and demonstrating the applicability of several biomolecules of archaeal origin in the biotechnology industry. Molecular biology has the potential both to overcome the limitations on enzyme

availability and to design new biocatalysts that can improve bioprocesses and facilitate the realization of innovative biotransformations. The great challenge will be to incorporate these new processes to improve the biotechnological applications of archaeal enzymes and biomolecules.

#### **ACKNOWLEDGEMENTS**

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ).

#### **RESUMO**

#### **Explorando as aplicações biotecnológicas do domínio archaea**

As arqueas representam uma considerável fração dos procariotos nos ecossistemas marinhos e terrestres, indicando que estes organismos devem possuir um grande impacto nos ciclos energéticos. A natureza extremófila de muitas arqueas tem estimulado intensos esforços para compreender sua adaptação fisiológica a ambientes extremos. Suas propriedades incomuns as tornam uma fonte valiosa no desenvolvimento de novos processos biotecnológicos e aplicações industriais como novos fármacos, cosméticos, suplementos nutricionais, sondas moleculares, enzimas e reagentes. Na presente mini-revisão, mostramos e discutimos algumas de suas características exclusivas correlacionando-as com seu potencial biotecnológico e aplicação industrial. Os tópicos são: características das arqueas, divisão filogenética, aplicações biotecnológicas, isolamento e cultivo de novos microrganismos, genoma e metagenoma.

**Palavras-chave:** Archaea, Biotecnologia, Enzimas extremófilas, Filogenia Molecular, Genômica.

#### **REFERENCES**

1. Almeida, R.V.; Alqueres, S.M.C.; Larentis, A.L.; Rossle, S.C.S.; Cardoso, A.M.; Almeida, W.I.; Bisch, P.M.; Alves, T.L.M.; Martins, O.B. (2006). Cloning, expression, partial characterization and structural modeling of a novel esterase from *Pyrococcus furiosus*. *Enz. Microb. Technol.*, 39, 1128-1136.
2. Amann, R.I.; Ludwig, W.; Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.*, 59(1), 143-169.
3. Auchtung, T.A.; Takacs-Vesbach, C.D.; Cavanaugh, C.M. (2006). 16S rRNA phylogenetic investigation of the candidate division "Korarchaeota". *Appl. Environ. Microbiol.*, 72(7), 5077-5082.
4. Bagai, R.; Madamwar, D. (1997). Continuous production of halophilic  $\alpha$ -amylase through whole cell immobilization of *Halobacterium salinarium*. *Appl. Biochem. Biotechnol.*, 62(2), 213-218.

5. Beja, O.; Suzuki, M.T.; Koonin, E.V.; Aravind, L.; Hadd, A.; Nguyen, L.P.; Villacorta, R.; Amjadi, M.; Garrigues, C.; Jovanovich, S.B.; Feldman, R.A.; DeLong EF. (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.*, 2(5), 516-529.
6. Brown, J.R.; Doolittle, W.F. (1997). Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.*, 61(4), 456-502.
7. Cardoso, A.M.; Clementino, M.B.M.; Martins, O.B.; Vieira, R.P.; Almeida, R.V.; Alquéres, S.M.C.; Almeida, W.I. (2003). Archaea: Potencial Biotecnológico. *Rev. Biotec. Cien. Des.*, 30, 71-77.
8. Cardoso, A.M.; Polycarpo, C.; Martins, O.B.; Soll, D. (2006). A non-discriminating aspartyl-tRNA synthetase from *Halobacterium salinarum*. *RNA Biology*, 3, 110-114.
9. Cavicchioli, R.; Siddiqui, K.S.; Andrews, D.; Sowers, K.R. (2002). Low-temperature extremophiles and their applications. *Curr. Opin. Biotechnol.*, 13(3), 253-261.
10. Chaban, B.; Ng, S.Y.; Jarrel, F.K. (2006). Archaeal habitats – from the extreme to the ordinary. *Can. J. Microbiol.*, 52, 73-116.
11. Clementino, M.M.; Fernandes, C.C.; Vieira, R.P.; Cardoso, A.M.; Polycarpo, C.R.; Martins O.B. (2007). Archaeal diversity in naturally occurring and impacted environments from a tropical region. *J. Appl. Microbiol.*, 103(1), 141-151.
12. Costerton, J.W.; Lewandowski, Z.; DeBeer, D.; Caldwell, D.; Korber, D.; James, G. (1994). Biofilms, the customized microniche. *J. Bacteriol.*, 176, 2137-2142.
13. DeLong, E.F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA*, 89, 5685-5689.
14. Dennis, P.P. (1997). Ancient ciphers: translation in Archaea. *Cell.*, 89, 1007-1010.
15. Eckburg, P.B.; Bik, E.M.; Bernstein, C.N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S.R.; Nelson, K.E.; Relman, D.A. (2005). Diversity of the human intestinal microbial flora. *Science*, 308 (5728), 1635-1638.
16. Eilers, H.; Perntaler, J.; Glockner, F.O.; Amann, R. (2000). Culturability and In situ abundance of pelagic bacteria from the North Sea. *Appl. Environ. Microbiol.*, 66, 3044-3051.
17. Hanzawa, S.; Hoaki, T.; Jannasch, H.W.; Maruyama, T. (1996). An extremely thermostable serine protease from a hyperthermophilic archaeon *Desulfurococcus* strain SY, isolated from a deep-sea hydrothermal vent. *J. Mar. Biotechnol.*, 4, 121-126.
18. Holmes, M.L.; Scopes, R.K.; Moritz, R.L.; Simpson, R.J.; Englert, C.; Pfeifer, F.; Dyall-Smith, M.L. (1997). Purification and analysis of an extremely halophilic  $\beta$ -galactosidase from *Haloferax alicantei*. *Biochim. Biophys. Acta*, 1337(2), 276-286.
19. Huber, H.; Hohn, M.J.; Rachel, R.; Fuchs, T.; Wimmer, V.C.; Stetter, K.O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, 417, 63-67.
20. Ikeda, M.; Clark, D.S. (1998). Molecular cloning of extremely thermostable esterase gene from hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli*. *Biotechnol. Bioeng.*, 57(5), 624-629.
21. Kim, J.; Dordick, J.S. (1997). Unusual salt and solvent dependence of a protease from an extreme halophile. *Biotechnol. Bioeng.*, 55(3), 471-479.
22. Koyama, K.; Yamaguchi, N.; Miyasaka, T. (1994). Antibody-mediated bacteriorhodopsin orientation for molecular device architectures. *Science*, 265, 762-764.
23. Ladenstein, R.; Antranikian, G. (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv. Biochem. Eng. Biotechnol.*, 61, 37-85.
24. Mayr, J.; Lupas, A.; Kellermann, J.; Eckerskorn, C.; Baumeister, W.; Peters, J. (1996). A hyperthermostable protease of the subtilisin family bound to the surface of the layer of the archaeon *Staphylothermus marinus*. *Curr. Biol.*, 6, 739-749.
25. Niehaus, F.; Bertoldo, C.; Kähler, M.; Antranikian, G. (1999). Extremophiles as a source of novel enzymes for industrial application. *Appl. Microbiol. Biotechnol.*, 51, 711-729.
26. Niehaus, F.; Peters, A.; Groudieva, T.; Antranikian, G. (2000). Cloning, expression and biochemical characterisation of a unique thermostable pullulan-hydrolysing enzyme from the hyperthermophilic archaeon *Thermococcus aggregans*. *FEMS Microbiol. Lett.*, 190(2), 223-229.
27. Olsen, G.J. (1994). Archaea, Archaea, Everywhere. *Nature*, 371, 657.
28. Olsen, G.J.; Woese, C.R. (1997). Archaeal genomics: an overview. *Cell.*, 89(7), 991-994.
29. Oren, A. (1999). Bioenergetic aspects of halophilism. *Microbiol. Mol. Biol. Rev.*, 63, 334-348.
30. Ovreas, L.; Forney, L.; Daae, F.L.; Torsvik, V. (1997). Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl. Environ. Microbiol.*, 63, 3367-3373.
31. Patel, G.B.; Zhou, H.; KuoLee, R.; Chen, W. (2004). Archaeosomes as adjuvants for combination vaccines. *J. Liposome Res.*, 14(3), 191-202.
32. Perler, F.B.; Kumar, S.; Kong, H. (1996). Thermostable DNA polymerases. *Adv. Protein Chem.*, 48, 377-435.
33. Polycarpo, C.; Sheppard, K.; Randau, L.; Ambrogelly, A.; Cardoso, A.M.; Fukai, S.; Herring, S.; Hohn, M.J.; Nakamura, Y.; Oshikane, H.; Palioura, S.; Salazar, J.C.; Yuan, J.; Nureki, O.; Soll, D. (2007). Features of Aminoacyl-tRNA Synthetase Unique to Archaea. In: Cavicchioli, R. (ed). *Archaea: Molecular and Cellular Biology*. Washington, D.C. - USA, p.198-208.
34. Pondexter, J.S.; Leadbetter, E.R. (1986). Enrichment cultures in microbial ecology. In: Poindexter, J.S.; Leadbetter, E.R. (eds). *Bacteria in Nature*. Plenum Press, New York, USA, p.229-260.
35. Robertson, C.E.; Harris, J.K.; Spear, J.R.; Pace, N.R. (2005). Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol.*, 8, 638-642.
36. Rondon, M.R.; August, P.R.; Bettermann, A.D.; Brady, S.F.; Grossman, T.H.; Liles, M.R.; Loiacono, K.A.; Lynch, B.A.; MacNeil, I.A.; Minor, C.; Tiong, C.L.; Gilman, M.; Osburne, M.S.; Clardy, J.; Handelsman, J.; Goodman, R.M. (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.*, 66(6), 2541-2547.
37. Russell, N.J. (2000). Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extrem.*, 4(2), 83-90.
38. Ryu, K.; Kim, J.; Dordick, J.S. (1994). Catalytic properties and potential of an extracellular protease from an extreme halophile. *Enzyme Microb. Technol.*, 16(4), 266-275.
39. Sandaa, R.A.; Enger, O.; Torsvik, V. (1999). Abundance and diversity of Archaea in heavy-metal-contaminated soils. *Appl. Environ. Microbiol.*, 65, 3293-3297.
40. Schiraldi, C.; Giuliano M.; de Rosa M. (2001). Perspectives on biotechnological applications of archaea. *Archaea*, 1, 75-86.
41. Schumacher, K.; Heine, E.; H. Hocker. 2001. Extremozymes for improving wool properties. *J. Biotechnol.*, 89: 281-288.
42. Skillman, L.C.; Evans, P.N.; Strompl, C.; Joblin, K.N. (2006). 16S rDNA directed PCR primers and detection of methanogens in the bovine rumen. *Let. Appl. Microbiol.*, 42(3), 222-228.
43. Sleytr, U.B.; Sara, M. (1997). Bacterial and archaeal S-layer proteins: structure-function relationships and their biotechnological applications. *Trends Biotechnol.*, 15(1), 20-26.
44. Southworth, M.W.; Kong, H.; Kucera, R.B.; Ware, J.; Jannasch, H.W.; Perler, F.B. (1996) Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on *Thermococcus sp.* 9 degrees N-7 and mutations affecting 3'-5'- exonuclease activity. *Proc Natl Acad Sci USA*, 93, 5281-5285.

45. Sprott, G.D.; Tolson, D.L.; Patel, G.B. (1997). Archaeosomes as novel antigen delivery systems. *FEMS Microbiol Lett.*, 154(1), 17-22.
46. Synowiecki, J.; Grzybowska, B.; Zdzieblo, A. (2006). Sources, properties and suitability of new thermostable enzymes in food processing. *Crit. Rev. Food. Sci. Nutr.*, 46(3), 197-205.
47. Vieira, R.P.; Clementino, M.M.; Cardoso, A.M.; Oliveira, D.N.; Albano, R.M.; Gonzalez, A.M.; Paranhos, R.; Martins, O.B. (2007). Archaeal Communities in a Tropical Estuarine Ecosystem: Guanabara Bay, Brazil. *Microb Ecol.*, DOI: 10.1007/s00248-007-9261-y
48. Voëlkl, P.; Markiewicz, P.; Stetter, K.O.; Miller, J.H. (1995). The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeon *Pyrobaculum aerophilum* reveals sites important to thermostability. *Protein Sci.*, 3, 1329-1340.
49. Voorhorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Platteeuw, C.; Siezen, R.J.; Vos, W.M. (1996). Isolation and characterization of the hyperthermostable serine protease, pyrolysin, and its gene from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.*, 271, 20426-20431.
50. Woese, C.R. (1998). The universal ancestor. *Proc. Natl. Acad. Sci. USA*, 95, 6854-6859.
51. Woese, C.R.; Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA*, 74, 5088-5090.
52. Yuan, J.; Palioura, S.; Salazar, J.C.; Su, D.; O'Donoghue, P.; Hohn, M.J.; Cardoso, A.M.; Whitman, W.B.; Soll, D. (2006). RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc Natl Acad Sci USA*, 103, 18923-1892.

## **REFERÊNCIAS BIBLIOGRÁFICAS/APENDICE**

## REFERÊNCIAS BIBLIOGRÁFICAS/APENDICE

---

- Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol.* 1990; 56(6):1919-25.
- Burton NP, Norris PR. Microbiology of acidic, geothermal springs of Montserrat: environmental rDNA analysis. *Extremophiles.* 2000; 4(5):315-320.
- Cardoso, AM.; Clementino, M. M. ; Martins, O. B. ; Vieira, R. P. ; Almeida, R. V. ; Alqueres, S. M. C. ; Almeida, W. I. . *Archaea: Potencial Biotecnológico. Biotecnologia ciência & desenvolvimento*, v. 30, p. 71-77, 2003.
- Carvalho CEV, Lacerda LD, & Gomes MP (1991) Heavy metals contamination of the marine biota along the Rio de Janeiro Coast, SE-Brazil. *Water Air Soil Pollut* 57–58:645–653
- Clementino MM, Fernandes CC, Vieira RP, Cardoso AM, Polycarpo CR, Martins OB. Archaeal diversity in naturally occurring and impacted environments from a tropical region. *J Appl Microbiol.* 2007 Jul;103(1):141-51.
- Clementino, M.M. Utilização de Marcadores Moleculares no Estudo da Taxonomia e da Biodiversidade de Organismos dos Domínios Archaea e Bacteria 2006 Tese de Doutorado em Química Biológica. CCS-Instituto de Bioquímica Médica. Rio de Janeiro.
- Cole, J. R, Chai, B., Marsh, T. L., Farris, R. J., Wang, Q., Kulam, S. A., Chandra, S., McGarrell, D. M., Schmidt, T. M., Garrity, G. M. & Tiedje, J. M. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* 31(1):442-3 (2003).
- DeLong EF. Microbial community genomics in the ocean. *Nat Rev Microbiol.* 2005 Jun;3(6):459-69
- Dopson M, Baker-Austin C, Hind A, Bowman JP, Bond PL. Characterization of *Ferroplasma* isolates and *Ferroplasma acidarmanus* sp. nov., extreme acidophiles from acid mine

drainage and industrial bioleaching environments. *Appl Environ Microbiol.* 2004; 70(4):2079-88.

Edwards KJ, Bond PL, Gihring TM, Banfield JF. An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. *Science.* 2000; 287(5459):1796-1799.

Fuhrman, J. A. & Campbell, L. Marine ecology: Microbial microdiversity. *Nature* 393: 410-411. (1998).

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. Genetic diversity in Sargasso Sea bacterioplankton nature. *Nature* 345, 60 - 63 (1990).

Golyshina OV, Pivovarova TA, Karavaiko GI, Kondrateva TF, Moore ER, Abraham WR, Lunsdorf H, Timmis KN, Yakimov MM, Golyshin PN. *Ferroplasma acidiphilum* gen. nov., sp. nov., an acidophilic, autotrophic, ferrous-iron-oxidizing, cell-wall-lacking, mesophilic member of the *Ferroplasmaceae* fam. nov., comprising a distinct lineage of the Archaea. *Int J Syst Evol Microbiol.* 2000; 50 Pt 3:997-1006.

Golyshina OV, Timmis KN. *Ferroplasma* and relatives, recently discovered cell wall-lacking archaea making a living in extremely acid, heavy metal-rich environments. *Environ Microbiol.* 2005; 7(9):1277-1288.

Gonzalez, A.M.; Paranhos, R.; Andrade, L. & Valentin, J.; 2000. Bacterial production in Guanabara Bay (Rio de Janeiro, Brazil) evaluated by <sup>3</sup>H-leucine incorporation. *Braz. Arch. Biol. Technol.* 43: 493-500.

Hallberg KB, Johnson DB. Biodiversity of acidophilic prokaryotes. *Adv Appl Microbiol.* 2001;49:37-84.

Heck Jr., K.L.; van Belle, G. & Simberloff, D.; 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology*, 56: 1459-1461.

Hurlbert, S. H. The nonconcept of species diversity: a critique and alternative parameters.

Ecology 52:577-586. (1971).

Kinnunen PH, Puhakka JA. High-rate ferric sulfate generation by a *Leptospirillum ferriphilum*-dominated biofilm and the role of jarosite in biomass retention in a fluidized-bed reactor.

Biotechnol Bioeng. 2004;85(7):697-705.

Kirchman, D.L.; K'nees, E. & Hodson, R.; 1985. Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. Appl. Environ. Microbiol. 49: 599-607.

Könneke, M., Bernhard, A.E., de la Torre J.R., Walker, C. B., Waterbury & Stahl, D. A. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437, 543-546 (2005).

Kurita, MH. & Pfeiffer, WC (1991) Heavy metals in sediment and biota of Sepetiba Bay, Rio de Janeiro—Brazil. Proc Inter Conf Heavy Metals in the Environment, vol 1, 19–23 September, Edinburgh. CEP Consultants, Edinburgh, pp 133–137.

Lewinsohn, TM & Prado, PI Quantas espécies há no Brasil? Megadiversidade 1(2005), 36 -42.

Martins, JL Amaral-Sobrinho, NMB Lins, U & Rosado, AS. Bacterial diversity in residues of a primary heavy metal producer plant: Characterization of cadmium resistance of Resistance of the isolated bacterial strains. XIX Congress of the Brazilian Society for Microscopy and Microanalysis. Acta Microscopica, Volume 12, Supplement B, September 2003, 269-270.

Molisani, MM; Marins, R.V. Machado, W. Paraquetti, H.H.M. Bidone, E.D. & Lacerda. L.D. Environmental changes in Sepetiba Bay, SE Brazil. Reg Environ Change (2004) 4:17–27.

Okibe N, Gericke M, Hallberg KB, Johnson DB. Enumeration and characterization of acidophilic microorganisms isolated from a pilot plant stirred-tank bioleaching operation. Appl Environ Microbiol. 2003; 69(4):1936-43.

Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA. Microbial ecology and evolution: a



- ribosomal RNA approach. *Annu Rev Microbiol.* 1986; 40:337-65.
- Palleroni NJ. Prokaryotic diversity and the importance of culturing. *Antonie Van Leeuwenhoek.* 1997; 72(1):3-19.
- Peixoto RS, da Costa Coutinho HL, Rumjanek NG, Macrae A, Rosado AS. Use of rpoB and 16S rRNA genes to analyse bacterial diversity of a tropical soil using PCR and DGGE. *Lett Appl Microbiol.* 2002;35(4):316-20.
- Pfeiffer WC, Lacerda LD, Fiszman M, Lima NRW (1985) Metais pesados no pescado da Baía de Sepetiba. *Ciência Cultura* 37:297–302
- Pivovarova TA, Kondrat'eva TF, Batrakov SG, Esipov SE, Sheichenko VI, Bykova SA, Lysenko AM, Karavaiko GI. Phenotypic features of *Ferroplasma acidiphilum* strains Yt and Y-2 *Mikrobiologija.* 2002; 71(6):809-18.
- Rappe, M. S. & Giovannoni, S. J. 2003. The uncultured microbial majority. *Annu. Rev. Microbiol.* 57:369-394.
- Rawlings DE. Heavy metal mining using microbes. *Annu Rev Microbiol.* 2002; 56:65-91.
- Rosado, Alexandre Soares; Duarte, Gabriela Frois; Seldin, Lucy; Elsas, Jan Dirk Van. *Molecular microbial ecology: a minireview* 1997 *Rev. microbiol*;28(3):135-47.
- Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425. (1987).
- Schleper, C., Jurgens, G., & Jonscheit, M. Genomic studies of uncultivated Archaea *Nature Reviews Microbiology* 3, 479-488 (2005)
- Silverman, MP & Lundgren, DG. Studies on the chemoautotrophic iron bacterium *Ferrobacillus ferrooxidans*. I. An improved medium and a harvesting procedure for securing high cell yields. *J Bacteriol.* 1959;77(5):642-7.
- Simberloff, D. Use of rarefaction and related methods in ecology, p. 150-165. In K. L. Dickson,

- J. J. Cairns, and R. J. Livingston (ed.), Biological data in water pollution assessment: quantitative and statistical analyses. American Society for Testing and Materials, Philadelphia, Pa. (1978).
- Somerville, C. C., Knight, I. T., Straube, W. L. & Colwell, R. R. Simple, rapid method for direct isolation of nucleic acids from aquatic environments. *Appl. Environ. Microbiol.* 55:548–554. (1989).
- Torsvik V, Salte K, Sorheim R, Goksoyr J. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. *Appl Environ Microbiol.* 1990 Mar;56(3):776-81.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature.* 2004; 428(6978):37-43.
- Vieira RP, Clementino MM, Cardoso AM, Oliveira DN, Albano RM, Gonzalez AM, Paranhos R, Martins OB. Archaeal communities in a tropical estuarine ecosystem: guanabara bay, Brazil. *Microb Ecol.* 2007a; 54(3):460-468.
- Vieira RP, Gonzalez AM, Cardoso AM, Oliveira DN, Albano RM, Clementino MM, Martins OB, Paranhos R. Relationships between bacterial diversity and environmental variables in a tropical marine environment, Rio de Janeiro. *Environ Microbiol.* 2007b (in press)
- Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol* 173(2), 697-703. (1991).
- Woese CR, Kandler O & Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A.* 1990 Jun;87(12):4576-4569.

Woese, CR. Bacterial evolution. *Microbiol Rev.* 1987; 51(2):221-271.

## **CURRICULUM VITAE**

Nome: WELINGTON INÁCIO DE ALMEIDA

Nascimento: 24 de janeiro de 1979.

Naturalidade: Rio de Janeiro

### Formação Acadêmica:

- Ciências Biológicas, pelo Centro Universitário Augusto Motta, 1998 – 2001.

- Mestrado em Ciências Biológicas (Química Biológica), Departamento de Bioquímica Médica, Instituto de Ciências Biomédicas, Universidade Federal do Rio de Janeiro, 2003.

- Doutorado em Ciências Biológicas (Química Biológica), Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro.

### Orientação de Estudantes de Iniciação Científica:

1. Rebeca Graça Costa
2. Caroline de Oliveira Barreiros

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)