

**UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE QUÍMICA**

**Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)**

**LEONOR MAGALHÃES PERES GALVÃO DE BOTTON**

**Estudos estruturais de proteínas de  
*Xanthomonas axonopodis pv citri*  
por Ressonância Magnética Nuclear**

São Paulo

Data do Depósito na SPG:

03/09/2007

LEONOR MAGALHÃES PERES GALVÃO DE BOTTON

**Estudos estruturais de proteínas de  
*Xanthomonas axonopodis* pv *citri*  
por Ressonância Magnética Nuclear**

*Tese apresentada ao Instituto de Química  
da Universidade de São Paulo para  
obtenção do Título de Doutor em Ciências  
(Bioquímica)*

*Orientador: Prof. Dr. Shaker Chuck Farah*

São Paulo  
2007

Leonor Magalhães Peres Galvão de Botton

Estudos estruturais de proteínas de *Xanthomonas axonopodis* pv *citri* por Ressonância Magnética Nuclear.

*Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutor em*

.....

Aprovado em: \_\_\_\_\_

**Banca Examinadora**

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

*Aos homens da minha vida:*

*aos meus filhos, André e Martim,  
que dentro de mim viveram intensamente  
os fracassos, os sucessos, as tristezas e as alegrias desta tese;*

*ao meu marido, Alexandre,  
por todo o tempo que esta tese lhe roubou.*

## AGRADECIMENTOS

Ao Prof. Dr. Chuck Farah pela orientação científica e por ter exigido sempre o melhor de mim.

À Dra. Thelma Pertinhez que, embora não oficialmente, foi muito mais do que uma co-orientadora. Foi uma Orientadora com “O” maiúsculo.

Ao Prof. Dr. Alberto Spisni pela oportunidade de trabalhar no seu grupo e pelas excelentes discussões científicas.

Aos Profs. Drs. Fábio Almeida e Ana Paula Valente pela oportunidade de colaboração e por terem sido os primeiros a me mostrar a beleza da RMN.

À Profa. Dra. Adelaide Faljoni Alário por me ter recebido de braços abertos quando eu cheguei ao Brasil e por sempre me ter apoiado nas minhas opções profissionais.

Ao Prof. Dr. Ronaldo Bento Quaggio, por ser o “tio científico” maravilhoso que é, sempre pronto para ouvir os seus sobrinhos em todos os momentos.

Ao Prof. Dr. Daniel Cicero pela coleta de alguns dos experimentos de RMN usando o equipamento Bruker Advance® operando a 700 MHz do Departamento de Ciências Químicas e Tecnologia da Universidade de Roma II.

À Ângela Katsuyama pela colaboração na clonagem e triagem de várias proteínas de *Xac*.

Aos Dr. Sérgio Oyama por toda a ajuda na construção do modelo de homologia da proteína ClpS de *Xac*.

Ao Dr. Maurício Sforça pela orientação durante o assinalamento das ressonâncias da proteína ClpS.

Ao Prof. Dr. Nilson Zanchin por ter cedido a proteína apo Sm14-M20(C62V) para os estudos de interação com a proteína ClpS.

Ao Dr. Aalim Weljie por me ensinar os princípios básicos do sistema Linux e dos programas NMRView e NMRPipe.

A Dra. Izaura Toma pela ajuda com o seqüenciamento dos vários clones.

Aos membros do ChuckLab, passados e presentes, por todo o apoio e ajuda nos experimentos, por todas as discussões científicas enriquecedoras e pela amizade: Dra. Áurea D. Souza, Dra. Adriana Paulucci, Dr. Luis Holtzhauten, Dr. Eric Vidsiunas, Ângela Katsuyama, Fernando Corrêa, Paola Capelletti, Diorge Souza, Maxuel Andrade, Pamela Zucker, Cláudia Pereda Rosa, Dra. Cássia Docena, Dra. Myriam Marrquin, Thaís Sousa, D. Cleuza, Alessandra Paiva e Ilda Costa.

Em especial aos meus quatro “anjos da guarda” do ChuckLab sem os quais este trabalho teria sido ainda mais difícil.

Ao Dr. Marcos Alegria por me ter trazido para o laboratório, por todas as dicas de biologia molecular e por ter sido um ombro amigo sempre que precisei.

À Dra. Luci Ribeiro, que sempre “segurou a barra” nos momentos mais difíceis. Tudo o que eu pudesse escrever aqui seria pouco para agradecer tudo o que ela fez por mim durante este doutorado.

À Cristiane Guzzo, por ter sido a melhor aluna de iniciação científica que alguém pode ter, por todas as vezes que foi o meu braço direito e esquerdo na bancada, por ter uma paixão por ciência contagiante e acima de tudo por toda a amizade.

Ao Dr. Roberto Salinas por ler, re-ler, criticar e discutir esta tese como se fosse sua, por ter o dom de sempre me fazer acreditar nas minhas capacidades e principalmente por tornar a RMN tão mais divertida.

A todos os membros dos laboratórios do Bloco Zero do Instituto de Química da USP, em especial à Dra. Elaine Favaro, à Dra. Cristina Juarez, ao Robson Francisco de Souza, à Dra. Rafaela Giorgi, à Dra. Juliana Luz, ao Carlos Francischini, à Juliana Fontanari e ao Alex Monteiro pelo empréstimo de reagentes e equipamentos e pelo ótimo convívio ao longo destes anos.

A todos os membros do laboratório dos Profs. Drs. Ana Paula Valente e Fábio Almeida na UFRJ, em especial ao (quase Dr!) Anderson Pinheiro, à Dra. Cristiane Anobom, à Dra. Fabiana Albernaz, à Dra. Luzineide Tinoco e ao Francisco Gomes Neto, pelo carinho com que me acolheram e por toda a ajuda durante a coleta dos dados de triagem.

A todos os membros do Centro de Biologia Molecular Estrutural do LNLS em especial ao Prof. Dr. Javier Francisco Medrano, ao Prof. Dr. Carlos Ramos, à Dra. Letícia Khater, à Dra. Marina Vanini, à Dra. Patrícia Moura, ao Dr. Sérgio Oyama, à Aline Oliveira e ao Leonardo Trabuco pelas discussões científicas e por tornarem as minhas estadias em Campinas tão mais agradáveis.

A todos os doutorandos e pós-doutorandos envolvidos no projeto SMOLBNet em especial ao Dr. Felipe Chamberg, à Dra. Juliana Luz e à Dra. Andréa Balan pela troca de experiências e pelas conversas noite a dentro no alojamento do LNLS.

Aos colegas e sócios do escritório de MOMSEN, LEONARDOS & CIA, por todo o apoio e compreensão durante os mais de dois anos em que conciliei as minhas obrigações perante escritório com as minhas obrigações enquanto estudante de doutorado.

A toda a minha família que, apesar de não entender o que eu fazia (ou porque eu o fazia!), sempre me apoiou e respeitou as minhas escolhas.

Ao meu marido, Alexandre, por todos os fins de semana passados no laboratório, por ter ficado ao meu lado durante todas as crises, por ter sido “pai solteiro” dos nossos filhos nos últimos meses e, acima de tudo, por nunca me ter deixado desistir.

Aos meus filhos, André e Martim, cujos sorrisos, abraços, beijos e mimos foram um alento para a minha alma nos momentos mais difíceis.

À minha mãe, Isabel, pelo seu amor incondicional, pelo apoio em todos os momentos e principalmente por ser o exemplo de mulher que é para mim.

Ao meu pai Cáci (*in memoriam*), por me ter adotado desde sempre como a sua filha do coração e por ter sido o grande responsável por um dia eu ter tido a vontade e a coragem de embarcar nesta longa viagem.

Ao meu pai Vasco que, mesmo sem poder envolver-se tanto quanto eu e ele muitas vezes gostaríamos, sempre me apoiou, respeitou as minhas escolhas e vibrou com as minhas conquistas.

Aos meus queridos avós, Maria Antonieta e Virgílio (*in memoriam*), por terem sido pilares essenciais na minha formação. Sem eles, eu não seria o que sou hoje e provavelmente não teria conseguido chegar até aqui.

Ao meu irmão Francisco, por toda a cumplicidade, por sempre me lembrar que a vida vai muito além de livros e estudos e por ter, ainda que a muito custo, cedido o seu urso Tótó ao meu filho André, o que permitiu uma diminuição significativa das noites mal dormidas durante este doutorado.

Aos meus sogros, André e Maria Victória, por me terem acolhido em sua casa no Rio de Janeiro durante os quase dois meses de trabalho experimental na UFRJ e por terem armazenado várias proteínas de *Xac* no seu congelador durante tanto tempo.

A todos os que me ajudaram com os meus filhos no decorrer desta tese, em especial à Lela e à Sila que me ajudaram a cuidar da minha casa e dos meus filhos com todo o carinho e dedicação, permitindo que eu conciliasse os meus papéis de mãe-trabalhadora-estudante durante tanto tempo.

Ao meu grupo de fórum por todo o apoio, pela amizade e pela troca experiências enriquecedora.



A todos os meus amigos, novos e velhos, de agora e de sempre, de longe e de perto, cientistas ou não, que dividiram comigo os bons e os maus momentos.

À Fundação para a Ciência e Tecnologia do Ministério da Ciência, Tecnologia e Ensino Superior de Portugal pela bolsa de estudos concedida.

À FAPESP e ao CNPq pelo suporte financeiro.

A todos aqueles que, direta ou indiretamente, consciente ou inconscientemente, possam ter contribuído para que um dia esta tese existisse.

Muito obrigada!

## O MILAGRE DA VIDA

Quando eu tinha oito anos o meu tio Luís foi diagnosticado com AIDS. Estávamos na década de 80, quando a discriminação e os preconceitos em relação à doença eram grandes e as ferramentas para a combater praticamente inexistentes. O seu sofrimento e de todos à sua volta foi algo que me marcou muito. Lembro-me de me revoltar com a idéia de um organismo tão simples ter o poder de causar tanta destruição. Mas ao mesmo tempo fascinava-me o milagre da vida, onde tudo parecia encaixar perfeitamente, e onde a peça errada no lugar errado eram suficientes para causar o desequilíbrio de todo um sistema. Quando ele finalmente descansou, prometi a mim mesma (mas nunca contei a ninguém!) que, se pudesse, um dia contribuiria para a descoberta da cura da AIDS. Nascia assim a minha vontade de pesquisar este milagre a que chamamos vida.

Mais tarde, no colégio, encontrei o meu primeiro mentor científico, o professor de Biologia Mr. Colm Moore. Foi ele, que com a sua barba ruiva, olhos azuis brilhantes e sotaque irlandês, me contou pela primeira vez sobre as experiências de Mendel e o mundo fascinante da genética. As suas aulas eram absolutamente apaixonantes e foi ele quem mais me estimulou a seguir a carreira de ciências biológicas.

Quando ingressei no curso de graduação em Bioquímica no Imperial College na Universidade de Londres, ainda sonhava em trabalhar com imunologia e o meu velho conhecido HIV. Mas durante o segundo ano, cursando a disciplina de “*Protein Science*”, descobri a beleza das proteínas e como elas eram a resposta, a nível molecular, para tantas questões biológicas. No meu primeiro estágio de iniciação

científica, no laboratório da Profa. Dra. Isabel Sá-Correia no Instituto Superior Técnico em Lisboa, aprendi, através dos mutantes de *Saccharomyces cerevisiae* que estudei, que a ausência de um único gene e da proteína por ele codificada, podia ter efeitos drásticos na vida de um organismo. Com o Prof. Dr. Gianfranco Gilardi, no Imperial College, durante a minha segunda iniciação científica, aprendi muito do pouco que hoje sei sobre estrutura e função de proteínas.

Durante os trabalhos para esta tese de doutorado no laboratório do Prof. Dr. Chuck Farah no Instituto de Química da Universidade de São Paulo solidifiquei os meus conhecimentos sobre proteínas. No laboratório dos Profs. Drs. Ana Paula Valente e Fábio Almeida na Universidade Federal do Rio de Janeiro, e especialmente com a Dra. Thelma Pertinhez no Laboratório Nacional de Luz Síncrotron em Campinas, aprendi o quanto a Ressonância Magnética Nuclear nos pode contar sobre elas.

Hoje, as proteínas, enquanto máquinas moleculares da vida que são, fascinam-me tanto ou mais quanto nos primeiros tempos. Mas foi com os meus filhos, e o seu amor inocente, que aprendi que o verdadeiro milagre da vida transcende genes, proteínas e organismos!

Leonor Galvão de Botton

*“O valor das nossas conquistas é medido  
por aquilo que abrimos mão para conquistá-las.”*

Mahatma Gandhi

## RESUMO

Galvão de Botton, L.M.P. **Estudos estruturais de proteínas de *Xanthomonas axonopodis* pv *citri* por Ressonância Magnética Nuclear**. 2007. 1v (paginação irregular). Tese - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

*Xanthomonas axonopodis* pv *citri* (*Xac*) é uma bactéria fitopatogênica que causa de cancro cítrico em plantações no mundo inteiro. Trinta e cinco proteínas alvo foram selecionadas para estudos de proteômica estrutural a partir do genoma de *Xac*. As proteínas foram clonadas, expressas e testadas usando uma nova metodologia de triagem de proteínas que permite que espectros de RMN 2D  $^{15}\text{N}$ -HSQC sejam coletados antes da purificação da proteína em estudo. Esta abordagem possibilitou determinar quais proteínas alvo melhor se adequavam para estudos estruturais futuros por RMN e/ou cristalografia de raios X de forma rápida e eficaz.

A proteína ClpS de *Xac*, descrita como moduladora da atividade da protease bacteriana ClpAP, foi uma das proteínas selecionadas para estudos estruturais por RMN usando esta metodologia. O assinalamento das ressonâncias da cadeia principal e das cadeias laterais desta proteína usando experimentos de tripla ressonância e dados de dinâmica e de troca H/D forneceram informações sobre a sua estrutura secundária. Um modelo tridimensional foi gerado por modelagem por homologia a partir de um homólogo de *E. coli* e foi validado por acoplamentos dipolares residuais ( $D_{\text{HN}}$ ) obtidos experimentalmente. Todos os dados RMN sugerem que a região N-terminal de ClpS se apresenta desestruturada. O mapeamento por RMN da interação de ClpS com a sua parceira ClpA é também apresentado.

**Palavras-chave:** *Xanthomonas axonopodis* pv *citri*; proteômica estrutural; ressonância magnética nuclear; ClpS.

## ABSTRACT

Galvão de Botton, L.M.P. **Structural studies of *Xanthomonas axonopodis* pv *citri* proteins using Nuclear Magnetic Resonance**. 2007. 1v (irregular pages). PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

*Xanthomonas axonopodis* pv *citri* (*Xac*) is a phytopathogenic bacterium that causes citrus canker around the world. Thirty-five target proteins for structural proteomics studies have been selected from the *Xac* genome. The target proteins were cloned, expressed and tested using a novel screening methodology that allows for 2D <sup>15</sup>N-HSQC NMR spectra to be collected prior to the purification of the target protein. This approach allowed us to determine which target proteins were amenable for future structural studies by NMR and/or X-ray crystallography in a fast and efficient manner.

The ClpS protein, which has been described as a modulator of substrate specificity of the bacterial protease ClpAP, was one of the proteins selected structural studies by NMR using this methodology. Backbone and side-chain assignment derived from 3D triple resonance NMR experiments and dynamic data from hydrogen-deuterium exchange NMR experiments have provided information on the secondary structure elements of this protein. A model for *Xac* ClpS was generated by homology modeling from an *E. coli* homologue and validated using experimentally obtained D<sub>HN</sub> residual dipolar couplings. All NMR data suggest that the N-terminal region of the protein ClpS is highly unstructured. NMR mapping of the interaction of ClpS with its partner protein ClpA is also presented.

**Keywords:** *Xanthomonas axonopodis* pv *citri*; structural proteomics; nuclear magnetic resonance; ClpS.

## LISTA DE FIGURAS

<b>Figura 1.</b> Infecção por cancro cítrico.....	7
<b>Figura 2.</b> Impacto de projetos de proteômica estrutural na biologia atual.....	16
<b>Figura 3.</b> Comparação de espectros 2D $^1\text{H}$ - $^{15}\text{N}$ -HSQC para uma proteína desordenada e uma proteína bem enovelada.....	21
<b>Figura 4.</b> Representação esquemática do vetor pET-3a.....	34
<b>Figura 5.</b> Fluxograma esquematizando o processo de seleção das 35 proteínas alvo de <i>Xac</i> para estudos estruturais.....	45
<b>Figura 6.</b> Marcação seletiva de proteínas heterólogas usando rifampicina.....	51
<b>Figura 7.</b> Espectros 1D de $^1\text{H}$ editados para $^{15}\text{N}$ obtidos para as 19 proteínas solúveis presentes nos lisados bacterianos e marcadas seletivamente com $^{15}\text{N}$ .....	55
<b>Figura 8.</b> Espectros 2D $^1\text{H}$ - $^{15}\text{N}$ -HSQC obtidos para as 19 proteínas solúveis presentes os lisados bacterianos e marcadas seletivamente com $^{15}\text{N}$ .....	56
<b>Figura 9.</b> Espectros 1D de $^1\text{H}$ 1D de $^1\text{H}$ editados para $^{15}\text{N}$ e 2D $^1\text{H}$ - $^{15}\text{N}$ -HSQC relativos às proteínas codificadas pelas ORFs XAC2775 e XAC0862.....	59
<b>Figura 10.</b> Espectros 2D $^1\text{H}$ - $^{15}\text{N}$ -HSQC obtidos para as proteínas codificadas pelas ORFs XAC2000, XAC2396, XAC3873 e XACb0070, marcadas com $^{15}\text{N}$ e purificadas.....	60
<b>Figura 11.</b> Dados estatísticos dos resultados de triagem.....	64
<b>Figura 12.</b> Representação esquemática da arquitetura dos complexos proteolíticos bacterianos dependentes de ATP.....	75
<b>Figura 13.</b> Comparação da arquitetura geral do complexo proteolítico ClpAP de <i>E. coli</i> e do proteasomo 26S de eucariotos.....	76
<b>Figura 14.</b> Estrutura tridimensional da proteína ClpP.....	79

<b>Figura 15.</b> Estrutura tridimensional da ATPase ClpA.....	84
<b>Figura 16.</b> Estrutura tridimensional da proteína ClpS de <i>E. coli</i> .....	89
<b>Figura 17.</b> Alinhamento da seqüência da proteína ClpS de <i>Xac</i> usando o programa Clustal W e localização genômica do gene <i>clps</i> de <i>Xac</i> .....	94
<b>Figura 18.</b> Espectros de RMN 1D de $^1\text{H}$ de proteínas.....	96
<b>Figura 19.</b> Esquematisação de um espectro 2D de RMN.....	97
<b>Figura 20.</b> Esquematisação de espectros COSY e TOCSY.....	99
<b>Figura 21.</b> Esquematisação de um espectro 3D de RMN.....	102
<b>Figura 22.</b> Resumo da estratégia geral utilizada na determinação de estruturas tridimensionais de proteínas por RMN.....	105
<b>Figura 23.</b> Ilustração de tipos de meio líquido cristalino.....	123
<b>Figura 24.</b> Pequenas regiões de espectros $^1\text{H}$ - $^{15}\text{N}$ HSQC da proteína ubiquitina coletados na ausência de desacoplamento de $^1\text{J}_{\text{NH}}$ na dimensão indireta para diferentes níveis de alinhamento molecular.....	124
<b>Figura 25.</b> Expressão e purificação da proteína ClpS e do N-terminal da proteína ClpA .....	150
<b>Figura 26.</b> Espectros de CD da proteína ClpS obtidos em diferentes concentrações de sal.....	151
<b>Figura 27.</b> Espectros de CD da proteína ClpS mediante variação de pH entre 3,0 e 9,0.....	153
<b>Figura 28.</b> Espectros de CD de ClpS obtidos a diferentes temperaturas.....	153
<b>Figura 29.</b> Estudos de desnaturação térmica da proteína ClpS.....	154
<b>Figura 30.</b> Espectro 2D $^{15}\text{N}$ -HSQC da proteína ClpS.....	157
<b>Figura 31.</b> Fatias selecionadas do espectro de HNCA da proteína ClpS.....	162



<b>Figura 32.</b> Fatias intercaladas selecionadas dos espectros de CACB(CO)NH e HNCACB da proteína ClpS.....	163
<b>Figura 33.</b> Espectro 2D $^1\text{H}$ - $^{15}\text{N}$ HSQC da proteína ClpS mostrando o assinalamento dos picos $^1\text{H}$ - $^{15}\text{N}$ da cadeia principal.....	165
<b>Figura 34.</b> Fatias selecionadas do espectro $^1\text{H}$ - $^{15}\text{N}$ TOCSY-HSQC de ClpS ilustrando o assinalamento dos prótons alifáticos dos resíduos 68 a 77.....	167
<b>Figura 35.</b> Regiões selecionadas de fatias de $^{13}\text{C}$ do espectro de HCCH-TOCSY da proteína ClpS.....	167
<b>Figura 36.</b> Espectro de $^1\text{H}$ - $^{13}\text{C}$ HSQC da proteína ClpS da região dos núcleos alifáticos.....	168
<b>Figura 37.</b> “Chemical Shift Index” calculado para a proteína ClpS de <i>Xac</i> .....	171
<b>Figura 38.</b> Comparação da estrutura secundária da proteína ClpS de <i>E.coli</i> com a estrutura secundária predita para a proteína ClpS de <i>Xac</i> .....	171
<b>Figura 39.</b> Constante de acoplamento $^3J_{\text{HNH}\alpha}$ em função do número de resíduo de ClpS.....	174
<b>Figura 40.</b> Predição dos ângulos diedros $\Phi$ e $\Psi$ da proteína ClpS de <i>Xac</i> usando o programa TALOS.....	176
<b>Figura 41.</b> Modelo tridimensional da proteína ClpS de <i>Xac</i> construído por modelagem por homologia com a proteína ClpS de <i>E. coli</i> .....	178
<b>Figura 42.</b> Diagrama de Ramachandran do modelo da proteína ClpS de <i>Xac</i> obtido usando o programa PROCHECK.....	179
<b>Figura 43.</b> Experimentos utilizados para calcular os acoplamentos dipolares residuais $D_{\text{NH}}$ da proteína ClpS de <i>Xac</i> .....	181
<b>Figura 44.</b> Acoplamentos dipolares residuais $D_{\text{NH}}$ da proteína ClpS de <i>Xac</i> em função da seqüência primaria.....	182

<b>Figura 45.</b> Correlação entre os acoplamentos dipolares residuais $D_{HN}$ observados e calculados para a proteína ClpS de <i>Xac</i> .....	184
<b>Figura 46.</b> Espectros $^1H$ - $^{15}N$ HSQC adquiridos após diferentes tempos de troca próton-deutério.....	186
<b>Figura 47.</b> Análise da cinética da troca próton-deutério da proteína ClpS.....	188
<b>Figura 48.</b> Parâmetros de relaxação de $^{15}N$ e fator B cristalográfico da proteína ClpS.....	192
<b>Figura 49.</b> Variações dos deslocamentos químicos dos grupos amídicos da proteína ClpS em função da seqüência protéica após a adição do N-terminal da proteína ClpA agregado e da proteína apo Sm14-M20(C62V) agregada.....	196
<b>Figura 50.</b> Elementos de estrutura secundária da proteína ClpS de <i>Xac</i> .....	200
<b>Figura 51.</b> Mapeamento dos resíduos V30, L31 e H64 no modelo tridimensional da proteína ClpS.....	205
<b>Figura 52.</b> Mapeamento das variações dos deslocamentos químicos compostos dos grupos amídicos ( $\Delta_{1H+15N}$ ) da proteína ClpS no modelo tridimensional da proteína após a adição do N-terminal da proteína ClpA agregado e da proteína apo Sm14-M20(C62V) agregada.....	207

## LISTA DE TABELAS

<b>Tabela 1.</b> Amplificação e clonagem das 35 ORFs de <i>Xac</i> : oligonucleotídeos e sítios de restrição utilizados.....	30
<b>Tabela 2.</b> Oligonucleotídeos utilizados no seqüenciamento das ORFs clonadas.....	35
<b>Tabela 3.</b> Características das 35 proteínas de <i>Xac</i> selecionadas para este projeto e o estágio alcançado por cada uma das 35 proteínas durante o processo de triagem.....	46
<b>Tabela 4.</b> Progresso dos trabalhos envolvendo as proteínas de <i>Xac</i> selecionadas para estudos estruturais utilizando a metodologia de triagem descrita neste trabalho.....	71
<b>Tabela 5.</b> Proteínas adaptadoras de <i>E.coli</i> e suas funções na adaptação da função de proteínas AAA+.....	86
<b>Tabela 6.</b> Deslocamento químico dos núcleos de hidrogênio dos 20 amino ácidos naturais em conformação randômica.....	107
<b>Tabela 7.</b> Experimentos 3D de tripla ressonância utilizados no assinalamento dos núcleos da cadeia principal.....	110
<b>Tabela 8.</b> Deslocamento químico médio dos núcleos de carbono e nitrogênio dos 20 amino ácidos naturais encontrados em proteínas.....	111
<b>Tabela 9.</b> Experimentos 3D de tripla ressonância utilizados no assinalamento dos núcleos das cadeias laterais.....	112
<b>Tabela 10.</b> Oligonucleotídeos utilizados na amplificação da região que codifica o N-terminal da proteína ClpA de <i>Xac</i> .....	132

<b>Tabela 11.</b> Número de pontos da matriz de dados dos experimentos 3D de tripla ressonância.....	138
<b>Tabela 12.</b> Número de pontos da matriz de dados dos experimentos 3D de dupla ressonância.....	139
<b>Tabela 13.</b> Interação da proteína ClpS com o N-terminal da proteína ClpA.....	142
<b>Tabela 14.</b> Resumo das correlações observadas nos experimentos 3D de tripla ressonância utilizados para assinalar os núcleos da cadeia principal da proteína ClpS.....	159
<b>Tabela 15.</b> Estatística da atribuição dos núcleos da cadeia principal da proteína..	165

## LISTA DE EQUAÇÕES

Equação 1.....	115
Equação 2.....	134
Equação 3.....	136
Equação 4.....	137
Equação 5.....	137
Equação 6.....	143
Equação 7.....	146
Equação 8.....	173
Equação 9.....	183
Equação 10.....	187
Equação 11.....	187
Equação 12.....	190
Equação 13.....	190

## LISTA DE ABREVIATURAS E SIGLAS

1D	Espectros de RMN unidimensionais
2D	Espectros de RMN bidimensionais
3D	Espectros de RMN tridimensionais
4D	Espectros de RMN quadridimensional
aa	Amino ácido
AAA+	ATPases associadas com várias atividades celulares
ADP	Adenosina 5'-difosfato
ATP	Adenosina 5'-trifosfato
ATPase	Adenosina trifosfatase
BMRB	“Biological Magnetic Resonance Bank”
CD	Dicroísmo circular
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COSY	“Correlation spectroscopy”
CSI	“Chemical shift index”
D <sub>2</sub> O	Água deuterada
DNA	Ácido desoxiribonucléico
DO	Densidade ótica
DOSY	“Diffusion Ordered Spectroscopy”
DTT	Dilitreititol
EDTA	Ácido etileno diamino tetracético
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
GFP	“Green Fluorescent Protein”
HSQC	“Heteronuclear Single Quantum Coherence”
IPTG	Isopropil-β-D-1-tiogalactopiranosídeo
kb	Quilobases
kDa	Quilodaltons
2xTY	Meio rico de cultura
LNLS	Laboratório Nacional de Luz Síncrotron
M9	Meio mínimo de cultura
MAD	“Multiple wavelegth Anomalous Dispersion”
NCBI	“National Center for Biotechnology Information”
NOE	“Nuclear Overhauser Effect”

NOESY	“Nuclear Overhauser Effect Spectroscopy“
ONSA	“Organization for Nucleotide Sequencing and Analysis“
ORF	“Open reading frame“
pb	Pares de bases
PCR	Reação de polimerase em cadeia
PDB	“Protein Data Bank“
pET	“Plasmid for Expression by T7 RNA polymerase“
pH	Potencial hidrogeniônico
PMSF	Fluoreto de fenilmetilsulfonila
ppm	Partes por milhão
pv	Patovar
RDC	Acoplamento dipolar residual
rf	Radiofrequência
RMN	Ressonância magnética nuclear
rms	Raiz quadrada da média dos quadrados
RMSD	Raiz quadrada do quadrado do desvio padrão médio
rpm	Rotações por minuto
SDS	Dodecilsulfato de sódio
SDS-PAGE	Eletroforese em gel de SDS-poliacrilamida
SMOLBNet	“Structural Molecular Biology Network“
TOCSY	“Total Correlation Spectroscopy“
Tris	Tris-(hidroximetil)aminometano
TROSY	“Transverse Relaxation Optimized Spectroscopy“
<i>Xac</i>	<i>Xanthomonas axonopodis</i> pv. <i>citri</i>

# SUMÁRIO

<b>CAPÍTULO 1. INTRODUÇÃO GERAL</b> .....	01
<b>CAPÍTULO 2. TRIAGEM DE ALVOS DE PROTEÔMICA ESTRUTURAL DE <i>Xanthomonas axonopodis</i> pv <i>citri</i> USANDO RMN</b> .....	05
<b>2.1. INTRODUÇÃO</b> .....	06
2.1.1. <i>Xanthomonas axonopodis</i> pv <i>citri</i> : a causa do cancro cítrico.....	06
2.1.2. Proteômica estrutural.....	10
2.1.3. O papel da RMN na era da proteômica estrutural.....	17
2.1.4. Triagem de proteínas alvo para estudos de proteômica estrutural....	22
<b>2.2. OBJETIVOS</b> .....	28
<b>2.3. MATERIAIS E MÉTODOS</b> .....	29
2.3.1. Seleção das proteínas alvo de <i>Xac</i> .....	29
2.3.2. Amplificação e Clonagem das ORFs alvo.....	29
2.3.2.1. Amplificação por PCR das ORFs alvo.....	29
2.3.2.2. Clonagem das ORFs alvo no vetor de expressão pET-3a.....	33
2.3.2.3. Seqüenciamento das ORFs alvo clonadas em pET-3a.....	34
2.3.3. Teste de expressão das proteínas alvo.....	35
2.3.4. Teste de solubilidade das proteínas alvo.....	36
2.3.5. Marcação seletiva das proteínas alvo com <sup>15</sup> N.....	36
2.3.6. Triagem das proteínas alvo marcadas com <sup>15</sup> N por RMN.....	37
2.3.7. Produção e purificação das proteínas XAC2000, XAC3873, XAC2396, XAC0862 e XACb0070.....	38
2.3.7.1. Expressão das proteínas em meio mínimo M9 contendo <sup>15</sup> N.....	38
2.3.7.2. Purificação da proteína XAC2000.....	39
2.3.7.3. Purificação da proteína XAC3873.....	40
2.3.7.4. Purificação da proteína XAC2396.....	41
2.3.7.5. Purificação das proteínas XAC0862 e XACb0070.....	42
<b>2.4. RESULTADOS</b> .....	43
2.4.1. Seleção das proteínas alvo.....	43
2.4.2. Amplificação e clonagem das ORFs alvo.....	50
2.4.3. Testes de expressão e solubilidade das proteínas alvo.....	50



2.4.4. Marcação seletiva com <sup>15</sup> N das proteínas alvo.....	50
2.4.5. Triagem por RMN das proteínas alvo.....	53
<b>2.5. DISCUSSÃO.....</b>	<b>62</b>

**CAPÍTULO 3. ESTUDOS ESTRUTURAIS POR RMN DA PROTEÍNA ClpS DE *Xanthomonas axonopodis* pv *citri*.....72**

<b>3.1. INTRODUÇÃO.....</b>	<b>73</b>
3.1.1. A proteína ClpS e a proteólise bacteriana.....	73
3.1.1.1. Complexos proteolíticos bacterianos.....	73
3.1.1.2. A protease ClpP.....	77
3.1.1.3. A ATPase ClpA.....	81
3.1.1.4. A proteína ClpS: uma proteína adaptadora.....	85
3.1.1.5. A proteína ClpS de <i>Xac</i> .....	92
3.1.2. A RMN na caracterização estrutural de proteínas.....	95
3.1.2.1. Assinalamento seqüencial.....	106
3.1.2.2. Identificação dos elementos de estrutura secundária.....	113
3.1.2.3. Obtenção de restrições para o cálculo de estrutura.....	117
3.1.2.4. Cálculo e refinamento da estrutura tridimensional.....	119
3.1.2.5. Análise de acoplamentos dipolares residuais.....	121
<b>3.2. OBJETIVOS.....</b>	<b>128</b>
<b>3.3. MATERIAIS E MÉTODOS.....</b>	<b>130</b>
3.3.1. Expressão e purificação das proteínas em estudo.....	130
3.3.1.1. Expressão da proteína ClpS de <i>Xac</i> .....	130
3.3.1.2. Purificação da proteína ClpS de <i>Xac</i> .....	131
3.3.1.3. Amplificação e clonagem de um fragmento do N-terminal da proteína CpA de <i>Xac</i> .....	132
3.3.1.4. Expressão e purificação do N-terminal da proteína ClpA de <i>Xac</i> .....	132
3.3.2. Experimentos de Dicroísmo Circular (CD).....	133
3.3.2.1. Variação do pH.....	134
3.3.2.2. Variação da força iônica.....	134
3.3.2.3. Desnaturação térmica.....	134
3.3.3. Experimentos de RMN.....	135

3.3.3.1. Preparo das amostras e otimização das condições de RMN.....	135
3.3.3.2. Cálculo do raio hidrodinâmico: experimento de difusão (DOSY).....	136
3.3.3.3. Experimentos tridimensionais de tripla ressonância.....	137
3.3.3.4. Experimentos de dupla ressonância.....	138
3.3.3.5. Cálculo dos acoplamentos dipolares residuais $^1\text{H}$ - $^{15}\text{N}$ .....	139
3.3.3.6. Experimento de troca próton-deutério.....	140
3.3.3.7. Experimentos de relaxação.....	141
3.3.3.8. Experimentos de interação da proteína ClpS de <i>Xac</i> com um o N-terminal da proteína ClpA de <i>Xac</i> .....	142
3.3.3.9. Experimentos de interação da proteína ClpS de <i>Xac</i> com a proteína apo Sm14-M20(C62V).....	143
3.3.4. Análise dos experimentos de RMN.....	144
3.3.4.1. Cálculo do “Chemical Shift Index” (CSI).....	144
3.3.4.2. Predição dos ângulos $\Phi$ e $\Psi$ usando o programa TALOS...	144
3.3.4.3. Análise dos acoplamentos dipolares residuais $^1\text{H}$ - $^{15}\text{N}$ usando o programa PALES.....	145
3.3.4.4. Cálculo dos fatores de proteção de troca dos prótons amidicos.....	145
3.3.5. Modelagem molecular.....	146
3.3.5.1. Recursos computacionais.....	146
3.3.5.2. Predição de estrutura secundária.....	146
3.3.5.3. Busca de estruturas de ClpS de <i>E. coli</i> no PDB e alinhamento de seqüências.....	147
3.3.5.4. Construção do modelo por homologia.....	147
3.3.5.5. Avaliação do modelo.....	148
<b>3.4. RESULTADOS</b> .....	<b>149</b>
3.4.1. Expressão e purificação da proteína em estudo.....	149
3.4.2. Experimentos de Dicroísmo Circular (CD).....	151
3.4.3. Estudos estruturais da proteína ClpS de <i>Xac</i> por RMN.....	155
3.4.3.1. Otimização das condições de RMN.....	155
3.4.3.2. Cálculo do raio hidrodinâmico.....	157
3.4.3.3. Assinalamento da proteína ClpS.....	158

3.4.3.3.1. Assinalamento da cadeia principal.....	159
3.4.3.3.2. Assinalamento das cadeias laterais.....	166
3.4.3.4. Identificação dos elementos de estrutura secundária.....	170
3.4.3.5. Validação do modelo estrutural da proteína ClpS de <i>Xac</i> por acoplamentos dipolares residuais (RDC).....	177
3.4.3.5.1. Cálculo do modelo estrutural da proteína ClpS de <i>Xac</i> usando modelagem por homologia.....	177
3.4.3.5.2. Validação do modelo da proteína ClpS de <i>Xac</i> por acoplamentos dipolares residuais (RDC).....	180
3.4.4. Estudos de dinâmica da proteína ClpS de <i>Xac</i> .....	185
3.4.4.1. Experimento de troca-próton deutério.....	185
3.4.4.2. Experimentos de relaxação.....	189
3.4.5. Estudos de interação da proteína ClpS de <i>Xac</i> com outras proteínas: resultados preliminares.....	193
<b>3.5. DISCUSSÃO.....</b>	<b>199</b>
3.5.1. Estudos estruturais da proteína ClpS de <i>Xac</i> por RMN.....	199
3.5.2. Estudos de dinâmica da proteína ClpS de <i>Xac</i> .....	204
3.5.3. Estudos de interação da proteína ClpS de <i>Xac</i> com outras proteínas.....	206
<b>CAPÍTULO 4. CONCLUSÕES.....</b>	<b>212</b>
<b>REFERÊNCIAS.....</b>	<b>215</b>
<b>APÊNDICE I.....</b>	<b>233</b>
<b>APÊNDICE II.....</b>	<b>250</b>
<b>LISTA DE ANEXOS.....</b>	<b>253</b>

**CAPÍTULO 1**  
**INTRODUÇÃO GERAL**

O projeto inicial desta tese nasceu no âmbito da rede SMOLBNet (“Structural Molecular Biology Network”), ambicioso projeto científico na área de biologia estrutural lançado pela FAPESP (Fundação de Amparo à pesquisa do Estado de São Paulo) em Julho de 2001. Esta rede, formada por dezesseis laboratórios do estado de São Paulo, entre os quais o laboratório do Prof. Dr. Shaker Chuck Farah onde foram realizados os trabalhos desta tese, tem como objetivo aumentar a participação brasileira na área de biologia estrutural, em especial no que diz respeito às proteínas codificadas pelos genomas já seqüenciados por cientistas brasileiros através da rede ONSA (Organization for Nucleotide Sequencing and Analysis), também financiada pela FAPESP.

Neste contexto, o interesse particular do nosso laboratório é o estudo das proteínas do fitopatógeno *Xanthomonas axonopodis* pv *citri* (*Xac*), agente causador do cancro cítrico. Desde o término do seqüenciamento do seu genoma, em 2002 (DA SILVA *et al.*), o nosso laboratório tem vindo a realizar diversos estudos funcionais e estruturais das proteínas de *Xac*.

Mais especificamente, para este projeto de doutorado foram selecionadas 35 proteínas alvo de *Xac* para estudos de proteômica estrutural, cujas novas estruturas tridimensionais pudessem vir a apontar possíveis funções biológicas para proteínas conservadas com funções desconhecidas e/ou revelar novos padrões de enovelamento (“folds”) de proteínas.

Na etapa inicial do projeto, 35 proteínas alvo de *Xac* com funções e estruturas desconhecidas foram selecionadas *in silico*. Uma vez que este tipo de seleção de alvos para estudos estruturais não permite uma estimativa precisa dos níveis de expressão, solubilidade e estabilidade conformacional da proteína a ser estudada,

nós nos propusemos a desenvolver uma nova metodologia para triagem de proteínas alvo para estudos estruturais.

Em colaboração com os Profs. Drs. Ana Paula Valente e Fábio Almeida da Universidade Federal do Rio de Janeiro, desenvolvemos uma nova metodologia para triagem de proteínas para estudos estruturais que permite avaliar os níveis de solubilidade, estabilidade e grau de enovelamento de uma dada proteína alvo, antes mesmo da purificação da mesma (GALVAO-BOTTON *et al.*, 2003). Esta abordagem possibilitou determinar quais as proteínas alvo que melhor se adequavam para estudos estruturais de uma maneira rápida e eficaz possibilitando assim a priorização de alvos de *Xac* para estudos estruturais futuros por RMN e/ou cristalografia de raios-X.

A triagem efetuada revelou várias proteínas como boas candidatas para estudos estruturais (i.e. as proteína apresentavam bons níveis de expressão, solubilidade e estabilidade). No entanto, não caberia no âmbito do meu trabalho de doutorado prosseguir com todas elas para as etapas posteriores de purificação e caracterização estrutural. Assim, os estudos estruturais foram focalizados na proteína XAC2000/ClpS de *Xac* enquanto as restantes proteínas estão sendo ou serão analisadas por outros membros do laboratório. A proteína XAC2000/ClpS foi então amplamente caracterizada por RMN em colaboração com a Dra. Thelma Pertinhez do Laboratório Nacional de Luz Síncrotron (LNLS) e diversos estudos estruturais, de dinâmica e de interação desta proteína são apresentados neste trabalho.

Dada a vasta gama de técnicas utilizadas e a disparidade dos objetivos das várias etapas dos trabalhos, esta tese foi dividida em quatro capítulos para facilitar a sua leitura. Neste primeiro capítulo, é feita uma breve introdução geral ao tema

desta tese. O Capítulo 2 descreve a triagem das 35 proteínas alvo de *Xac* para estudos de proteômica estrutural usando RMN. No capítulo 3, é apresentada a caracterização estrutural por RMN da proteína XAC2000/ClpS de *Xac*. Os Capítulos 2 e 3, que formam o cerne deste trabalho contêm, cada um, uma respectiva seção de Introdução, Materiais e Métodos, Resultados e Discussão. Finalmente, no Capítulo 4 são apresentadas as conclusões finais deste trabalho.

## **CAPÍTULO 2**

### **TRIAGEM DE ALVOS DE PROTEÔMICA ESTRUTURAL DE *Xanthomonas axonopodis* pv *citri* USANDO RMN**



## 2.1. INTRODUÇÃO

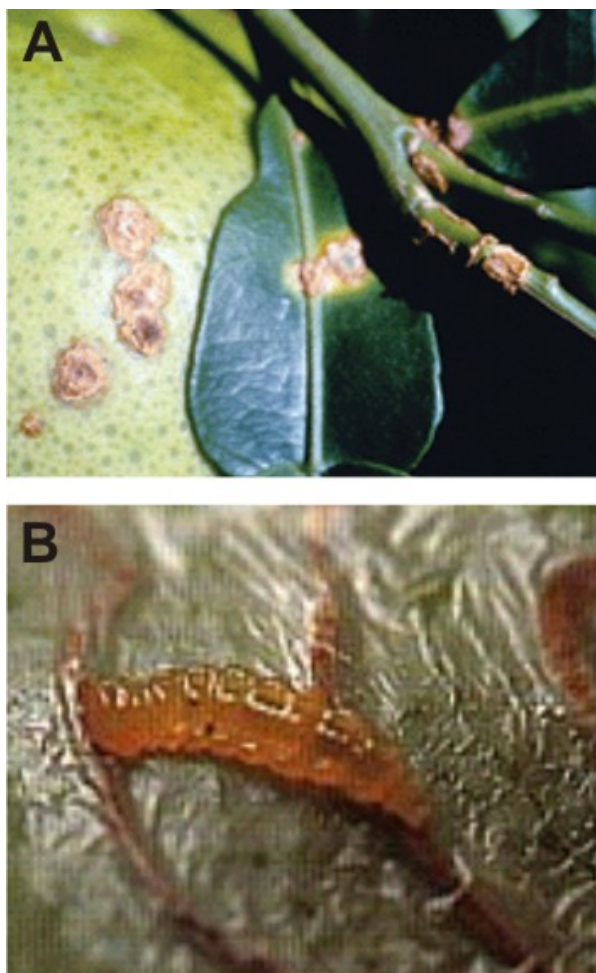
### 2.1.1. *Xanthomonas axonopodis* pv *citri*: a causa do cancro cítrico

O Brasil produz aproximadamente um terço dos frutos cítricos mundiais e responde atualmente por nada menos que 80% do suco de laranja concentrado negociado mundialmente (BEVAN, 2000). Com uma área cultivada de 820 mil hectares, os frutos cítricos (*Citrus spp*) representam 49% de toda a produção brasileira de frutas (ESTANISLAU *et al.*, 2001). Apesar do grande potencial produtivo, a citricultura brasileira é alvo constante de inúmeras pragas e doenças que são capazes de causar danos irreversíveis na planta e na fruta que resultam na queda da produtividade e na qualidade dos frutos. Dentre as principais doenças com grande impacto econômico para a citricultura brasileira destacam-se a tristeza dos citros (BOVÉ e AYRES, 2007) a morte súbita dos citros (RÓMAN *et al.*, 2004; BOVÉ e AYRES, 2007), a clorose variegada dos citros (CARVALHO *et al.*, 1995; BOVÉ e AYRES, 2007) e o cancro cítrico (BRUNINGS e GABRIEL, 2003).

O cancro cítrico, um dos mais graves problemas fitossanitários da citricultura atual, é uma doença causada pela bactéria *Xanthomonas axonopodis* pv *citri* (*Xac*) que provoca lesões nos frutos tornando-os inviáveis para consumo e comercialização (HAYWARD, 1993; BRUNINGS e GABRIEL, 2003). *Xac* é uma bactéria Gram-negativa, baciliforme de respiração aeróbica pertencente ao gênero *Xanthomonas* do qual fazem parte um vasto número de fitopatógenos que atacam uma grande variedade de plantas. Infecções causadas por espécies de *Xanthomonas* ocorrem em pelo menos 68 famílias de plantas monocotiledôneas e em mais de 240 gêneros. A gama de hospedeiros inclui várias espécies agrícolas economicamente relevantes como os frutos cítricos (*Xac*), o arroz (*X. oryzae* pv.

*oryzae*), o feijão (*X. axonopodis* pv. *phadeoli*) e os cereais e gramíneas (*X. translucens*) (HAYWARD, 1993).

*Xac* é altamente contagiosa e resistente e não necessita de um vetor específico de transmissão, podendo ser facilmente transferida entre vários cultivares cítricos através do vento, da chuva ou do próprio homem, penetrando a planta através de aberturas naturais nos tecidos. Todos os tecidos da planta com exceção da raiz são susceptíveis de contaminação mas tecidos jovens em crescimento ativo são alvos mais sensíveis (Figura 1).



**Figura 1.** Infecção por cancro cítrico.

A. Lesões características do cancro cítrico (BROWN, 2001). B. Larva mineradora dos citros.

(Fonte: <http://www.fundecitrus.com.br>, 2001).

Para que penetre em folhas, frutos e ramos mais velhos é preciso haver ferimentos, causados normalmente por material de colheita, pelo vento e, principalmente, por *Phyllocnistis citrella* ou larva mineradora dos citros cujas lesões que provoca na planta contribuem grandemente para a entrada da bactéria (BRUNINGS e GABRIEL, 2003) (Figura 1B). Os primeiros sintomas aparecem uma semana a dois meses após a infecção como lesões nas folhas, frutos e ramos. As lesões são pequenas, circulares e de cor amarela. Com o tempo as lesões tornam-se maiores e salientes, podendo se expandir até 9-10mm de diâmetro, de textura mais grossa e apresentam cor castanha (Figura 1A). Estas lesões formam o inóculo para futuras contaminações em outros tecidos ou plantas e provocam na maioria dos casos a queda precoce dos frutos, inviabilizando o seu consumo (STALL e CIVEROLO, 1993; BRUNINGS e GABRIEL, 2003).

A mais antiga constatação do cancro cítrico foi feita em folhas de *Citrus medica* coletadas na Índia entre 1827-1831. No entanto, a doença foi descrita pela primeira vez em 1915 quando esta foi introduzida nos Estados Unidos (FAWCETT e JENKINS, 1993).

Desde então esta a bactéria causa o cancro cítrico em plantações do mundo inteiro, especialmente nos trópicos onde as condições climáticas de alta pluviosidade e temperaturas elevadas favorecem a sua proliferação (STALL e CIVEROLO, 1993). No Brasil, sua primeira constatação ocorreu no município paulista de Presidente Prudente em 1957 (ROSSETTI, 1977) e espalhou-se rapidamente na chamada região nobre da citricultura que inclui grande parte do estado de São Paulo e o sul do estado de Minas Gerais (FERCHTNBERGER *et al.*, 1997). Posteriormente tornou-se endêmico no Mato Grosso, Rio Grande do Sul, Santa Catarina e, mais recentemente em Rondônia (NASCIMENTO *et al.*, 2003).

A indústria cítrica da Flórida nos Estados Unidos, a segunda maior do mundo depois do Brasil, está ameaçada por uma epidemia de *Xac* espalhada pelo tornado de 1996 (BEVAN, 2000). Os programas de erradicação da doença nos EUA já custaram mais de US\$ 200 milhões de dólares e estima-se que US\$ 100 milhões sejam desperdiçados por ano em fruta estragada (BROWN, 2001).

No estado de São Paulo, o número de árvores cítricas afetadas pela doença tem crescido exponencialmente na última década. Em 1992, o número de árvores contaminadas era inferior a 1000 enquanto que em 1999 foram registrados 300.000 novos casos da doença. Em 2006, foram erradicadas quase 2 milhões de plantas devido à ocorrência de cancro cítrico (Fonte: [http://www.fundecitrus.com.br/est\\_cancro06\\_br.html](http://www.fundecitrus.com.br/est_cancro06_br.html), 2007).

O principal problema do cancro cítrico, é que uma vez identificados os sintomas iniciais, não só as árvores contaminadas, bem como todas as árvores num raio de 30m, precisam ser destruídas. Assim, atualmente a bactéria afeta mais de 30% dos laranjais brasileiros e causa prejuízos anuais de aproximadamente R\$ 300 milhões (Fonte: <http://www.iapar.br>, 2004).

Dado o impacto econômico das doenças causadas por *Xac*, bem como a possibilidade desta bactéria ser usada como modelo biológico para estudos de outras doenças causadas por membros do gênero *Xanthomonas*, esta bactéria tem sido alvo de grande interesse científico. O genoma de *Xac* foi recentemente seqüenciado pela ONSA, um consórcio público de laboratórios brasileiros (DA SILVA *et al.*, 2002).

O genoma de *Xac* apresenta um único cromossomo circular de 5.2 Mb e dois plasmídeos adicionais de 34 e 65 Kb. A anotação do genoma revelou 4313 ORFs (“Open Reading Frames”) putativas. Aproximadamente metade destas ORFs teve a

sua função putativa determinada baseada na sua homologia com outros genes da base de dados do NCBI (“National Center for Biotechnology Information”). As restantes ORFs, ou não apresentaram homologia com nenhuma seqüência conhecida, ou apresentaram homologia com seqüências que codificam proteínas de função desconhecida (proteínas hipotéticas conservadas e ORFs sem categoria definida). Espera-se que os resultados gerados pelo seqüenciamento e pela análise funcional deste genoma forneçam dados de extrema importância para o entendimento deste fitopatógeno a nível molecular e para a erradicação do cancro cítrico.

Além do genoma de *Xac* mesmo consórcio seqüenciou também o genoma de *Xanthomonas campestris* pv *campestris* (DA SILVA *et al.*, 2002) sendo que estes foram os primeiros genomas do gênero *Xanthomonas* a serem seqüenciados. Desde então vários genomas deste gênero têm vindo a ser seqüenciados tais como *Xanthomonas campestris* pv. *vesicatoria* (THIEME *et al.*, 2005) e *Xanthomonas oryzae* pv. *oryzae* (OCHIAI *et al.*, 2005).

### **2.1.2. Proteômica estrutural**

O seqüenciamento de genomas é uma nova e poderosa ferramenta que vem sendo aplicada a uma grande variedade de organismos. Desde a publicação do primeiro genoma completo, o de *Haemophilus influenzae* em 1995 (FLEISCHMANN *et al.*, 1995), o crescimento do número de genomas completos seqüenciados tem sido exponencial. Até à data, 1800 genomas virais, 480 genomas bacteriais, 40 genomas de arquea e 27 genomas de eucariotos foram seqüenciados (Fonte: <http://www.ncbi.nlm.nih.gov/>, 2007). Entre estes encontram-se importantes patógenos de plantas e animais, o genoma humano e outros organismos

freqüentemente utilizados em estudos científicos como modelos biológicos. Além das bactérias do gênero *Xanthomonas*, outros fitopatógenos foram também recentemente seqüenciados incluindo *Xyllela fastidiosa* (SIMPSON *et al.*, 2000), *Agrobacterium tumefaciens* (GOODNER *et al.*, 2001), *Pseudomonas syringae* pv. *tomato* (BUELL *et al.*, 2003); *Pseudomonas syringae* pv. *syringae* (FEIL *et al.*, 2005), *Pseudomonas syringae* pv. *phaseolicola* (JOARDAR *et al.*, 2005), *Ralstonia solanacearum* (SALANOUBAT *et al.*, 2002) e *Erwinia carotovora* (BELL *et al.*, 2004 - resultados não publicados).

A seqüência completa do genoma de um organismo é um passo fundamental para o entendimento deste organismo a nível molecular. No entanto, esta informação é apenas a primeira peça-chave de um grande quebra-cabeça. O verdadeiro desafio atual está em entender as funções dos genes seqüenciados e das proteínas por eles codificadas numa escala genômica. Para tal, vários programas de genômica funcional e proteômica decorrem mundialmente numa tentativa de decifrar as bases moleculares da vida bem como as características que tornam cada organismo já seqüenciado num ser único.

A genômica funcional tem como objetivo principal associar funções a genes numa escala genômica (PHIZICKY *et al.*, 2003; TYERS e MANN, 2003). Novos métodos que analisam vários genes em uma escala genômica incluem a mutação de todo o genoma para testar o efeito da ausência de cada gene no organismo em estudo (GLAEVER *et al.*, 2002) e “microarrays” de DNA, SAGE (“Serial Analysis of Gene Expression”) e projetos EST (“Expressed Sequence Tags”) que nos fornecem informações sobre os níveis de expressão de cada gene nas mais variadas condições biológicas (SCHOOLNIK, 2002).

A proteômica por seu turno, visa o estudo de todas as proteínas codificadas por um genoma. O termo proteômica foi originalmente usado para descrever a análise e identificação de perfis de expressão protéica numa escala genômica utilizando eletroforese em gel de duas dimensões (WILKINS *et al.*, 1996). Atualmente o termo engloba todo o tipo de análises de pós-genômica de proteínas realizadas em larga-escala, como a análise do proteoma de uma célula por eletroforese em gel de duas dimensões ou cromatografia líquida acoplada a espectrometria de massa (GYGI e AEBERSOLD, 2000; DUTT e LEE, 2000; TYERS e MANN, 2003), “chips” e “microarrays” de proteínas (ZHU *et al.*, 2001; PHIZICKY *et al.*, 2003; TYERS e MANN, 2003), triagens para identificar interações proteína-proteína usando ensaios de duplo-híbrido (UETZ *et al.*, 2000; PHIZICKY *et al.*, 2003), TAP-TAG (“Tandem Affinity Purification Tag”) ou imunoprecipitação acoplada a espectrometria de massa (RIGAUT *et al.*, 1999; BÜRCKSTÜMMER *et al.*, 2006) e determinação da localização celular por fusões GFP (PHIZICKY *et al.*, 2003).

Todos estes métodos se adequam a estudos em larga-escala provendo o pesquisador com uma quantidade formidável de dados sobre as proteínas codificadas pelos vários genomas já seqüenciados. No entanto, o entendimento da função de uma proteína a nível molecular requer informação estrutural detalhada das proteínas codificadas pelo genoma em estudo. Surge assim a genômica ou proteômica estrutural que tem como objetivo a obtenção de modelos tridimensionais de proteínas em uma escala genômica (MITTL e GRÜTTER, 2001; STEVENS, 2000; BRENNER, 2001; O'TOOLE *et al.*, 2003).

O objetivo final da proteômica estrutural é fornecer uma descrição tridimensional de todas as proteínas existentes (MITTL e GRÜTTER, 2001; MONTELIONE, 2001; VITKUP *et al.*, 2001). Uma vez que pelas técnicas atuais é praticamente inviável

determinar a estrutura tridimensional de todas as proteínas existentes, muitas destas estruturas serão obtidas por modelagem por homologia usando as estruturas de um grupo de proteínas alvo determinadas experimentalmente. Assim, a seleção de proteínas alvo é talvez o passo estratégico mais importante de um projeto de proteômica estrutural e é crucial para o sucesso do projeto em relação aos seus objetivos específicos. De um modo geral a seleção de alvos para estudos de proteômica estrutural é normalmente guiada por um de dois critérios: preenchimento de espaço conformacional ou elucidação de função biológica (BRENNER, 2000; MITTL e GRÜTTER, 2001).

No caso de uma abordagem conformacional, visa-se descrever um novo padrão de enovelamento (“fold”) protéico nunca antes determinado contribuindo assim para aumento do conhecimento do chamado espaço conformacional (“fold space”) i.e. o número total de conformações possíveis de proteínas existentes na natureza, que está estimado entre 1,000 a 10,000 (BURLEY, 2000). Neste tipo de abordagem, a função e relevância biológica dos alvos selecionados são de menor importância pois o foco é selecionar seqüências que possam apresentar novos enovelamentos e que portanto não possuam homólogos cujas estruturas tridimensionais já tenham sido determinadas.

O racional por trás desta estratégia baseia-se na teoria de que determinando uma estrutura tridimensional representativa para cada família de proteínas existente, será possível no futuro descrever a estrutura tridimensional de qualquer proteína por modelagem por homologia (“homology modelling”) e deste modo contribuir para aumentar ainda mais o conhecimento sobre um dado organismo ou família de proteínas. De um modo geral, quanto maior a homologia entre duas seqüências protéicas melhor será a qualidade do modelo gerado (VITKUP *et al.*, 2001). Assim,



estima-se que para cada nova estrutura tridimensional depositada, aproximadamente 100 seqüências protéicas sem conhecimento estrutural prévio possam ter um modelo tridimensional de qualidade razoável construído por modelagem por homologia (BAKER e SALI, 2001). Desta forma seriam necessárias aproximadamente 15,000 a 20,000 proteínas alvo bem selecionadas cujas estruturas tridimensionais determinadas experimentalmente permitiriam construir modelos atômicos da grande maioria das proteínas existentes possibilitando a cobertura de quase todo o espaço conformacional (BRENNER, 2000; MONTELLIONE, 2001; VITKUP *et al.*, 2001; O'TOOLE *et al.*, 2004).

Exemplos do emprego desta abordagem incluem os estudos de proteômica estrutural dos genomas de *Mycoplasma genitalium* e *Mycoplasma pneumoniae* desenvolvidos pelo “Berkeley Structural Genomics Center” (GREGORIEV e CHOI, 2002; KIM *et al.*, 2003; KIM *et al.*, 2005). Nestes estudos, as proteínas alvo foram selecionadas com base na probabilidade destes alvos poderem potencialmente fornecer novas informações estruturais e/ou funcionais. Os resultados obtidos aumentaram substancialmente a cobertura do “fold space” de *M. pneumoniae* e *M. genitalium* sendo que aproximadamente metade das estruturas protéicas de seqüências “únicas” revelaram novos enovelamentos (KIM *et al.*, 2005).

No caso de uma abordagem biológica busca-se um maior entendimento sobre relação estrutura-função de todas as proteínas de um determinado organismo ou de uma família de proteínas muitas vezes associada com processos biológicos importantes do organismo em estudo (ex. proteínas associadas com patogenezidade). Exemplos recentes deste tipo de abordagem incluem uma vasta gama de estudos tais como a caracterização estrutural de todas as proteínas codificadas pelo genoma do coronavírus causador da doença SARS (Síndrome

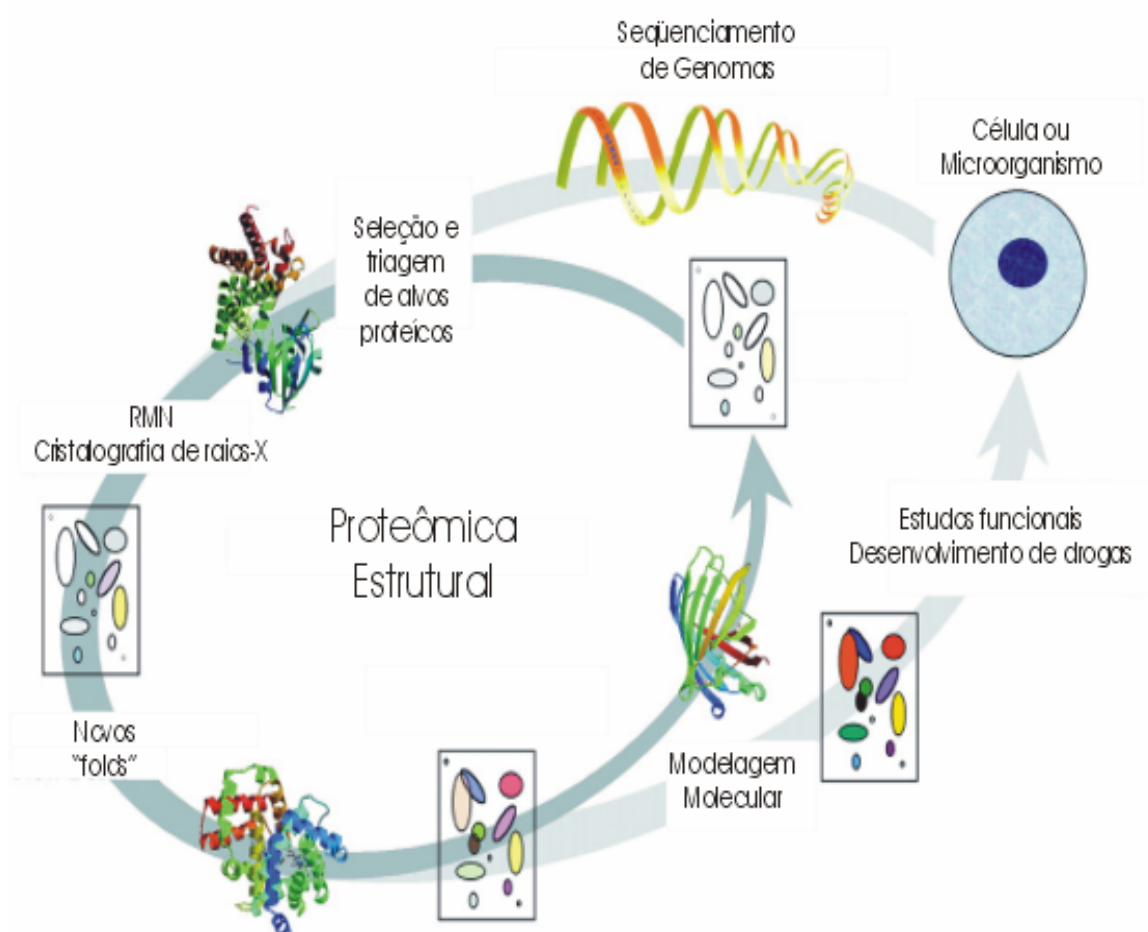
Respiratória Aguda Grave) (BARTLAM *et al.*, 2007) e a caracterização estrutural das proteínas de *Escherichia coli* (*E. coli*) envolvidas da via de biossíntese de histidina (MATTE *et al.*, 2007).

As duas abordagens implicam em uma transformação da pesquisa tradicional baseada na formulação e confirmação de hipóteses. Ao contrário dos estudos de biologia estrutural clássicos em que a determinação da estrutura de uma proteína é habitualmente o último passo de uma vasta gama de estudos bioquímicos e funcionais da proteína de interesse, os estudos de proteômica estrutural são efetuados muitas vezes conhecendo-se apenas a seqüência da proteína em estudo e a estrutura tridimensional resultante é por vezes o primeiro dado experimental que se obtém sobre a proteína.

Desta forma, a estrutura obtida para a proteína em estudo poderá fornecer um importante leque de informações funcionais sobre a mesma. De fato, vários estudos recentes de proteômica estrutural demonstraram que é possível obter informações funcionais a partir da estrutura tridimensional de uma proteína cuja estrutura não seja previamente conhecida (JACKSON e RUSSEL, 2001; BHATTACHARYYA *et al.*, 2002; CHRISTENDAT *et al.*, 2002; ZHANG e KIM, 2003; KIM *et al.*, 2005; GUZZO *et al.*, 2007; HERMANN *et al.*, 2007).

Apesar de distintas, as duas abordagens são complementares e ocorrem muitas vezes em simultâneo, contribuindo para um melhor entendimento de estrutura e função protéica bem como do organismo em estudo a nível molecular (Figura 2).

O principal impacto da nova era da proteômica estrutural será, no entanto, no processo de desenho de drogas para uso humano, veterinário e agrônomico (RUSSEL e EGGLESTON, 2000; MITTL e GRÜTTER, 2001; HANASH, 2003).



**Figura 2.** Impacto de projetos de proteômica estrutural na biologia atual.

Projetos de proteômica estrutural contribuirão para um melhor entendimento dos tipos de conformações protéicas e famílias de proteínas existentes e das relações entre estrutura e função de proteínas. Os painéis retangulares representam o universo descrito de seqüências de famílias de proteínas com estruturas conhecidas (círculos coloridos) ou desconhecidas (círculos não-coloridos). À medida que mais estruturas são determinadas mais círculos são coloridos e mais tipos de enovelamentos são conhecidos. As novas estruturas contêm também informações importantes que permitem modelagem por homologia de outras seqüências protéicas, estudos funcionais ou desenvolvimento de novas drogas (adaptado de STEVENS *et al.*, 2001).

Atualmente, a maioria das drogas ainda é descoberta pelos processos tradicionais de testar uma grande biblioteca de compostos químicos contra a proteína alvo na esperança de encontrar algum composto que provoque a ação desejada. Este processo, além de extremamente laborioso e caro, é altamente ineficiente. A proteômica estrutural, ao produzir estruturas de proteínas de patógenos de interesse, experimentalmente ou por homologia, permitirá alterar a perspectiva do processo de desenvolvimento de drogas para um processo de desenho de drogas racional (“rational drug design”) ou desenho de drogas baseado em estrutura (“structure based design”). Com a estrutura tridimensional da proteína alvo em mãos, é possível gerar racionalmente um pequeno número de compostos que poderiam provocar as ações desejadas, aumentando grandemente a eficiência do processo de desenho de drogas (RUSSEL E EGGLESTON, 2000).

Uma abordagem de proteômica estrutural em grande escala não resolverá todas as questões da biologia estrutural. Primeiro, nem todas as proteínas podem ter a sua estrutura determinada pelas técnicas atuais. Segundo, o universo de estruturas de proteínas existentes é verdadeiramente imenso, incluindo alterações conformacionais, diferentes modos de agregação e diferenças chave entre proteínas relacionadas. No entanto, a nova era da genômica estrutural promete capitalizar nos inúmeros avanços da ciência até à data, determinando milhares de estruturas protéicas experimentalmente e possivelmente milhões delas por homologia, alterando para sempre a nossa compreensão dos sistemas biológicos (HOL, 2000).

### **2.1.3. O papel da RMN na era da proteômica estrutural**

A espectroscopia de RMN e a cristalografia por difração de raios-X são atualmente as duas técnicas que podem revelar a estrutura tridimensional de

proteínas e de outras biomoléculas com detalhes atômicos. A cristalografia é no entanto considerada como o “cavalo de força” da proteômica estrutural, pois uma vez que se obtém um cristal com bom padrão de difração, a determinação da estrutura tridimensional é relativamente simples e rápida, e os procedimentos utilizados nas etapas de cristalização, coleta de dados e análise de dados estão cada vez mais automatizados (YEE *et al.*, 2002; SAVCHENKO *et al.*, 2003). O resultado disso é que o número de estruturas de proteínas depositadas no PDB (Protein Data Bank) determinadas pela técnica de cristalografia excede em aproximadamente cinco vezes o número de estruturas depositadas determinadas por RMN (BERMAN *et al.*, 2003). Apesar disso, existem várias razões que permitem sugerir que a RMN terá um papel importante em vários projetos de proteômica estrutural. As principais contribuições da RMN para estes projetos devem incluir três abordagens principais: determinação de estruturas tridimensionais de proteínas, elucidação de relações estrutura-função, e caracterização preliminar de alvos para futuros estudos estruturais (MONTELIONE, 2000; PRESTGARD *et al.*, 2001; KENNEDY *et al.*, 2002; AL-HASHIMI e PATEL, 2002).

Em primeiro lugar a RMN constitui uma importante técnica complementar de determinação de estruturas tridimensionais de proteínas pois a cristalografia, apesar da sua posição de destaque como técnica rápida e eficaz de determinação estrutural, apresenta ainda um importante fator limitante: a obtenção de cristais, um processo imprevisível que pode demorar alguns dias, meses, ou pode mesmo nunca acontecer (MONTELIONE *et al.*, 2002; PRESTGARD *et al.*, 2001; KENNEDY *et al.*, 2002). A RMN por seu turno, encontra fatores limitantes no limite de tamanho da molécula em estudo que é de aproximadamente 40 kDa e nos longos tempos de coleta e análise de dados mas apresenta a enorme vantagem de ser realizada em

solução. Assim, a RMN torna-se a única técnica atual disponível para a determinação de estruturas de proteínas de difícil cristalização, desde que estas se enquadrem dentro do limite de tamanho da técnica de RMN.

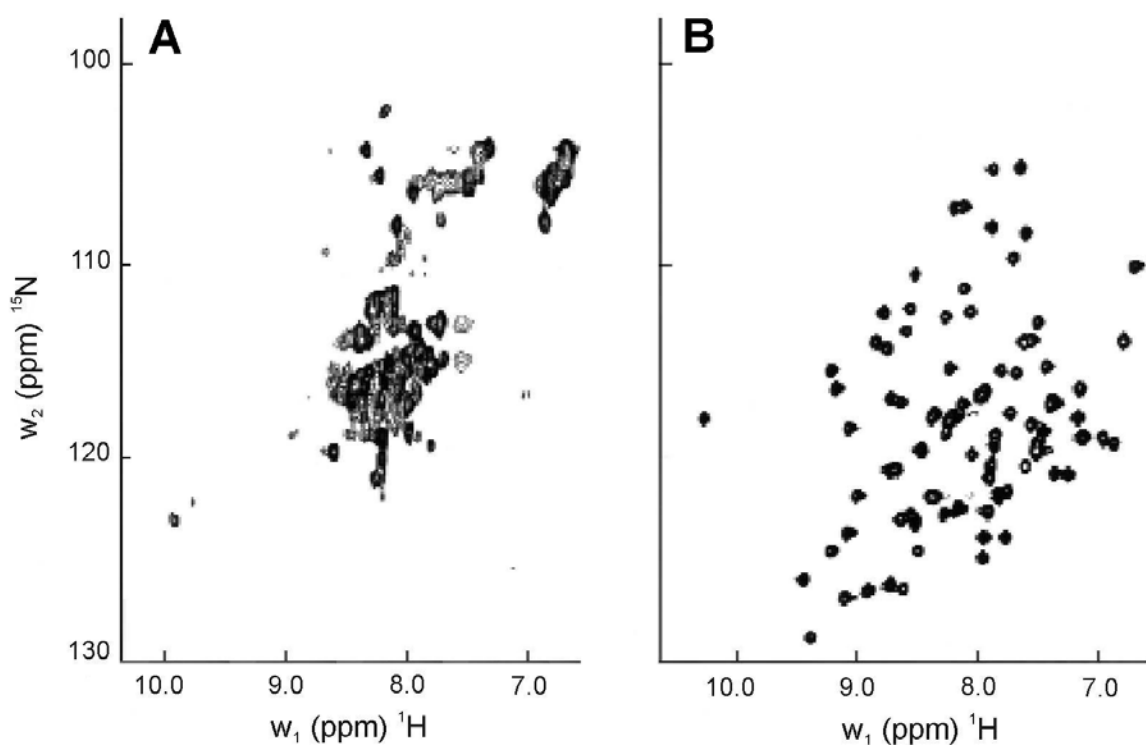
O maior valor da RMN para a proteômica estrutural será no entanto como ferramenta para a elucidação de relações estrutura-função de proteínas (MONTELIONE *et al.*, 2000). Por um lado, técnicas de perturbação de deslocamentos químicos permitem validar funções bioquímicas propostas através da identificação de possíveis ligantes e do mapeamento das respectivas superfícies de interação (ZUIDERWEG, 2002). Por outro lado, a dinâmica interna das proteínas em estudo, que muitas vezes está diretamente relacionada com a sua função, também pode ser facilmente avaliada através de medidas de relaxação nuclear (WAGNER, 1995; KAY, 1998; PALMER, 2001).

Além disso, a RMN pode também ser utilizada em projetos de proteômica estrutural para caracterizar rapidamente uma dada proteína antes dos estudos estruturais por RMN e/ou cristalografia (MONTELIONE *et al.*, 2000; PRESTGARD *et al.*, 2001; REHM *et al.*, 2002; AL-HASHIMI e PATEL, 2002.; YOKOYAMA, 2002). Tradicionalmente as técnicas de caracterização biofísica de proteínas mais utilizadas são o dicroísmo circular (CD) para caracterização do conteúdo de estrutura secundária, o espalhamento de luz para avaliar agregação heterogênea e a fluorescência para avaliar perturbações na estrutura terciária (PRESTGARD *et al.*, 2001; PAGE *et al.*, 2005). A RMN é uma alternativa poderosa neste tipo de caracterização pois permite a caracterização simultânea do estado conformacional, estabilidade e solubilidade da proteína em estudo e estes critérios podem ser utilizados para selecionar proteínas para estudos estruturais futuros.

Há muito que é reconhecido que o deslocamento químico observado em experimentos de RMN de proteínas contém informações estruturais importantes (WISHART *et al.*, 1991; WISHART e SYKES, 1994; REHM e HUBER, 2002). É também sabido que a análise combinada da dispersão dos deslocamentos químicos e das larguras das linhas observadas em um experimento de RMN pode fornecer informações quanto ao estado conformacional, grau de enovelamento e grau de agregação de uma proteína em solução, (MONTELLIONE *et al.*, 2000; REHM E HUBER, 2002; YEE *et al.*, 2002). Espectros de RMN de proteínas bem enoveladas apresentam linhas estreitas e dispersão dos deslocamentos químicos considerável enquanto proteínas desenoveladas ou agregadas apresentam linhas largas e pequena dispersão dos deslocamentos químicos. Teoricamente, todos os experimentos de RMN de proteínas fornecem este tipo de informação e até mesmo o mais simples experimento, o espectro unidimensional (1D) de  $^1\text{H}$ , é bastante informativo e vários trabalhos têm feito uso deste tipo de experimento para a caracterização inicial das proteínas em estudo (PETI *et al.*, 2004; SCHEICH *et al.*, 2004; PAGE *et al.*, 2005). No entanto, o experimento mais utilizado na caracterização de proteínas é o espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (“Heteronuclear Single Quantum Correlation Spectroscopy”) (Figura 3).

Este tipo de espectro, que requer marcação isotópica com  $^{15}\text{N}$  da proteína em estudo, é muitas vezes denominado de “impressão digital” de uma proteína, uma vez que nele se observa um pico de correlação  $^1\text{H}$ - $^{15}\text{N}$  para cada grupo amídico da proteína em estudo. Neste tipo de espectro, além de se analisar o padrão apresentado pelos picos (dispersão do deslocamento químico ( $\delta$ ), largura da linha e intensidade), analisa-se também o número de picos observados. Comparando o número de picos observado experimentalmente com o número de picos esperado a

partir da seqüência primária da proteína em estudo, pode-se aferir a ocorrência de mais de uma conformação (excesso de picos) ou a presença de regiões que apresentam troca conformacional ou são pouco estáveis (ausência de picos). Assim, o espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC revela informações únicas sobre a solubilidade, o grau de enovelamento e a conformação da proteína em estudo e permite que se testem de forma simples e rápida diferentes condições de concentração, pH, sal e temperatura nas quais a proteína seria mais estável.



**Figura 3.** Comparação de espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC para uma proteína desordenada e uma proteína bem enovelada (adaptado de MONTELLIONE, 2000).

A. Espectro do domínio C-terminal da proteína Par 1 da *Drosophila melanogaster*, um domínio que é predominantemente desordenado nas condições de coleta.

B. Espectro da proteína COG272 de *Thermus thermophilus*, uma proteína cuja estrutura tridimensional em solução é bem definida.



A Figura 3A mostra o espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da proteína Par1 de *Drosophila melanogaster*, cujo espectro é tipicamente de uma proteína desordenada pois apresenta pouca dispersão de deslocamento químico e alargamento das linhas. A Figura 3B mostra o espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da proteína COG72 de *Thermus thermophilus* cuja estrutura tridimensional em solução é bem definida e cujo espectro apresenta uma pouca dispersão de deslocamento químico e linhas estreitas e definidas.

Apesar do custo acrescido devido à marcação isotópica necessária, estas características fazem com que espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC sejam cada vez mais a técnica de eleição para a caracterização preliminar de alvos desconhecidos de projetos de proteômica estrutural (MONTELLIONE *et al.*, 2000; FOLKERS *et al.*, 2004).

#### **2.1.4. Triagem de proteínas alvo para estudos de proteômica estrutural**

O crescimento exponencial de projetos de proteômica estrutural que se tem observado nos últimos anos está diretamente relacionado com o surgimento de novas técnicas que facilitam a determinação de estruturas protéicas em larga escala (BURLEY *et al.*, 2000; BRENNER, 2001; MITTL e GRÜTTER, 2001; STEVENS *et al.*, 2001). Várias técnicas têm sido desenvolvidas para clonagem, expressão, marcação seletiva e purificação das ORFs alvo em larga escala (CAI *et al.*, 1998; DIECKMAN *et al.*, 2002; BOETTNER *et al.*, 2002; LESLEY *et al.*, 2002; PHIZICKY *et al.*, 2003) Avanços importantes têm também sido alcançados em procedimentos que visam a automatização dos processos de cristalização de proteínas (LESLEY *et al.*, 2002; VILLASENOR *et al.*, 2002), análise de dados de cristalografia de raios-X (LAMZIN E PERRAKIS, 2000) e determinação de estrutura por RMN (SZYPERSKI *et*

*al.*, 2002; BHAVESH *et al.*, 2001; PRESTEGARD *et al.*, 2001; ZWECKSTETTER e BAX, 2001).

No entanto, ainda existem passos limitantes de extrema importância que necessitam ser melhorados para que o processo de conversão de seqüência gênica em estrutura protéica tridimensional se torne mais rápido e eficiente (SERVICE, 2002; YOKOYAMA, 2003; NORIN e SUNDSTROM, 2002). Um dos principais problemas associados com projetos de proteômica estrutural é que grande parte das ORFs existentes na natureza codifica proteínas que não são apropriadas para estudos estruturais pelas técnicas atuais de cristalografia de raios-X e RMN devido a problemas relacionados com tamanho, solubilidade, estabilidade conformacional da proteína e com a obtenção de cristais (BRENNER, 2000).

Os processos de seleção de proteínas alvo para projetos de proteômica estrutural visam aumentar as chances de se selecionar proteínas cuja estrutura possa ser resolvida facilmente pelas técnicas atuais. No entanto, a seleção feita *in silico* não permite uma estimativa precisa dos níveis de expressão da proteína recombinante, solubilidade e estabilidade conformacional da proteína e por isso muitas das proteínas alvo selecionadas são invariavelmente eliminadas no decorrer do processo por não apresentarem as características desejáveis para estudos estruturais. Este tipo de fenômeno em que proteínas alvo selecionadas para estudos de proteômica estrutural são progressivamente eliminadas no decorrer do processo é conhecido como “efeito funil” e foi descrito por YEE *et al.*, 2002. Neste trabalho, das 513 ORFs selecionadas como alvos para estudos de proteômica estrutural, apenas 94 das proteínas codificadas apresentavam características que viabilizavam estudos estruturais e apenas 12 tiveram as suas estruturas tridimensionais resolvidas. Estudos mais recentes demonstraram que, de um modo geral, um alvo

de proteômica estrutural progredirá ao longo de cada uma das etapas de clonagem, expressão, purificação e cristalização com uma probabilidade de aproximadamente 45% (O'TOOLE *et al.*, 2004b). Além disso, aproximadamente 80% das proteínas alvo que apresentam dados de difração produzirão uma estrutura cristalina enquanto 20% dos alvos que possuem um espectro 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC produzirão uma estrutura por RMN (O'TOOLE *et al.*, 2004b).

Para que os projetos de proteômica estrutural se tornem mais produtivos e eficientes, é desejável que se desenvolvam metodologias de triagem que permitam avaliar de maneira rápida e eficaz se uma dada proteína alvo é adequada para estudos estruturais. Os experimentos de RMN 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC freqüentemente utilizados para selecionar proteínas alvo que apresentam as características desejáveis para estudos estruturais são muito informativos. No entanto, estes experimentos são normalmente realizados utilizando proteína pura, o que torna o procedimento caro e extremamente demorado para ser utilizado em projetos de larga escala. Assim, recentemente vários métodos têm sido desenvolvidos para selecionar proteínas alvo com as características desejadas rapidamente e antes da purificação das mesmas. Estes incluem a expressão de proteínas heterólogas contendo fusões específicas, (HUTH *et al.*, 1997; WALDO *et al.*, 1999), espectroscopia de RMN “in-cell” (SERBER *et al.*, 2001a), sistemas de expressão “cell-free” (KIGAWA *et al.*, 1999; GUIGNARD *et al.*, 2002) entre outros.

Em 1997, HUTH e colaboradores desenvolveram vetores de expressão específicos para preparação de amostras para triagem por RMN. Estes vetores codificam o domínio imunoglobulina da proteína G de *Streptococcus* (Domínio GB1) ligados ao N-terminal da proteína de interesse. A proteína GB1 é pequena, estável, bem enovelada e apresenta elevados níveis de expressão em bactérias. Estas

características permitem que se realize a análise por RMN das proteínas de fusão diretamente, i.e. sem a necessidade de purificação prévia, e portanto em um período de tempo relativamente curto.

WALDO e colaboradores (1999) por sua vez, desenvolveram um vetor que pode ser utilizado como repórter de enovelamento. Neste vetor, a proteína de interesse é expressa como uma fusão N-terminal da proteína GFP (“Green Fluorescent Protein”). Após a expressão protéica, a fluorescência da proteína GFP nas células de *E. coli* é analisada por microscopia de fluorescência. Neste tipo de fusão, o enovelamento correto da proteína GFP depende do enovelamento adequado da proteína de interesse, e portanto a fluorescência das células estará diretamente relacionada com o enovelamento da proteína em estudo. Assim, é possível determinar rapidamente quais proteínas serão provavelmente enoveladas corretamente quando expressas em *E. coli*.

Apesar das vantagens evidentes das metodologias que fazem uso de fusões específicas para caracterizar proteínas de interesse, estas técnicas apresentam a grande desvantagem não permitir a expressão direta de proteínas para estudos estruturais. Uma dada proteína de interesse, que seja inicialmente analisada em fusão com a proteína GB1 ou a proteína GFP e subseqüentemente selecionada para estudos estruturais, precisará ser clivada ou re-clonada em um outro vetor que não contenha as referidas fusões, o que torna o procedimento demorado e dispendioso para uso em projetos de proteômica estrutural.

Outra metodologia promissora para analisar rapidamente proteínas candidatas para estudos estruturais é a RMN “in cell” (SERBER *et al.*, 2001a; RECKEL *et al.*, 2005; SELENKO e WAGNER, 2007). Recentemente vários grupos de pesquisa, utilizando diferentes técnicas de RMN e sistemas biológicos distintos,

demonstraram que é possível estudar macromoléculas por RMN dentro de células vivas (SERBER *et al.*, 2001a; SERBER *et al.*, 2001b; WIERUSZESKI *et al.*, 2001; SERBER *et al.*, 2006). Neste tipo de análise as proteínas de interesse são expressas e analisadas diretamente dentro da célula permitindo uma análise rápida das suas características em um ambiente muito próximo ao seu ambiente natural. Embora este tipo de análise não permita estudos estruturais detalhados, devido a problemas de baixa concentração da proteína em estudo e alargamento dos sinais de RMN, os resultados obtidos por ambos os grupos mostram-se promissores para estudos de dinâmica, interações com ligantes *in vivo*, e triagem de alvos para estudos estruturais.

Alternativamente, métodos de síntese de proteínas *in vitro*, também denominados “cell-free”, vêm também sendo freqüentemente utilizados (KIGAWA *et al.*, 1999; GUIGNARD *et al.*, 2002). Neste tipo de abordagem as proteínas são produzidas *in vitro* usando extratos celulares, que contêm polimerases e ribossomos, mas eliminam a maioria da maquinaria celular. Células intactas muitas vezes não conseguem lidar com grandes quantidades de proteína recombinante devido a problemas de toxicidade. O método “cell-free”, permite que sejam produzidas proteínas em grandes quantidades e de forma bastante pura, uma vez que o único DNA presente na mistura é o DNA que codifica a proteína de interesse. Por outro lado, a técnica também permite a adição de co-fatores químicos e outras enzimas para promover o enovelamento adequado das proteínas em estudo. Estas características tornam a técnica de produção de proteínas “in-cell”, na mais promissora metodologia para produção de proteínas forma “high-throughput” em projetos de proteômica estrutural (SERVICE, 2002). Esta técnica é no entanto

dispendiosa para ser utilizada como a metodologia de triagem de alvos para estudos estruturais de seleção.

Em 2001, ALMEIDA e colaboradores, desenvolveram uma metodologia que permite marcar seletivamente com um isótopo de interesse (ex.  $^{15}\text{N}$ ), uma proteína heteróloga em *E. coli*. Esta metodologia faz uso do antibiótico rifampicina, um inibidor da polimerase bacteriana e de vetores que expressam proteínas heterólogas utilizando promotores de RNA polimerase do fago T7, como no caso dos vetores pET (STUDIER *et al.*, 1990). Durante a indução com IPTG, rifampicina e o isótopo de interesse são também adicionados ao meio de cultura. Desta forma, a expressão de todas as proteínas bacterianas será inibida na presença de rifampicina enquanto as proteínas heterólogas clonadas sob controle de promotores de T7, não terão a sua expressão inibida e serão portanto seletivamente marcadas com o isótopo de interesse. O fato de apenas a proteína de interesse estar marcada com  $^{15}\text{N}$  significa que se pode obter o primeiro espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da mesma antes da sua purificação. Ao omitir a etapa de purificação, este protocolo constitui um método rápido, barato e eficaz para analisar características conformacionais de proteínas. Apesar desta metodologia ter sido originalmente testada usando uma proteína de função e estrutura conhecida, esta seria potencialmente uma metodologia promissora para uso em estudos de triagem de proteínas alvo em projetos de proteômica estrutural.

Nas seções que se seguem é apresentado um estudo de triagem de alvos de proteômica estrutural em que a metodologia descrita acima foi utilizada para selecionar proteínas de *Xac* com características adequadas para estudos estruturais por RMN e/ou difração de raios-X de cristais protéicos.

## 2.2. OBJETIVOS

O trabalho descrito neste capítulo teve por objetivo a seleção *in silico* e triagem de 35 proteínas alvo de *Xac* para estudos estruturais utilizando uma metodologia inicialmente desenvolvida pelos Prof. Drs. Ana Paula Valente e Fábio Almeida da Universidade Federal do Rio de Janeiro (ALMEIDA *et al.*, 2001) que permite marcar seletivamente uma proteína heteróloga com  $^{15}\text{N}$  em *E. coli* e obter o espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da mesma antes da sua purificação.

No presente estudo, esta metodologia foi utilizada pela primeira vez para analisar um conjunto grande de proteínas para as quais não existia qualquer informação funcional ou estrutural. Esta abordagem possibilitou determinar quais as proteínas alvo que melhor se adequavam para estudos estruturais de uma maneira rápida e eficaz possibilitando assim a priorização de alvos de *Xac* para estudos estruturais futuros por RMN e/ou cristalografia de raios-X (GALVAO-BOTTON *et al.*, 2003).

## 2.3. MATERIAIS E MÉTODOS

### 2.3.1. Seleção das proteínas alvo de *Xac*

As 35 proteínas alvo de *Xac* foram selecionadas *in silico* usando os seguintes critérios:

- a) Proteínas de tamanho inferior a 300 amino ácidos;
- b) Proteínas cuja localização celular predita usando o programa PSORT (NAKAI e HORTON, 1999) fosse o citóssol;
- c) Proteínas cujo conteúdo de metionina fosse superior a 1.1%;
- d) Proteínas de função desconhecida ou que apresentassem baixa similaridade com proteínas de função conhecida e que fossem altamente conservadas em vários genomas (“proteínas hipotéticas conservadas”);
- e) Proteínas que não apresentassem homologia com nenhuma estrutura depositada no PDB (BERMAN *et al.*, 2003), identificadas pelo programa PSI-BLAST (ALTSCHUL *et al.*, 1997), segundo o método de predição de homologia estrutural descrito por HUYNEN *et al.* (1998).

### 2.3.2. Amplificação e Clonagem das ORFs alvo

#### 2.3.2.1. Amplificação por PCR das ORFs alvo

A amplificação das ORFs selecionadas foi feita por PCR (Reação de Polimerase em Cadeia) segundo SAMBROOK e colaboradores (2001), a partir de DNA genômico de *Xac* utilizando oligonucleotídeos desenhados especificamente para cada ORF conforme descrito na Tabela 1.



**Tabela 1.** Amplificação e clonagem das 35 ORFs de *Xac*: oligonucleotídeos e sítios de restrição utilizados. Os sítios de restrição utilizados na clonagem encontram-se sublinhados na seqüência dos oligonucleotídeos.

ORF	Oligo	Seqüência	Sítios de clonagem	Vetor
XAC 0021	0021 (+)	5' CATGCCATGGCTCATATGTCCAAGTCTCTGTACAAAGTCACCT 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	0021 (-)	5' GGAATTCAAGCTTTTCAGCGTGGCTTGATCGGC 3'		
XAC 0400	0400 (+)	5' CATGCCATGGCTCATATGATCCGTGCGATTTCTCCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	0400 (-)	5' CGAATTCAAGCTTTTATGGGCGAACCTCCTGAGC 3'		
XAC 0404	0404 (+)	5' CATGCCATGGCTCATATGCGCATCCTGCCGGT 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	0404 (-)	5' CGAATTCAAGCTTTTAGACAACCTCGATGCTGATCTGG 3'		
XAC 0862	0862 (+)	5' CATGCCATGGCTCATATGCAAGATGATCC GCGCTACC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	0862 (-)	5' CGAATTCAAGCTTTTAGTGCAGCGTCCTGGGCA 3'		
XAC 1110	1110 (+)	5' CATGCCATGGCTCATATGCGTGAAACATTGCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	1110 (-)	5' GGAATTCAAGCTTTTCAGAACGGGAGCTTCATG 3'		
XAC 1223	1223 (+)	5' CATGCCATGGCTCATATGCGCGTGACCACTCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	1223 (-)	5' GGAATTCAAGCTTTTCAGGCAGGCAACAAGCC 3'		
XAC 1352	1352 (+)	5' CATGCCATGGCTCATATGGACTCTTCGCTACCTC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	1352 (-)	5' GGAATTCAAGCTTTTTCAGCGATCGGTGGACAG 3'		
XAC 1670	1670 (+)	5' CATGCCATGGCTCATATGGCTTACGCATTACCCAACGC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	1670 (-)	5' CGAATTCAAGCTTTTAGAGCGCCGTCTGTACGATGC 3'		
XAC 1747	1747 (+)	5' CATGCCATGGCTCATATGTTGAATGTCCGCAAGAGC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	1747 (-)	5' GGAATTCAAGCTTTTACTTGGGCAGGTGCAACAC 3'		
XAC 1757	1757 (+)	5' CATGCCATGGCTCATATGACCGACGCTGTATTGAACTC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	1757 (-)	5' CGAATTCAAGCTTTTACTGTTTGGCGTGCGCCT 3'		
XAC 1883	1883 (+)	5' CATGCCATGGCTCATATGGAAGCCACCGTTGCCG 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	1883 (-)	5' CGAATTCAAGCTTTTATGGGTGCAACGGGTCTCCG 3'		
XAC 1884	1884 (+)	5' CATGCCATGGATATGGCTAGCCTGGATTCCCTCGTCCCTGCT 3'	<i>Nhe</i> I e <i>Hind</i> III	pET3a
	1884 (-)	5' CGAATTCAAGCTTTTAGGGGGCGGGTTTGGG 3'		
XAC 2000	2000 (+)	5' CATGCCATGGCTCATATGCCTCGGAATACCTCACAC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2000 (-)	5' GGAATTCAAGCTTTTCAGGCCTGTTCCATCGTG 3'		

ORF	Oligo	Seqüência	Sítios de clonagem	Vetor
XAC 2218	2218 (+)	5' CATGCCATGGCTCATATGAATTCCA <del>CT</del> CTCCCTCCGG 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2218 (-)	5' GGAATTC <del>CAAGCTT</del> CTACCATCCGACATAGTTCAGCACG 3'		
XAC 2237	2237 (+)	5' CATGCCATGGCTCATATGGCAACCACCACGGCAC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2237 (-)	5' GGAATTC <del>CAAGCTT</del> CTACGCCGCCTGGGGTTC 3'		
XAC 2355	2355 (+)	5' CATGCCATGGCTCATATGACCACCTCCCCCTTCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2355 (-)	5' GGAATTC <del>CAAGCTT</del> TTCACTGCTGCGCGCG 3'		
XAC 2375	2375 (+)	5' CATGCCATGGCTCATATGAGCCAATCCAGCGCCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2375 (-)	5' GGAATTC <del>CAAGCTT</del> TCAGTCCGCGCCGCTG 3'		
XAC 2396	2396 (+)	5' CATGCCATGGCTCATATGGCCCTCACCGCCA 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2396 (-)	5' GGAATTC <del>CAAGCTT</del> TCATTCGGCCGGGGCT 3'		
XAC 2462	2462 (+)	5' CATGCCATGGCTCATATGAACGCACCCGCACATCG 3'	<i>Nde</i> I e <i>Eco</i> R I	pET3a
	2462 (+)	5' <del>CGAATTC</del> CAAGCTTTTATGCCGCAGGTTCCGGCA 3'		
XAC 2573	2573 (+)	5' CATGCCATGGCTCATATGCATCCGACCGCCGC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2573 (-)	5' <del>CGAATTC</del> CAAGCTTTTAGCGTGCCAGCAGCGTCG 3'		
XAC 2775	2775 (+)	5' CATGCCATGGCTCATATGAAATGCCGACTCATCGC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2775 (-)	5' GGAATTC <del>CAAGCTT</del> TTATGCGCGGTGATACGGA 3'		
XAC 2942	2942 (+)	5' CATGCCATGGCTCATATGTTGTTGCCCATCCCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2942 (-)	5' GGAATTC <del>CAAGCTT</del> TCAGACATCGGCCCATG 3'		
XAC 2956	2956 (+)	5' CATGCCATGGCTCATATGAGTGTTCCGCAGTTGCCG 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	2956 (-)	5' <del>CGAATTC</del> CAAGCTTTTACGAGCCGCGATCCTCAAG 3'		
XAC 3069	3069 (+)	5' CATGCCATGGCTCATATGCGCCTGACCCGTTCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3069 (-)	5' GGAATTC <del>CAAGCTT</del> CTATAGGTCCCCGAACCGC 3'		
XAC 3146	3146 (+)	5' CATGCCATGGCTCATATGACCATCGAATCCCCATTCC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3146 (-)	5' <del>CGAATTC</del> CAAGCTTTTAGTAATCGGATTCTGGGGTTTC 3'		
XAC 3151	3151 (+)	5' CATGCCATGGCTCATATGGGTAGAGGTCCCTCGATC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3151 (-)	5' GGAATTC <del>CAAGCTT</del> TTAGGCGCCAAGCGCC 3'		
XAC 3410	3410 (+)	5' CATGCCATGGCTCATATGAAGTCCGAACCGGACACG 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3410 (-)	5' GGAATTC <del>CAAGCTT</del> CTATTCCATCGCCAGCAACACTT 3'		

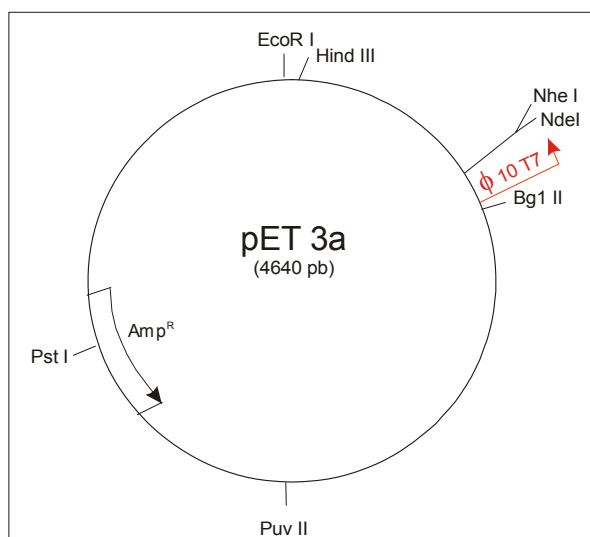
ORF	Oligo	Seqüência	Sítios de clonagem	Vetor
XAC 3671	3671 (+)	5' CATGCCATGGCTCATATGCCTTCCTTCGACGTTGT 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3671 (-)	5' GGAATTCAAGCTTTCAATCGCGGAAGTTGTCTG 3'		
XAC 3725	3725 (+)	5' CATGCCATGGCTCATATGGCGATCAAGACTGTCTGAA 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3725 (-)	5' GGAATTCAAGCTTTCAAGCGGCCTTGCTG 3'		
XAC 3726	3726 (+)	5' CATGCCATGGCTCATATGTTTCATGCACAACAAGCGT 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3726 (-)	5' GGAATTCAAGCTTTAACGCTGGATATCGCCG 3'		
XAC 3873	3873 (+)	5' CATGCCATGGCTCATATGTCCATGAAGCAGCAGCTC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	3873 (-)	5' GGAATTCAAGCTTTCAGCCTGCCAGCTGCT 3'		
XAC 4145	4145 (+)	5' CATGCCATGGCTCATATGGCGTTTCGACATGCATATCAAG 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	4145 (-)	5' CGAATTCAAGCTTTTAGGCCTTGCCACACCTTG 3'		
XAC 4147	4147 (+)	5' CATGCCATGGCTCATATGGCCAGCAGTCAGAAGTTCATC 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	4147 (-)	5' CGAATTCAAGCTTTTAGGTTCCCTCCGGTTCGGA 3'		
XAC 4330	4330 (+)	5' CATGCCATGGCTCATATGCCAGCGAACACTTGCCACAA 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	4330 (-)	5' CGAATTCAAGCTTTAATGAAATGTCGGAGCGCTGGT 3'		
XACb 0070	b0070 (+)	5' CATGCCATGGCTCATATGAATACCGTGCCTGGAACAT 3'	<i>Nde</i> I e <i>Hind</i> III	pET3a
	b0070 (-)	5' CGAATTCAAGCTTTAATGCTCACGCGCCATTG 3'		

Os oligonucleotídeos da fita “sense” (+) foram desenhados para conter sítios de restrição de *Nco* I e *Nde* I antes do códon de iniciação com exceção do oligonucleotídeos da ORF XAC1884 que foi desenhado para conter sítios de restrição de *Nco* I e *Nhe* I. Os oligonucleotídeos da fita “anti-sense” (-) foram desenhados para conter os sítios de restrição de *Hind* III e *EcoR* I depois do códon de terminação. Todos os oligonucleotídeos foram desenhados de forma que a temperatura de desnaturação estimada fosse de aproximadamente 60 °C. A reação de amplificação foi feita da seguinte forma: 5 µL de tampão ThermoPol (200 mM Tris-HCl, 100 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100 mM KCl, 20 mM MgSO<sub>4</sub>, 1 % Triton, pH 8,8 - New England BioLabs); 0,5 µL de Deep Vent DNA Polimerase (2000 U/mL - New England BioLabs); 200 µM de mistura de dNTPs; 1 µL oligonucleotídeo sense (+) (10 pmol); 1 µL de solução 10 pmol oligonucleotídeo anti-sense (-); 6 ng de DNA genômico de *Xac*; água deionizada para um volume final de 50 µL. Utilizando-se o termociclador Applied Biosystems GeneAmp® 9700, foram ajustadas as seguintes temperaturas e tempos de reação: 96 °C por 2 minutos (1 ciclo); 96 °C por 1 minuto, 53°C por 30 segundos, 72 °C por 1 minuto (5 ciclos); 96 °C por 1 minuto, 60 °C por 30 segundos, 72 °C por 1 minuto (30 ciclos); 72 °C por 10 minutos. Em alguns casos a adição de 5% de DMSO permitiu uma melhor especificidade de reação. Os produtos da reação de PCR foram analisados em gel de agarose 1% e purificados utilizando o sistema “Concert Nucleic Acid Purification System” (GIBCO BRL).

#### **2.3.2.2. Clonagem das ORFs alvo no vetor de expressão pET-3a**

As ORFs amplificadas foram clonadas no vetor pET-3a (STUDIER *et al.*, 1990). (Figura 4). As ORFs amplificadas e purificadas, assim como o vetor pET-3a (STUDIER *et al.*, 1990) foram digeridos com pares de enzimas de restrição

específicas para cada caso conforme descrito na Tabela 1. As digestões foram realizadas usando as seguintes condições: ~ 5 µg de DNA; 2 µL de tampão no. 2 para digestão (500 mM NaCl, 100 mM Tris-HCl, 100 mM MgCl<sub>2</sub>, 1 mM ditioneitol, pH 7,9 - New England BioLabs); 1 µL de enzima 1 (*Nde* I ou *Nhe* I) (2000 U/ml); 1 µL de enzima 2 (*Hind* III ou *EcoR* I (2000 U/ml); água deionizada para um volume final de 20 µL. O DNA foi digerido por 6h a 37 °C e as enzimas foram inativadas a 65 °C por 20 minutos.



**Figura 4.** Representação esquemática do vetor pET-3a (STUDIER *et al.*, 1990).

### 2.3.2.3. Seqüenciamento das ORFs alvo clonadas em pET-3a

O seqüenciamento dos plasmídeos contendo as ORFs clonadas foi feito pelo método de SANGER e colaboradores (1977), utilizando os oligonucleotídeos flanqueadores do inserto “T7 forward primer” (+) e “*EcoR* I reverse primer” (–) (Tabela 2).

**Tabela 2.** Oligonucleotídeos utilizados no seqüenciamento das ORFs clonadas.

Oligo	Seqüência
T7 forward primer (+)	5' CACTATAGGGAGACCACAACGGTTTCC 3'
<i>EcoR</i> I reverse primer (-)	5' CATTAACCTATAAAAATAGGCG 3'

A reação de seqüenciamento foi preparada utilizando-se o “kit” ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit III (Applied Biosystems) seguindo o protocolo: 5 µL de DNA de mini-preparação; 1 µL de solução contendo 4 pmol de oligonucleotídeo sense (+) ou 1 µL de solução contendo 10 pmol de oligonucleotídeo anti-sense (-); 6 µL de tampão (200 mM Tris-HCl pH 9,0; 5 mM MgCl<sub>2</sub>); 2 µL de Big Dye terminator; água deionizada para um volume final de 20 µL. As reações de seqüenciamento foram realizadas em um termociclador Applied Biosystems GeneAmp® 9700 de acordo com o seguinte protocolo: 96 °C por 30 segundos, 50 °C por 5 segundos, 60 °C por 4 minutos (40 ciclos). Terminada a reação, o DNA foi precipitado com isopropanol 75% e seqüenciado utilizando-se o seqüenciador automático ABI 377 de Perkin-Elmer Applied Biosystems.

### 2.3.3. Teste de expressão das proteínas alvo

Células da cepa BL21(DE3)pLysS (STUDIER *et al.*, 1990) foram transformadas com os plasmídeos contendo as ORFs em estudo pelo método de CaCl<sub>2</sub> (SAMBROOK *et al.*, 2001). As colônias obtidas de cada clone foram inoculadas em 4 mL de meio 2xTY (16g/L triptona; 10 g/L extrato de levedo; 5 g/L NaCl; pH 7,4) contendo 200 µg/mL de ampicilina e 200 µg/mL de cloranfenicol e a cultura foi crescida a 37 °C e 200 rpm. Quando a densidade ótica a 600 nm (D.O.<sub>600nm</sub>) se encontrava entre 0,8 e 1,0 recolheu-se 1 mL de amostra não induzida e procedeu-se à indução da expressão da ORF clonada adicionando IPTG

(isopropil- $\beta$ -tiogalactopiranosídeo) para uma concentração final de 400  $\mu$ M. Após 3,5 h de indução recolheu-se 1 mL de amostra induzida. As amostras foram centrifugadas a 14000 rpm a temperatura ambiente por 20 minutos e lisadas pela ressuspensão em 200  $\mu$ L de tampão de amostra (10% p/v glicerol; 2% SDS p/v; 50 mM Tris-HCl pH 6,8; 0.1% Bromofenol blue, 100 mM  $\beta$ -mercaptoetanol) e incubação a 100 °C por 5 minutos. As amostras foram analisadas por SDS-PAGE (Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis) em géis de 15% acrilamida e 0.4% bisacrilamida para verificação da indução da proteína em estudo.

#### **2.3.4. Teste de solubilidade das proteínas alvo**

Células da cepa BL21(DE3) pLysS (STUDIER *et al.*, 1990) transformadas com os plasmídeos contendo as ORFs em estudo foram crescidas e induzidas conforme descrito em 2.3.3, mas inoculando um volume de 250 mL de meio 2xTY. Após 3h de indução as células foram centrifugadas por 20 min, a 4 °C e 5000 rpm e armazenadas a -70 °C. As células congeladas foram ressuspendidas em 500  $\mu$ L de 20mM fosfato de sódio pH 7.0, 10 mM  $\beta$ -mercaptoetanol, 40  $\mu$ g/mL PMSF e lisadas por sonicação. Os lisados foram clarificados por centrifugação a 4000 rpm e 8°C. O grau de solubilidade das proteínas em estudo foi monitorado por SDS-PAGE em géis de 15% acrilamida e 0.4% bisacrilamida, aplicando-se alíquotas do sobrenadante e do precipitado obtidos após a centrifugação do lisado e comparando a intensidade das respectivas bandas.

#### **2.3.5. Marcação seletiva das proteínas alvo com $^{15}\text{N}$**

Utilizando o método de marcação seletiva de proteínas heterólogas usando rifampicina descrito por ALMEIDA e colaboradores (2001) as 31 proteínas alvo de

Xac que apresentaram bom nível de expressão foram seletivamente marcadas com  $^{15}\text{N}$ . Células de *E. coli* da cepa BL21(DE3)pLysS foram transformadas com os plasmídeos contendo as ORFs em estudo pelo método de  $\text{CaCl}_2$  (SAMBROOK *et al.*, 2001). As colônias obtidas foram pré-inoculadas em 2 mL de meio 2xTY contendo 100  $\mu\text{g/mL}$  de ampicilina e 100 $\mu\text{g/mL}$  de cloranfenicol. Após 8 h de crescimento a 37 °C e 200 rpm, 1 mL desta cultura foi utilizado para inocular 25 mL de meio mínimo M9 contendo  $^{14}\text{N}$  (6 g/L de  $\text{Na}_2\text{HPO}_4$  pH 7,35, 3 g/L de  $\text{KH}_2\text{PO}_4$ , 0,5 g/L  $\text{NaCl}$ , 2 mM  $\text{MgSO}_4$ , 0,4% glicose, 100  $\mu\text{M}$   $\text{CaCl}_2$ , 10 mg tiamina, 1g/L  $^{14}\text{NH}_4\text{Cl}$ ) e suplementado com 200  $\mu\text{g/mL}$  de ampicilina e 200  $\mu\text{g/mL}$  de cloranfenicol. Após mais 16h de crescimento foi feito o inóculo de 250 mL de meio mínimo M9 contendo  $^{14}\text{NH}_4\text{Cl}$  a partir de 12,5 mL do pré-inóculo. Quando a D.O.<sub>600nm</sub> atingiu 0,8 procedeu-se à indução da expressão protéica através da adição de 1 mM de IPTG. Após 15 minutos de incubação foi adicionado o antibiótico rifampicina (inibidor da polimerase bacteriana) para uma concentração final de 200  $\mu\text{g/mL}$ . Após uma segunda incubação de 15 minutos, as células foram centrifugadas por 20 minutos, a 4 °C e 5000 rpm e ressuspendidas em 100 mL de meio mínimo M9 contendo  $^{15}\text{NH}_4\text{Cl}$  e suplementado com 1 mM de IPTG, 200  $\mu\text{g/mL}$  de rifampicina, 200  $\mu\text{g/mL}$  de ampicilina e 200  $\mu\text{g/mL}$  de cloranfenicol. Após 3h de indução as células foram centrifugadas por 20 min, a 4 °C e 5000 rpm e armazenadas a -70 °C. O nível de expressão foi verificado através de SDS-PAGE em géis de 15% acrilamida e 0.4% bisacrilamida.

### **2.3.6. Triagem das proteínas alvo marcadas com $^{15}\text{N}$ por RMN**

As células congeladas (derivadas de 100 mL de cultura) contendo as proteínas alvo seletivamente marcadas com  $^{15}\text{N}$  foram ressuspendidas em 500  $\mu\text{L}$



de 20mM fosfato de sódio pH 7.0, 10 mM  $\beta$ -mercaptoetanol, 40  $\mu$ g/mL PMSF e lisadas por sonicação. Os lisados foram clarificados por centrifugação a 4,000 rpm, 8°C e a fração solúvel foi concentrada para 300  $\mu$ L usando uma membrana Centricom com um “cutoff” de 3 kDa.

Os lisados contendo as proteínas solúveis marcadas com  $^{15}\text{N}$  foram então analisados por RMN após a adição de 10%  $\text{D}_2\text{O}$ . No caso das proteínas puras usadas como controle foram também adicionados 10 %  $\text{D}_2\text{O}$ . Foram adquiridos espectros 1D de  $^1\text{H}$ , 1D de  $^1\text{H}$  editado para  $^{15}\text{N}$  e 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC a 303 K (30°C) em um equipamento Bruker® Advance DRX operando a 600 MHz usando sondas de 5mm de tripla ressonância. Para a aquisição e processamento dos dados foi utilizado o programa Bruker® XWINNMR.

### **2.3.7. Produção e purificação das proteínas XAC2000, XAC3873, XAC2396, XAC0862 e XACb0070**

#### **2.3.7.1. Expressão das proteínas em meio mínimo M9 contendo $^{15}\text{N}$**

A expressão das proteínas marcadas com  $^{15}\text{N}$  foi realizada em meio mínimo M9 contendo  $^{15}\text{NH}_4\text{Cl}$ . Células de *E. coli* da cepa BL21(DE3)pLysS foram transformadas com os plasmídeos contendo as ORFs em estudo pelo método de  $\text{CaCl}_2$  (SAMBROOK *et al.*, 2001). As colônias obtidas foram pré-inoculadas em 50 mL de meio 2xTY (16g/L triptona; 10 g/L extrato de levedo; 5 g/L NaCl; pH 7,4). suplementado com 200  $\mu$ g/mL de ampicilina e 200  $\mu$ g/mL de cloranfenicol. Após o crescimento a 37 °C e 200 rpm até a  $\text{DO}_{600\text{nm}} = 0,8-1,2$ , 5 mL desta cultura foram utilizados para inocular 500 mL de meio mínimo M9 contendo  $^{15}\text{NH}_4\text{Cl}$  e suplementado com 200  $\mu$ g/mL de ampicilina e 200  $\mu$ g/mL de cloranfenicol. As

células foram novamente incubadas a 37 °C, e 200 rpm até D.O.<sub>600nm</sub> = 0,8 quando foram induzidas pela adição de 1 mM de ITPG. Após 8 h de indução, as células foram centrifugadas por 20 minutos, a 4 °C e 5000 rpm, ressuspensas em 25 mL de solução de lise (1 mM EDTA, 25% sacarose, 50 mM Tris-HCl pH 8,0, 14 mM de β-mercaptoetanol e 100 μM de PMSF) e -20 °C. O nível de expressão foi verificado através de SDS-PAGE em géis de 15% acrilamida e 0,4% bisacrilamida, aplicando-se as alíquotas do controle e das amostras induzidas.

### **2.3.7.2. Purificação da proteína XAC2000**

As células de *E. coli* ressuspensas em tampão de lise e armazenadas a -20 °C foram lisadas utilizando French Press<sup>®</sup> com uma pressão de 16000 psi. O lisado obtido foi centrifugado a 5000 rpm a 4 °C por 1 hora. Ao sobrenadante obtido foi adicionado 50% de sulfato de amônio e a mistura agitada em gelo por 30 minutos seguido de centrifugação a 10000 rpm por 30 minutos a 4 °C. O precipitado obtido após a centrifugação foi ressuspensado em tampão 50 mM Tris-HCl pH 7,0; 1 mM EDTA; 14 mM β-mercaptoetanol e dialisado em membranas com um “cutoff” de 6-8 kDa contra o mesmo tampão. Após a diálise, foi adicionada uréia sólida para uma concentração final de 8 M, e a solução foi aplicada em uma coluna de troca iônica Q-Sepharose-FF (sistema Äkta, Amersham Biosciences). Proteínas ligadas à coluna foram eluídas utilizando um gradiente de eluição de zero a 300 mM NaCl, em tampão 8 M uréia; 50 mM Tris-HCl pH 7,0; 1 mM EDTA; 14 mM β-mercaptoetanol, em 12 volumes de coluna. As frações obtidas foram analisadas por SDS-PAGE utilizando géis de 15% acrilamida e 0.4% bisacrilamida, e aquelas que continham a proteína XAC2000 foram dialisadas em membranas com um “cutoff” de 12-16 kDa contra 50 mM Tris-HCl pH 7,0; 0,5 mM EDTA; 14 mM β-mercaptoetanol e

precipitadas com 80% de sulfato de amônio. O precipitado obtido foi ressuspendido em 5mL de 50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7,0; 100 mM KCl; 0,5 mM EDTA; 1 mM DTT e aplicado em uma coluna de gel filtração Superdex-75 prepgrade (sistema Äkta, Amersham Biosciences) e eluído utilizando 1,5 volumes de coluna do mesmo tampão. As frações puras da proteína XAC2000, verificadas em géis SDS-PAGE, foram dialisadas em membranas com um “cutoff” de 12-16 kDa contra 10 mM NH<sub>4</sub>CO<sub>3</sub> e 14 mM β-mercaptoetanol, congeladas a -80 °C e liofilizadas. O liofilizado obtido foi armazenado a 4 °C.

### **2.3.7.3. Purificação da proteína XAC3873**

As células de *E. coli* ressuspendidas em tampão de lise e armazenadas a -20 °C foram lisadas utilizando French Press<sup>®</sup> com uma pressão de 16000 psi. O lisado obtido foi centrifugado a 5000 rpm a 4 °C por 1 hora. Ao sobrenadante obtido foi adicionada uréia sólida para uma concentração final de 8 M. Esta solução foi aplicada em uma coluna de troca iônica Q-Sepharose-FF (sistema Äkta, Amersham Biosciences). Proteínas ligadas à coluna foram eluídas utilizando um gradiente de eluição de zero a 300 mM NaCl, em tampão 8 M uréia, 50 mM Tris-HCl pH 7,0, 1 mM EDTA, 14 mM β-mercaptoetanol, em 12 volumes de coluna. As frações obtidas foram analisadas por SDS-PAGE utilizando géis de 15% acrilamida e 0.4% bisacrilamida, e aquelas que continham a proteína de interesse foram dialisadas em membranas com um “cutoff” de 12-16 kDa contra 50 mM Tris-HCl pH 7,0; 0,5 mM EDTA; 14 mM β-mercaptoetanol e precipitadas com 80% de sulfato de amônio. O precipitado obtido foi ressuspendido em 5mL de 50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7,0, 100 mM KCl, 0,5 mM EDTA, 1 mM DTT e aplicado em uma coluna de gel filtração Superdex-75 prepgrade (sistema Äkta, Amersham Biosciences) e eluído utilizando 1,5 volumes

de coluna do mesmo tampão. As frações puras da proteína XAC3873, verificadas em géis SDS-PAGE, foram dialisadas em membranas com um “cutoff” de 12-16 kDa contra 10 mM  $\text{NH}_4\text{CO}_3$  e 14 mM  $\beta$ -mercaptoetanol, congeladas a  $-80\text{ }^\circ\text{C}$  e liofilizadas. O liofilizado obtido foi armazenado a  $4\text{ }^\circ\text{C}$ .

#### **2.3.7.4. Purificação da proteína XAC2396**

As células de *E. coli* ressuspendidas em tampão de lise e armazenadas a  $-20\text{ }^\circ\text{C}$  foram lisadas utilizando French Press<sup>®</sup> com uma pressão de 16000 psi. O lisado obtido foi centrifugado a 5000 rpm a  $4\text{ }^\circ\text{C}$  por 1 hora. O sobrenadante obtido foi precipitado com 75% etanol, deixado em gelo por duas horas e centrifugado novamente por mais 15 min. O novo sobrenadante foi novamente precipitado com 75% etanol e 14 mM  $\beta$ -mercaptoetanol e centrifugado nas mesmas condições por mais 20 min. O precipitado final obtido foi ressuspendido em 50mL de 50 mM Tris-HCl pH 7,5, 1M KCl, 14 mM  $\beta$ -mercaptoetanol e 0,1 mM de PMSF, submetido a banho de gelo sob agitação lenta por 4 horas e centrifugado novamente a 5000 rpm a  $4\text{ }^\circ\text{C}$  por 20 min. O precipitado obtido foi ressuspendido em 50 mM Tris-HCl pH 7, 14 mM  $\beta$ -mercaptoetanol, 1 mM EDTA e 8M uréia e aplicado em uma coluna de troca iônica Q-Sepharose-FF (sistema Äkta, Amersham Biosciences). Proteínas ligadas à coluna foram eluídas utilizando um gradiente de eluição de zero a 300 mM NaCl, em tampão 8 M uréia, 50 mM Tris-HCl pH 7,0, 1 mM EDTA, 14 mM  $\beta$ -mercaptoetanol, em 12 volumes de coluna. As frações obtidas foram analisadas por SDS-PAGE utilizando géis de 15% acrilamida e 0.4% bisacrilamida, e aquelas que continham a proteína de interesse foram dialisadas em membranas com um “cutoff” de 12-16 kDa contra 50 mM Tris-HCl pH 7,0, 1 mM EDTA, 14 mM  $\beta$ -mercaptoetanol e precipitadas com 80% de sulfato de amônio. O precipitado obtido foi

ressuspendido em 5mL de 50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7,0, 100 mM KCl, 0,5 mM EDTA, 1 mM DTT e aplicado em uma coluna de gel filtração Superdex-75 pregrade (sistema Äkta, Amersham Biosciences) e eluído utilizando 1,5 volumes de coluna do mesmo tampão. As frações puras da proteína XAC2396, analisadas por SDS-PAGE utilizando géis de 15% acrilamida e 0.4% bisacrilamida, foram dialisadas em membranas com um “cutoff” de 12-16 kDa contra 50 mM Tris-HCl pH 7,0 e foram armazenadas a -20 °C. Esta proteína foi purificada em colaboração com a aluna Cristiane Rodrigues Guzzo.

#### **2.3.7.5. Purificação das proteínas XAC0862 e XACb0070**

As proteínas XAC0862 e XACb0070 foram purificadas pela aluna Ângela Mika Katsuyama utilizando métodos convencionais de lise por French Press®, precipitação por sulfato de amônio, cromatografia de troca iônica e gel filtração de forma análoga ao descrito para as demais proteínas.

## **2.4. RESULTADOS**

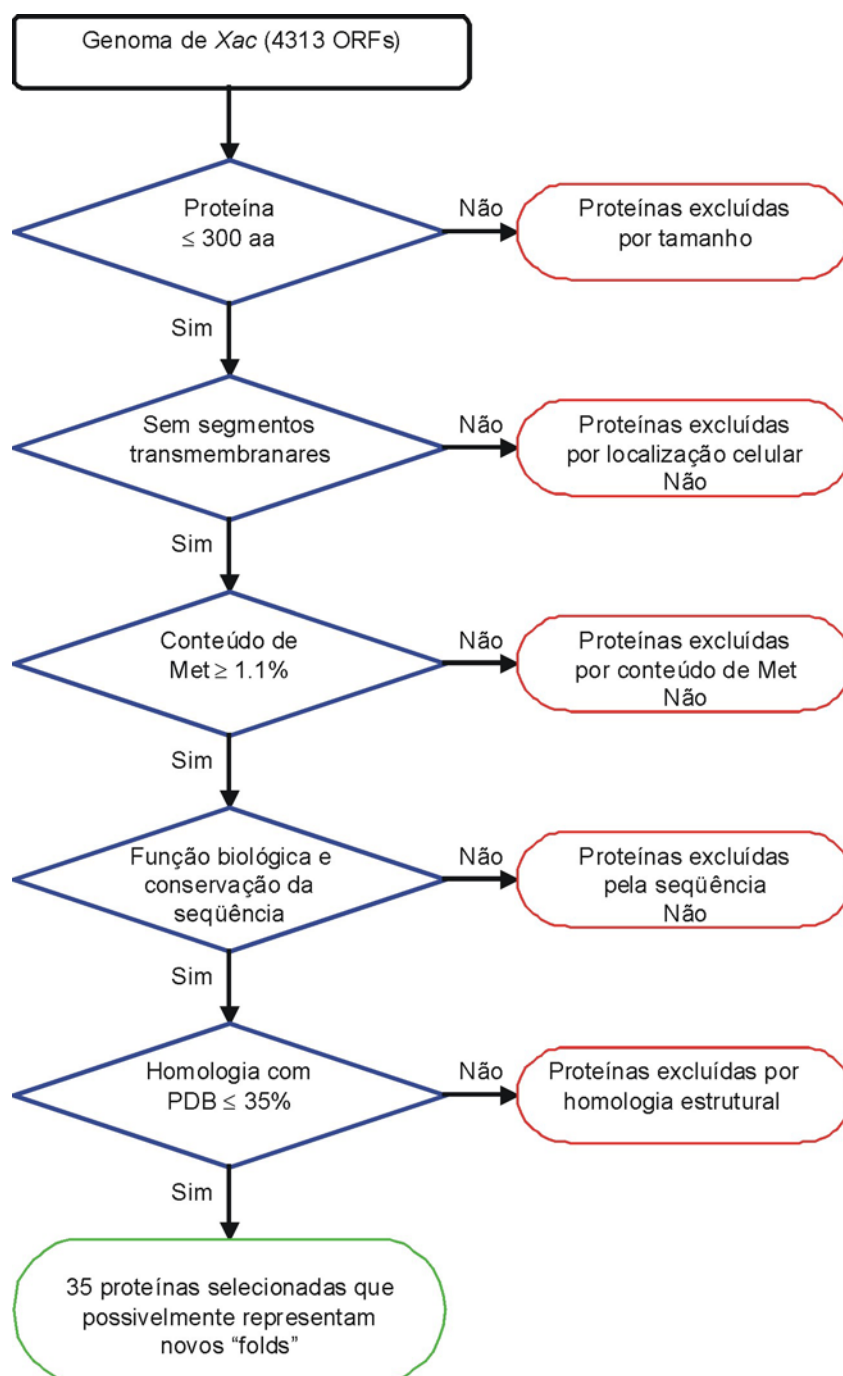
### **2.4.1. Seleção das proteínas alvo**

A seleção de proteínas alvo é um passo estratégico em um projeto de proteômica estrutural pois a seleção das proteínas alvo apropriadas é crucial para o sucesso do projeto. O nosso objetivo foi selecionar 35 ORFs de *Xac* que codificassem proteínas hipotéticas conservadas (função e estrutura desconhecidas) cujas estruturas pudessem fornecer informações sobre funções de famílias de proteínas conservadas, revelar novos padrões de enovelamento protéico (“folds”) e contribuir para o entendimento deste fitopatógeno a nível molecular. O processo de seleção de proteínas alvo para um projeto de proteômica estrutural é um processo laborioso que requer uma análise detalhada de todo o genoma. Em primeiro lugar, é necessário selecionar um número considerável de proteínas alvo para se obter um número reduzido de estruturas tridimensionais, pois muitas das proteínas selecionadas são eliminadas no decorrer do processo por não se mostrarem apropriadas. Em geral, nos casos descritos na literatura, observa-se uma “taxa de sucesso” de obtenção de estrutura tridimensional em projetos de proteômica estrutural de aproximadamente 10% a 20% (MITTL E GRÜTTER, 2001; YEE *et al.*, 2002; LESLEY *et al.*, 2002; PETI *et al.*, 2004; PAGE *et al.*, 2005; Ó'TOOLE *et al.*, 2004b). Assim, ao selecionarmos 35 proteínas alvo estávamos aumentando as chances de obtermos pelo algumas estruturas tridimensionais para as proteínas selecionadas.

Foi também necessário certificarmos-nos de que durante o processo de seleção não seriam escolhidas proteínas que fossem ser impossíveis de estudar por cristalografia de raios-X ou RMN, como por exemplo proteínas transmembranares, ou proteínas muito grandes (BRENNER, 2000). Assim, foram selecionadas apenas

proteínas com tamanho inferior a 300 amino ácidos e cuja localização celular predita fosse o citóssol (ausência de hélices transmembranares e peptídeos sinal). Para facilitar uma possível análise futura por cristalografia de raios-X, selecionamos também proteínas cujo conteúdo de metioninas fosse superior a 1.1% permitindo assim a eventual expressão de proteínas contendo selenometionina e resolução das fases por MAD (“Multiple wavelegth Anomalous Dispersion”).

Uma vez que ambicionávamos descrever novos padrões de enovelamento protéica e desta forma preencher novo espaço conformacional, priorizamos a escolha de proteínas de função e estruturas desconhecidas, e que fossem também altamente conservadas em vários genomas. O fato de serem proteínas de função e estrutura desconhecidas aumentava grandemente as chances e obtermos uma nova conformação protéica. Por outro lado, o fato de ser uma proteína conservada aumenta grandemente as chances de se tratar de uma proteína de elevada relevância biológica. Vale salientar também que, se a proteína estiver conservada em vários genomas e apresentar homologia  $\geq 35\%$  com outras proteínas, é provável que a estrutura obtida permita estimar a estrutura dos seus homólogos noutras espécies por modelagem por homologia (“homology modelling”) contribuindo assim para o aumento do número total de estruturas protéicas conhecidas. A Figura 5 resume os critérios utilizados para a seleção das proteínas alvo deste projeto. Aproximadamente 200 das proteínas codificadas pelas 4313 ORFs de *Xac* atendiam aos requisitos descritos acima. A escolha específica das 35 proteínas alvo para estudos estruturais visou uma representação bastante heterogênea em termos de conservação entre espécies e tamanho das proteínas codificadas. A Tabela 3 apresenta um resumo das características das proteínas selecionadas para este projeto.



**Figura 5.** Fluxograma esquematizando o processo de seleção das 35 proteínas alvo de *Xac* para estudos estruturais.



**Tabela 3.** Características das 35 proteínas de *Xac* selecionadas para este projeto e o estágio alcançado por cada uma das 35 proteínas durante o processo de triagem. Na coluna intitulada “Estágio da Triagem” os códigos referem-se às seguintes características apresentadas durante as etapas de triagem: proteína não expressou nas condições testadas (NX); proteína expressou mas apresentou-se insolúvel nas condições testadas (XI); proteína expressou, apresentou-se solúvel nas condições testadas e foi testada por RMN (XS-RMN). Nas que nas colunas que descrevem proteínas homólogas cujas estruturas se encontram depositadas no PDB apenas o homólogo de maior identidade é mostrado.

ORF	Estágio da Triagem	No. a.a	No. Met.	Organismo do homólogo mais próximo (2001)	% Id. Homólogo	Função Biológica Preditada (2001)	Identidade com proteínas depositadas no PDB (em 2001)	Identidade com proteínas depositadas no PDB (em 2007)
XAC0021	XI	106	2	<i>X. fastidiosa</i>	80	Proteína hipotética conservada	N/A	Carbamato Quinase de <i>Enterococcus faecalis</i> (31%)
XAC0400	XI	271	6	<i>X. campestris</i> pv. <i>vesicatoria</i>	76	Proteína associada com patogenicidade (hpa).	N/A	N/A
XAC0404	NX	204	3	<i>X. campestris</i> pv. <i>glycines</i>	98	Proteína de função desconhecida associada como operon hrp.	N/A	$\beta$ -secretase humana (27%)
XAC0862	XS-RMN	127	3	<i>X. fastidiosa</i>	73	Proteína do operon pdxA-DsgA-apaG-apaH	N/A	ApaG de <i>Xac</i> (100%) *Estrutura determinada por membros do nosso grupo
XAC1110	XS-RMN	106	8	<i>X. fastidiosa</i>	88	Proteína hipotética conservada	N/A	YbaB de <i>E. coli</i> (57%)
XAC1223	XS-RMN	165	3	<i>X. fastidiosa</i>	70	Proteína hipotética conservada	N/A	Fosforibosilpirofosfato Sintetase de <i>Bacillus subtilis</i> (40%)
XAC1352	XI	166	6	<i>E. coli</i>	60	Proteína hipotética conservada	N/A	Uronate isomerase (Tm0064) de <i>Thermotoga marítima</i> (29%)
XAC1670	NX	124	6	<i>Caulobacter</i> sp	44	Proteína com domínio regulador de resposta.	“Switch” fosforilado de transdução de sinal bacteriano (29%)	Domínio ativado ef3receptor de PhoB (34%)

ORF	Estágio da Triagem	No. a.a	No. Met.	Organismo do homólogo mais próximo (2001)	% Id. Homólogo	Função Biológica Predita (2001)	Identidade com proteínas depositadas no PDB (em 2001)	Identidade com proteínas depositadas no PDB (em 2007)
XAC1747	XI	233	4	<i>Synechocystis</i> sp.	52	Proteína hipotética conservada	N/A	YhhW de <i>E. coli</i> (45%)
XAC1757	XS-RMN	160	3	<i>X. fastidiosa</i>	72	Proteína co-migratoria de bacterioferritina	Triparedoxina peroxidase de <i>C. fasciculata</i> (31%)	Triparedoxina peroxidase (Txnpx) de <i>Trypanosoma cruzi</i> (31%)
XAC1883	XI	82	2	<i>P. abyssi</i>	40	Proteína de função desconhecida. Associada com um cluster de virulência.	Regulador da transcrição ABRB de <i>B. subtilis</i> (35%)	Domínio N-terminal do fator de transcrição AbrB de <i>E. coli</i> (36%)
XAC1884	NX	136	5	<i>X. fastidiosa</i>	40	Proteína associada com o cluster rpf.	N/A	AmpT de <i>Thermus thermophilus</i> (36%)
XAC2000	XS-RMN	106	5	<i>X. fastidiosa</i>	71	Proteína hipotética conservada	N/A	ClpS de <i>E. coli</i> (54%)
XAC2218	XS-RMN	130	4	<i>X. fastidiosa</i>	70	Proteína hipotética conservada	N/A	Quinohemoproteína Alcool desidrogenase Adhiig de <i>Pseudomonas putida</i> (30%)
XAC2237	XS-RMN	233	4	<i>X. fastidiosa</i>	72	Proteína hipotética conservada	N/A	Xantina desidrogenase de <i>Rhodobacter capsulatus</i> (33%)
XAC2355	XS-RMN	145	3	<i>X. fastidiosa</i>	80	Proteína hipotética conservada	N/A	N/A
XAC2375	XI	229	4	<i>X. fastidiosa</i>	65	Proteína hipotética conservada	N/A	N/A
XAC2396	XS-RMN	182	7	<i>E. coli</i>	40	Proteína hipotética conservada	N/A	YaeQ de <i>Xac</i> (100%) *Estrutura determinada por membros do nosso grupo
XAC2462	XS-RMN	328	5	<i>X. fastidiosa</i>	77	Proteína que se liga a ATP. Proteína de transporte de maltose.	MALK – Proteína que liga açúcar (17%)	Domínio central da ATPase Swi2SNF2 de <i>Sulfolobus solfataricus</i> (24%)

ORF	Estágio da Triagem	No. a.a	No. Met.	Organismo do homólogo mais próximo (2001)	% Id. Homólogo	Função Biológica Predita (2001)	Identidade com proteínas depositadas no PDB (em 2001)	Identidade com proteínas depositadas no PDB (em 2007)
XAC2573	NX	153	3	<i>Mesorhizobium sp</i>	44	Proteína associada com o operon gum.	N/A	Proteína hipotética Q8u9w de <i>A. tumefaciens</i> (46%)
XAC2775	XS-RMN	156	6	<i>X. fastidiosa</i>	74	Proteína hipotética conservada	N/A	YbeA de <i>E. coli</i> (50%)
XAC2942	XI	298	3	<i>B. bronchiseptica</i>	64	Proteína hipotética conservada	N/A	Fator de Iniciação de replicação de DNA (23%)
XAC2956	XS-RMN	233	2	<i>X. fastidiosa</i>	70	Proteína relacionada à replicação. Proteína que se liga a DNA.	<i>T. maritime</i> RUVBA156S (mutante) (16%)	<i>T. maritime</i> RUVBA156S (mutante) (16%)
XAC3069	XS-RMN	244	3	<i>X. fastidiosa</i>	61	Proteína hipotética conservada	N/A	N/A
XAC3146	XS-RMN	152	6	<i>X. fastidiosa</i>	73	Proteína associada com o cluster tol.	4-hydroxibenzoil coA tioesterase de <i>Pseudomonas sp</i> (21%)	Hypothetical Protein Ec709 From <i>Escherichia Coli</i> 44
XAC3151	XI	242	8	<i>X. fastidiosa</i>	78	Proteína hipotética conservada	N/A	YebC de <i>E. coli</i> (56%)
XAC3410	XS-RMN	147	5	<i>X. fastidiosa</i>	75	Proteína hipotética conservada	N/A	Proteína hipotética conservada de <i>Pseudomonas Syringae</i> Psr62 (51%)
XAC3671	XS-RMN	161	2	<i>P. aeruginosa</i>	54	Proteína hipotética conservada	N/A	YajQ de <i>E. coli</i> (46%)
XAC3725	XS-RMN	168	2	<i>S. typhimurium</i>	69	Proteína hipotética conservada	N/A	YciF de <i>E. coli</i> (65%)
XAC3726	XI	299	14	<i>P. aeruginosa</i>	67	Proteína hipotética conservada	N/A	Manganese catalase de <i>Lactobacillus plantarum</i> (33%)
XAC3872	XS-RMN	148	9	<i>X. fastidiosa</i>	72	Proteína hipotética conservada	N/A	YqeY de <i>E. coli</i> (37%)

ORF	Estágio da Triagem	No. a.a	No. Met.	Organismo do homólogo mais próximo (2001)	% Id. Homólogo	Função Biológica Predita (2001)	Identidade com proteínas depositadas no PDB (em 2001)	Identidade com proteínas depositadas no PDB (em 2007)
XAC4145	XI	167	2	<i>P. aeruginosa</i>	47	Proteína de função desconhecida em um cluster de 7 proteínas hipotéticas conservadas	N/A	Proteína co-regulada com hemolisina de <i>P. aeruginosa</i> (41%)
XAC4147	XI	169	6	<i>P. aeruginosa</i>	73	Proteína de função desconhecida em um cluster de 7 proteínas hipotéticas conservadas	N/A	N/A
XAC4330	XI	184	6	<i>N. meningitis</i>	33	Proteína hipotética conservada	N/A	Rop4 <i>E. coli</i> (30%)
XACb0070	XS-RMN	79	4	<i>X. fastidiosa</i>	73	Proteína hipotética conservada	N/A	N/A

### 2.4.2. Amplificação e clonagem das ORFs alvo

Todas as 35 ORFs de *Xac* selecionadas como alvos para estudos estruturais foram amplificadas por PCR a partir de DNA genômico e clonadas com sucesso no vetor de expressão pET-3a conforme descrito na seção 2.3.2.

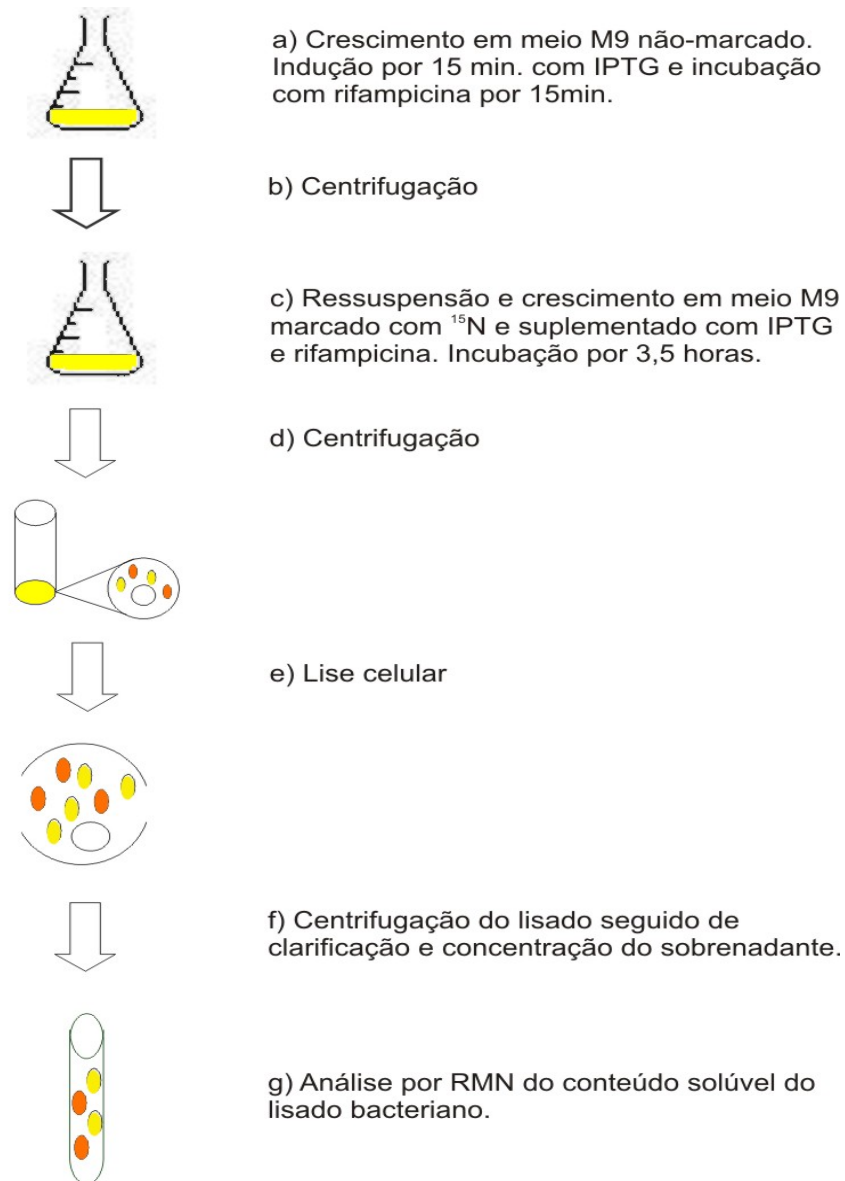
### 2.4.3. Testes de expressão e solubilidade das proteínas alvo

Tal como esperado, os níveis de expressão observados variaram entre as diferentes ORFs e apenas 31 das 35 ORFs selecionadas expressaram proteínas de tamanho esperado em células de *E. coli* BL21(DE3)pLysS crescidas em meio rico 2xTY. Das 31 proteínas expressas, apenas 19 permaneceram na fração solúvel do lisado. Na Tabela 3 pode ser vista uma listagem das 35 proteínas de *Xac* selecionadas para este estudo contendo uma descrição quanto à expressão e solubilidade apresentadas nestes testes.

### 2.4.4. Marcação seletiva com $^{15}\text{N}$ das proteínas alvo

As 19 proteínas solúveis (vide Tabela 3) foram marcadas seletivamente com  $^{15}\text{N}$  utilizando uma adaptação do protocolo descrito por ALMEIDA e colaboradores (2001) conforme descrito na seção 2.3.5. Esta metodologia faz uso do antibiótico rifampicina (inibidor da polimerase bacteriana) de forma que apenas proteínas clonadas em vetores, tais como os vetores pET (STUDIER *et al.*, 1990) que expressam proteínas heterólogas utilizando promotores de RNA polimerase do fago T7 sejam marcadas com o isótopo de interesse (Figura 6).

Os vetores da família pET (“Plasmid for Expression by T7 RNA polymerase”) possibilitam a clonagem de DNAs alvo sob controle do promotor  $\Phi 10$  do bacteriófago T7.



**Figura 6.** Marcação seletiva de proteínas heterólogas usando rifampicina.

Células de *E.coli* transformadas com o clone de interesse são crescidas em meio M9 não marcado até  $\text{D.O.}_{600\text{nm}} = 0,8$ , quando são induzidas com IPTG por 15 minutos seguido de incubação com rifampicina por 15 minutos (a). Após centrifugação (b), as células são ressuspensas em meio M9 contendo  $^{15}\text{N}$  e suplementado com rifampicina e IPTG. As células são induzidas por 3,5 horas (c) e novamente centrifugadas (d). Após lise celular por sonicação (e), o lisado é centrifugado e o sobrenadante obtido clarificado e concentrado por filtração utilizando um filtro centricom (f). O conteúdo solúvel do lisado bacteriano contendo proteínas de *E. coli* não marcadas (círculos amarelos) e a proteína alvo marcada com  $^{15}\text{N}$  (círculos laranja) é analisado por RMN

A RNA polimerase de T7 é altamente seletiva para promotores específicos que são raramente encontrados em DNA não relacionado com o fago T7. Assim, o gene clonado em um vetor sob controle do promotor  $\Phi 10$  de T7, só é expresso na presença de RNA polimerase de T7. Para expressar genes clonados nestes vetores são utilizadas cepas de *E. coli* como a cepa BL23(DE3) (STUDIER *et al.*, 1990), um lisógeno que contém o gene da RNA polimerase de T7 controlado pelo promotor induzível *lacUV5*. Esta cepa possibilita a regulação da expressão dos genes de interesse através da adição de IPTG que induz expressão da RNA polimerase de T7 e portanto a expressão dos genes clonados sob o controle do promotor  $\Phi 10$  em vetores pET. A rifampicina, por seu turno, é um forte inibidor de polimerases bacterianas que não inibe a polimerase do bacteriófago T7. Assim, proteínas heterólogas clonadas sobre controle de promotores de T7, como no caso dos vetores pET, não terão a sua expressão inibida na presença de rifampicina. Deste modo, a indução com IPTG da expressão de um gene clonado em um vetor pET na presença de rifampicina, resultará na expressão seletiva da proteína heteróloga.

No protocolo utilizado neste trabalho, as células de *E. coli* transformadas com a ORF de interesse clonada no vetor pET-3a são inicialmente crescidas em meio mínimo M9 contendo  $^{14}\text{N}$ . Ao atingir D.O.<sub>600nm</sub> = 0.8 as células são induzidas pela adição de IPTG que promove a síntese da RNA polimerase de T7. Após 15 minutos de indução é adicionado o antibiótico rifampicina que inibe a síntese de todas as proteínas de *E. coli* mas permite a síntese da proteína heteróloga de interesse. Após mais 15 minutos de indução, as células são centrifugadas e ressuspendidas em meio mínimo M9 contendo  $^{15}\text{NH}_4\text{Cl}$ . A este meio são também adicionados a rifampicina e o IPTG. Ao adicionar estes dois compostos ao meio marcado, observa-se assim uma continuação da inibição da síntese das proteínas de *E. coli*, enquanto

que a expressão da proteína de interesse matem-se ativada. Desta forma, apenas a proteína de interesse e possivelmente alguns amino ácidos livres serão marcados com o recém adicionado  $^{15}\text{N}$ . Desta forma, ao coletar um espectro de RMN 1D de  $^1\text{H}$  editado para  $^{15}\text{N}$  ou 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC das proteínas solúveis presentes no lisado bacteriano, este corresponderá apenas à proteína de interesse, uma vez que esta é a única proteína presente na mistura marcada com  $^{15}\text{N}$ . Deste modo, a marcação seletiva com rifampicina permite que o primeiro espectro de RMN seja obtido imediatamente após crescimento, lise celular e filtração, de forma eficiente e barata, sem a necessidade de recorrer aos processos laboriosos de purificação (Figura 6).

#### 2.4.5. Triagem por RMN das proteínas alvo

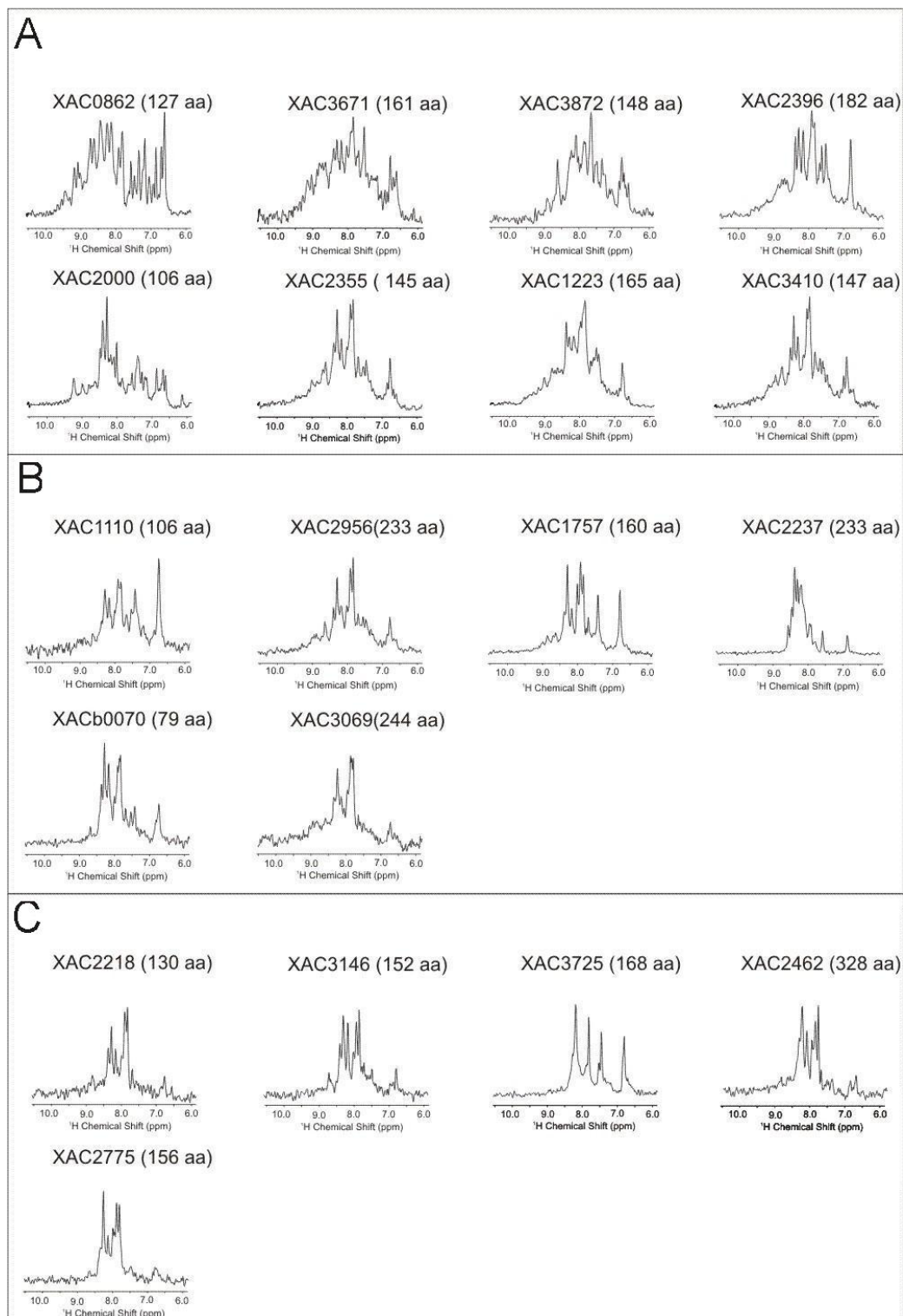
As frações solúveis dos 19 lisados bacterianos contendo as proteínas heterólogas de interesse marcadas seletivamente com  $^{15}\text{N}$  foram rapidamente analisadas por RMN através de espectros 1D de  $^1\text{H}$  editados para  $^{15}\text{N}$  e 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC. A Figura 7 mostra os espectros 1D de  $^1\text{H}$  editados para  $^{15}\text{N}$  obtidos para todas as 19 proteínas de *Xac* analisadas neste estudo. As proteínas foram agrupadas de acordo com a qualidade dos espectros obtidos: espectros “bons”, “promissores” e “pobres”. Os espectros foram classificados como “bons” quando apresentavam picos estreitos, intensos e com boa dispersão do deslocamento químico, característicos de proteínas bem estruturadas e estáveis (Figura 7A). Espectros que apresentavam picos mais largos, menos intensos e uma dispersão média de deslocamento químico foram classificados como “promissores” (Figura 7B). Este tipo de espectro pode estar relacionado com proteínas com pouca diversidade estrutural ou até heterogeneidade conformacional relacionada com processos dinâmicos. Espectros “pobres” foram caracterizados por um aglomerado



de picos de linhas largas no centro do espectro (baixa dispersão do deslocamento químico) (Figura 7C). Estes espectros provavelmente representam proteínas desenoveladas, de conformações instáveis ou agregadas.

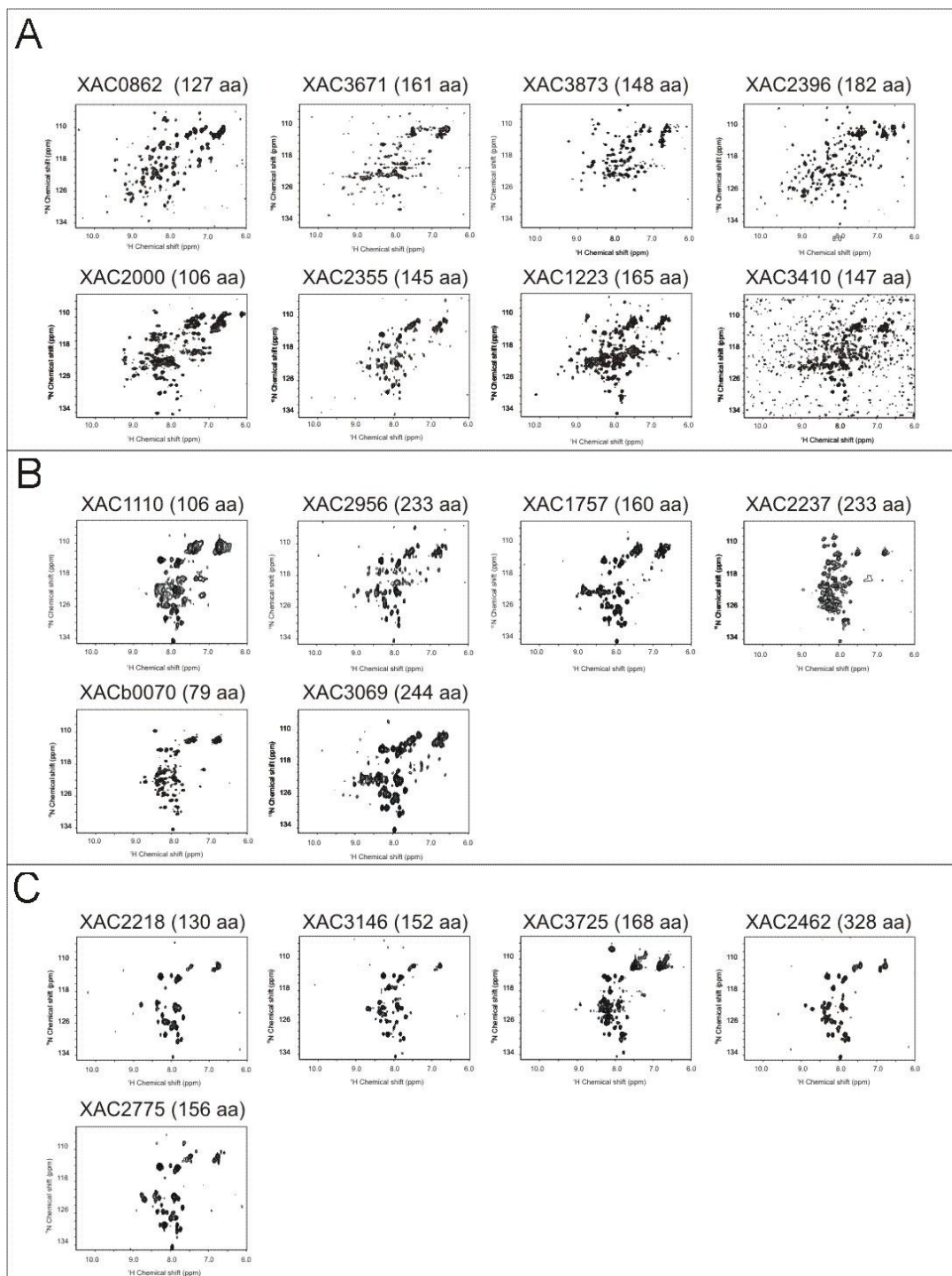
Os espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC das 19 proteínas solúveis são apresentados na Figura 8 e de novo ordenados de acordo com a qualidade dos espectros. As conclusões derivadas dos espectros 2D são consistentes com aquelas obtidas nos experimentos 1D. No entanto, os espectros 2D nos fornecem uma informação adicional importante, o número de picos esperado em um espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC para cada proteína de acordo com a sua seqüência primária. Como os espectros 2D são mais informativos que os espectros 1D, a classificação final dos espectros apresentada nas Figuras 7 e 8 foi baseada principalmente nos espectros 2D.

A Figura 9 compara os espectros 1D de  $^1\text{H}$ , 1D de  $^1\text{H}$  editado para  $^{15}\text{N}$  e 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC dos lisados bacterianos após expressão de duas proteínas: XAC2775 (Figura A,D,G) e XAC0862 (Figura B,E,H) que apresentaram espectros “*pobres*” e “*bons*” respectivamente. A Figura também mostra os espectros 1D de  $^1\text{H}$ , 1D de  $^1\text{H}$  editado para  $^{15}\text{N}$  e 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da proteína XAC0862 purificada (Figura 9 C,F,I). Os espectros 1D  $^1\text{H}$  das duas amostras dos lisados (Figura 9 A,B) são muito similares. Este resultado é esperado uma vez que estes espectros mostram todas as ressonâncias de  $^1\text{H}$  de todos os componentes solúveis presentes no lisado bacteriano que devem ser os mesmos (à exceção da proteína heteróloga) em ambos os casos.



**Figura 7.** Espectros 1D de  $^1\text{H}$  editados para  $^{15}\text{N}$  obtidos para as 19 proteínas solúveis presentes nos lisados bacterianos e marcadas seletivamente com  $^{15}\text{N}$ .

Os espectros obtidos foram agrupados de acordo com a qualidade dos mesmos: espectros “bons” (A), “promissores” (B) e “pobres” (C). A classificação é baseada na análise da dispersão do deslocamento químico e largura das linhas. O número acima de cada espectro identifica a ORF que codifica a proteína expressa heterologicamente enquanto o número entre parênteses mostra o número de resíduos de amino ácidos da proteína (adaptado de GALVÃO-BOTTON *et al.*, 2003).



**Figura 8.** Espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC obtidos para as 19 proteínas solúveis presentes nos lisados bacterianos e marcadas seletivamente com  $^{15}\text{N}$ .

Os espectros obtidos foram agrupados de acordo com a qualidade dos mesmos: espectros “bons” (A), “promissores” (B) e “pobres” (C). A classificação é baseada na análise da dispersão do deslocamento químico, largura das linhas e número de picos observados. Os espectros são apresentados na mesma ordem que na Figura 7. O número acima de cada espectro identifica a ORF que codifica a proteína expressa heterologicamente enquanto o número entre parênteses mostra o número de resíduos de amino ácidos da proteína (adaptado de GALVÃO-BOTTON *et al.*, 2003).

No entanto, observando os espectros 1D de  $^1\text{H}$  editados para  $^{15}\text{N}$  dos dois lisados, os espectros apresentam picos em menor quantidade e mais bem definidos que correspondem apenas aos prótons amídicos das proteínas heterólogas marcadas seletivamente com  $^{15}\text{N}$  (Figura 9 D,E). É importante notar que os espectros 1D de  $^1\text{H}$  editados para  $^{15}\text{N}$  das duas proteínas presentes nos lisados bacterianos são claramente diferentes um do outro, refletindo características estruturais distintas das duas proteínas. O espectro da proteína XAC2775 (Figura 9 D) apresenta baixa dispersão de deslocamento químico com picos  $\text{H}_\text{N}$  entre 7.5 e 8.5 ppm, o que é característico de proteínas desenoveladas. Por outro lado, a proteína XAC0862 (Figura 9 E) apresenta uma vasta dispersão de deslocamento químico (6.5-9.5 ppm), indicativo de uma proteína bem enovelada. Os espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC dos dois lisados (Figura 9 G,H) confirmam estas observações. Especificamente, o espectro 2D do lisado contendo a proteína XAC2775 (Figura 9 G) contém um pequeno número de picos largos e sobrepostos enquanto espectro 2D do lisado contendo a proteína XAC0862 (Figura 9 H) contém picos estreitos e definidos. Além disso, o número de picos observado no espectro 2D da proteína XAC0862 (143 picos) equivale ao número de picos esperado com base na seqüência primária desta proteína.

Analisando os espectros da proteína XAC0862 presente no lisado bacteriano (Figura 9 B,E,H) e purificada (Figura 9 C,F,I) podemos observar que estes são bastante semelhantes entre si. Assim, torna-se evidente que o antibiótico rifampicina permite de fato a marcação seletiva de proteínas heterólogas conforme demonstrado anteriormente (ALMEIDA *et al.*, 2001), e que os espectros dos lisados bacterianos contendo proteínas heterólogas marcadas com  $^{15}\text{N}$  contêm informações importantes sobre a conformação da proteína de interesse.

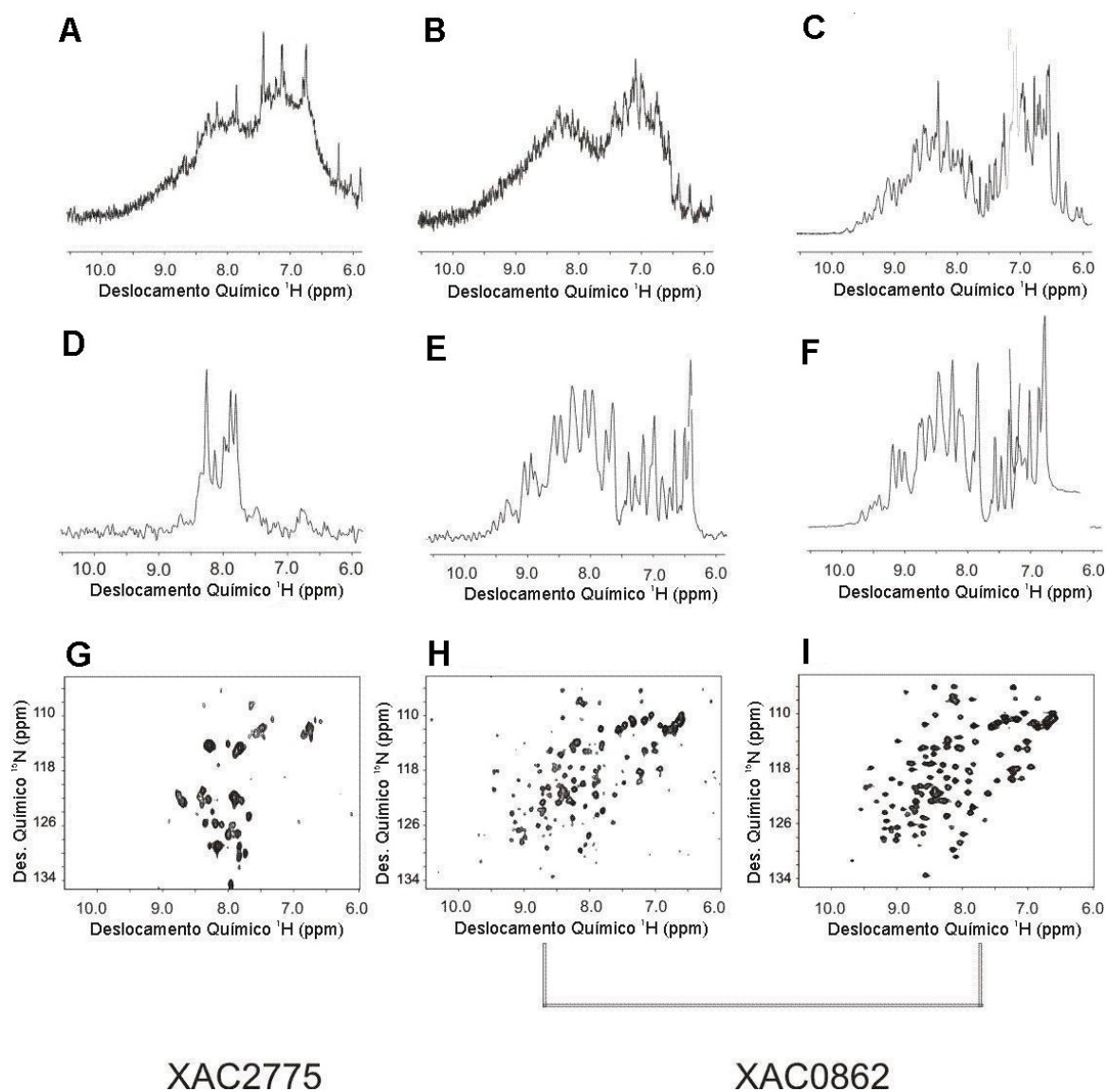
Além da proteína XAC0862, outras três proteínas classificadas como “boas” candidatas para estudos estruturais e uma proteína classificada como “promissora” foram marcadas com  $^{15}\text{N}$  e purificadas. Os espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC das quatro proteínas purificadas são mostrados na Figura 10.

Conforme esperado, os espectros das proteínas XAC2000, XAC2396 e XAC3873 purificadas (Figura 10), classificadas como “boas” durante a triagem, são muito semelhantes aos espectros correspondentes dos lisados bacterianos (Figura 8). No caso das proteínas XAC2000 e XAC2396 o número de picos observado no espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da proteína purificada aproxima-se bastante ao número de picos esperado a partir da seqüência primária destas proteínas. Os espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da proteína XAC3873 presente no lisado e purificada também são bastantes similares e apresentam boa dispersão de picos.

No entanto, neste caso apenas 102 dos 164 picos esperados são observados tanto no espectro do lisado quanto no espectro da proteína purificada.

Este tipo de espectro, com linhas estreitas e boa dispersão de deslocamento químico mas número de picos  $H_N$  inferior ao esperado, pode ser indicativo de sobreposição de picos ou a efeitos dinâmicos de troca conformacional que levam ao alargamento das linhas nesta região.

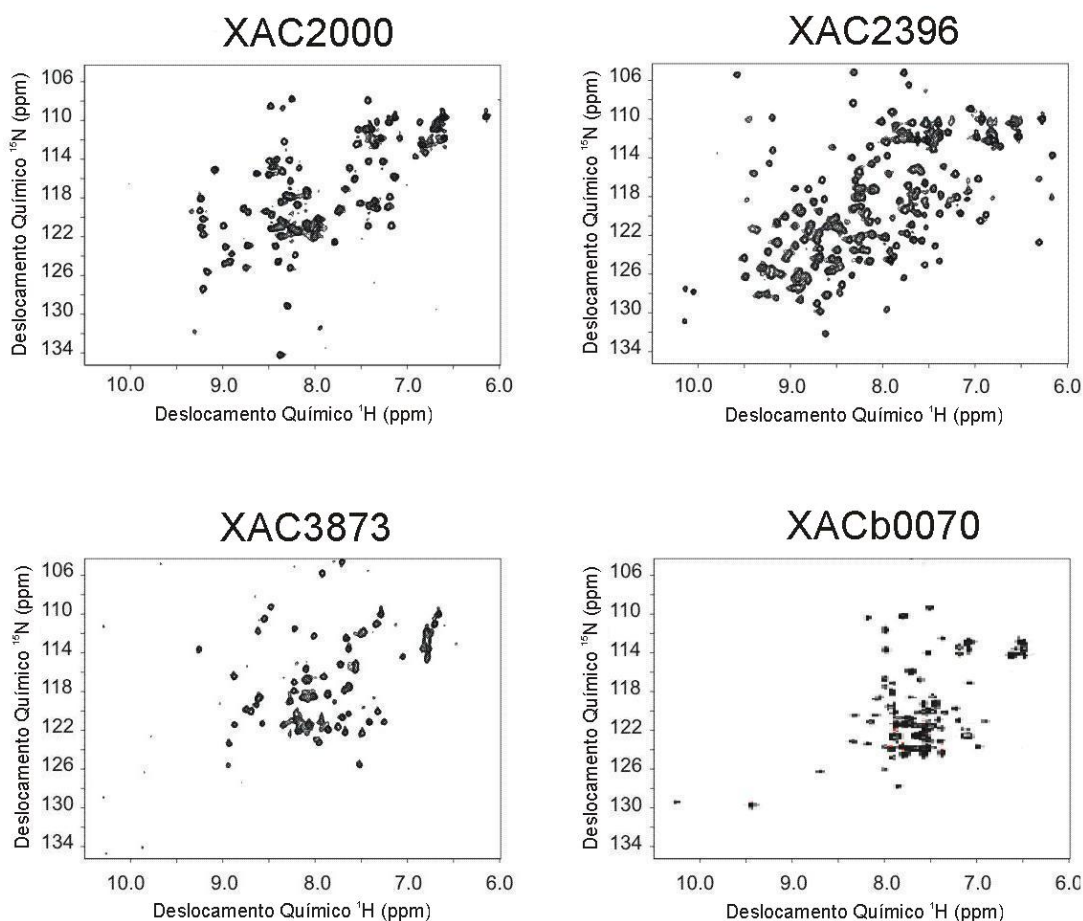
No caso de ambigüidade, proteínas bem enoveladas poderiam ser distinguidas de proteínas parcialmente enoveladas através seqüências de pulso que identificam prótons amídicos presentes em regiões não-estruturadas, que não formam pontes de hidrogênio e apresentam troca rápida com o solvente aquoso (DALVIT e HOMMEL, 1995; HWANG *et al.*, 1997; HWANG *et al.*, 1998).



**Figura 9.** Espectros 1D de  $^1\text{H}$  (A,B,C), 1D de  $^1\text{H}$  editados para  $^{15}\text{N}$  (D,E,F) e 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (G,H,I) relativos às proteínas codificadas pelas ORFs XAC2775 e XAC0862.

Os espectros A,D e G correspondem aos lisados solúveis contendo a proteína codificada pela ORF XAC2775 marcada seletivamente com  $^{15}\text{N}$ . Os espectros B,E e H correspondem aos lisados solúveis contendo a proteína codificada pela ORF XA0862 marcada seletivamente com  $^{15}\text{N}$ . Os espectros C,F e I correspondem à proteína codificada pela ORF XA0862 marcada com  $^{15}\text{N}$  e purificada (adaptado de GALVÃO-BOTTON *et al.*, 2003).

No caso da proteína “*promissora*” XACb0070, que apresentou linhas estreitas e baixa dispersão do deslocamento químico, o espectro da proteína purificada (Figura 10) também é similar ao espectro do lisado e o número de picos observado é próximo ao número esperado (Figura 8). A baixa dispersão do deslocamento químico observada é provavelmente devido a pouca diversidade estrutural. De fato, estudos posteriores desta proteína revelaram um elevado conteúdo helicoidal da mesma (KATSUYAMA, comunicação pessoal) confirmando o caráter promissor desta proteína para estudos estruturais.



**Figura 10.** Espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC obtidos para as proteínas codificadas pelas ORFs XAC2000, XAC2396, XAC3873 e XACb0070, marcadas com  $^{15}\text{N}$  e purificadas. Os espectros foram adquiridos a  $25^\circ\text{C}$  e pH 7,0 (adaptado de GALVÃO-BOTTON *et al.*, 2003)

Estes resultados são consistentes com a hipótese de que a maioria das proteínas classificadas por esta metodologia como “boas”, e possivelmente também as “promissoras”, podem ser de fato proteínas bem estruturadas no lisado bacteriano que mantêm o seu estado enovelado durante os procedimentos de purificação. De fato, várias das proteínas aqui descritas foram purificadas usando uréia o que demonstra que estas proteínas além de serem bem enoveladas, re-enovelam facilmente uma vez que os espectros do lisado e das proteínas puras correspondentes são muito semelhantes.



## 2.5. DISCUSSÃO

Uma vez que nem todas as proteínas são adequadas para estudos estruturais devido a problemas de solubilidade e estabilidade conformacional, é desejável que se desenvolvam metodologias que aumentem a eficácia do processo de seleção de proteínas alvo. A metodologia de triagem aqui proposta, combina marcação seletiva com  $^{15}\text{N}$  seguida de triagem rápida por RMN e permite avaliar os níveis de solubilidade, estabilidade e grau de enovelamento de uma dada proteína alvo, antes mesmo da purificação da mesma.

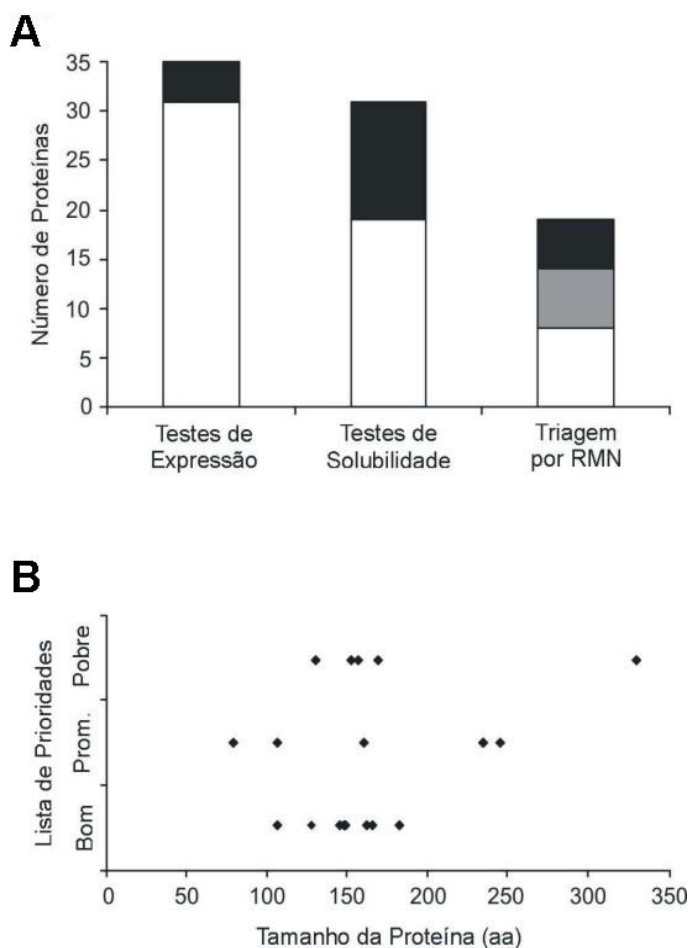
A marcação seletiva de proteínas heterólogas com  $^{15}\text{N}$  usando rifampicina foi originalmente descrita e testada utilizando proteínas cujas estruturas e comportamentos em solução eram previamente conhecidos (ALMEIDA *et al.*, 2001). Não se sabia se esta metodologia seria eficiente na triagem de alvos de projetos de proteômica estrutural desconhecidos. Para determinar a eficácia desta metodologia para realizar a triagem de alvos de proteômica estrutural, nós analisamos 35 proteínas de *Xac* que nunca haviam sido caracterizadas. Destas, 4 não foram expressas e 12 mostraram-se insolúveis (Figura 11 A). Das 19 proteínas solúveis testadas por RMN, 42% apresentaram espectros “bons” (branco), 32% apresentaram espectros “promissores” (cinza) e 26% apresentaram espectros “pobres” (preto) (Figura 11 A).

O objetivo deste protocolo de triagem foi criar uma lista de prioridades para ser usada como guia na escolha de proteínas apropriadas para estudos estruturais em projetos de proteômica estrutural de média ou larga escala. Uma proteína classificada como “boa” ou “promissora” durante a triagem tem grandes chances de ser uma proteína bem comportada após a purificação da mesma e ser portanto adequada para estudos estruturais (solúvel, enovelada, não-agregada). Por outro

lado, uma proteína classificada como “*pobre*” durante a triagem, será provavelmente uma má candidata para estudos estruturais.

Apesar disso, vale lembrar no entanto, que existem vários exemplos na literatura que descrevem grandes alterações de deslocamentos químicos de proteínas induzidos pela ligação de co-fatores inorgânicos ou orgânicos (YEE *et al.*, 2002; ZARTLER *et al.*, 2001; JAREN *et al.*, 2002). Estes fatores estariam presentes no lisado bacteriano mas poderiam perder-se durante o processo de purificação. Assim, proteínas classificadas como “*boas*” ou “*promissoras*” no lisado, poderiam se revelar más candidatas para estudos estruturais caso o co-fator necessário para o seu enovelamento apropriado fosse removido durante as etapas de purificação. Da mesma forma, proteínas classificadas como “*pobres*” durante a triagem podem de fato ser bem comportadas após a purificação. Este seria o caso de proteínas que interagem com componentes celulares no lisado de forma que o sinal do RMN da amostra do lisado é bastante piorado. Além disso, deve também ser notado que durante a triagem nenhuma otimização das etapas de expressão (cepa de *E. coli*, temperatura, meio de cultura) e de lise (pH, temperatura, força iônica) foi realizada antes da análise por RMN. Assim, proteínas classificadas como “*pobres*” não devem ser necessariamente descartadas. Ou seja, este protocolo serve apenas como uma indicação se uma dada proteína é ou não uma potencial boa candidata para estudos estruturais, sendo que uma análise detalhada da mesma ainda requer a sua purificação.

Uma das nossas preocupações durante a triagem foi que dado o limite de tamanho existente para a RMN, os melhores espectros obtidos correspondessem às menores proteínas testadas enquanto que os piores espectros correspondessem às maiores proteínas testadas.



**Figura 11.** Dados estatísticos dos resultados de triagem.

A. Histograma do número de proteínas expressas (branco), não-expressas (preto), solúveis (branco), não-solúveis (preto) e com espectros de RMN *bons* (branco), *promissores* (cinza) e *pobres* (preto).

B. Correlação entre classificação dos espectros obtidos (*bons*, *promissores* e *pobres*) e o tamanho das proteínas em estudo (aa = número de resíduos de amino ácidos; prom. = promissor) (adaptado de GALVÃO-BOTTON *et al.*, 2003).

No entanto, encontramos exemplos de proteínas menores e maiores produzindo todos os tipos de espectros (“bons”, “promissores” e “pobres”) (Figura 11 B). Isto demonstra que o nosso método é adequado para criar uma lista de prioridades para alvos de proteômica estrutural, independentemente do seu tamanho, desde que estes sejam suficientemente pequenos para análise por RMN. Além disso, estes dados sugerem que o tamanho de uma dada proteína não deve ser o único critério a ser considerado ao selecioná-la para estudos estruturais por RMN.

O trabalho aqui descrito foi publicado na revista *FEBS Letters* em Agosto de 2003 (GALVÃO-BOTTON *et al.*, 2003). Desde então, as restantes proteínas que haviam sido classificadas como “boas” candidatas mas ainda não haviam sido purificadas, foram também expressas em larga escala e purificadas por membros do nosso laboratório. Assim, todas as oito proteínas classificadas como “boas” candidatas, e uma proteína classificada como “promissora” prosseguiram para as etapas de purificação e estudos estruturais.

A Tabela 4 resume o progresso dos trabalhos envolvendo as proteínas de *Xac* selecionadas para estudos estruturais utilizando a metodologia de triagem aqui proposta. É interessante notar que embora estas proteínas fossem de função e estruturas desconhecidas quando foram selecionadas para este trabalho em Julho de 2001, algumas delas tiveram no decorrer do projeto, a sua anotação modificada baseada em proteínas homólogas cuja função e/ou estrutura já havia sido descrita (vide Tabela 3). Assim, algumas das proteínas cuja nomenclatura era apenas numérica e baseada na localização da ORF correspondente no genoma de *Xac*, agora têm uma nomenclatura baseada na sua função descrita e/ou na nomenclatura de seus homólogos.

Das nove proteínas selecionadas para estudos estruturais (Tabela 4), apenas uma (XAC3873) foi descartada para estudos estruturais futuros por apresentar problemas de instabilidade. Apesar desta proteína apresentar um bom espectro 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC quando purificada, ela degradava facilmente e não se mantinha estável em solução por longos períodos de tempo. Por outro lado, é interessante notar que uma outra proteína, XAC1883 que se apresentou insolúvel em nosso estudo e não foi portanto submetida à nossa triagem por RMN (vide Tabela 3), foi posteriormente purificada e está tendo a sua estrutura tridimensional determinada por RMN por membros do nosso grupo. O interesse nesta proteína surgiu devido ao seu provável envolvimento na regulação de “quorum sensing” em *Xanthomonas*.

Todas as outras proteínas selecionadas para estudos estruturais prosseguiram para etapas futuras de estudos estruturais por RMN e/ou cristalografia de raios-X, e pelo menos cinco delas tiveram ou deverão ter a sua estrutura tridimensional descrita num futuro próximo por uma das duas técnicas por membros do nosso grupo e seus colaboradores.

A proteína XAC0862 (ApaG), apresenta elevada homologia com a proteína ApaG, uma proteína de 14 kDa altamente conservada em bactérias gram-negativas cujo gene *apaG* está localizado em um operon multifuncional (*pdxA-ksgA-apaG-apaH*). Este operon codifica uma proteína envolvida na biossíntese de piridoxal fosfato (PdxA), uma 16S RNA  $m_2$  metiltransferase (KsgA), uma diadenosina (AppppA) tetrafosfatase (ApaH) além de ApaG (ROA *et al.*, 1989). Embora a função de ApaG permaneça desconhecida, estudos recentes sugeriram que ApaG e ApaH estão funcionalmente relacionadas e que esta proteína está envolvida na ligação de pirofosfatos ou nucleotídeos fosfatos (LIU *et al.*, 2003). Estudos estruturais por RMN desta proteína desenvolvidos pelo nosso grupo revelaram que esta proteína

apresenta uma estrutura formada por sete fitas  $\beta$  antiparalelas (KATSUYAMA *et al.*, 2004; CICERO *et al.*, 2007). Apesar da presença de um motivo putativo de ligação a pirofosfato (GXGXXG) a proteína ApaG de *Xac* não liga ATP ou GTP *in vitro* (CICERO *et al.*, 2007). No entanto, esta proteína adota um enovelamento de fibronectina tipo III (Fn3), o que é consistente com a hipótese de esta proteína esteja envolvida na mediação de interações proteína-proteína (CICERO *et al.*, 2007).

A proteína YaeQ (XAC2396) é um membro de uma família de proteínas conservada em vários patógenos Gram-negativos (WONG *et al.*, 1998). Estudos de YaeQ *Salmonella typhimurium* sugeriram que esta proteína está envolvida na regulação da transcrição de genes que codificam fatores de virulência tais como hemolisina e enzimas envolvidas na produção de lipopolisacarídeos (WONG *et al.*, 1998). No entanto, estudos recentes da proteína YaeQ de *E. coli* não apresentaram os mesmos resultados (VICARI e ARTSIMOVITCH, 2004). A estrutura tridimensional de YaeQ de *Xac* determinada por cristalografia de raios-X usando a técnica de MAD por membros do nosso grupo constitui a primeira estrutura tridimensional descrita de um membro da família YaeQ (GUZZO *et al.*, 2007). Uma comparação da estrutura de YaeQ com outras estruturas protéicas indica que esta proteína representa um novo enovelamento protéico (“fold”) que constitui uma variação do motivo de nucleases PD-(D/E)XK encontrado em endonucleases do tipo II e em enzimas envolvidas na replicação, reparo e recombinação de DNA sugerindo que esta proteína pode estar envolvida em processos biológicos semelhantes (GUZZO *et al.*, 2007).

A proteína SufE (XAC2355) é um membro de uma família de proteínas conservada em procariotos e eucariotos. Em *E. coli*, o cluster sufABCDSE é expresso em condições de estresse oxidativo e limitação de ferro, sendo que uma

das suas funções é a restauração de clusters Fe-S destruídos (NACHIN *et al.*, 2001; NACHIN *et al.*, 2003). Em *Xac* o cluster de genes XAC2935-XAC2936-XAC2937-XAC2938 codifica os homólogos de SufB, SufC, SufD e SufS, respectivamente, embora não tenham sido anotados como tal durante o seqüenciamento do seu genoma (DA SILVA *et al.*, 2002). Os genes que codificam os homólogos de SufA (XAC1619) e SufE (XAC2355) estão localizados em outras regiões do genoma. O nosso grupo cristalizou SufE de *Xac* com sucesso (GUZZO *et al.*, 2006) e a sua estrutura tridimensional obtida por cristalografia de raios-X está no momento sendo validada (GUZZO, comunicação pessoal).

A proteína YajQ (XAC3671) está presente em muitos genomas bacterianos e ausente em eucariotos (SAVEANU *et al.*, 2002). A estrutura tridimensional da proteína YajQ de *E. coli*, determinada por RMN, revelou uma proteína composta por dois domínios, cada um centrado em torno de uma fita  $\beta$ , que estão conectados por dois segmentos helicoidais (SAVEANU *et al.*, 2002). Até ao momento, a qualidade dos cristais obtidos de YajQ não possibilitam a resolução da sua estrutura. Espera-se que o refinamento das condições de cristalização da proteína YajQ de *Xac* possibilitem estudos estruturais futuros desta proteína por cristalografia de raios-X de forma a contribuir para o conhecimento sobre a relação estrutura-função desta proteína (GUZZO, comunicação pessoal).

A proteína XACb0070 permanece sendo de função e estrutura desconhecidas. Espera-se que os estudos estruturais desta proteína por RMN atualmente em andamento (CICERO, comunicação pessoal) possam fornecer informações a respeito da sua função à semelhança do que vem acontecendo com outros alvos de estudos de proteômica estrutural (JACKSON e RUSELL, 2001; BHATTACHARYYA *et al.*, 2002; CHRISTENDAT *et al.*, 2002; GUZZO *et al.*, 2007).

A proteína XAC2000 é hoje denominada ClpS pois no decorrer dos trabalhos desta tese, três artigos foram publicados descrevendo pela primeira vez a função (DOUGAN *et al.*, 2002a) e a estrutura cristalográfica (GUO *et al.*, 2002a; ZETH *et al.*, 2002) de uma proteína homóloga de *E. coli*. A proteína XAC2000/ClpS de *Xac* foi amplamente caracterizada por RMN em colaboração com a Dra. Thelma Pertinhez do LNLS e diversos estudos estruturais, de dinâmica e de interação desta proteína são apresentados detalhadamente no Capítulo 3.

Os resultados de estudos estruturais obtidos para as diversas proteínas selecionadas neste estudo demonstram que o nosso método é adequado para criar uma lista de prioridades para alvos de proteômica estrutural, e que a metodologia de triagem proposta é um método alternativo promissor para selecionar rápida e eficazmente alvos para estudos estruturais de alta resolução por RMN ou cristalografia de raios-X.

Recentemente, dois trabalhos sugeriram também que a metodologia aqui descrita (e publicada em GALVÃO-BOTTON *et al.*, 2003) poderia ser usada de forma ainda mais abrangente em projetos de proteômica estrutural se utilizada em conjunto com outras técnicas de RMN (MOREAU *et al.*, 2006; ETEZADY-ESFARJANI *et al.*, 2006). MOREAU e colaboradores (2006), do grupo dos Prof. Drs. Ana Paula Valente e Fábio Almeida, desenvolveram uma nova metodologia que permite prever a quantidade de estrutura secundária de uma dada proteína não-assinalada a partir dos seus espectros 2D  $^{15}\text{N}$  e/ou  $^{13}\text{C}$  HSQC usando um software chamado PASSNMR (“Prediction of the Amount of Secondary Structure by Nuclear Magnetic Resonance”). Se o PASSNMR for usado em conjunto com a marcação seletiva por rifampicina, será possível prever a quantidade de estrutura secundária de uma dada proteína, a partir dos seus espectros 2D  $^{15}\text{N}$  e/ou  $^{13}\text{C}$  HSQC, antes



mesmo da sua purificação possibilitando assim melhorar a triagem e seleção alvos de estudos de projetos proteômica estrutural.

ETEZADY-ESFARJANI e colaboradores (2006), do grupo do Prof. Dr. Kurt Wüthrich, descreveram uma metodologia que permite determinar por RMN, de forma automatizada, a estrutura de uma proteína marcada com  $^{13}\text{C}$  e  $^{15}\text{N}$  presente no extrato celular sem a respectiva purificação da mesma. Neste método, inicialmente, as ressonâncias dos vários núcleos da proteína são atribuídas usando experimentos de tripla ressonância e baseando-se na alta intensidade dos sinais das ressonâncias da proteína heteróloga quando comparados à baixa intensidade dos sinais das ressonâncias das demais proteínas presentes no lisado bacteriano. Posteriormente, as restrições de distancia são obtidas através da análise automatizada dos sinais de NOE no espectro 3D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC-NOESY usando do programas ATNOS e CANDID (HERRMANN *et al.*, 2002a,b) que permitem diferenciar os sinais de NOE da proteína de interesse dos demais sinais presentes no espectro. Esta metodologia promete ser uma poderosa ferramenta para a determinação de estruturas de proteínas presentes em extratos celulares. Os autores deste trabalho, sugeriram que aliando este método ao método de marcação seletiva por rifampicina e triagem da proteína de interesse descrito nesta tese, seria possível não só selecionar alvos para estudos de proteômica estrutural como também possivelmente determinar a estrutura tridimensional dos mesmos antes da sua purificação (ETEZADY-ESFARJANI *et al.*, 2006).

**Tabela 4.** Progresso dos trabalhos envolvendo as proteínas de *Xac* selecionadas para estudos estruturais utilizando a metodologia de triagem descrita neste trabalho.

Proteína	Responsável	Técnica utilizada	Progresso dos trabalhos
XAC2000 (ClpS)	Leonor G. de Botton	RMN	<ul style="list-style-type: none"> <li>Assinalamento concluído.</li> <li>Modelo tridimensional calculado e validado usando acoplamentos residuais dipolares obtidos experimentalmente e uma estrutura homóloga determinada por cristalografia de raios X (GUO <i>et al.</i>, 2002a).</li> <li>Experimentos de interação em andamento.</li> </ul>
XAC3873	Leonor G. de Botton	RMN	<ul style="list-style-type: none"> <li>Estudos preliminares por RMN concluídos.</li> <li>Abandonada devido a problemas de instabilidade.</li> </ul>
XAC0862 (ApaG)	Ângela Katsuyama Daniel O. Cícero Thelma Pertinhez Alberto Spisni	RMN	<ul style="list-style-type: none"> <li>Assinalamento concluído (KATSUYAMA <i>et al.</i>, 2004).</li> <li>Estrutura tridimensional determinada por RMN (CICERO <i>et al.</i>, 2007)</li> </ul>
XACb0070	Ângela Katsuyama Daniel O. Cícero Thelma Pertinhez Alberto Spisni	RMN	<ul style="list-style-type: none"> <li>Coleta de dados e assinalamento em andamento.</li> </ul>
XAC2396 (YaeQ)	Cristiane R. Guzzo	Cristalografia Técnica MAD	<ul style="list-style-type: none"> <li>Cristalização e análise preliminar por raios-X concluídas (GUZZO <i>et al.</i>, 2005)</li> <li>Estrutura tridimensional determinada (GUZZO <i>et al.</i>, 2007)</li> <li>Estudos funcionais em andamento.</li> </ul>
XAC2355 (SufE)	Cristiane R. Guzzo	Cristalografia Substituição Molecular	<ul style="list-style-type: none"> <li>Cristalização e análise preliminar por raios-X concluídas. (GUZZO <i>et al.</i>, 2006)</li> <li>Refinamento da estrutura tridimensional em andamento</li> </ul>
XAC3671 (YajQ)	Cristiane R. Guzzo	Cristalografia	<ul style="list-style-type: none"> <li>Refinamento das condições de cristalização.</li> </ul>
XAC1223	Cristiane R. Guzzo	Cristalografia	<ul style="list-style-type: none"> <li>Ensaio de cristalização sem sucesso.</li> </ul>
XAC3410	Cristiane R. Guzzo Lucleide Ribeiro	Cristalografia	<ul style="list-style-type: none"> <li>Ensaio de cristalização sem sucesso.</li> </ul>

## **CAPÍTULO 3**

### **ESTUDOS ESTRUTURAIS POR RMN DA PROTEÍNA ClpS DE *Xanthomonas axonopodis* pv *citri***

### 3.1. INTRODUÇÃO

A introdução deste capítulo está dividida em duas partes. Na primeira parte, descrita na seção 3.1.1, é feita uma introdução à proteína ClpS e ao seu envolvimento na proteólise bacteriana. Na segunda parte, descrita na seção 3.1.2, é feita uma introdução à RMN enquanto técnica utilizada na caracterização estrutural de proteínas.

#### 3.1.1. A proteína ClpS e a proteólise bacteriana

##### 3.1.1.1. Complexos proteolíticos bacterianos

A proteólise é um mecanismo preciso de regulação de um largo espectro de processos celulares. Primeiramente, este mecanismo constitui uma importante ferramenta de controle de qualidade da síntese protéica. Além disso, a proteólise tem também um papel dinâmico e vital na regulação de diferentes processos metabólicos e das respostas celulares a diferentes estímulos ambientais através da regulação precisa dos níveis das enzimas metabólicas (PORANKIEWICZ *et al.*, 1999; SAUER *et al.*, 2004). A proteólise controla ainda a remoção do ambiente celular de proteínas danificadas irreversivelmente (GOTTESMAN, 1999; WICKNER *et al.*, 1999). O surgimento destas proteínas anormais ocorre continuamente através de eventos como a desnaturação espontânea, erros biossintéticos ou mutações cumulativas e a remoção destas proteínas não-funcionais é essencial para um funcionamento celular adequado.

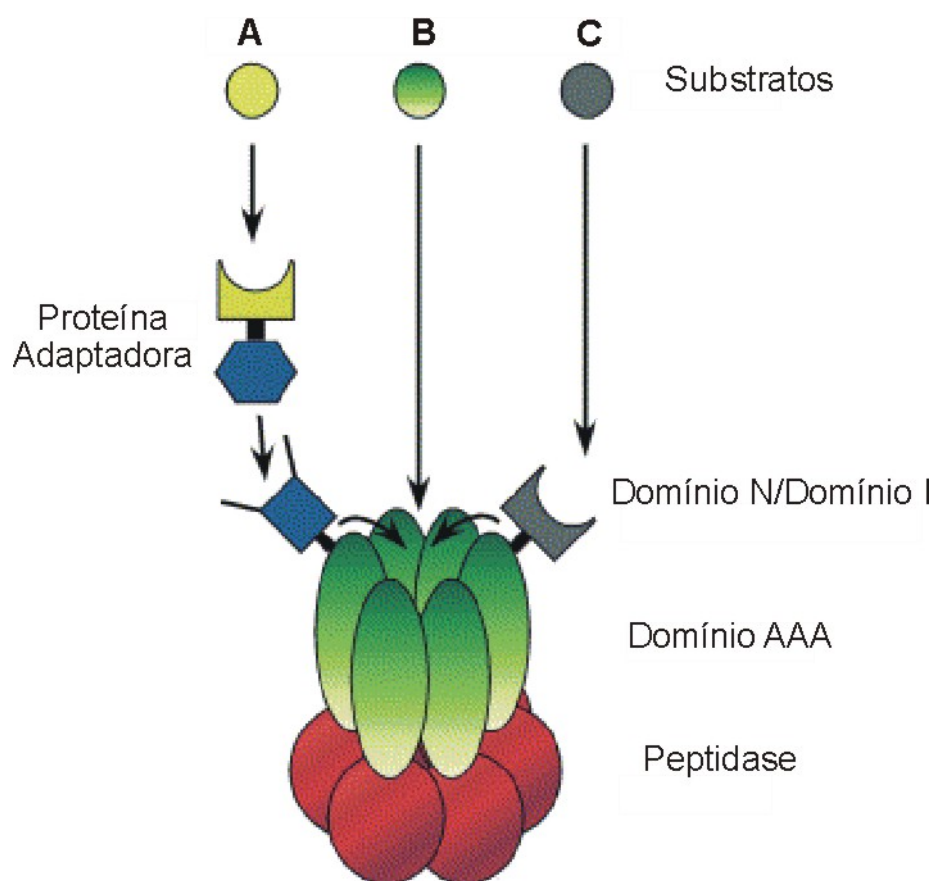
A proteólise em procariotos e eucariotos é realizada majoritariamente por proteases multiméricas dependentes de energia (“energy-dependent proteases”) (PORANKIEWICZ *et al.*, 1999). Estes complexos proteolíticos são capazes de clivar um amplo espectro de ligações peptídicas. No entanto, apenas proteínas anormais

ou proteínas nativas com tempos de vida curtos são degradadas (LUPAS *et al.*, 1997). Esta regulação específica e precisa da proteólise depende de componentes reguladores que se associam aos complexos proteolíticos e que podem incluir até 6 ATPases diferentes e outras proteínas acessórias. Estes componentes podem estar envolvidos na regulação da proteólise de diversas formas (GOTTESMAN *et al.*, 1997; LUPAS *et al.*, 1997; WICKNER *et al.*, 1999):

- i. Eles podem ligar-se a substratos específicos levando-os para as proximidades dos complexos proteolíticos;
- ii. Eles podem desenovelar o substrato para que este possa passar através dos canais de acesso dos complexos proteolíticos;
- iii. Eles podem participar ativamente da translocação do substrato para a cavidade proteolítica;
- iv. Eles podem afetar alostericamente os sítios ativos proteolíticos de forma a alterar a velocidade e/ou especificidade de clivagem.

Em células eucarióticas, a principal via de degradação protéica dependente de energia envolve a conjugação de ubiquitina às proteínas alvo a serem degradadas. Estas proteínas ubiquitinadas são então degradadas pelo complexo protéico proteasomo 26S (MAURIZI *et al.*, 1990a; SCHMIDT *et al.*, 1999; PICKART e COHEN, 2004). Em bactérias, a degradação protéica é realizada por pelo menos cinco complexos proteolíticos dependentes de adenosina trifosfato (ATP): ClpAP, ClpXP, HslUV (ClpYQ), Lon e FtsH (HOSKINS *et al.*, 2000; DOUGAN *et al.*, 2002b). Estas máquinas proteolíticas são normalmente compostas por uma protease/peptidase (ClpP ou HslV) que possui atividade proteolítica e que forma dois anéis hexa- ou heptaméricos sobrepostos em forma de barril e por uma chaperona ATPase da família AAA+ (ClpA, ClpX ou HslU) que forma um anel

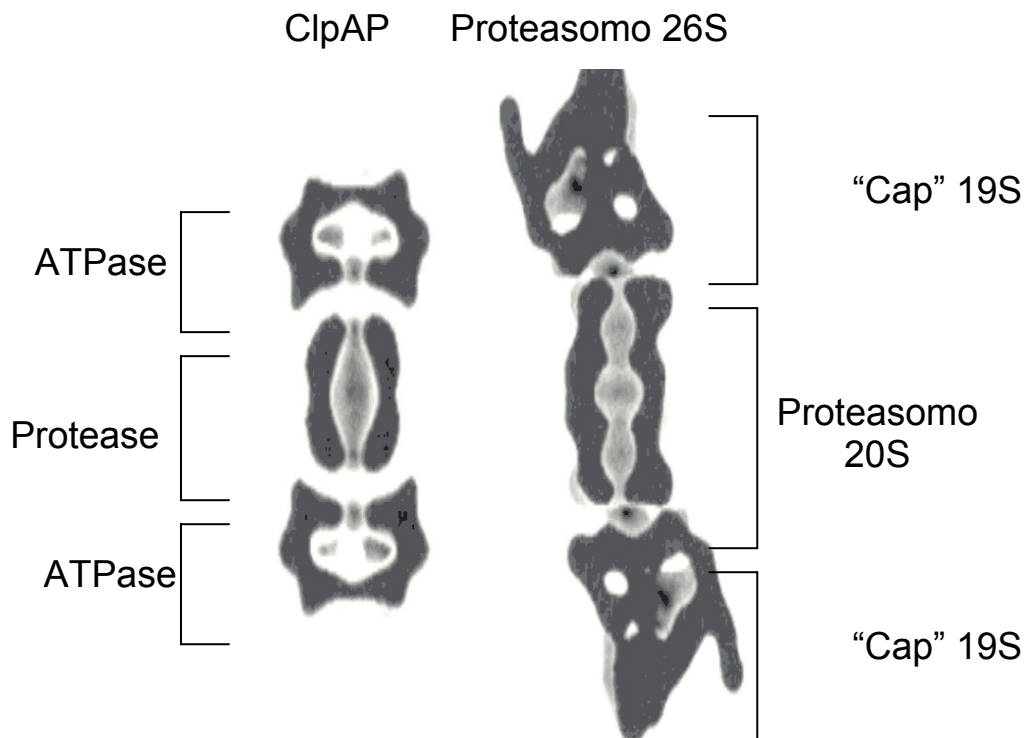
hexamérico nas duas extremidades do barril proteolítico e promove o desenovelamento do substrato antes da etapa proteolítica regulando assim a atividade da protease. Em alguns casos, como Lon e FtsH, a chaperona e a protease estão localizadas em uma única cadeia polipeptídica. Além disso, em alguns casos, os complexos protease-chaperona muitas vezes ligam-se também a outras proteínas acessórias que contribuem para a regulação precisa da proteólise (vide Figura 12) (MOGK *et al.*, 2004; BEWLEY *et al.*, 2006).



**Figura 12.** Representação esquemática da arquitetura dos complexos proteolíticos bacterianos dependentes de ATP (MOGK *et al.*, 2004).

- Os substratos protéicos podem ser inicialmente reconhecidos por proteínas adaptadoras.
- Os substratos protéicos podem ser diretamente reconhecidos pelo domínio N-terminal da ATPase AAA+.
- Os substratos protéicos podem ser reconhecidos por um domínio extra (domínio I ou domínio N) na ATPase AAA+.

As proteases bacterianas multiméricas melhor caracterizadas até à presente data são as proteases ClpP e HslV (ClpQ) de *E. coli*. Estas proteases não partilham semelhanças na sua seqüência primária nem no seu modo de ação (LARSEN e FINLEY, 1997). A peptidase ClpP associa-se a uma de duas ATPases, ClpA ou ClpX formando os complexos ClpAP e ClpXP (GOTTESMAN *et al.*, 1993; SZYK e MAURIZI, 2006) enquanto HslV (ClpQ) se associa a ATPase HslU (ClpY) formando o complexo HslUV (ClpYQ) (ROHRWILD *et al.*, 1996). ClpP é uma serina-protease que possui homólogos identificados em uma vasta gama de bactérias bem como em diversas plantas e mamíferos. HslV (ClpQ) por seu turno possui um resíduo ativo de treonina no seu sítio ativo e os seus homólogos são encontrados no citóssol de eucariotos e arqueabactérias (PORANKIEWICZ *et al.*, 1999).



**Figura 13.** Comparação da arquitetura geral do complexo proteolítico ClpAP de *E. coli* e do proteasomo 26S de eucariotos (GOTTESMAN, 1997).

Apesar destas diferenças, a arquitetura geral dos dois complexos proteolíticos ClpAP e ClpXP e do proteasomo 26S é bastante similar, sugerindo que estes complexos proteolíticos bacterianos podem partilhar um mecanismo comum de degradação de proteínas com o proteasomo 26S de eucariotos (KESSEL *et al.*, 1995; WANG *et al.*, 1998; HORWICH *et al.*, 1999) (vide Figura 13).

As seções que se seguem versão sobre os componentes do complexo proteolítico ClpAP, sem dúvida o complexo proteolítico bacteriano melhor descrito na literatura até à data, e sobre a proteína acessória reguladora deste complexo, a proteína ClpS.

### 3.1.1.2. A protease ClpP

O conhecimento que se detém hoje sobre a proteína ClpP provém, principalmente, de estudos bioquímicos e estruturais nos últimos anos sobre a proteína ClpP de *E.coli*. A proteína ClpP de *E.coli* é sintetizada inicialmente como uma pró-proteína de 207 resíduos que é auto-proteoliticamente processada durante o enovelamento da cadeia para produzir uma proteína madura com 193 resíduos (MAURIZI *et al.*, 1990b; WANG *et al.*, 1997).

A estrutura cristalina de ClpP revelou uma nova família estrutural dentro da superfamília de serina-proteases e definiu um novo enovelamento do tipo  $\alpha/\beta$  (WANG *et al.*, 1997; SZYK e MAURIZI, 2006). Cada subunidade da proteína ClpP consiste de seis unidades repetidas de uma unidade  $\alpha/\beta$  com uma unidade  $\alpha/\beta$  adicional protuberante.

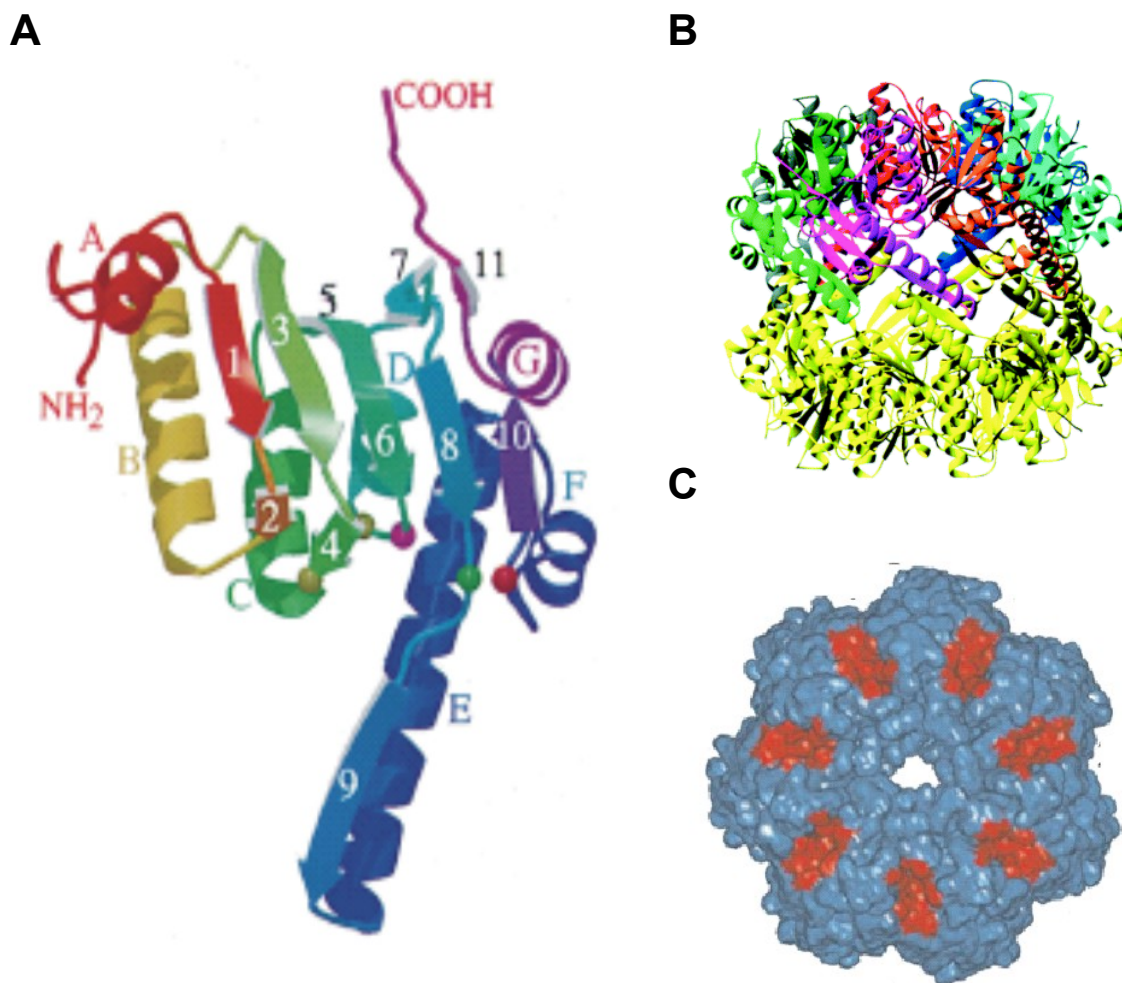
Em uma projeção, o monômero ClpP assemelha-se a uma larga cabeça de um machado (domínio “cabeça”) que possui um pequeno cabo (domínio “cabo”). Os resíduos 28 a 120 e 160 a 188 formam o domínio “cabeça” enquanto o domínio “cabo” é formado pelos resíduos 125 a 130 e 132 a 157 (WANG *et al.*, 1997; SZYK e



MAURIZI, 2006). Os resíduos amino (11 a 27) e carboxiterminais (189 a 193) encontram-se em extremos opostos do domínio “cabeça”, projetando-se para fora deste domínio (vide Figura 14 A).

O sítio ativo desta serina-protease compreende a tríade catalítica formada pelos resíduos Ser-97, His-122 e Asp-171 e está localizado em uma reentrância na junção entre os domínios “cabeça” e “cabo” (WOO *et al.*, 1989; WANG *et al.*, 1997; SZYK e MAURIZI, 2006). Esta disposição define uma nova família de serina-proteases que é estruturalmente distinta das demais serina-proteases anteriormente descritas (famílias da quimotripsina, tripsina, substilina e proteases de citomagalovirus) porém as características essenciais da arquitetura da tríade catalítica Ser-Asp-His da proteína ClpP mantêm-se em relação às demais serinas-proteases.

Na proteína ClpP nativa, 14 monômeros idênticos da cadeia de ClpP de 193 resíduos juntam-se para formar um tetradecâmero estável de aproximadamente 294 kDa composto por dois anéis heptaméricos sobrepostos e intercalados (FLANAGAN *et al.*, 1995; WANG *et al.*, 1997; SZYK e MAURIZI, 2006) (Figura 14). O tetradecâmero de ClpP tem um formato cilíndrico medindo aproximadamente 90Å em altura e 50Å de diâmetro e possuindo um canal central oco (KESSEL *et al.*, 1995; WANG *et al.*, 1997; WANG *et al.*, 1998). No oligômero, o amino-terminal de cada subunidade de ClpP está localizado dentro desta cavidade perto do eixo de simetria enquanto o carboxi-terminal está localizado na periferia do cilindro. A interface monômero-monômero em cada anel heptamérico é predominantemente hidrofóbica e dá-se através de interações entre os domínios “cabeça” dos vários monômeros. O tetradecâmero é formado pela intercalação dos dois anéis heptaméricos via os seus domínios “cabo” (WANG *et al.*, 1997).



**Figura 14.** Estrutura tridimensional da proteína ClpP.

A. O monômero de ClpP apresenta um enovelamento do tipo  $\alpha/\beta$ . Hélices e fitas estão representadas do amino terminal para o carboxi-terminal em uma coloração do tipo arco íris. Hélices estão identificadas com letras de A a G enquanto fitas estão numeradas de 1 a 8. Os resíduos que formam a tríade catalítica estão representados como círculos nas cores magenta (S97), verde (H122) e vermelho (D171). Em círculos verde claros estão representados os resíduos que estabilizam o intermediário oxianión (G68 e M98) (WANG *et al.*, 1997).

B. Representação tridimensional do tetradecâmico da proteína ClpP visto pelo eixo da junção dos dois anéis heptaméricos. No anel superior cada monômero apresenta uma coloração distinta. No anel inferior todos os monômeros estão coloridos em amarelo (PORANKIEWICZ *et al.*, 1999).

C. Representação tridimensional de um anel hexamérico da proteína ClpP visto pelo eixo da cavidade central mostrando os sete centros catalítico como pequenas aberturas circulares (WANG, 1998).

Os 14 sítios ativos proteolíticos estão localizados no interior do cilindro do tetradecâmero numa câmara esférica de 51Å formada pela associação dos dois anéis heptaméricos (KESSEL *et al.*, 1995). Desta forma, os sítios ativos encontram-se inacessíveis a proteínas enoveladas presentes no ambiente celular.

O acesso ao centro catalítico dá-se por meio de dois poros de 12Å de diâmetro em lados opostos do anel (WANG *et al.*, 1997; SZYK e MAURIZI, 2006). O tamanho de entrada dos poros permite a passagem apenas de uma fita  $\beta$  ou de uma hélice  $\alpha$  implicando que os substratos protéicos entram na câmara proteolítica em uma conformação essencialmente desenovelada (WANG *et al.*, 1998). Na câmara proteolítica as tríades catalíticas estão espaçadas a cada 26Å, de forma que um substrato hepta ou octapeptídeo pode facilmente ser estendido entre dois sítios ativos.

Desta forma, os substratos desenovelados podem estender-se simultaneamente por vários sítios ativos, permitindo a clivagem simultânea e rápida da cadeia polipeptídica em segmentos de 8 a 10 resíduos de amino ácidos (WANG *et al.*, 1998). Além disso, a arquitetura do tetradecâmero permite que as vastas superfícies hidrofóbicas da câmara proteolítica, que são essenciais para manter o substrato desenovelado facilitando a degradação do mesmo, permanecem protegidas das substâncias hidrofílicas do meio celular (SZYK e MAURIZI, 2006).

Assim, a compartimentalização dos sítios ativos no interior da câmara proteolítica provavelmente serve pelo menos três funções (WANG *et al.*, 1998). Primeiramente, uma vez que os poros de entrada na câmara são pequenos (aproximadamente 12Å) excluindo a maioria das proteínas citoplásmicas enoveladas, a compartimentalização dos sítios ativos fornece um controle espacial e temporal da proteólise. Em segundo lugar, a compartimentalização cria uma

concentração local elevada de sítios ativos permitindo uma velocidade de clivagem de ligações peptídicas também elevada. Por último, a compartimentalização fornece uma forma adequada de tornar acessíveis ao substrato vastas superfícies hidrofóbicas (WANG *et al.*, 1997; KESSEL *et al.*, 1995).

A arquitetura do oligômero ClpP (tamanho de poro, distribuição dos sítios ativos, e a superfície externa final de conservação) fornece uma base estrutural para se entender o mecanismo de ação da ClpP sozinha e na protease ClpP (WANG e col., 1997; SZYK e MAURIZI, 2006). A proteína ClpP possui em sua estrutura primária os elementos necessários para catalisar a clivagem de uma ligação peptídica. No entanto, a proteína ClpP sozinha demonstra atividade proteolítica limitada pois é capaz de hidrolisar apenas peptídeos de até 30 resíduos de amino ácidos (WOO *et al.*, 1989; THOMPSON e MAURIZI, 1994). De fato, para se obter uma atividade proteolítica completa de ClpP, é necessária a presença de componentes reguladores dependentes de ATP que desenovelam o substrato e conseqüentemente ativam a atividade proteolítica de ClpP (GOTTESMANN *et al.*, 1997; MOGK *et al.*, 2004). Além disso, estes componentes regulam a especificidade de ClpP que sozinha possui atividade proteolítica para uma vasta gama de substratos (GOTTESMANN *et al.*, 1997).

### 3.1.1.3. A ATPase ClpA

Os componentes reguladores de ClpP dependentes de ATP, ClpA, ClpX e ClpC, foram inicialmente descritos como membros da família de chaperonas moleculares Clp/Hsp100, um grupo de ATPases que se associa na forma de anéis (SCHIRMER *et al.*, 1996). Posteriormente, NEUWALD e colaboradores (1999) descreveram estas proteínas como pertencendo também à superfamília de proteínas AAA+ (ATPases associadas com várias atividades celulares).

As proteínas AAA+ estão presentes em todos os organismos e estão envolvidas numa grande variedade de processos celulares vitais incluindo regulação do ciclo celular, biogênese de organelas e degradação protéica. Estas proteínas partilham um domínio ATPase comum (domínio AAA) composto por uma seqüência conservada de 230 a 250 resíduos de amino ácidos e a sua atividade enzimática depende de ATP (NEUWALD *et al.*, 1999; GUO *et al.*, 2002a; MOGK *et al.*, 2004). Cada domínio AAA é composto por um domínio ATPase central contendo os motivos Walker A e Walker B e um domínio helicoidal C-terminal (MOGK *et al.*, 2004). Os domínios AAA mediam a oligomerização de proteínas AAA+, a ligação a ATP e a hidrólise de ATP. As proteínas AAA+ são divididas em dois subgrupos de acordo com o número de domínios AAA: proteínas da Classe I possuem dois domínios AAA enquanto que proteínas da Classe II possuem apenas um único domínio AAA (WEIBEZAHN e col, 2004).

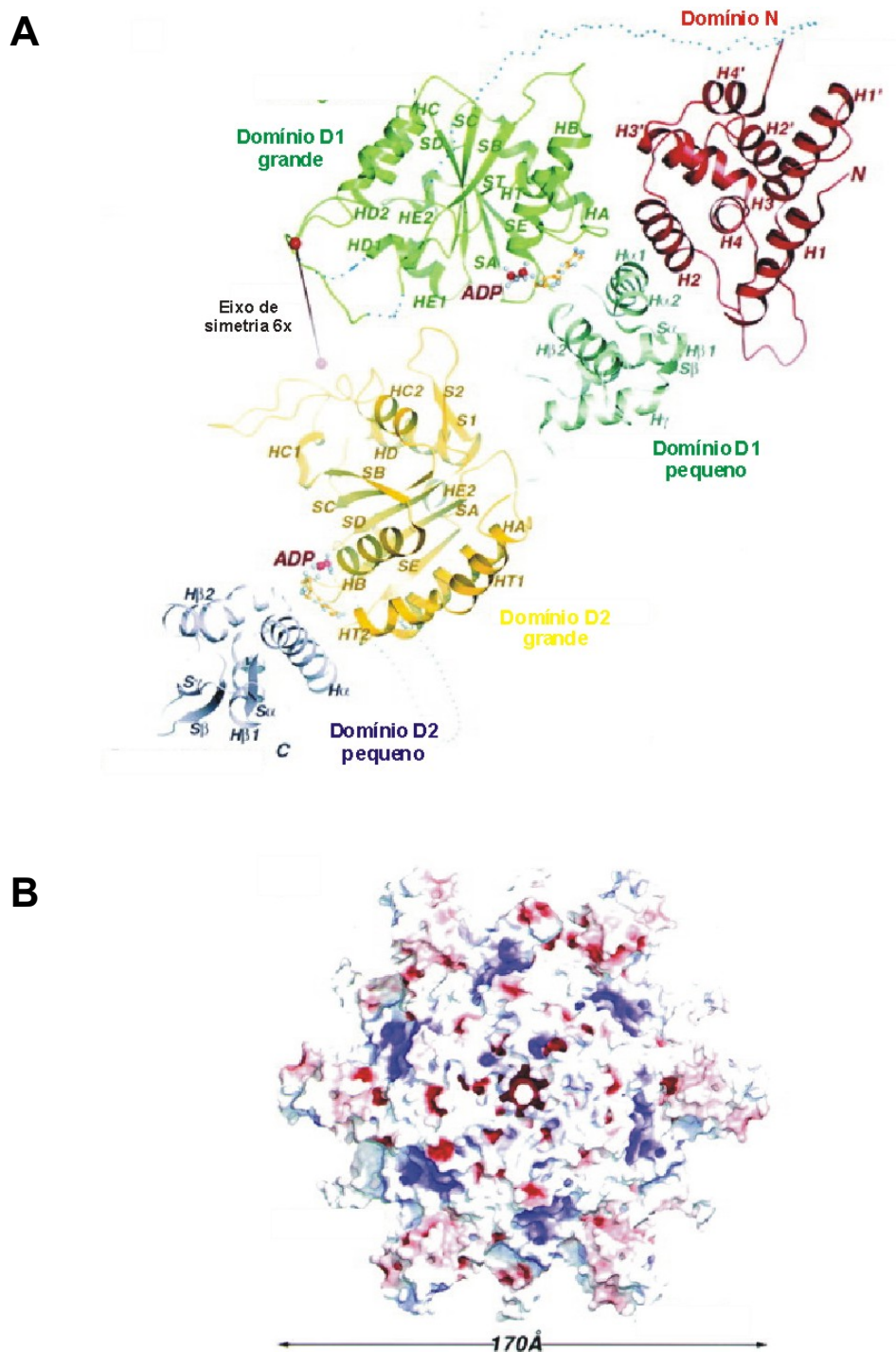
As proteínas ClpA, ClpX e ClpC, que pertencem à Classe I de proteínas AAA+, regulam a proteólise através da sua interação com ClpP e atuam também como chaperonas em várias reações celulares tais como a ativação da proteína RepA do fago P1 e a montagem da transposase Mu (WANG *et al.*, 1997; WEIBEZAHN e col, 2004). A capacidade destas subunidades Clp ATPases de participar tanto no enovelamento de proteínas como chaperonas quanto em reações de degradação protéica sugere que estas proteínas têm um importante papel no controle de qualidade celular (MOGK *et al.*, 2004).

Em *E. coli* e *Xac*, ClpA e ClpX são as proteínas reguladoras de ClpP enquanto em outras bactérias ClpC é o equivalente evolucionário de ClpA (WANG *et al.*, 1997; MOGK *et al.*, 2004; ISHIKAWA *et al.*, 2004). Na presença de ClpX, ClpP degrada proteínas envolvidas no controle da replicação, transcrição e tradução,

como as proteínas  $\lambda$ -O, Phd e o fator  $\sigma$  RpoS (WANG *et al.*, 1997). Na presença de ClpA, ClpP degrada proteínas fusionadas à proteína  $\beta$ -galactosidase através de um mecanismo que depende diretamente da natureza do amino ácido N-terminal (TOBIAS *et al.*, 1991).

Como parte do complexo ClpAP, a proteína ClpA exerce múltiplas funções: reconhecimento do substrato, hidrólise de ATP, desenovelamento do substrato, translocação do substrato para o centro proteolítico e ativação alostérica da atividade da peptidase ClpP (XIA *et al.*, 2004). ClpA reconhece seqüências específicas no N e no C terminal dos seus substratos (GOTTESMAN *et al.*, 1998; HOSKINS *et al.*, 2000). Uma vez reconhecido, o substrato é desenovelado e translocado para a cavidade proteolítica de ClpP (WEBER-BAN, 1999; HOSKINS *et al.*, 2000; ISHIKAWA *et al.*, 2001). Apesar dos mecanismos de translocação e degradação dos substratos pelo complexo ClpAP estarem bem estudados, o reconhecimento e desenovelamento do substrato por ClpAP ainda permanecem pouco definidos (ZETH *et al.*, 2002).

O monômero da proteína ClpA de *E. coli* é uma proteína de 758 amino ácidos composta por dois domínios AAA. A estrutura cristalina do monômero de ClpA revela um domínio N-terminal e dois módulos AAA<sup>+</sup> não-idênticos (D1 e D2) cada um consistindo de um subdomínio pequeno e grande com um sítio de ligação a ADP em uma junção de subdomínios. (ISHIKAWA *et al.*, 2001; GUO *et al.*, 2002b) (Figura 15). O domínio N-terminal (resíduos 1 a 153) está envolvido no reconhecimento e ligação do substrato (HINNERWISCH *et al.*, 2005b). Já os domínios D1 e D2 ligam e hidrolisam ATP e estão relacionados com a formação do hexâmero de ClpA, a interação com ClpP e o desenovelamento do substrato (GOTTESMAN *et al.*, 1997; HOSKINS *et al.*, 2000; GUO *et al.*, 2002b; HINNERWISCH *et al.*, 2005a).



**Figura 15.** Estrutura tridimensional da ATPase ClpA.

A. Representação em fita dos cinco domínios estruturais do monômero de ClpA: domínio N (vermelho), sub-domínios grande e pequeno do domínio D1 (verde escuro e verde claro, respectivamente) e sub-domínios grande e pequeno do domínio D2 (amarelo e roxo, respectivamente). As duas moléculas de ADP localizadas na interface dos sub-domínios estão representadas como bolas vermelhas (GUO *et al.*, 2002b).

B. Modelo hexamérico da proteína ClpA. Observa-se o anel hexamérico ao longo do eixo central. As projeções grandes e pequenas referem-se aos domínios N e D2, respectivamente. O diâmetro máximo do hexâmero é de 170 Å (GUO *et al.*, 2002b).

Não existem estruturas de alta resolução de hexâmeros de ClpA ou de complexos ClpAP. Porém estudos de microscopia eletrônica caracterizaram as interações de ClpAP com um substrato modelo, a proteína P1 do bacteriófago RepA (ISHIKAWA *et al.*, 2001). Além disso, estudos recentes da formação do complexo ClpAP em tempo real por fluorescência e espalhamento de luz forneceram evidências quanto à estrutura deste complexo (KRESS *et al.*, 2007).

Seis monômeros idênticos de ClpA interagem através dos seus sub-domínios AAA+ para formar um anel hexamérico. No complexo ClpAP os dois anéis hexaméricos de ClpA interagem com os dois anéis heptaméricos de ClpP, sendo que uma das superfícies do anel de ClpA está envolvida na interação com ClpP enquanto que a outra superfície está envolvida na interação com o substrato. Recentemente foi também demonstrado que, apesar de formar um hexâmero, a proteína ClpA possui um único sítio de ligação de alta afinidade de cadeia peptídicas por hexâmero (PISZCZEK *et al.*, 2005).

#### **3.1.1.4. A proteína ClpS: uma proteína adaptadora**

Uma das formas de garantir a especificidade e a diversidade funcional das proteínas AAA+ é através de proteínas adaptadoras (DOUGAN *et al.*, 2002b; MOGK *et al.*, 2004; FARRELL *et al.*, 2005) (vide Figura 12). As proteínas adaptadoras constituem uma nova classe de proteínas que modulam a especificidade de ligação e a atividade de chaperona de proteínas AAA+. Estas proteínas não são relacionadas seqüencialmente ou estruturalmente e variam consideravelmente de tamanho, embora tendam a ser pequenas na maioria dos casos (DOUGAN *et al.*, 2002b). As proteínas adaptadoras podem exercer os seus efeitos ligando-se diretamente aos seus substratos ou formando complexos proteína AAA+/proteína



adaptadora, (MOGK *et al.*, 2004; DOUGAN *et al.*, 2002b). As proteínas adaptadoras constituem assim uma forma simples, porém eficaz, de modular a atividade de chaperonas. Neste contexto, a própria regulação das proteínas adaptadoras pode ser uma forma de responder rápida e especificamente a alterações ambientais na célula, redirecionando assim a atividade das proteínas AAA+ (DOUGAN *et al.*, 2002b; FARRELL *et al.*, 2005).

Várias proteínas adaptadoras foram caracterizadas nos últimos anos em procariotos e em organismos superiores (DOUGAN *et al.*, 2002b; FARRELL *et al.*, 2005). A Tabela 5 descreve as proteínas adaptadoras descritas em *E. coli*. De um modo geral as proteínas adaptadoras de *E. coli* podem ser classificadas em duas categorias: aquelas que modulam a ligação do substrato à proteína AAA+ e aquelas que são absolutamente necessárias para o reconhecimento e subsequente degradação de substratos específicos.

**Tabela 5.** Proteínas adaptadoras de *E. coli* e suas funções na adaptação da função de proteínas AAA+ (adaptado de DOUGAN *et al.*, 2002b).

Proteína Adaptadora	Proteína AAA+	Função
ClpS	ClpA	Inibe a degradação de proteínas com o “tag” SsrA pelo complexo ClpAP. Promove a degradação de proteínas agregadas pelo complexo ClpAP. Modula a degradação de proteínas da via “N-end-rule” (DOUGAN <i>et al.</i> , 2002a; ERBSE <i>et al.</i> , 2006)
RssB	ClpX	Necessária para a degradação de $\sigma^S$ mediada por ClpXP MUFFLER <i>et al.</i> , 1996; ZHOU <i>et al.</i> , 2001)
SspB	ClpX	Liga-se ao “tag” SsrA aumentando a degradação de proteínas com este “tag” por ClpXP e inibindo a degradação destas proteínas por ClpAP (FLYNN <i>et al.</i> , 2001; WAH <i>et al.</i> , 2002)
UmuD	ClpX	Liga-se à proteína UmuD' e entrega-a para degradação por ClpXP (NEHER <i>et al.</i> , 2003).

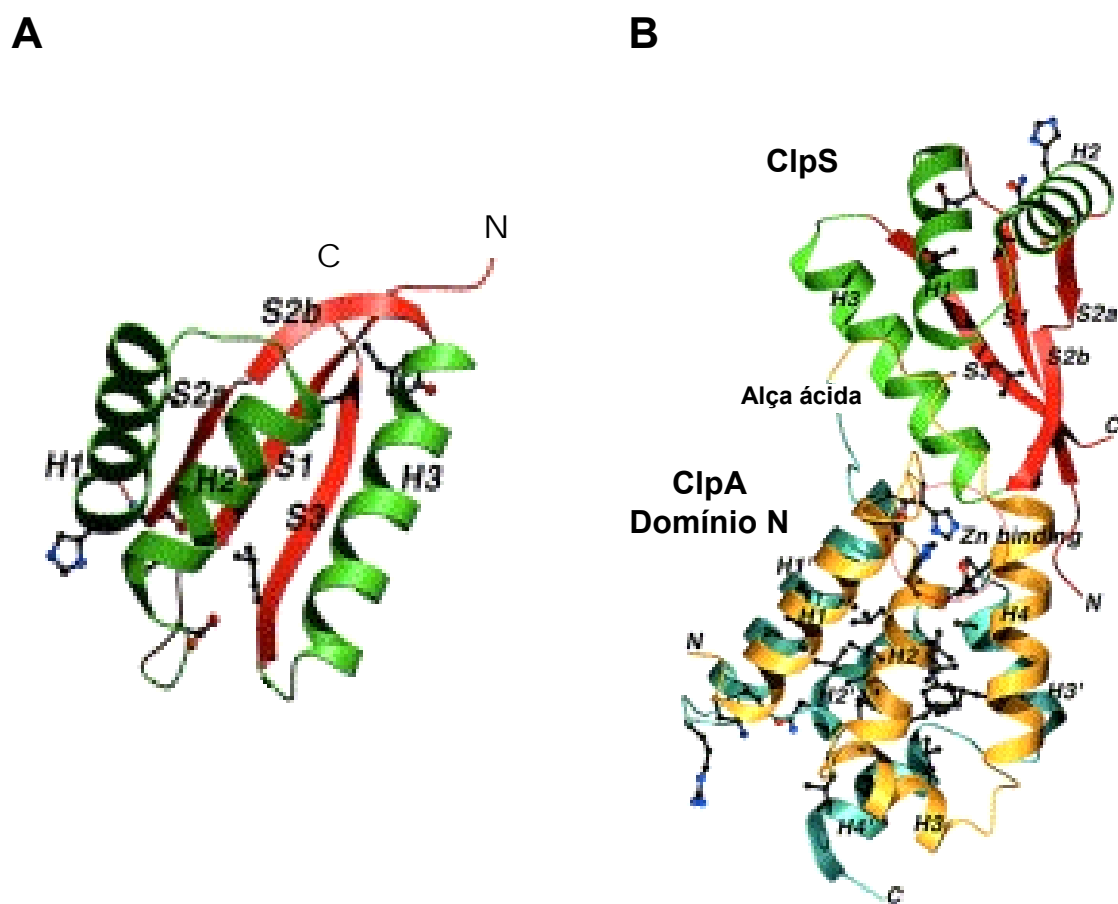
No caso da protease ClpX, três proteínas adaptadoras modulam a sua atividade: RssB (MUFFLER *et al.*, 1996; ZHOU *et al.*, 2001), SspB (FLYNN *et al.*, 2001; WAH *et al.*, 2002) e UmuD (NEHER *et al.*, 2003). A proteína RssB é essencial para a degradação do fator sigma S (“starvation”),  $\sigma^S$ , pelo complexo ClpXP. A proteína SspB reconhece a seqüência específica “tag” SsrA (um peptídeo de 11 amino ácidos – AANDENYALAA) adicionada a produtos de tradução incompletos. Ao contrário de RssB, SspB não altera radicalmente a especificidade de ClpXP, porém altera consideravelmente a cinética de degradação de proteínas marcadas com o “tag” SsrA (LEVCHENKO *et al.*, 2000). A proteína UmuD liga-se ao componente UmuD’ da DNA polimerase V e entrega-o para degradação por ClpXP (NEHER *et al.*, 2003).

No caso da proteína ClpA, apenas a proteína adaptadora ClpS foi identificada até à data como sendo moduladora da atividade de ClpAP (DOUGAN *et al.*, 2002a). O gene *clps* foi descrito como um gene localizado em um potencial operon com ClpA que codifica uma proteína de 106 amino ácidos (DOUGAN *et al.*, 2002a). Análises bioinformáticas deste gene revelaram que existem dois grupos de homólogos de ClpS distintos em bactérias (LUPAS e KORETKE, 2003). No primeiro grupo, ClpS está presente no contexto genômico de ClpA, e encontra-se universalmente distribuída em proteobactérias e em organismos superiores como fungos e animais. No segundo grupo, a proteína ClpS está presente em organismos que não possuem ClpA tais como actinobactérias, cianobactérias e cloroplastos, sugerindo que esta proteína deve possuir atividades moduladoras de outras proteínas ainda não descritas.

Inicialmente, DOUGAN e colaboradores (2002a) descreveram esta proteína como mediadora da degradação de substratos com “tag” SsrA e de proteínas

agregadas. ClpS forma complexos estáveis com ClpA e também com ClpAP de forma ATP-dependente. Ao se ligar a ClpA, esta proteína previne a degradação de proteínas com “tag” SsrA e promove o desligamento destes substratos da proteína ClpA. Foi também demonstrado que ClpS aumenta a taxa de degradação de proteínas agregadas possivelmente através da estabilização da ligação entre ClpA e o substrato agregado ou modulando a especificidade de ligação de ClpA ao substrato (DOUGAN *et al.*, 2002a). Apesar da proteína ClpS ser necessária para a degradação eficiente de proteínas agregadas pelo complexo ClpAP, ela não é essencial para que tal degradação ocorra *in vivo*. DOUGAN e colaboradores (2002a) demonstraram ainda que a proteína ClpS está também envolvida na inibição da auto-degradação de ClpA pelo complexo ClpAP. Ou seja, ao interagir com ClpA, ClpS afeta a ligação de ClpA aos seus substratos, redirecionando a atividade de ClpA da degradação de proteínas com “tag” SsrA e auto-degradação para a degradação de proteína agregadas (DOUGAN *et al.*, 2002a).

Estudos estruturais por cristalografia da proteína ClpS de *E. coli* em complexo com o domínio N-terminal da proteína ClpA realizados por dois grupos independentes revelaram uma proteína  $\alpha/\beta$  composta por três  $\alpha$ -hélices conectadas a três fitas- $\beta$  antiparalelas em um arranjo  $\beta\alpha\alpha\beta\alpha\beta$  (GUO *et al.*, 2002a; ZETH *et al.*, 2002) (Figura 16). A proteína ClpS tem formato de um cone, com uma camada curva de três  $\alpha$ -hélices antiparalelas sobre uma folha- $\beta$  antiparalela deformada. O enovelamento da proteína assemelha-se ao enovelamento do domínio C-terminal da proteína ribossomal L7/L12 (WAHL *et al.*, 2000). É também interessante notar que em ambos os estudos não foi possível determinar a estrutura da região N-terminal (resíduos 1 a 20) da proteína ClpS de *E. coli*.



**Figura 16.** Estrutura tridimensional da proteína ClpS de *E. coli*.

A. Representação em fita da proteína ClpS. As hélices- $\alpha$  H1, H2 e H3 estão representadas em verde e as fitas- $\beta$  S1, S2 e S3 estão representadas em vermelho (XIA *et al.*, 2004).

B. Representação em fita do domínio N da proteína ClpA interagindo com a proteína ClpS. As partes N-terminal e C-terminal do domínio N estão representadas em laranja e turquesa respectivamente. A proteína ClpS está colorida tal conforme descrito acima (XIA *et al.*, 2004).

Em ambas as estruturas cristalinas o domínio N-terminal da proteína ClpA interage com uma única molécula de ClpS para formar um complexo heterodimérico (GUO *et al.*, 2002a; ZETH *et al.*, 2002) corroborando os resultados de DOUGAN e colaboradores (2002a) que demonstraram que são necessários seis monômeros de ClpS para se observar a maior variação na degradação mediada por ClpAP.

A interface entre o N-terminal de ClpA e ClpS é primordialmente de natureza iônica e é composta de duas pontes salinas e duas pontes de hidrogênio. Na

proteína ClpA os resíduos que fazem contatos com ClpS estão localizados nas hélices  $\alpha 2$  e  $\alpha 5$ . Já no caso de ClpS os resíduos de contato estão localizados em uma única hélice, a hélice  $\alpha 3$ . Os resíduos E82 e K 84 de ClpS formam pontes salinas com os resíduos R23 e E23 de ClpA, respectivamente. E79 de ClpS forma duas pontes de hidrogênio com os resíduos E28 e T81 de ClpA (GUO *et al.*, 2002a; ZETH *et al.*, 2002).

Usando mutagênese, foi demonstrado que dois resíduos de ClpS envolvidos no contato com ClpA (E79 e K84) são essenciais não apenas para a formação do complexo ClpAS mas também para a própria degradação do substrato (ZETH *et al.*, 2002). Os resíduos correspondentes não estão presentes na chaperona ClpB, uma chaperona que interage com Hsp70/DnaK para formar uma bi-chaperona que está envolvida na reativação de proteínas fortemente agregadas *in vivo* e *in vitro* (KIM *et al.*, 2000; KEDZIERSKA *et al.*, 2005). A ausência destes resíduos em ClpB fornece um racional explicar porque a proteína ClpS modula apenas a atividade de ClpA apesar de elevada similaridade entre ClpA e ClpB.

Estudos de mutagênese demonstraram ainda que na ausência do N-terminal (resíduos 1 a 17), a proteína ClpS liga-se a ClpA, porém não modula a sua atividade sugerindo que o N-terminal de ClpS pode estar envolvido em interações importantes com o substrato (DOUGAN *et al.*, 2002a).

Recentemente, a proteína ClpS foi descrita como um componente essencial da via de degradação da regra do N-terminal (“N-end rule”) em *E. coli* (ERBSE *et al.*, 2006). Esta regra é derivada da observação que a meia vida de uma proteína é determinada pela natureza dos seus resíduos amino-terminais (VARSHAVSKY *et al.*, 1996). Eucariotos e procariotos fazem uso de resíduos do N-terminal como um sinal para direcionar proteínas para degradação. Em eucariotos, a ligase E3 N-

recognina, reconhece os substratos da regra do N-terminal, media a sua ubiquitinação e a sua degradação pelo proteasomo (VARSHAVSKY *et al.*, 1996). Em *E. coli* os substratos da regra do N-terminal são degradados pelo complexo ClpAP porém pouco se conhece sobre o mecanismo de reconhecimento dos substratos.

ERBSE e colaboradores (2006) demonstraram que a proteína ClpS tem uma afinidade elevada por substratos da regra do N-terminal, ligando-se diretamente aos resíduos desestabilizadores do N-terminal através da sua superfície de ligação ao substrato que se encontra distante da superfície de ligação ClpS-ClpA (ZETH *et al.*, 2002; XIA *et al.*, 2004). Uma vez ligada ao substrato da regra do N-terminal, a proteína ClpS redireciona-o para degradação pelo complexo ClpAP, definido assim a especificidade desta máquina proteolítica. Este trabalho demonstrou ainda que, na presença de ClpS, a degradação dos substratos da regra do N-terminal aumenta significativamente enquanto que a degradação de outros substratos é inibida. Além disso, foi demonstrado que mutantes de *E. coli* que não possuem o gene *clps* ( $\Delta clps$ ) não degradam substratos da regra do N-terminal e que a co-expressão de ClpS nestes mutantes restaura esta atividade, demonstrando que a proteína ClpS é essencial para a degradação de substratos da regra do N-terminal pelo complexo ClpAP.

Posteriormente, o grupo de WANG e colaboradores (2007) publicou um artigo controverso que contradiz o trabalho de ERBSE e colaboradores (2006) ao afirmar que ClpS de fato modula a via de degradação da regra do N-terminal, porém esta proteína adaptadora não seria essencial para que a referida degradação ocorra conforme proposto originalmente. Este grupo demonstrou que o complexo ClpAP degrada substratos da regra do N-terminal na ausência de ClpS tanto *in vitro* quanto

*in vivo*. Além disso, este trabalho demonstrou ainda que a ClpS aumenta a degradação de substratos da regra do N-terminal por ClpAP quando estes se encontram em concentrações baixas (4  $\mu$ M). Em concentrações altas de substrato (40  $\mu$ M), no entanto, a presença de ClpS suprime a degradação. Desta forma, dependendo das condições, a proteína ClpS pode funcionar tanto como um fator de inibição ou com um fator de aumento da degradação de substratos de “N-end” por ClpAP (WANG *et al.*, 2007). Os resultados contraditórios obtidos pelos dois grupos podem ser potencialmente explicados pelo fato do grupo de ERBSE e colaboradores (2006) terem realizado poucos experimentos de degradação e em uma única concentração baixa (500 nM). Nestas condições, dá-se a maior dependência de ClpS, não sendo portanto surpreendente que este grupo tenha observado uma forte dependência da presença de ClpS na degradação de substratos da regra do N-terminal.

#### **3.1.1.5. A proteína ClpS de *Xac***

A proteína XAC2000 de *Xac* foi selecionada, em julho de 2001, como uma das proteínas-alvo para os estudos de triagem descritos no Capítulo 2. Apesar de, na época em que foi selecionada esta proteína ter estrutura e função desconhecidas, a proteína XAC2000 apresentava um elevado grau de conservação entre vários genomas o que nos levou a postular que esta proteína estaria provavelmente envolvida em processo biológicos importantes e evolutivamente significativos.

Em 2002, DOUGAN e colaboradores descreveram o gene *clps* de *E.coli* que está presente no operon de *clpA* em várias bactérias e codifica uma proteína de 106 amino ácidos. Esta proteína apresenta 56% identidade e 84% de similaridade com a

proteína hipotética de *Xac* XAC2000 por nós selecionada para estudos de triagem, indicando que a proteína XAC2000 é o homólogo de *Xac* da proteína ClpS. A Figura 17 mostra o alinhamento da proteína ClpS de *Xac* com a proteína ClpS de *E. coli* (Figura 17A) e com outras proteínas ClpS presentes em vários genomas bacterianos (Figura 17B).

A Figura 17C mostra uma representação esquemática da localização do gene *clps* no mapa genômico de *Xac*. É interessante notar que o gene *clps* de *Xac* encontra-se localizado no mesmo contexto genômico que o gene *clps* de *E. coli*. O gene *clps* de *Xac* codifica uma proteína de 106 amino ácidos e está localizado no mesmo operon que o gene *clpa*. Os demais genes presentes no mesmo contexto genômico codificam proteínas envolvidas na tradução, enovelamento e degradação de proteínas e outros processos celulares.



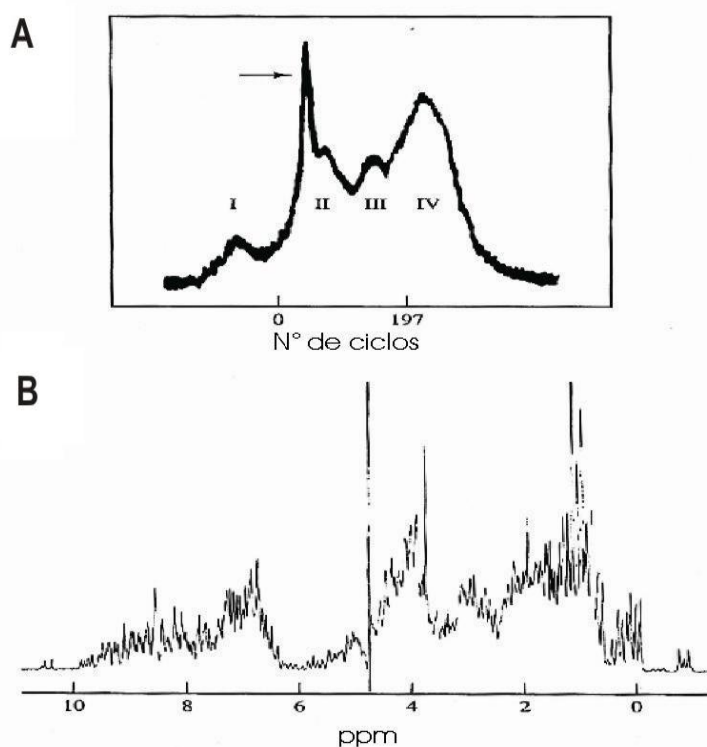


### 3.1.2. A RMN na caracterização estrutural de proteínas

O fenômeno de ressonância magnética nuclear (RMN) foi detectado pela primeira vez em 1946, independentemente, pelos grupos de Felix Bloch e Carl Purcell (GIL e GERALDES, 1987). Por muito tempo a técnica de RMN foi utilizada somente por químicos e físicos para estudar moléculas pequenas. Porém, as exigências da química, e principalmente da bioquímica, estimularam a evolução da técnica. O avanço nas técnicas de preparo das amostras e o desenvolvimento de magnetos com campos cada vez mais altos permitiram que a RMN viesse a ter aplicações tão variadas como a elucidação estrutural de proteínas e outras macromoléculas a nível atômico, acompanhamento da cinética de reações químicas e a obtenção de imagens internas de amostras intactas. Atualmente, a técnica de RMN permite a determinação da estrutura de uma proteína em solução com resolução atômica bem como estudos de dinâmica e de interação da mesma, o que a torna uma técnica de extrema importância para estudos de proteínas (WÜTHRICH, 1986; EVANS, 1995; CLORE e GRONENBORN, 1994; ARROWSMITH e WU, 1998; PRESTGARD *et al.*, 2001).

Desde a publicação da primeira estrutura completa de uma proteína em solução (WILLIAMSON *et al.*, 1985) vários avanços tecnológicos têm permitido que a RMN alcance um papel de destaque na biologia estrutural. Apesar da cristalografia de raios X ainda ser a técnica de eleição para a maioria dos estudos estruturais de proteínas, inovações nos campos magnéticos, nas seqüências de pulso, nos métodos de análise e nos métodos de marcação isotópica, têm colaborado para que a RMN se tenha transformado numa poderosa técnica para o estudo de proteínas e outras macromoléculas em solução a nível atômico (EVANS, 1995; CLORE e GRONENBORN, 1998).

O primeiro espectro de RMN unidimensional (1D) de  $^1\text{H}$  de uma proteína foi coletado na década de 50, em um espectrômetro de 40 MHz utilizando a enzima ribonuclease (SAUNDERS *et al.*, 1957). Posteriormente, o desenvolvimento de magnetos com campos cada vez mais altos os permitiu que espectros de proteínas sejam coletados usando instrumentos de RMN que vão de 500 MHz a 1 GHz. Como pode ser observado na Figura 18, apesar dos espectros unidimensionais de  $^1\text{H}$  conterem informações importantes como a largura dos picos e a dispersão do deslocamento químico ( $\delta$ ) que permitem aferir as condições em que a proteína em estudo se encontra em solução, estes espectros são insuficientes para a determinação estrutural da proteína em estudo.



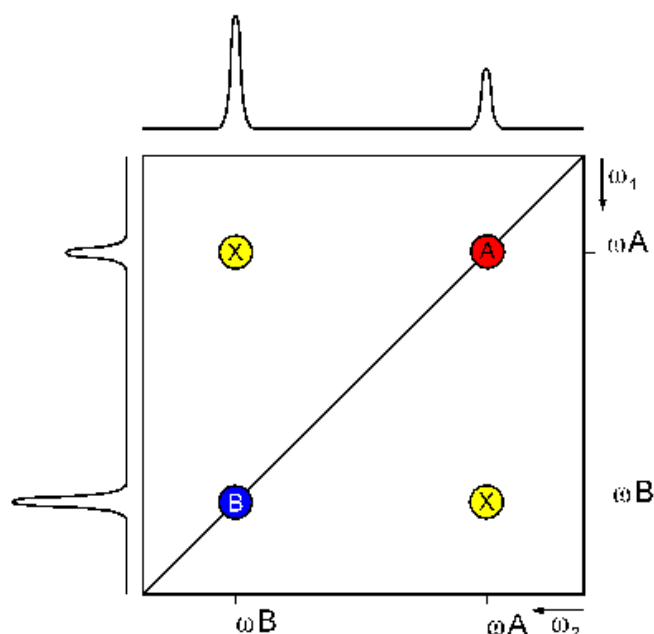
**Figura 18.** Espectros de RMN 1D de  $^1\text{H}$  de proteínas.

A. Primeiro espectro de RMN 1D de  $^1\text{H}$  de RMN de uma proteína, a enzima ribonuclease, coletado usando um instrumento operando a 40 MHz em 1957 (SAUNDERS *et al.*, 1957).

B. Espectro de RMN 1D de  $^1\text{H}$  da enzima lisozima coletado usando um instrumento operando a 750MHz (EVANS, 1995).

O principal problema em se utilizar a RMN unidimensional (1D) de  $^1\text{H}$  para estudar macromoléculas tais como proteínas deve-se ao fato destas moléculas conterem um grande número de núcleos de  $^1\text{H}$  o que leva à sobreposição dos vários sinais obtidos no espectro de RMN, tornando o seu assinalamento inviável. Era portanto desejável que se desenvolvessem metodologias que permitissem resolver os sinais sobrepostos dos espectros 1D de RMN.

Na década de 80, o tamanho das estruturas macromoleculares que puderam ser resolvidas por RMN aumentou significativamente devido à introdução de uma vasta gama de experimentos de RMN bidimensionais (2D) desenvolvidas por WÜTHRICH e colaboradores que culminou com a determinação de várias estruturas de proteínas pequenas (WÜTHRICH, 1986).



**Figura 19.** Esquematização de um espectro 2D de RMN.

A linha diagonal (sinais A e B) apresenta os sinais que seriam observados em um simples experimento 1D. Os sinais simétricos a esta diagonal (sinais X) são chamados de picos cruzados. A diagonal resulta de contribuições de magnetização que não foi alterada durante a etapa de mixagem. Os sinais cruzados resultam de núcleos que trocaram magnetização durante a etapa de mixagem. Estes picos apresentam frequências de cada um dos núcleos em cada uma das dimensões e refletem a interação entre os dois núcleos (SCHIRRA, 1996).

Os experimentos bidimensionais tornaram-se assim indispensáveis para a determinação estrutural de proteínas por RMN pois com a introdução de mais uma dimensão (frequência) aos espectros, a análise dos mesmos pode ser simplificada e mais informações podem ser obtidas (CLORE e GRONENBORN, 1998).

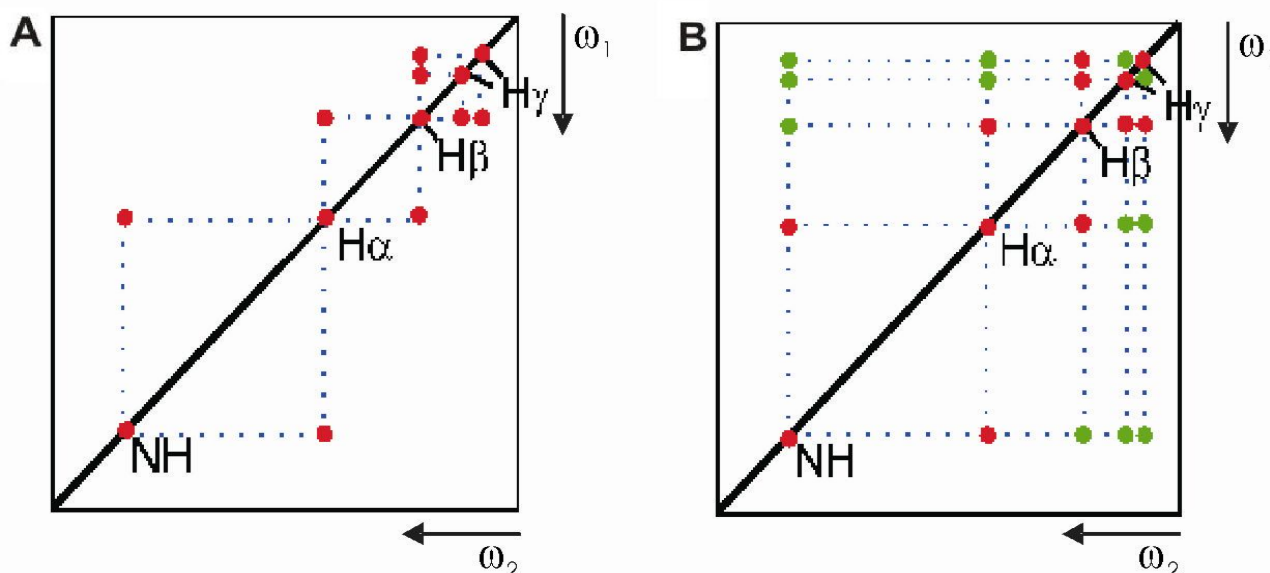
Como pode ser observado na Figura 19 um espectro bidimensional de RMN possui dois eixos de frequência. Os picos na diagonal do espectro contêm informações sobre o deslocamento químico dos núcleos. Os picos cruzados (i.e. fora da diagonal) fornecem informações sobre a correlação entre os vários núcleos (WÜTHRICH, 1986; EVANS, 1995).

Os espectros bidimensionais podem ser homonucleares, i.e. têm sinais obtidos a partir da correlação do mesmo núcleo (geralmente  $^1\text{H}$ ), ou heteronucleares, i.e. mostram a transferência de magnetização entre  $^1\text{H}$  e heteronúcleos como  $^{15}\text{N}$  e  $^{13}\text{C}$ . Além disso, a transferência de magnetização entre os vários spins em um experimento de RMN de duas dimensões, pode ocorrer por dois mecanismos: via acoplamento escalar ou via NOE (Efeito Nuclear Overhauser ou “Nuclear Overhauser Effect”) (WÜTHRICH, 1990; EVANS, 1995; ARROWSMITH e WU, 1998).

O acoplamento escalar faz a transferência de magnetização via ligação química e é usado em experimentos como COSY (“COrrrelation SpectroscopY”) e TOCSY (“TOtal Correlation SpectroscopY”). Em um COSY, apenas acoplamentos de núcleos que estão a menos de três ligações químicas são observados. Já em um TOCSY, são observados acoplamentos entre núcleos separados por mais do que três ligações (WÜTHRICH, 1990; EVANS, 1995).

Assim, para um dado amino ácido um espectro COSY mostrará apenas a interação entre núcleos de hidrogênio separados por 2 ou 3 ligações químicas. Um

espectro TOCSY, do mesmo amino ácido a mostrará a interação entre todos núcleos de hidrogênio pois a magnetização é transferida para todo o sistema de spins do amino ácido (EVANS, 1995). O resultado de ambos os espectros são picos cruzados em uma mesma linha vertical, sendo cada uma destas linhas representa um sistema de spins pertencente a um resíduo de amino ácido (WÜTHRICH, 1990; EVANS, (vide Figura 20).



**Figura 20.** Esquematização de espectros COSY e TOCSY.

A. Esquematização de um espectro COSY. A linha diagonal apresenta os sinais correspondentes a cada um dos núcleos de hidrogênio do amino ácido em estudo. Os picos cruzados representam as interações entre núcleos de hidrogênio separados por 2 ou 3 ligações químicas.

B. Esquematização de um espectro TOCSY. A linha diagonal apresenta os sinais correspondentes a cada um dos núcleos de hidrogênio do amino ácido em estudo. Os picos cruzados representam as interações entre todos os núcleos de hidrogênio do mesmo amino ácido. Em vermelho estão representados os picos que seriam observados em um experimento COSY e em verde estão representados os picos que seriam apenas observados em um experimento TOCSY. Cada linha vertical de picos cruzados em representa um sistema de spins de todos os núcleos de hidrogênio pertencentes a um resíduo de amino ácido (adaptado de SCHIRRA, 1996).

Outro mecanismo de transferência de magnetização utilizado em experimentos de RMN é via NOE, em que a transferência ocorre não através da ligação química mas através do espaço por uma relaxação cruzada via interação dipolar. Se considerarmos um sistema constituído apenas por um par de núcleos de hidrogênio  $i$  e  $j$ , acoplados entre si via acoplamento dipolar, o efeito de NOE é a mudança na magnetização do hidrogênio  $i$  após a perturbação da magnetização do hidrogênio  $j$ . Picos de correlação entre as ressonâncias  $i$  e  $j$  são observados quando os hidrogênios transferem magnetização entre si via relaxação cruzada (GIL e GERALDES, 1987).

A transferência de magnetização entre os dois spins nucleares depende da distância entre eles, e o sinal (NOE) é observado apenas se estes estiverem a menos de 5Å de distância (WÜTHRICH, 1986; WÜTHRICH, 1990). Este limite resulta do fato de que o sinal de NOE ser proporcional ao inverso da distância entre os núcleos elevada a sexta potência. Por esta razão, a intensidade do NOE diminui muito rapidamente com o aumento da distância entre o par de núcleos. Assim, uma proteína que apresenta vários sinais de NOE fortes possui um grande número núcleos muito próximos um dos outros (i.e. esta proteína encontra-se enovelada).

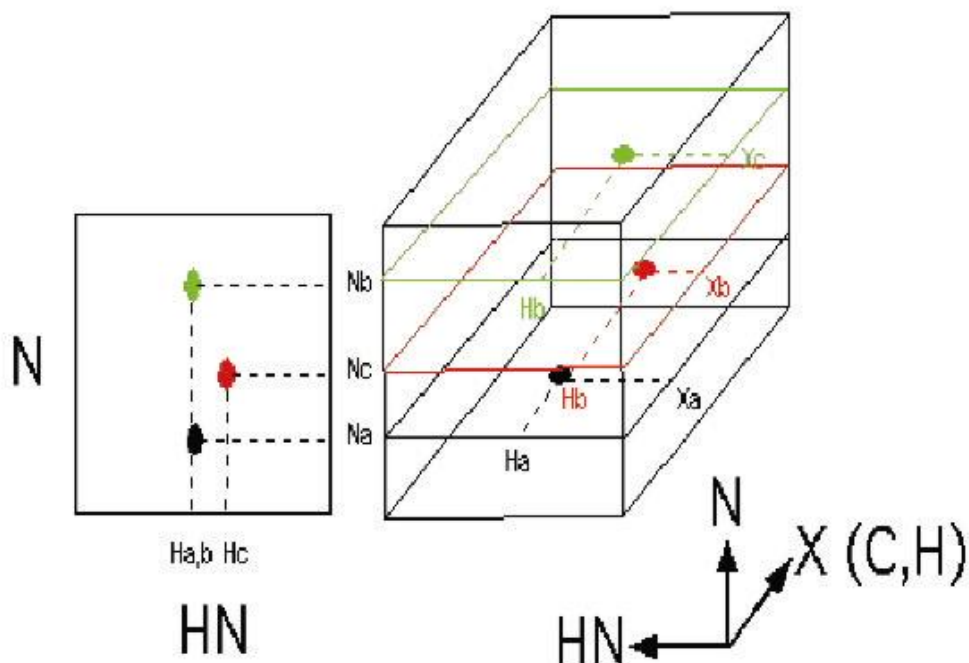
Os NOEs de uma proteína podem ser analisados utilizando um experimento chamado NOESY (“Nuclear Overhauser Effect SpectroscopY”) que correlaciona todos os NOEs de uma proteína e os exprime num espectro bidimensional (WÜTHRICH, 1990; ARROWSMITH e WU, 1998). Desta forma, os picos da linha diagonal representam os sinais dos núcleos de hidrogênio da proteína em estudo enquanto os picos cruzados devem-se às interações do tipo NOE entre núcleos de hidrogênio que se encontram a uma distância de até 5Å. Neste espectro são observadas tanto correlações entre núcleos do mesmo amino ácido (tal como no

espectro TOCSY) quanto correlações entre núcleos de amino ácidos diferentes. O espectro NOESY é portanto essencial para a determinação da estrutura de uma proteína pois é capaz de correlacionar spins que estão em amino ácidos seqüencialmente distantes mas próximos na estrutura terciária da proteína (WÜTHRICH, 1986; EVANS, 1995; WÜTHRICH, 1990; ARROWSMITH e WU, 1998).

Outros experimentos bidimensionais de extrema importância para o estudo de proteínas por RMN são os experimentos bidimensionais heteronucleares. Estes experimentos também fazem uso da transferência de magnetização via ligação química e correlacionam os átomos de hidrogênio da molécula em estudo com outros núcleos que possuam sinal de RMN presentes na molécula. Os núcleos mais utilizados neste tipo de experimento são  $^{15}\text{N}$  e  $^{13}\text{C}$ . Estes núcleos têm abundância natural muito baixa e são menos sensíveis à ressonância que  $^1\text{H}$ , porém o enriquecimento de proteínas com estes núcleos usando marcação isotópica permite que estes sejam freqüentemente utilizados em experimentos de RMN.

O experimento bidimensional heteronuclear mais utilizado é o HSQC descrito na secção 2.1.3. Este experimento correlaciona o deslocamento químico do  $^{15}\text{N}$  de um agrupamento amídico com o deslocamento químico do hidrogênio que está diretamente ligado a ele ( $^1\text{H}/^{15}\text{N}$ -HSQC), podendo correlacionar também hidrogênios que estão ligados a  $^{13}\text{C}$  ( $^1\text{H}/^{13}\text{C}$ -HSQC). Os experimentos de RMN bidimensionais foram uma das mais importantes descobertas para o estudo de proteínas por RMN. Porém, à medida que o interesse por proteínas maiores foi aumentando, rapidamente apareceram problemas com a sobreposição de picos nos espectros 2D. Assim, o espectro com três dimensões – a espectroscopia de RMN tridimensional (3D) - surgiu no final de década de 80 para simplificar a análise de espectros 2D com muitas sobreposições (WÜTHRICH, 1990; EVANS, 1995).





**Figura 21.** Esquemática de um espectro 3D de RMN.

A introdução de uma terceira dimensão permite desdobrar o espectro 2D em vários planos de acordo com a frequência de um terceiro núcleo, eliminando assim eventuais sobreposições no espectro 2D.

Exemplos deste tipo de experimento são os experimentos TOCSY-HSQC e NOESY-HSQC editados para  $^{15}\text{N}$  ou  $^{13}\text{C}$ . O espectro obtido assemelha-se a um cubo que contém o deslocamento químico do  $^1\text{H}$  em duas dimensões e uma terceira dimensão contendo os deslocamentos químicos do  $^{15}\text{N}$  ou do  $^{13}\text{C}$ . O que diferencia estes espectros dos espectros de correlação homonuclear em duas dimensões é que a informação contida no espectro bidimensional homonuclear está distribuída em vários planos na terceira dimensão que correspondem aos deslocamentos químicos do  $^{15}\text{N}$  e/ou  $^{13}\text{C}$ . Desta forma, é possível resolver eventuais sobreposições do espectro 2D, permitindo o estudo de proteínas de dimensões maiores (WÜTHRICH, 1990; EVANS, 1995) (Figura 21).

Outros experimentos de RMN 3D muito utilizados são os experimentos de tripla ressonância. Estes experimentos fazem uso de amostras duplamente marcadas com  $^{15}\text{N}$  e  $^{13}\text{C}$  e correlacionam spins como o hidrogênio amídico ( $^1\text{H}^{\text{N}}$ ), nitrogênio ( $^{15}\text{N}$ ), hidrogênio alfa ( $^1\text{H}^{\alpha}$ ), carbono alfa ( $^{13}\text{C}^{\alpha}$ ) e carbono carbonila ( $^{13}\text{CO}$ ) da cadeia principal, e hidrogênio beta ( $^1\text{H}^{\beta}$ ) e carbono beta ( $^{13}\text{C}^{\beta}$ ) da cadeia lateral de proteínas, via acoplamento escalar.

Estes experimentos de tripla ressonância podem ser utilizados para analisar correlações intra-residuais e correlações seqüenciais eliminando possíveis erros no assinalamento provenientes da confusão entre picos seqüenciais e não-seqüenciais, resultantes da proximidade estrutural dos amino ácidos da proteína (EVANS, 1995; FERENTZ e WAGNER, 2000).

Mais recentemente métodos como a perdeuteração das proteínas em estudo, a técnica de TROSY (“Transverse Relaxation Optimized Spectroscopy”), medidas de acoplamento dipolar residual (RDC) e até mesmo a RMN em 4D vêm possibilitando a determinação estrutural de proteínas cada vez maiores (FERENTZ e WAGNER, 2000; ARROWSMITH e WU, 1998).

Como pode ser visto acima, a gama de experimentos de RMN disponíveis para o estudo estrutural de proteínas é vastíssima. Os experimentos específicos a serem realizados em cada caso vão depender, entre outros fatores, do tamanho da proteína de interesse, da possibilidade de marcação isotópica da mesma e do tipo de equipamento de RMN disponível. No entanto, quaisquer que sejam os experimentos adotados, existem quatro etapas fundamentais comuns a qualquer estratégia de determinação estrutural de proteínas por RMN que devem ser preenchidas:

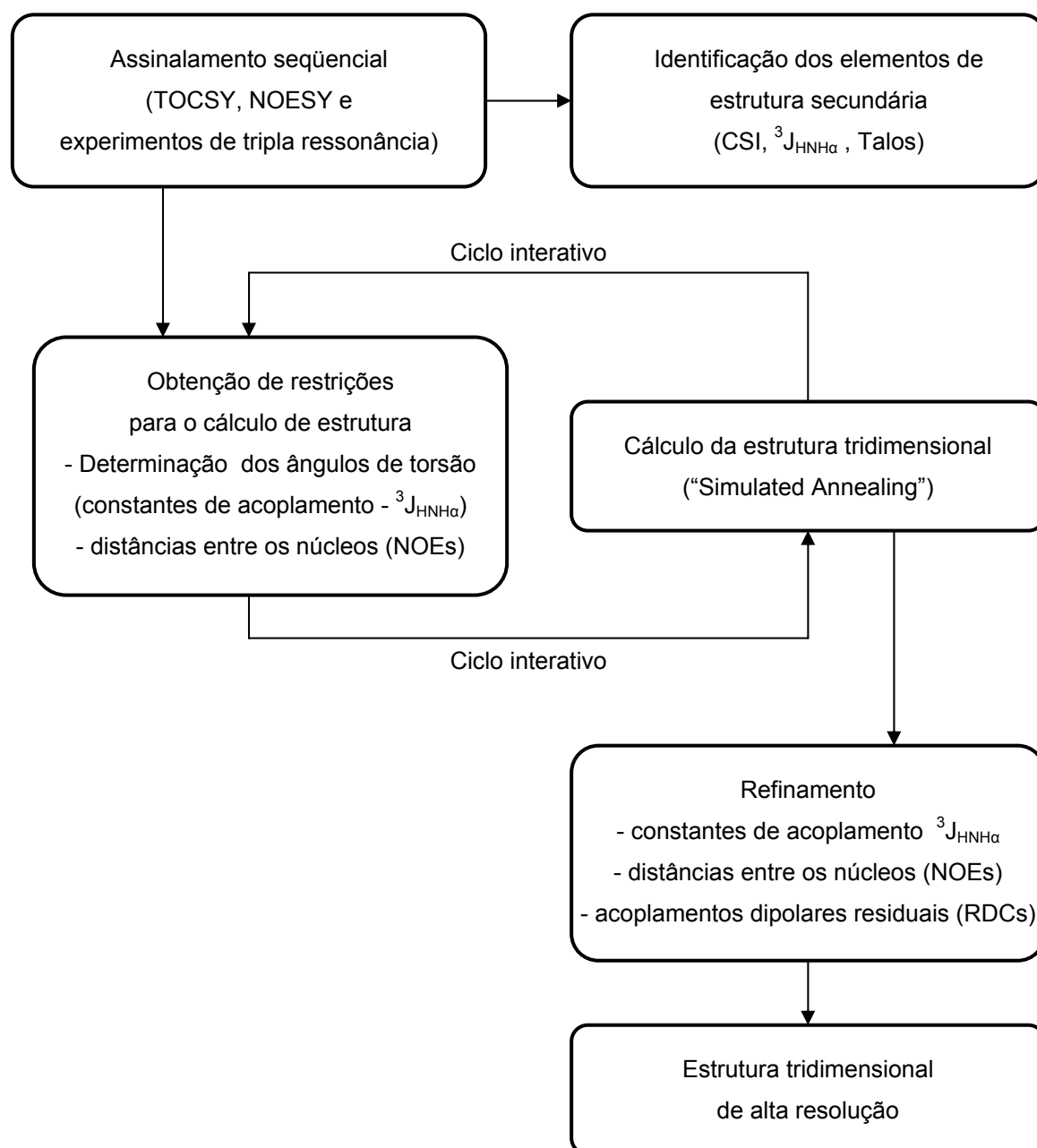
i) assinalamento seqüencial das ressonâncias de todos os resíduos de amino ácidos presentes na molécula utilizando informações de correlações através da ligação química e/ou através do espaço;

ii) identificação dos elementos de estrutura secundária;

iii) obtenção de restrições para o cálculo de estrutura como ângulos de torção, distância entre os núcleos e acoplamentos dipolares residuais;

iv) cálculo e refinamento da estrutura tridimensional com base nas restrições experimentais de distâncias e ângulos, utilizando um ou mais algoritmos de dinâmica molecular.

Assim, os experimentos selecionados devem permitir a realização das várias etapas necessárias à determinação estrutural por RMN. A Figura 22 resume as várias etapas envolvidas na determinação estrutural de proteínas por RMN e como estas se relacionam entre si. Estas etapas serão descritas detalhadamente nas seções seguintes.



**Figura 22.** Resumo da estratégia geral utilizada na determinação de estruturas tridimensionais de proteínas por RMN (adaptado de CLORE e GRONENBORN, 1998).

### 3.1.2.1. Assinalamento seqüencial

O assinalamento seqüencial não-ambíguo de todos os amino ácidos presentes na molécula é a primeira e mais importante etapa no estudo estrutural de uma proteína por RMN. As várias técnicas de assinalamento disponíveis atualmente partem de uma mesma premissa: cada amino ácido em uma proteína está separado de seus vizinhos através de uma ligação peptídica. Devido à natureza eletrônica desta, os amino ácidos constituem sistemas de spins nucleares distintos e isolados. Assim, o assinalamento seqüencial envolve a identificação de cada um destes sistemas de spins e das conectividades entre eles.

Os processos convencionais de assinalamento seqüencial desenvolvidos na década de 80, fazem uso da análise combinada dos espectros 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY e  $^1\text{H}$ - $^1\text{H}$  NOESY (WÜTHRICH,1986). Inicialmente o espectro de correlação homonuclear  $^1\text{H}$ - $^1\text{H}$  TOCSY é analisado. Este experimento permite a identificação de grupos de spins isolados (i.e. amino ácidos) através das possíveis correlações via ligação química. Uma vez que cada um dos 20 amino ácidos comumente encontrados na natureza produzem padrões e valores de deslocamento químico dos hidrogênios distintos, é possível identificar o sistema de spins de cada amino ácido da proteína em estudo (Tabela 6).

Uma vez identificados os sistemas de spin presentes na proteína, e com base na seqüência primária da proteína em estudo, inicia-se a etapa de assinalamento seqüencial dos vários sistemas de spins de amino ácidos vizinhos. Esta etapa faz uso das correlações obtidas no espectro  $^1\text{H}$ - $^1\text{H}$  NOESY, que podem ser entre amino ácidos seqüenciais ou não-seqüenciais, e identifica as conectividades entre os vários sistemas de spins identificados na etapa anterior.

**Tabela 6.** Deslocamento químico dos núcleos de hidrogênio dos 20 amino ácidos naturais em conformação randômica (adaptado de WÜTHRICH, 1986).

Resíduo	$^1\text{H}^{\text{N}}$	$^1\text{H}^{\alpha}$	$^1\text{H}^{\beta}$	$^1\text{H}^{\beta}$	Outros
Gly	8,39	3,97			
Ala	8,25	4,35	1,39		
Val	8,44	4,18	2,13		$\gamma\text{CH}_3$ 0,97 0,94
Ile	8,19	4,23	1,90		$\gamma\text{CH}_2$ 1,48 1,19
					$\gamma\text{CH}_3$ 0,95
Leu	8,42	4,38	1,65	1,65	$\gamma\text{H}$ 1,64
Pro	4,44	2,28	2,02		$\gamma\text{CH}_2$ 2,03 2,03
					$\delta\text{CH}_2$ 3,68 3,65
Ser	8,38	4,50	3,88	3,88	
Thr	8,24	4,35	4,22		$\gamma\text{CH}_3$ 1,23
Asp	8,41	4,76	2,84	2,75	
Glu	8,37	4,29	2,09	1,97	$\gamma\text{CH}_2$ 2,31 2,28
Lys	8,41	4,36	1,85	1,76	$\gamma\text{CH}_2$ 1,45 1,45
					$\delta\text{CH}_2$ 1,70 1,70
					$\epsilon\text{CH}_2$ 3,02 3,02
					$\epsilon\text{NH}^3$ 7,52
Arg	8,27	4,38	1,89	1,79	$\gamma\text{CH}_2$ 1,70 1,70
					$\delta\text{CH}_2$ 3,32 3,32
					NH 7,17 6,62
Asn	8,75	4,75	2,83	2,75	$\gamma\text{NH}_2$ 7,59 6,91
					$\delta\text{NH}_2$ 2,38
Gln	8,41	4,37	2,13	2,01	$\gamma\text{CH}_2$ 2,38 2,38
					$\delta\text{CH}_2$ 6,87 7,59
Met	8,42	4,52	2,15	2,01	$\gamma\text{CH}_2$ 2,64 2,64
					$\epsilon\text{CH}_3$ 2,13
Cys	8,31	4,69	3,28	2,96	
Trp	8,09	4,70	3,23	3,19	2H 7,24
					4H 7,65
					5H 7,17
					6H 7,24
					7H 7,50
					NH 10,22
Phe	8,23	4,66	3,22	2,99	2,6H 7,30
					3,5H 7,39
					4H 7,34
Tyr	8,18	4,60	3,13	2,92	2,6H 7,15
					3,5H 6,86
His	8,41	4,63	3,26	3,20	2H 8,12
					4H 7,14

Desta forma, todos os sinais de RMN oriundos dos vários prótons da proteína são devidamente assinalados. Esta metodologia foi empregue com sucesso em peptídeos de até 60 resíduos de amino ácidos (FERENTZ e WAGNER, 2000). Em polipeptídeos maiores é possível aplicar a mesma estratégia de assinalamento seqüencial porém fazendo uso dos espectros TOCSY e NOESY em 3D editados para  $^{15}\text{N}$  ou  $^{13}\text{C}$  de forma a minimizar a sobreposição de picos. É comum, no entanto, que mesmo utilizando espectros 3D as ambigüidades permaneçam pois torna-se difícil identificar os vários amino ácidos vizinhos usando apenas as conectividades de NOE. Tornou-se então necessário adotar uma estratégia de assinalamento seqüencial que não dependesse das conectividades de NOE.

Os experimentos de tripla ressonância usam acoplamentos escalares de uma ou duas ligações para correlacionar até três núcleos diferentes. A magnetização inicia-se no núcleo do hidrogênio amídico e é transferida sucessivamente através das ligações covalentes dos átomos fornecendo informações sobre os núcleos de hidrogênio, nitrogênio ( $^{15}\text{N}$ ) e carbono ( $^{13}\text{C}$ ) da cadeia principal. A nomenclatura dos experimentos indica os caminhos de correlação entre os núcleos sendo que HN, N, HA, CA, HB, CB e CO, representam os spins hidrogênio amídico ( $^1\text{H}^{\text{N}}$ ), nitrogênio amídico ( $^{15}\text{N}$ ), hidrogênio alfa ( $^1\text{H}^{\alpha}$ ), carbono alfa ( $^{13}\text{C}^{\alpha}$ ), hidrogênio beta ( $^1\text{H}^{\beta}$ ), carbono beta ( $^{13}\text{C}^{\beta}$ ) e carbono da carbonila ( $^{13}\text{CO}$ ), respectivamente.

Núcleos através dos quais a coerência é apenas transferida durante a seqüência de pulso são indicados entre parênteses (IKURA *et al.*, 1990; KAY *et al.*, 1990). Assim, o experimento HNCO, por exemplo, correlaciona o carbono da carbonila (CO) com o próton (HN) e nitrogênio amídicos (N) do amino ácido seguinte. Já os experimentos HNCA e HN(CO)CA correlacionam o próton (HN) e nitrogênio amídicos (N) com o carbono alfa (CA) do próprio amino ácido e do amino

ácido anterior ou apenas do amino ácido anterior respectivamente (FERENTZ e WAGNER, 2000).

A Tabela 7 mostra os caminhos de transferência de magnetização entre os vários núcleos em alguns dos experimentos de tripla ressonância comumente utilizados para o assinalamento dos núcleos  $H^N$ , N,  $C\alpha$ ,  $H\alpha$  e CO da cadeia principal. É importante notar que em alguns destes experimentos específicos para o assinalamento da cadeia principal, os núcleos  $C\beta$  também são correlacionados.

Dependendo das características da proteína em estudo, nem sempre é necessário coletar todos os experimentos descritos acima para assinalar completamente os núcleos da cadeia principal da mesma. Ao analisar diversos experimentos de tripla ressonância simultaneamente, é possível assinalar vários sistemas de spins (i.e. amino ácidos) conectados entre si.

Os segmentos de spins conectados entre si e identificados na primeira etapa são então mapeados na seqüência da proteína. Este mapeamento é possível pois cada tipo de amino ácido apresenta deslocamentos químicos específicos para os diversos núcleos de nitrogênio e de carbono conforme descrito na Tabela 8. Assim, analisando os deslocamentos químicos obtidos é possível identificar os vários amino ácidos conectados entre si na seqüência primária.

Uma vez assinalados os núcleos da cadeia principal, os núcleos das cadeias laterais de cada amino ácido são também assinalados. O assinalamento das cadeias laterais é possível utilizando experimentos 3D de tripla ressonância específicos para o efeito. O experimento mais utilizado é o HCCH-TOCSY, mas experimentos semelhantes tais como os descritos na Tabela 9 podem também ser utilizados (FERENTZ e WAGNER, 2000).



**Tabela 7.** Experimentos 3D de tripla ressonância utilizados no assinalamento dos núcleos da cadeia principal (adaptado de FERENTZ e WAGNER, 2000).

Experimento	Transferência de Magnetização
HNCA	
HN(CO)CA	
HNCO	
HN(CA)CO	
HNCACB	
HN(COCA)CB	
CBCA(CO)NH	
CBCANH	

**Tabela 8.** Deslocamento químico médio dos núcleos de carbono e nitrogênio dos 20 amino ácidos naturais encontrados em proteínas (adaptado do BIOLOGICAL MAGNETIC RESONANCE BANK – BMRB Database, 2001).

Resíduo	$^{13}\text{C}^{\text{O}}$	$^{13}\text{C}^{\alpha}$	$^{13}\text{C}^{\beta}$	$^{13}\text{C}^{\gamma}$	$^{13}\text{C}^{\delta}$	$^{13}\text{C}^{\epsilon}$	$^{13}\text{C}^{\zeta}$	$^{15}\text{N}^{\text{H}}$	$^{15}\text{N}$
Gly	173.96	45.42	-	-	-	-	-	109.71	-
Ala	177.76	53.21	19.04	-	-	-	-	123.19	-
Val	175.75	62.55	32.78	21.59 21.43	-	-	-	121.64	-
Ile	175.85	61.63	38.84	28.02 17.70	13.82	-	-	121.62	-
Leu	176.98	55.63	42.45	26.90	24.84 24.18	-	-	122.34	-
Pro	176.69	63.28	32.03	27.44	50.39	-	-	35.53	-
Ser	174.56	58.61	63.85	-	-	-	-	116.42	-
Thr	174.59	62.21	69.55	21.71	-	-	-	115.79	-
Asp	176.49	54.62	40.94	177.00	-	-	-	120.71	-
Glu	176.92	57.52	30.19	36.17	182.25	-	-	120.68	-
Lys	176.53	56.87	32.97	25.16	29.06	42.00	-	121.10	-
Arg	176.49	56.92	30.87	27.42	43.21	-	158.98	120.66	-
Asn	175.19	53.44	38.85	176.60	-	-	-	119.12	112.75
Gln	176.39	29.27	33.79	179.11	-	-	-	119.89	111.66
Met	176.17	56.22	33.06	32.04	17.44	-	-	120.03	-
Cys	174.89	57.57	33.89	-	-	-	-	119.98	-
Trp	176.19	57.62	30.54	109.76	125.39 127.43	137.41 119.18	113.02 120.48 122.05	121.96	-
Phe	175.48	58.21	40.16	138.15	130.75 130.68	129.40 129.42	128.18	120.67	-
Tyr	175.44	58.00	39.41	129.32	131.88 131.17	117.57 117.42	151.68	121.62	-
His	175.19	56.35	30.03	129.45	119.40	135.22	-	119.64	187.00 181.00

**Tabela 9.** Experimentos 3D de tripla ressonância utilizados no assinalamento dos núcleos das cadeias laterais (adaptado de FERENTZ e WAGNER, 2000).

Experimento	Transferência de Magnetização
HCCH-TOCSY	
H(CC)NH-TOCSY	
(H)C(C)NH-TOCSY	
(H)C(C-CO)NH-TOCSY	
H(CC-CO)NH-TOCSY	

O experimento HCCH-TOCSY correlaciona os spins dos núcleos de carbono e de próton das cadeias laterais com carbono  $\alpha$  e com o próton  $\alpha$  da cadeia principal enquanto os demais experimentos correlacionam os núcleos de carbono e de próton das cadeias laterais com o grupo amídico da cadeia principal. Outros experimentos freqüentemente utilizados para o assinalamento dos núcleos das cadeias laterais são os experimentos de dupla ressonância 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TOCSY, 3D  $^1\text{H}$ - $^{13}\text{C}$  HSQC-NOESY e 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC.

Independentemente de quais forem os experimentos de RMN utilizados, o assinalamento completo de todas as ressonâncias é uma etapa fundamental no estudo de proteínas por RMN. Uma vez obtido o assinalamento completo da proteína não só é possível prosseguir para as demais etapas de estudos estruturais por RMN, como também é possível realizar toda uma vasta gama de outros estudos por RMN tais como estudos de dinâmica e estudos de interação da proteína de interesse com outras moléculas.

### 3.1.2.2. Identificação dos elementos de estrutura secundária

Uma vez obtido o assinalamento seqüencial de todos os amino ácidos presentes na molécula, e os respectivos valores de deslocamentos químicos dos vários núcleos, tem-se acesso imediato a informação estrutural muito valiosa. Há muito que é reconhecido que os valores dos deslocamentos químicos dos núcleos da cadeia principal contêm informações sobre a estrutura secundária da proteína em estudo. De fato, estruturas secundárias regulares de proteínas tais como hélices  $\alpha$  e fitas  $\beta$  apresentam valores de deslocamentos químicos característicos quando comparados com estruturas randômicas. Assim, os valores dos deslocamentos químicos de certos núcleos, como  $^1\text{H}^\alpha$ ,  $^{13}\text{C}^\alpha$  e  $^{13}\text{C}^\beta$ , podem ser utilizados como um indicativo de estrutura secundária. Em estruturas helicoidais, os núcleos  $^1\text{H}^\alpha$  e  $^{13}\text{C}^\beta$

apresentam deslocamentos químicos menores do que os valores esperados para um determinado aminoácido em conformação randômica, enquanto os núcleos  $^{13}\text{C}^\alpha$  apresentam valores maiores. Em fitas  $\beta$  observa-se o fenômeno oposto. Mudanças similares são também observadas para outros núcleos, como:  $^{13}\text{CO}$ ,  $^1\text{H}^\beta$ ,  $^1\text{H}^\text{N}$  e  $^{15}\text{N}$ .

Este fenômeno possibilitou a criação do chamado Índice de Deslocamento Químico ou “Chemical Shift Index” (CSI) (WISHART, *et al.*, 1992; WISHART e SYKES, 1994). O índice de deslocamento químico nada mais é do que a diferença entre os valores de deslocamento químico para os núcleos da cadeia principal de um dado amino ácido obtidos experimentalmente a partir do assinalamento seqüencial da proteína e os valores de deslocamento químico médios para o mesmo amino ácido em conformação randômica. Desta forma as regiões da proteína que apresentarem diferenças negativas para  $^1\text{H}^\alpha$  e  $^{13}\text{C}^\beta$  e, positivas para  $^{13}\text{C}^\alpha$  em quatro ou mais resíduos de amino ácidos consecutivos indicam a presença de uma conformação em  $\alpha$ -hélice. Já as regiões da proteína que apresentarem diferenças negativas para  $^1\text{H}^\alpha$  e  $^{13}\text{C}^\beta$  e, positivas para  $^{13}\text{C}^\alpha$  em quatro ou mais resíduos de amino ácidos consecutivos indicam a presença de uma conformação de fita  $\beta$ . Caso o índice de deslocamento químico esteja próximo de zero para uma dada região da proteína, esta região provavelmente não apresenta estrutura secundária regular (WUTRICH, 1986). Desta forma, assim que as ressonâncias da cadeia principal de uma dada proteína estão atribuídas, é possível predizer de forma rápida e simples o conteúdo de estrutura secundária da mesma.

Um outro método freqüentemente utilizado em estudos de RMN para obter informações sobre a estrutura secundária de uma proteína envolve o cálculo da constante de acoplamento escalar  $^3J_{\text{H}^\text{N}\text{H}^\alpha}$ . O acoplamento escalar ou acoplamento spin-spin é o fenômeno pelo qual o sinal de RMN de um dado núcleo A é afetado

pela interação com um núcleo B ao qual A está unido por ligação química. O acoplamento escalar, é representado pela constante de acoplamento  ${}^nJ_{ab}$ , onde n é o número de ligações covalentes separando os núcleos A e B. O acoplamento escalar pode ser evidenciado no espectro de RMN como um pico composto, denominado de multipeto (sinal dividido em dois ou mais componentes) em que os sinais estão separados por uma distância correspondente a J (Hz). Esta constante independe do campo magnético aplicado e a sua unidade é dada em Hz. Assim, a constante de acoplamento escalar  ${}^3J_{\text{HNH}\alpha}$  representa o acoplamento escalar a 3 ligações entre os núcleos  ${}^1\text{HN}$  e  ${}^1\text{H}^\alpha$  de um dado amino ácido. Em sistemas simples, a constante de acoplamento pode ser medida pela separação de uma ressonância específica em um espectro 1D ou 2D. Para sistemas maiores, onde a largura de linha dos sinais excede o valor do acoplamento, fica difícil extrair uma medida exata utilizando esta estratégia. Para solucionar este problema, surgiram uma série de experimentos que possibilitam a realização de medidas quantitativas de diversas constantes de acoplamento tais como o experimento 3D de dupla ressonância HNHA que fornece informações sobre a constante de acoplamento escalar  ${}^3J_{\text{HNH}\alpha}$ , nos quais o valor de J está relacionado à intensidade dos sinais em questão (VUISTER e BAX, 1993).

Uma vez obtidas as constantes de acoplamento experimentalmente, os ângulos diedros  $\Phi$  e  $\Psi$  da proteína em estudo podem ser facilmente calculados através da relação de Karplus (KARPLUS, 1959) que define uma relação geométrica entre a constante de acoplamento  ${}^3J_{\text{HNH}\alpha}$  e o ângulo diedro  $\Phi$ :

$${}^3J_{\text{HNH}\alpha} = 6.4 \cos^2(\Phi - 60) - 1.4 \cos(\Phi - 60) + 1.9 \quad [\text{Equação 1}]$$

em que  ${}^3J_{\text{HNH}\alpha}$  é dado em Hz (PARDI *et al.*, 1984).

Em proteínas os ângulos diedros  $\Phi$  para todos os resíduos de amino ácidos, com exceção das glicinas, estão concentrados na faixa de  $-30^\circ$  a  $-180^\circ$  (WÜTHRICH, 1986). Além disso, em regiões de estrutura secundária, os valores de  $\Phi$  são conhecidos:  $^3J_{\text{HNH}\alpha}$  varia de 4-6 Hz em  $\alpha$ -hélices e de 8-10 Hz em regiões de fita- $\beta$  (VUISTER *et al.*, 2002).

Assim, conhecendo o valor de  $^3J_{\text{HNH}\alpha}$  obtido experimentalmente, é possível identificar as estruturas secundárias regulares presentes na proteína em estudo. Segmentos de três ou mais resíduos de amino ácidos consecutivos que apresentam valores de  $^3J_{\text{HNH}\alpha} < 6,0$  Hz indicam conformação em  $\alpha$ -hélice. Os mesmos segmentos com valores de  $^3J_{\text{HNH}\alpha} > 8,0$  Hz indicam conformação em fita- $\beta$  (WÜTHRICH, 1986; VUISTER *et al.*, 2002).

Atualmente é também possível prever os valores de ângulos diedros, e conseqüentemente o tipo de estrutura secundária presente na proteína em estudo, baseando-se apenas nos valores de deslocamento químico dos núcleos da cadeia principal obtidos experimentalmente. Este método foi descrito por CORNILESCU e colaboradores (1999) e foi denominado TALOS (“Torsion Angle Likelihood Obtained from Shift and Sequence Similarity”). Este programa baseia-se em uma base de dados de 20 proteínas para as quais os valores de deslocamentos químicos bem como as estruturas cristalinas de alta resolução são conhecidos. Usando os deslocamentos químicos dos núcleos  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^{13}\text{CO}$ ,  $^1\text{H}^\alpha$  e  $^{15}\text{N}$  obtidos experimentalmente para a proteína em estudo, o programa procura na base de dados pelos resultados mais próximos para cada tripleto de amino ácidos baseado na similaridade dos deslocamentos químicos e da seqüência primária. Se os resíduos centrais nos dez melhores “hits” apresentarem ângulos  $\Phi$  e  $\Psi$  semelhantes entre si, então a média destes valores pode ser usada como predição dos valores

dos ângulos diedros da proteína em estudo. Ângulos diedros preditos por TALOS são freqüentemente incorporados como restrições para o cálculo da estrutura da proteína (veja por exemplo KOPKE SALINAS *et al.*, 2005; KATSUYAMA *et al.*, 2007). Recentemente WANG e LIU (2007) descreveram um novo método muito semelhante para a predição de ângulos  $\Phi$  e  $\Psi$  que também faz uso da seqüência primária e dos deslocamentos químicos dos núcleos da cadeia principal de segmentos tripeptídicos. Esta nova abordagem baseia-se na probabilidade do resíduo central apresentar uma determinada conformação, baseado nos deslocamentos químicos de segmentos tripeptídicos de estruturas conhecidas e apresenta predições tão boas quanto o programa TALOS (WANG e LIU, 2007).

### 3.1.2.3. Obtenção de restrições para o cálculo de estrutura

Com o assinalamento em mãos, pode-se iniciar a coleta de restrições para o cálculo da estrutura tridimensional. As duas principais fontes de restrições para o cálculo de estrutura são as restrições dos ângulos de torção e as restrições de distância entre os núcleos.

Conforme descrito acima, os ângulos de torção podem ser derivados das medidas de constantes de acoplamento escalar a cada três ligações ( $^3J$ ) ou preditos usando o programa TALOS (CORNILESCU *et al.*, 1999). Estas constantes podem ser utilizadas como parâmetros de restrição para o cálculo da estrutura tridimensional da proteína. Além disso, restrições dos ângulos entres vetores de ligações químicas vêm sendo utilizadas freqüentemente em cálculos de estrutura.

Apesar da importância das restrições dos ângulos de torção, a principal fonte de informação geométrica usada na determinação da estrutura de proteínas reslide porém nas distancias entre os vários núcleos da molécula em estudo. Os experimentos de RMN que nos fornecem informações sobre a distância entre os



núcleos se baseiam no Efeito Nuclear Overhauser (NOE). Conforme descrito anteriormente, a intensidade do sinal de NOE resultante da interação espacial entre dois núcleos é inversamente proporcional à distância entre os núcleos elevada à sexta potência. Desta forma, as medidas de NOE podem ser utilizadas para calcular as distâncias entre os vários núcleos e conseqüentemente para determinar a estrutura tridimensional da proteína em estudo. De fato, as distâncias derivadas das medidas de NOE são muito importantes para o cálculo de estrutura pois são altamente restritivas para a conformação da proteína, particularmente quando envolvem resíduos que não estão próximos na seqüência primária (WÜTHRICH, 1986).

O passo inicial na determinação da conformação de uma proteína consiste no assinalamento de todos os picos cruzados do espectro de NOESY, utilizando para isso os deslocamentos químicos obtidos através o assinalamento seqüencial das ressonâncias de todos os núcleos  $^1\text{H}$  da proteína. Para proteínas pequenas é possível atribuir os vários NOEs fazendo uso do experimento convencional 2D  $^1\text{H}$ - $^1\text{H}$ -NOESY. Porém, à medida que o tamanho da proteína em estudo aumenta, a sobreposição de picos torna-se inevitável, e o assinalamento dos vários picos de NOE requer a coleta de um experimento 3D heteronuclear  $^{15}\text{N}$ - ou  $^{13}\text{C}$ -NOESY-HSQC (FERENTZ e WAGNER, 2000). Ainda assim, é comum que várias ambigüidades permaneçam no assinalamento dos vários picos de NOE pois um mesmo pico de correlação pode ser atribuído a dois ou mais pares de prótons. Estas ambigüidades só podem ser resolvidas após a determinação inicial de uma estrutura de baixa resolução. Em geral, as intensidades dos picos cruzados dos espectros de NOESY são classificadas em três categorias diferentes: picos fortes, médios e fracos correspondentes a distâncias de 1,8 a 2,8 Å, 1,8 a 3,4 Å e 1,8 a 5,0

Á, respectivamente. As intensidades dos NOEs são comumente calibradas contra alguma referência interna de distância fixa. Deve-se assegurar que o grupo funcional escolhido como referência de distância não está presente em uma região de movimento local da proteína, pois isto pode levar a resultados errôneos (WÜTHRICH, 1986). As restrições de distância inter-prótons obtidas através do espectro NOESY são então utilizadas em algoritmos computacionais para calcular a estrutura tridimensional.

#### 3.1.2.4. Cálculo e refinamento da estrutura tridimensional

A determinação de estruturas tridimensionais de proteínas por RMN faz uso de restrições estruturais tais como distâncias entre núcleos e ângulos de torção obtidas a partir de diversos experimentos de RMN. No entanto, estas restrições sozinhas muitas vezes não são suficientes para calcular a estrutura de uma proteína e são portanto suplementadas com dados conhecidos sobre a estereoquímica de proteínas tais como comprimentos e ângulos de ligações e interações de van der Waals (EVANS, 1995; GÜNTERT, 1998).

Existem diversos métodos para varrer o espaço conformacional em busca estruturas compatíveis com as restrições de RMN. Atualmente, o método mais utilizado é a dinâmica molecular. Os diferentes protocolos existentes combinam etapas de dinâmica molecular com etapas “simulated annealing” no espaço cartesiano ou no espaço de ângulos torsionais e estão implementados em programas computacionais como o CNS (BRUNGER *et al.*, 1998) e o DYANA (GÜNTERT *et al.*, 1997).

Todos os algoritmos de cálculo de estrutura assumem que as restrições experimentais estão corretamente assinaladas. No entanto, nos estágios iniciais de cálculo da estrutura, existem muitas vezes deslocamentos químicos e NOEs mal-

assinalados devido a artefatos experimentais e erros humanos. Desta forma, após o cálculo inicial de estrutura devem-se reavaliar as restrições experimentais de forma a remover potenciais ambigüidades ou erros nas mesmas, promovendo assim a convergência e o refinamento das estruturas calculadas. Este processo é repetido iterativamente de forma a identificar precisamente mais e mais restrições, removendo ambigüidades e obtendo assim estruturas mais precisas. Uma vez que, usualmente várias conformações são compatíveis com os dados experimentais, o resultado final obtido é uma família de estruturas que obedecem às restrições experimentais (EVANS, 1995; GÜNTERT, 1998).

A precisão das estruturas obtidas depende diretamente do número e da qualidade de restrições experimentais. Quanto mais restrições puderem ser obtidas, mais precisa será a família de estruturas. A precisão das estruturas obtidas é avaliada pela raiz quadrada do desvio padrão médio (RMSD) das coordenadas atômicas das varias estruturas. Quando a família de estruturas converge para um único modelo, o RMSD é pequeno. A qualidade das estruturas é também avaliada através dos ângulos  $\Phi$  e  $\Psi$  observados no diagrama de Ramachandran. Para uma estrutura de boa qualidade, os ângulos  $\Phi$  e  $\Psi$  devem constar nas regiões permitidas do diagrama. Para fazer essa análise existem programas como AQUA/PROCHECK NMR (LASKOWSKI e col., 1996) e WHAT IF (VRIEND e SANDER, 1991) que avaliam ainda outros parâmetros de qualidade estereoquímica como os ângulos  $\omega$  e  $\chi$  e o número de sobreposições de átomos.

A estrutura final depositada no PDB é um conjunto de 20 estruturas possíveis de menor energia (minimização da estrutura) que obedecem a todas as restrições experimentais e cujo desvio médio RMSD entre os seus átomos da cadeia principal é normalmente inferior a 1 Å (EVANS, 1995; GÜNTERT, 1998).

### 3.1.2.5. Análise de acoplamentos dipolares residuais (RDCs)

Em estudos estruturais clássicos de proteínas por RMN, as informações estruturais são derivadas de informações tais como distâncias, e ângulos diedros, (WÜTHRICH, 1986). Estas informações estruturais são então compiladas para construir o modelo tridimensional da molécula em estudo. Apesar de eficaz, esta abordagem apresenta limitações importantes. Em primeiro lugar, em moléculas grandes (>25 kDa) é muitas vezes difícil analisar dados de NOE devido ao alargamento das linhas e ao aumento complexidade espectral. Por outro lado, estudos de proteínas de múltiplos domínios e complexos proteína-proteína ou proteína-ácido nucléico são também dificultados pela falta de NOEs suficientes nas regiões de interface entre os vários domínios/moléculas.

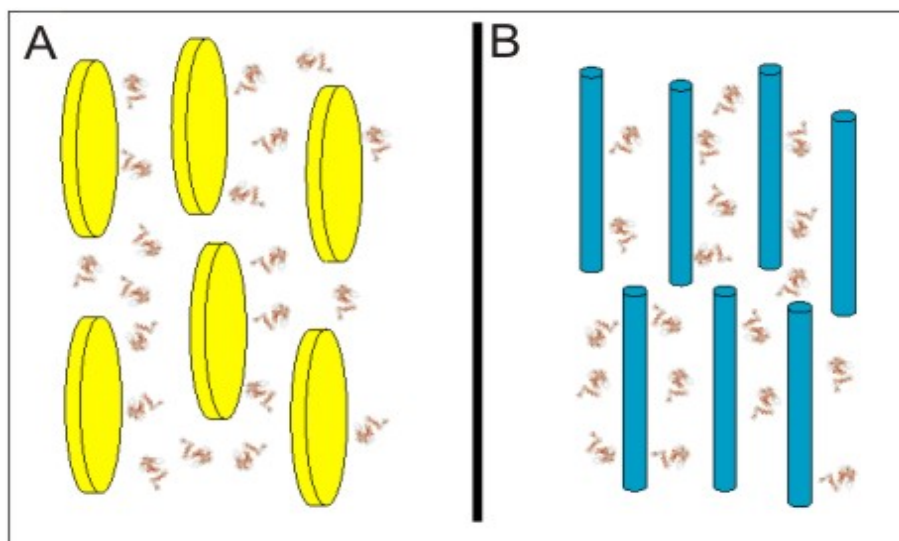
Os acoplamentos dipolares residuais, ao fornecerem informações de orientação dos núcleos da molécula em estudo, permitem abordagens totalmente novas para a biologia estrutural em solução (CLORE e GRONENBORN, 1998; TJANDRA, 1999; PRESTGARD, 2001; TOLMAN, 2001; PRESTGARD *et al.*, 2004; SIMON *et al.*, 2005).

O acoplamento dipolar é uma interação entre dois spins nucleares que ocorre através do espaço e, tal como no acoplamento escalar (J), produz um desdobramento do sinal das ressonâncias no espectro de RMN. Acoplamentos dipolares no entanto, diferem bastante de acoplamentos escalares uma vez que o acoplamento dipolar médio é igual a zero nos casos em que a molécula tomba isotropicamente, ou seja, assume todas as orientações possíveis relativas ao campo magnético com igual probabilidade. Assim, uma vez que na RMN em solução as moléculas em estudo se encontram em condições de tombamento isotrópico, os acoplamentos dipolares não costumam ser observados. No entanto, ao se

estabelecer um pequeno grau de alinhamento das moléculas em solução, pequenos desdobramentos residuais resultantes destas interações podem ser observados no espectro, denominados acoplamentos dipolares residuais ou RDCs. Os acoplamentos dipolares residuais contêm informações estruturais valiosas pois estão diretamente relacionados à orientação dos vetores nucleares relativamente a uma referência padrão global que é determinada pela natureza do alinhamento molecular (TJANDRA, 1999; TOLMAN, 2001; PRESTGARD *et al.*, 2004).

Assim, a obtenção de pequenos níveis de alinhamento (aproximadamente 1 molécula em 1000) é fundamental para a coleta de dados de acoplamentos dipolares residuais de uma molécula em solução. No caso de moléculas paramagnéticas e DNA um pequeno grau de alinhamento é naturalmente observado na presença de campos magnéticos fortes. Este método, apesar de eficaz é bastante limitado uma vez que pode ser aplicado apenas em sistemas específicos e requer campos magnéticos muito altos. Um método alternativo de produzir um pequeno alinhamento de uma macromolécula é pela adição de uma fase líquida cristalina ao sistema (TJANDRA e BAX, 1997).

Esta fase líquida cristalina, devido à sua anisotropia na susceptibilidade magnética, orienta-se com o campo magnético. A orientação da fase líquida cristalina com o campo induz o alinhamento residual das moléculas em estudo presentes no soluto. Atualmente existem vários tipos de meios que podem ser utilizados com este objetivo tais como bicelas lipídicas (SANDERS e SCHWONEK, 1992; SANDERS *et al.*, 1994), o fago filamentoso Pf1 (HANSEN *et al.*, 1998), fragmentos de membrana roxa (KOENIG *et al.*, 1999) e partículas virais (CLORE *et al.*, 1998) (Figura 23).



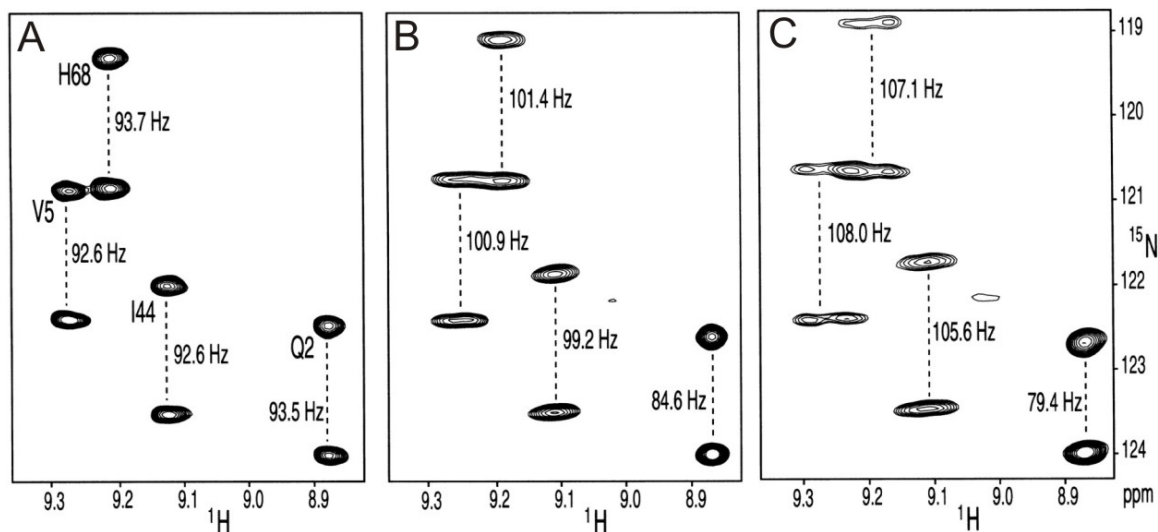
**Figura 23.** Ilustração de tipos de meio líquido cristalino (TJANDRA, 1999).

A. Os discos grandes representam bicelas lipídicas ou fragmentos de membrana roxa que se orientam ortogonalmente relativamente ao campo magnético.

B. Os cilindros longos representam partículas virais que se orientam com o seu eixo longo paralelo ao campo magnético.

Tipicamente procede-se à coleta de dados de RMN em condições isotrópicas e anisotrópicas. O acoplamento dipolar  $D$  entre dois núcleos A e B ( $D_{AB}$ ) pode ser facilmente determinado como uma variação no desdobramento do sinal das ressonâncias no estado alinhado ( $J+D$ ) relativo ao estado isotrópico ( $J$ ) (TJANDRA, 1999; AL-HASHIMI e PATEL, 2002) (Figura 24).

A magnitude de  $D_{AB}$  pode ser positiva ou negativa (Figura 24). Para uma molécula totalmente orientada o acoplamento dipolar do grupo amídico  $^1\text{H}-^{15}\text{N}$  ( $D_{\text{HN}}$ ) seria da ordem de 20 kHz. No entanto o grau de alinhamento aplicado é tão pequeno que os acoplamentos dipolares  $^1\text{H}-^{15}\text{N}$  ou  $^1\text{H}-^{13}\text{C}$  observados são normalmente da ordem de 5 Hz (HANSEN *et al.*, 1998). A magnitude de  $D$  está relacionada ao tensor de alinhamento do vetor nuclear em estudo (ex. vetor NH) e fornece informações da orientação deste vetor no espaço cartesiano.



**Figura 24.** Pequenas regiões de espectros  $^1\text{H}$ - $^{15}\text{N}$  HSQC da proteína ubiquitina coletados na ausência de desacoplamento de  $^1\text{J}_{\text{NH}}$  na dimensão indireta para diferentes níveis de alinhamento molecular. (BAX, 2003).

A. Espectro isotrópico. Um dubleto é observado para cada par de núcleos  $^1\text{H}$ - $^{15}\text{N}$ . Os acoplamentos escalares medidos ( $^1\text{J}_{\text{NH}}$ ) estão indicados em Hz.

B. Espectro coletado em meio líquido cristalino contendo 4,5% (p/v) de bicelas (proporção molar de 30:10:1 DMPC, DHPC e CTAB).

C. Espectro coletado em meio líquido cristalino contendo 8% (p/v) de bicelas (proporção molar de 30:10:1 DMPC, DHPC e CTAB).

Os desdobramentos de sinal medidos nos painéis B e C correspondem à soma do acoplamento escalar  $^1\text{J}_{\text{NH}}$  e do acoplamento dipolar  $\text{D}_{\text{NH}}$ . O alargamento dos sinais na dimensão de  $^1\text{H}$  é causado por acoplamento dipolares  $^1\text{H}$ - $^1\text{H}$ .

Existem diversos métodos que podem ser usados para medir acoplamentos dipolares residuais mas os utilizados são experimentos do tipo SCE-HSQC, E.COSY e IPAP (“in-phase/anti-phase”) (TJANDRA, 1999).

Ao fornecerem importantes restrições de orientação, os acoplamentos dipolares residuais têm-se tornado ferramentas cada vez mais fundamentais para o estudo de proteínas por RMN. Os acoplamentos dipolares residuais têm sido incorporados como restrições adicionais para cálculo de estrutura (LIPSITZ e TJANDRA, 2004). A inclusão destas restrições geralmente melhora a qualidade das

famílias de estruturas geradas e reduz significativamente o número de resíduos em regiões não permitidas do diagrama de Ramachandran. Além disso, os acoplamentos dipolares residuais permitem calcular estruturas de proteínas que apresentam um número reduzido de NOEs, o que além de aumentar a precisão das mesmas diminui imensamente o tempo necessário para a sua determinação. Atualmente, os acoplamentos dipolares residuais são comumente utilizados no refinamento de estruturas por RMN e a literatura está repleta de exemplos onde isso ocorre (veja por exemplo KOPKE SALINAS *et al.*, 2005 e CICERO *et al.*, 2007).

Os acoplamentos dipolares residuais permitem também fazer comparações das estruturas obtidas por RMN com homólogos estruturais disponíveis, possibilitando assim a validação rápida das estruturas calculadas inicialmente e subsequente refinamento das mesmas (SASS *et al.*, 2001; KIKUCHI *et al.*, 2002).

Não há dúvida que os acoplamentos dipolares residuais têm permitido melhorar significativamente a qualidade das estruturas determinadas por RMN. No entanto, a principal vantagem dos acoplamentos dipolares residuais reside no fato destes poderem ser medidos rotineiramente imediatamente após o assinalamento da cadeia principal da proteína, ao contrário das medidas de distância de NOE que requerem o assinalamento completo das várias ressonâncias seguido da identificação não-ambígua das correlações de NOE. Esta característica abre um sem número de possibilidades para o uso de dados de acoplamentos dipolares residuais em estudos estruturais por RMN (TIAN *et al.*, 2001; PRESTGARD *et al.*, 2004; LIPSITZ e TJANDRA, 2004).

Os acoplamentos dipolares residuais podem também ser utilizados para determinar se a estrutura cristalina de uma proteína apresenta a mesma conformação em solução (LUKIN *et al.*, 2003) ou ainda para caracterizar sistemas



que sofreram alterações conformacionais de forma rápida e eficaz (CHOU *et al.*, 2000). Os acoplamentos dipolares residuais são ainda muito úteis no estudo de proteínas de múltiplos domínios para determinar a orientação relativa dos vários domínios (FISCHER *et al.*, 1999; PRESTGARD *et al.*, 2004).

Recentemente foi também sugerido que nos casos em que bastantes acoplamentos dipolares residuais podem ser medidos ao longo da molécula na presença de dois tipos de meios de alinhamento, a estrutura da proteína poderá, em teoria, ser determinada usando apenas dados de acoplamentos dipolares residuais. O modelo estrutural pode ser construído através da junção de unidades peptídicas depositadas no PDB, cuja orientação é conhecida e não-ambígua, e que reproduzem os dados experimentais (DELAGLIO *et al.*, 2000; CHAMPIER *et al.*, 2002). Alternativamente a estrutura pode ser calculada *de novo* usando a abordagem MECCANO, um algoritmo que calcula a estrutura enovelada da cadeia principal de um polipeptídeo a partir da orientação de cada unidade peptídica obtida via acoplamentos dipolares residuais (HUS *et al.*, 2001; BÉRAUD *et al.*, 2002).

Os acoplamentos dipolares residuais podem ainda ser utilizados para validar modelos estruturais e prever enovelamentos de proteínas desconhecidas (TJANDRA *et al.*, 1997; TOLMAN, 2001; PRESTGARD *et al.*, 2004). Nestes casos, os valores de acoplamentos dipolares residuais experimentais medidos para a molécula em estudo são comparados com valores de acoplamentos dipolares residuais teóricos calculados a partir de uma estrutura conhecida. Se os valores teóricos e experimentais forem próximos, então as duas estruturas, i.e. modelo conhecido e molécula em estudo, são necessariamente próximas. Esta metodologia foi proposta pela primeira vez por ANNILA e colaboradores (1999) em um estudo da proteína caleritina de *Saccharopolyspora erythracea* que foi identificada como tendo

uma estrutura muito similar a duas proteínas sarcoplasmáticas que se ligam a cálcio de duas espécies diferentes. Posteriormente, foi desenvolvido o programa DIPO COUP que usando dados de deslocamentos químicos e acoplamentos dipolares residuais como dados de entrada, faz uma busca por estruturas homologas em um subgrupo do PDB (MEILER *et al.*, 2000). Este tipo de estudo poderá ser particularmente útil em projetos de proteômica estrutural, onde é desejável avaliar se o enovelamento de uma dada proteína é novo, e portanto alvo desejável para determinação estrutural detalhada por cristalografia de raios-X ou RMN.

### 3.2. OBJETIVOS

A proteína XAC2000 de *Xac* foi selecionada, em julho de 2001, como uma das proteínas-alvo para os estudos de triagem descritos no Capítulo 2. Esta proteína apresentou um excelente espectro 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC e revelou-se uma excelente candidata para estudos estruturais.

Apesar de ter estrutura e função desconhecidas, esta proteína apresentava um elevado grau de conservação entre vários genomas o que a tornava um alvo interessante para estudos estruturais mais detalhados.

Em 2002, DOUGAN e colaboradores descreveram a proteína ClpS de *E.coli*, uma proteína acessória do complexo proteolítico ClpAP. Posteriormente, a estrutura cristalográfica da proteína ClpS de *E.coli* foi determinada em complexo com o N-terminal da proteína ClpA por dois grupos independentes, revelando uma proteína  $\alpha/\beta$  composta por três  $\alpha$ -hélices conectadas a três fitas- $\beta$  antiparalelas em um arranjo  $\beta\alpha\alpha\beta\alpha\beta$  (GUO *et al.*, 2002a; ZETH *et al.*, 2002).

Esta proteína apresenta 56% identidade e 84% de similaridade com a proteína hipotética XAC2000 por nós selecionada, e cujos estudos estruturais por RMN já haviam iniciado. Assim, os objetivos do nosso trabalho, que originalmente visavam a determinação da estrutura tridimensional da proteína XAC2000/ClpS por RMN, passaram a focalizar principalmente a complementação dos dados obtidos por GUO e colaboradores (2002a) e ZETH e colaboradores (2002) em seus estudos por cristalografia.

Desta forma, em colaboração com a Dra. Thelma Pertinhez do Laboratório Nacional de Luz Síncrotron, os trabalhos descritos neste capítulo tiveram os seguintes objetivos:

- a) Comparação da estrutura tridimensional de ClpS de *E. coli* com a estrutura ClpS de *Xac*;
- b) Análise da estrutura da proteína ClpS em solução;
- c) Obtenção de informações estruturais sobre a região N-terminal - resíduos 1 a 20 da proteína ClpS ausentes nas estruturas cristalinas;
- d) Análise da dinâmica da proteína ClpS em solução;
- e) Estudos de interação da proteína ClpS com o domínio N-terminal da proteína ClpA em solução.

### 3.3. MATERIAIS E MÉTODOS

#### 3.3.1. Expressão e purificação das proteínas em estudo

##### 3.3.1.1. Expressão da proteína ClpS de Xac

A expressão da proteína ClpS não-marcada foi realizada em meio rico 2xTY (16g/L triptona, 10 g/L extrato de levedo, 5 g/L NaCl, pH 7,4). Células de *E. coli* da cepa BL21(DE3)pLysS foram transformadas com o gene *clps* clonado no vetor pET3a pelo método de CaCl<sub>2</sub> (SAMBROOK *et al.*, 2001) conforme descrito na seção 2.3.7.1. Uma das colônias obtidas foi pré-inoculada em 50 mL de meio 2xTY suplementado com 200 µg/mL de ampicilina e 200 µg/mL de cloranfenicol. Após o crescimento a 37 °C e 200 rpm até D.O.<sub>600nm</sub> = 0,8 a 1,2, 5 mL desta cultura foram utilizados para inocular 500 mL de meio 2xTY suplementado com 200 µg/mL de ampicilina e 200 µg/mL de cloranfenicol. As células foram novamente incubadas a 37 °C, e 200 rpm até D.O.<sub>600nm</sub> = 0,8 quando foram induzidas pela adição de 1 mM de IPTG. Após 4h de indução, as células foram centrifugadas por 20 minutos, a 4 °C e 5000 rpm, ressuspendidas em 100 mL de solução de lise (1 mM EDTA, 25% sacarose, 50 mM Tris-HCl pH 8,0, 14 mM de β-mercaptoetanol e 100 µM de PMSF) e armazenadas a -20 °C. O nível de expressão foi verificado através de SDS-PAGE em géis de 15% acrilamida e 0,4% bisacrilamida, aplicando-se as alíquotas do controle e das amostras induzidas.

A expressão da proteína ClpS marcada com <sup>15</sup>N foi realizada em meio mínimo M9 contendo <sup>15</sup>N (6 g/L de Na<sub>2</sub>HPO<sub>4</sub> pH 7,35, 3 g/L de KH<sub>2</sub>PO<sub>4</sub>, 0,5 g/L NaCl, 2 mM MgSO<sub>4</sub>, 0,4% glicose, 100 µM CaCl<sub>2</sub>, 10 mg tiamina, 1g/L <sup>15</sup>NH<sub>4</sub>Cl). Células de *E. coli* da cepa BL21(DE3)pLysS foram transformadas com o gene *clps* clonado no

vetor pET3a pelo método de  $\text{CaCl}_2$  (SAMBROOK *et al.*, 2001). Uma das colônias obtidas foi pré-inoculada em 50 mL de meio 2xTY suplementado com 200  $\mu\text{g/mL}$  de ampicilina e 200  $\mu\text{g/mL}$  de cloranfenicol. Após o crescimento a 37 °C e 200 rpm até a  $\text{DO}_{600} = 0,8-1,2$ , 5 mL desta cultura foram utilizados para inocular 500 mL de meio mínimo M9 contendo  $^{15}\text{NH}_4\text{Cl}$  suplementado com 200  $\mu\text{g/mL}$  de ampicilina e 200  $\mu\text{g/mL}$  de cloranfenicol. As células foram novamente incubadas a 37 °C, e 200 rpm até  $\text{DO}_{600} = 0,8$  quando foram induzidas pela adição de 1mM de ITPG. Após 8h de indução, as células foram centrifugadas por 20 minutos, a 4 °C e 5000 rpm, ressuspensas em 25 mL de solução de lise (1 mM EDTA, 25% sacarose, 50 mM Tris-HCl pH 8,0, 14 mM de  $\beta$ -mercaptoetanol e 100  $\mu\text{M}$  de PMSF) e armazenadas a -20 °C. O nível de expressão foi verificado através de SDS-PAGE em géis de 15% acrilamida e 0,4% bisacrilamida, aplicando-se as alíquotas do controle e das amostras induzidas.

A expressão da proteína ClpS duplamente marcada com  $^{15}\text{N}$  e  $^{13}\text{C}$  foi realizada em meio mínimo M9 contendo com  $^{15}\text{N}$  e  $^{13}\text{C}$  (6 g/L de  $\text{Na}_2\text{HPO}_4$  pH 7,35, 3 g/L de  $\text{KH}_2\text{PO}_4$ , 0,5 g/L NaCl, 2 mM  $\text{MgSO}_4$ , 0,4%  $^{13}\text{C}$ -glicose, 100  $\mu\text{M}$   $\text{CaCl}_2$ , 10 mg Tiamina, 1g/L  $^{15}\text{NH}_4\text{Cl}$ ) suplementado com 200  $\mu\text{g/mL}$  de ampicilina e 200  $\mu\text{g/mL}$  de cloranfenicol e utilizando a metodologia descrita acima para expressão com  $^{15}\text{N}$ .

### 3.3.1.2. Purificação da proteína ClpS de *Xac*

A proteína ClpS, não marcada, marcada com  $^{15}\text{N}$  e duplamente marcada com  $^{15}\text{N}$  e  $^{13}\text{C}$ , foi purificada utilizando métodos convencionais de lise por French Press®, precipitação por sulfato de amônio, cromatografia de troca iônica e gel filtração conforme descrito na seção 2.3.7.2.

### 3.3.1.3. Amplificação e clonagem de um fragmento do N-terminal da proteína ClpA de *Xac*

A seqüência de DNA que codifica a região N-terminal (resíduos 1-143) da proteína ClpA foi amplificada por PCR a partir de DNA genômico de *Xac* utilizando oligonucleotídeos desenhados especificamente para o efeito descritos na Tabela 10. Uma vez amplificado, o DNA de interesse foi clonado no vetor pET3a (Studier *et al.*, 1990) usando os sítios de restrição *Nde* I e *Hind* III e seqüenciado conforme descrito nas secções 2.3.2.1 a 2.3.2.3.

**Tabela 10.** Oligonucleotídeos utilizados na amplificação da região que codifica o N-terminal da proteína ClpA de *Xac*.

Oligo	Seqüência
ClpA (+)	5' CATGCCATGCATATGTTTCAGCAAAGACC 3'
ClpA (-)	5' ACACTCGAGAAGCTTGGTACTTGGCAATGCG3'

### 3.3.1.4. Expressão e purificação do N-terminal da proteína ClpA de *Xac*

O N-terminal da proteína ClpA de *Xac* foi expresso em meio não-marcado 2xTY conforme descrito na secção 3.3.1.1. O N-terminal da proteína ClpA de *Xac* foi purificado utilizando métodos convencionais. As células de *E. coli* ressuspendidas em tampão de lise e armazenadas a -20 °C foram lisadas utilizando French Press® com uma pressão de 16000 psi. O lisado foi centrifugado a 5000 rpm a 4 °C por 1 hora. Ao sobrenadante obtido foi adicionado 2U de DNase (Sigma) e 4 mM de cloreto de magnésio e a mistura foi agitada lentamente em banho de gelo por 60 min. A amostra tratada com DNase foi aplicada em uma coluna de troca iônica Q-Sepharose-FF (sistema Äkta, Amersham Biosciences). Utilizou-se um gradiente de eluição de zero a 400 mM NaCl, em tampão 50 mM Tris-HCl pH 8,0, 1 mM EDTA, 14 mM β-mercaptoetanol, em 10 volumes de coluna. As frações obtidas foram

analisadas utilizando géis de 15% SDS-PAGE, e aquelas que continham a proteína de interesse foram concentradas usando um concentrador Amicon com um “cutoff” de 10 kDa até um volume final de 5mL. A amostra foi então aplicada em uma coluna de gel filtração Superdex-75 prepgrade (sistema Äkta, Amersham Biosciences) e eluída utilizando 1,5 volumes de coluna de tampão 50 mM Tris-HCl pH 8,0, 1 mM EDTA, 14 mM  $\beta$ -mercaptoetanol e 0,5 mM azida sódica com um fluxo de 1,5 mL por min. As frações puras da proteína, verificadas em géis SDS-PAGE, foram dialisadas em membranas com um “cutoff” de 10 kDa contra 10 mM  $\text{NH}_4\text{CO}_3$  e 14 mM  $\beta$ -mercaptoetanol, congeladas a  $-80\text{ }^\circ\text{C}$  e liofilizadas. O liofilizado obtido foi armazenado a  $4\text{ }^\circ\text{C}$ .

### 3.3.2. Experimentos de Dicroísmo Circular (CD)

Visando uma melhor caracterização estrutural, a proteína ClpS foi analisada por CD utilizando um espectropolarímetro Jasco J-810 CD (Jasco Corp., Japão). As amostras foram preparadas utilizando a proteína não-marcada liofilizada, ressuspendida no tampão de interesse. A concentração das amostras foi determinada através da leitura da sua absorvância a 280 nm, utilizando-se um espectrofotômetro Beckman Coulter DU640<sup>®</sup>, e de seu coeficiente de absorvidade molar ( $\epsilon$ ) teórico calculado a partir da seqüência primária da proteína através do servidor EXPASY (Fonte: <http://expasy.org/>, 2003). A normalização dos dados de CD para elipticidade molar  $[\theta]$  ( $\text{degree} \times \text{cm}^2 \times \text{dmol}^{-1}$ ) em função da concentração e tamanho da proteína e das condições de leitura é dada pela equação 2:



$$[\theta] = \theta / 10 \times C \times l \times N_A \quad \text{[Equação 2]}$$

onde:  $\theta$  é a elipticidade (milidegree); C, a concentração molar; l, o caminho ótico em centímetros e  $N_A$  o número de amino ácidos da proteína.

### 3.3.2.1. Variação do pH

Para testar a estabilidade da proteína ClpS em diferentes condições de pH, foi utilizada uma amostra de 20  $\mu$ M de proteína em tampão 10 mM  $\text{NaH}_2\text{PO}_4$  pH 7. O pH da amostra foi variado entre pH 3,0 e pH 9,0 por titulações com HCl ou NaOH e foram adquiridos espectros de CD a 20 °C no UV distante entre 190 e 260 nm. Os espectros foram obtidos em intervalos de 0,5 nm usando um tempo de resposta de 4 s e uma velocidade de 50 nm/min. A fenda utilizada foi de 1 nm. Um total de 4 varreduras foi adquirido para cada ponto no espectro. É importante notar que, em todas as titulações, foram considerados os efeitos da diluição da amostra nos cálculos de elipticidade molar  $[\theta]$ .

### 3.3.2.2. Variação da força iônica

Para testar a estabilidade da proteína ClpS em diferentes forças iônicas, utilizou-se a proteína a 10  $\mu$ M, em 10 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8 titulando-a com, 1 mM  $\text{MgCl}_2$ , 2 mM  $\text{MgCl}_2$  e 5 mM  $\text{MgCl}_2$ , 10 mM NaCl, 50 mM NaCl e 100 mM NaCl. Os espectros de CD foram adquiridos conforme descrito em 3.3.2.1.

### 3.3.2.3. Desnaturação térmica

Para testar a estabilidade térmica da proteína ClpS foi utilizada uma amostra de 20  $\mu$ M de proteína em 10 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8. Para a obtenção das curvas de

desnaturação, primeiramente foram adquiridos os espectros de varredura no UV distante de 190 a 260 nm, variando a temperatura a cada 5 °C, entre 15 °C e 90 °C. Para se obter a curva de desnaturação térmica e estabelecer a temperatura de transição da desnaturação, foi adotado o comprimento de onda fixo de 222 nm e a temperatura foi variada em 1 °C/min, entre 15 °C e 93 °C. A reversibilidade da desnaturação térmica foi acompanhada pelo resfriamento da amostra de 93 °C para 15 °C em 1 °C/min. Como controle da amostra, foram adquiridos espectros de CD conforme descrito em 2.3.1.1, antes da desnaturação (15 °C), após a desnaturação (93 °C) e após o retorno (15 °C).

### 3.3.3. Experimentos de RMN

A maioria dos experimentos de RMN foram realizados a 298 K (25 °C) em equipamentos Varian Inova® operando a 500 MHz ou 600 MHz do Centro de Biologia Estrutural Molecular do LNILS. Alguns espectros de dupla ressonância foram adquiridos em um instrumento Bruker Advance® operando a 700 MHz, em colaboração com o Dr. Daniel Cícero do Departamento de Ciências Químicas e Tecnologia da Universidade de Roma II.

Todos os espectros de RMN foram processados utilizando o programa NMRPipe (DELAGLIO *et al.*, 1995). Para a análise dos espectros foi utilizado o programa NMRView 5.0 (JOHNSON e BLAVINS, 1994).

#### 3.3.3.1. Preparo das amostras e otimização das condições de RMN

Para testar a influência da concentração protéica no sinal de RMN da proteína ClpS, foram adquiridos espectros 2D <sup>1</sup>H-<sup>15</sup>N-HSQC em diferentes concentrações da proteína. Inicialmente, partiu-se de uma amostra marcada com <sup>15</sup>N preparada a uma

concentração de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ . Esta amostra inicial foi diluída para 400  $\mu\text{M}$ , 300  $\mu\text{M}$ , e 100  $\mu\text{M}$ , utilizando o mesmo tampão e novos espectros foram adquiridos. Além disso foram testadas diferentes condições de tampão fosfato que variaram de 10 a 40 mM. Para testar a estabilidade da proteína, as amostras foram mantidas a 298 K por 1 semana quando a coleta de espectros foi repetida. Todos os espectros 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC foram adquiridos utilizando um equipamento Varian Inova® operando a 500 MHz e uma matriz de dados de 1024 x 256 pontos.

### 3.3.3.2. Cálculo do raio hidrodinâmico: experimento de difusão (DOSY)

Para a obtenção do coeficiente de difusão, que permite calcular o raio hidrodinâmico de proteínas, foi utilizada a técnica de DOSY (“Diffusion Ordered Spectroscopy”) (SOUZA e LAVERDE JR., 2002). Foi utilizada uma amostra de 500  $\mu\text{M}$  de proteína ClpS em uma solução 100%  $\text{D}_2\text{O}$ , 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 50 mM  $\text{NaCl}$ , 0,5%  $\text{NaN}_3$  e 0,1% dioxano e os espectros foram adquiridos utilizando um equipamento Varian Inova® operando a 500 MHz.

Neste experimento, a diminuição do sinal em função do aumento do gradiente de campo está relacionada à difusão da molécula no meio. O coeficiente de difusão pode ser definido pela equação de Stokes-Einstein:

$$D = \frac{kT}{6\pi R\eta} \quad [\text{Equação 3}]$$

onde D é o coeficiente de difusão, T é a temperatura, k é a constante de Boltzman, R é o raio hidrodinâmico e  $\eta$  é a viscosidade da solução (SOUZA e LAVERDE JR., 2002).

Os valores de difusão são calculados pelo ajuste de equações gaussianas para cada sinal de ressonância da molécula. É possível, desta forma, identificar no espectro bidimensional DOSY os picos referentes à água residual (~ 4,8 ppm) e ao dioxano (3,75 ppm), por exemplo. O dioxano atua como um padrão interno da solução. Uma vez que o raio hidrodinâmico do dioxano é conhecido, os coeficientes de difusão da proteína em estudo e do dioxano na mesma solução, podem ser usados para calcular o raio hidrodinâmico da proteína utilizando-se uma regra de três simples (CHEN *et al.*, 1995):

$$R_H^{\text{proteína}} = (D_{\text{referência}} / D_{\text{proteína}}) \times R_H^{\text{referência}} \quad [\text{Equação 4}]$$

onde  $D$  é o coeficiente de difusão e  $R_H$  é o raio hidrodinâmico.

Para o cálculo do coeficiente de difusão da proteína ( $D_{\text{proteína}}$ ) empregou-se a média dos valores de coeficiente de difusão obtidos para os prótons dos resíduos aromáticos (~ 6,0 a 7,0 ppm) e para os prótons dos grupos metila (~1,0 ppm) da proteína. O raio hidrodinâmico obtido experimentalmente foi então comparado com o valor teórico calculado usando a seguinte equação (WILKINS *et al.*, 1999):

$$R_H = (4,75 \pm 1,11) N^{0,29 \pm 0,02} \quad [\text{Equação 5}]$$

onde  $N$  é o número de amino ácidos da proteína (i.e.  $N = 106$ ).

### 3.3.3.3. Experimentos tridimensionais de tripla ressonância

Visando o assinalamento seqüencial dos núcleos da cadeia principal da proteína ClpS, os seguintes experimentos de 3D de tripla ressonância foram adquiridos: HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH e HNCO. Os experimentos

foram adquiridos em um equipamento Varian Inova® operando a 600 MHz utilizando uma amostra de proteína duplamente marcada com  $^{15}\text{N}$  e  $^{13}\text{C}$  preparada para uma concentração de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ .

**Tabela 11.** Número de pontos da matriz de dados dos experimentos 3D de tripla ressonância.

Experimento	Matriz de Dados (No. Pontos)		
	HN	N	C
HNCA	1024	128	64
HN(CO)CA	1024	128	64
HNCACB	1024	128	64
CBCA(CO)NH	1024	128	128
HNCO	1024	128	64

O número de pontos da matriz de dados de cada experimento pode ser visto na Tabela 11. Periodicamente, nos intervalos entre cada espectro 3D, foram adquiridos espectros 2D  $^{15}\text{N}$ -HSQC, para o controle da estabilidade da amostra.

#### 3.3.3.4. Experimentos de dupla ressonância

Visando o assinalamento dos núcleos das cadeias laterais da proteína ClpS, os seguintes experimentos de dupla ressonância foram adquiridos: 3D HCCH-COSY, 3D HCCH-TOCSY, 3D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC-TOCSY, 3D  $^1\text{H}$ - $^{13}\text{C}$  NOESY-HSQC e 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC. Estes experimentos foram adquiridos usando um equipamento Bruker Advance® operando a 700 MHz, em colaboração com o Dr. Daniel Cícero do Departamento de Ciências Químicas e Tecnologia da Universidade de Roma II.

Um espectro 3D HNHA foi adquirido com o intuito de calcular a constante de acoplamento  $^3J_{\text{HNH}\alpha}$ . Para a análise dos sinais de NOEs da proteína ClpS foi

adquirido o espectro 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY. Estes experimentos foram adquiridos usando um equipamento Varian Inova® operando a 600 MHz.

Todos os experimentos de dupla ressonância foram adquiridos utilizando uma amostra de proteína duplamente marcada com  $^{15}\text{N}$  e  $^{13}\text{C}$  preparada para uma concentração de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ . O número de pontos da matriz de dados de cada experimento pode ser visto na Tabela 12.

**Tabela 12.** Número de pontos da matriz de dados dos experimentos 3D de dupla ressonância.

Experimento	Matriz de Dados (No. pontos)		
	HN/HC	N/C	C/H
$^1\text{H}$ CCCH-COSY	512	128	128
CCCH-TOCSY	512	128	128
$^1\text{H}$ - $^{15}\text{N}$ HSQC-TOCSY	512	128	128
$^1\text{H}$ - $^{13}\text{C}$ NOESY-HSQC	512	128	128
$^{13}\text{C}$ -HSQC	-	512	2048
$^1\text{H}$ - $^{15}\text{N}$ HNHA	2048	96	64
$^1\text{H}$ - $^{15}\text{N}$ HSQC-NOESY	1024	128	128

Periodicamente, nos intervalos entre cada espectro 3D, foram adquiridos espectros 2D  $^{15}\text{N}$ -HSQC, para o controle da estabilidade da amostra.

### 3.3.3.5. Cálculo dos acoplamentos dipolares residuais $^1\text{H}$ - $^{15}\text{N}$

Os acoplamentos dipolares residuais  $^1\text{H}$ - $^{15}\text{N}$  ( $D_{\text{HN}}$ ) foram medidos usando uma amostra de ClpS marcada com  $^{15}\text{N}$  preparada para uma concentração de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ . O alinhamento residual da proteína foi induzido usando o fago filamentoso Pf1 (Asla Labs, HANSEN *et al.*, 1998) em uma concentração de 20 mg/mL. Espectros IPAP

(“in-phase/anti-phase”)  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (OTTIGER *et al.*, 1998) foram adquiridos na ausência e na presença de fago utilizando um equipamento Varian Inova® operando a 500 MHz. Os espectros foram adquiridos com 1024 e 256 pontos complexos nas dimensões de aquisição e indireta, respectivamente. De forma a obter suficiente resolução digital, os espectros foram processados estendendo-se as dimensões de aquisição e indireta até 2048 e 512 pontos respectivamente por “zero filling” seguido de multiplicação por uma função coseno. Os acoplamentos dipolares residuais foram calculados a partir da diferença dos acoplamentos escalares  $^1J_{\text{NH}}$  medidos na ausência e presença de fago.

### 3.3.3.6. Experimento de troca próton-deutério

Uma amostra de ClpS marcada com  $^{15}\text{N}$  e preparada para uma concentração final de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  de tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 0,025% de  $\text{NaN}_3$  foi liofilizada. Antes de iniciar o experimento, todos os parâmetros necessários para a coleta de dados foram estabelecidos no equipamento Varian Inova® operando a 500 MHz utilizando-se uma outra amostra de ClpS em solução marcada com  $^{15}\text{N}$  e preparada para uma concentração final de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  do mesmo tampão. A amostra liofilizada foi então ressuspensa em 600  $\mu\text{L}$  de 100%  $\text{D}_2\text{O}$ . Imediatamente, a amostra foi colocada no tubo de RMN, inserida no equipamento e, em seguida, iniciou-se a coleta de uma série de espectros  $^1\text{H}$ - $^{15}\text{N}$ -HSQC. Todo o processo desde a ressuspensão da amostra até o início da obtenção do primeiro espectro  $^1\text{H}$ - $^{15}\text{N}$ -HSQC demorou aproximadamente 20 minutos. Os espectros  $^1\text{H}$ - $^{15}\text{N}$ -HSQC foram adquiridos a 298 K utilizando uma matriz de dados de 2048 x 256 pontos complexos nas dimensões de aquisição e indireta, respectivamente. De forma a obter suficiente resolução digital, os espectros foram processados a estendendo a dimensão indireta

até 1024 pontos por “zero filling” e multiplicando-a por uma função seno deslocada em  $0.38 \pi$  e multiplicando a dimensão de aquisição por uma função gauseana.

### 3.3.3.7. Experimentos de relaxação

Os experimentos de relaxação foram realizados usando uma amostra de ClpS marcada com  $^{15}\text{N}$  preparada para uma concentração de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ .

As medidas de  $^{15}\text{N}$   $T_1$ ,  $^{15}\text{N}$   $T_2$  e  $^1\text{H}$ - $^{15}\text{N}$  NOE foram realizadas a 303 K em um equipamento Varian Inova® operando a 500 MHz usando as seqüências de pulso descritas por KAY e colaboradores (1989) e BARBATO e colaboradores (1992). As medidas de  $T_1$  e  $T_2$  foram realizadas com tempos de espera de evolução de 1, 21, 41, 61, 81, 101, 121, 141 e 161 ms para  $T_1$  e 1, 5, 9, 13, 17, 21 25 e 29 ms para  $T_2$ . Para medir  $^1\text{H}$ - $^{15}\text{N}$  NOE (NOE heteronuclear) foram adquiridos espectros com e sem saturação de  $^1\text{H}$ , empregando um tempo de espera entre “scans” de 5s nos dois experimentos. Todos os espectros foram adquiridos utilizando uma matriz de 1024 x 256 pontos complexos nas dimensões de aquisição e indireta, respectivamente. De forma a obter suficiente resolução digital, os espectros foram processados multiplicando a dimensão de aquisição por uma função gauseana. A dimensão indireta foi estendida por 64 pontos usando predição linear, seguido de “zero filling” até 1024 pontos e finalmente multiplicada por uma função seno deslocada em  $0.38 \pi$ . Os volumes dos picos foram calculados usando a rotina de integração Rate Analysis do programa NMRView 5.0 (JOHNSON e BLAVINS, 1994) e a velocidade do decaimento do volume dos picos foi ajustada a uma função exponencial simples. Os valores de  $^1\text{H}$ - $^{15}\text{N}$  NOE foram determinados a partir da razão dos volumes dos picos com e sem saturação de  $^1\text{H}$ .



### 3.3.3.8. Experimentos de interação da proteína ClpS de *Xac* com um o N-terminal da proteína ClpA de *Xac*

Os experimentos de interação da proteína ClpS de *Xac* com o N-terminal da proteína ClpA (resíduos 1-143) de *Xac* foram conduzidos analisando a perturbação dos deslocamentos químicos observados no espectro  $^1\text{H}$ - $^{15}\text{N}$  HSQC da proteína ClpS na presença do N-terminal da proteína ClpA.

A amostra de ClpS marcada com  $^{15}\text{N}$  foi preparada para uma concentração de 783  $\mu\text{M}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,95, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ . A amostra não-marcada do N-terminal de ClpA foi preparada para uma concentração de 137  $\mu\text{M}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,95, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ . Para analisar a interação entre as duas proteínas foram preparadas amostras contendo diferentes proporções de ClpS e ClpA conforme descrito na Tabela 13.

Foram adquiridos espectros de  $^1\text{H}$ - $^{15}\text{N}$  HSQC das várias amostras. Todos foram adquiridos utilizando um equipamento Varian Inova® operando a 500 MHz e uma matriz de dados de 512 x 256 pontos.

**Tabela 13.** Interação da proteína ClpS com o N-terminal da proteína ClpA.

Proporção ClpA:ClpS	Concentração de Clp A (N-terminal)	Concentração de ClpS
ClpS sozinha	-	100 $\mu\text{M}$
1:6	16.6 $\mu\text{M}$	100 $\mu\text{M}$
1:5	20 $\mu\text{M}$	100 $\mu\text{M}$
1:4	25 $\mu\text{M}$	100 $\mu\text{M}$
1:3	33.3 $\mu\text{M}$	100 $\mu\text{M}$
1:2	50 $\mu\text{M}$	100 $\mu\text{M}$
1:1	100 $\mu\text{M}$	100 $\mu\text{M}$

As variações dos deslocamentos químicos compostos do grupo amídico de cada resíduo,  $\Delta_{1H+15N}$ , foram calculados usando a Equação 6 descrita por JACOBS e colaboradores (2003).

$$\Delta_{1H+15N} = \sqrt{((\Delta_{1H})^2 + (0.17 \times \Delta_{15N})^2)} \quad \text{[Equação 6]}$$

### 3.3.3.9. Experimentos de interação da proteína ClpS de *Xac* com a proteína apo Sm14-M20(C62V)

Os experimentos de interação da proteína ClpS de *Xac* com a proteína apo Sm14-M20(C62V), um mutante da proteína Sm14 de *Schistosoma mansoni* (PERTINHEZ *et al.*, 2004), foram conduzidos analisando a perturbação dos deslocamentos químicos observados no espectro  $^1\text{H}$ - $^{15}\text{N}$  HSQC da proteína ClpS na presença da proteína apo Sm14-M20(C62V).

Foi preparada uma amostra contendo 100  $\mu\text{M}$  da proteína ClpS marcada com  $^{15}\text{N}$  e aproximadamente 100  $\mu\text{M}$  da proteína apo Sm14-M20(C62V) agregada em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,95, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ . Foram adquiridos espectros de  $^1\text{H}$ - $^{15}\text{N}$  HSQC das várias amostras. Todos foram adquiridos utilizando um equipamento Varian Inova® operando a 500 MHz e uma matriz de dados de 512 x 256 pontos. As variações dos deslocamentos químicos compostos ( $\Delta_{1H+15N}$ ) foram calculadas conforme descrito na secção 3.3.3.8.

### 3.3.4. Análise dos experimentos de RMN

#### 3.3.4.1. Cálculo do “Chemical Shift Index” (CSI)

O Índice de Deslocamento Químico ou “Chemical Shift Index” (CSI) foi calculado para a proteína ClpS usando o método descrito por WISHART e colaboradores (1992). O método calcula a diferença entre os valores de deslocamento químico obtidos experimentalmente para os núcleos da cadeia principal da proteína em estudo e os valores de deslocamento químico descritos na literatura para os mesmos amino ácidos quando estes se encontram em conformações randômicas.

Inicialmente, o emprego dos deslocamentos químicos como um parâmetro indicativo da estrutura secundária foi estabelecido utilizando-se apenas os valores de deslocamento químico dos núcleos  $^1\text{H}^\alpha$ . Posteriormente, este método foi extrapolado para todos os demais núcleos da cadeia principal ( $^{13}\text{CO}$ ,  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^1\text{HN}$  e  $^{15}\text{N}$ ).

#### 3.3.4.2. Predição dos ângulos $\Phi$ e $\Psi$ usando o programa TALOS

Os ângulos de torção  $\Phi$  e  $\Psi$  da proteína ClpS foram preditos usando o programa TALOS (CORNILESCU *et al.*, 1999) a partir dos valores de deslocamento químico obtidos experimentalmente para os núcleos da cadeia principal ( $^{13}\text{CO}$ ,  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^1\text{HN}$ ,  $^\alpha\text{H}$  e  $^{15}\text{N}$ ). Os deslocamentos químicos obtidos experimentalmente foram comparados com a base de dados do programa TALOS composta por 20 proteínas para as quais os valores de deslocamentos químicos bem como as estruturas cristalinas de alta resolução são conhecidos. Posteriormente, os valores dos ângulos

$\Phi$  e  $\Psi$  preditos para proteína ClpS de *Xac* foram também comparados com os valores de  $\Phi$  e  $\Psi$  da proteína ClpS de *E.coli*.

### 3.3.4.3. Análise dos acoplamentos dipolares residuais $^1\text{H}$ - $^{15}\text{N}$ usando o programa PALES

Os acoplamentos dipolares residuais calculados experimentalmente foram analisados usando o modulo BestFit do programa PALES (ZWECKSTETTER e BAX, 2000). Inicialmente, um total de 54 acoplamentos dipolares residuais obtidos experimentalmente para a proteína ClpS de *Xac* foram analisados. Posteriormente, apenas os 44 acoplamentos dipolares residuais que se encontram em regiões de estrutura secundária da proteína foram analisados.

### 3.3.4.4. Cálculo dos fatores de proteção de troca dos prótons amídicos

Os volumes dos picos dos espectros  $^1\text{H}$ - $^{15}\text{N}$ -HSQC foram calculados usando a rotina de integração “Rate Analysis” do programa NMRView 5.0 (JOHNSON e BLAVINS, 1994). As constantes de velocidade de troca experimentais ( $k_{\text{obs}}$ ) foram calculadas ajustando-se a curva do decaimento do volume dos picos em função do tempo a uma função exponencial simples usando três parâmetros (intensidade no tempo zero, velocidade de decaimento e um fator de deslocamento) pelo algoritmo Levenberg-Marquardt de ajuste não-linear de mínimos quadrados implementado no programa GNUPLOT 3.7 (Fonte: <http://www.gnuplot.info>, 2005).

As constantes de velocidade de troca intrínsecas ( $k_{\text{int}}$ ) foram preditas usando o programa SPHERES (Fonte: <http://www.fccc.edu/research/labs/roder/sphere>, 2005).

O fator de proteção para cada próton amídico foi calculado usando a Equação 7:

$$P = (k_{\text{int}}/k_{\text{obs}}) \quad \text{[Equação 7]}$$

conforme descrito por BAI e colaboradores (1993).

### **3.3.5. Modelagem molecular**

#### **3.3.5.1. Recursos computacionais**

As etapas de construção por homologia, refinamento e avaliação do modelo foram realizadas em um multi-processador Compaq AlphaServer ES40. Os procedimentos de visualização e manipulação das estruturas foram realizados através da interface gráfica gerada pelo programa InsightII (Accelrys Inc.), instalado em uma estação gráfica Octane2 (Silicon Graphics Inc.). As imagens finais foram obtidas pelo programa WebLab Viewer 3.7 (MOLECULAR SIMULATIONS, 2000), rodando em ambiente Windows. Todas as etapas de modelagem molecular foram realizadas em colaboração com o Dr. Sérgio Oyama Jr. do Laboratório Nacional de Luz Síncrotron.

#### **3.3.5.2. Predição de estrutura secundária**

A estrutura secundária da proteína CLpS de Xac foi predita usando o programa PSIPRED (JONES, 1999) que utiliza um algoritmo baseado em técnicas de redes neurais e alimentado por bancos de dados contendo informações estruturais e de seqüências.

### 3.3.5.3. Busca de estruturas de ClpS de *E. coli* no PDB e alinhamento de seqüências

O programa BLASTp (ALTSCHUL *et al.*, 1990) foi utilizado no banco de dados PDB de modo a alinhar a seqüência primária da proteína ClpS de *Xac* com a seqüência da proteína ClpS de *E. coli* e de forma a selecionar a estrutura de ClpS de *E.coli* com melhor resolução depositada no PDB. O alinhamento preliminar obtido através deste programa foi posteriormente aprimorado de forma manual, levando-se em consideração as informações obtidas através da predição de estrutura secundária.

### 3.3.5.4. Construção do modelo por homologia

Um modelo tridimensional da proteína ClpS de *Xac* foi construído fazendo uso do programa MODELLER (SALI e BLUNDELL, 1993) e da estrutura de ClpS de *E.coli* depositada no PDB com o código 1MBX (GUO *et al.*, 2002a) como estrutura-molde. Baseando-se em uma lista de restrições obtidas a partir do alinhamento das duas seqüências, o programa realiza várias etapas de dinâmica molecular, baseada na técnica de "simulated annealing", de modo a gerar possíveis modelos que satisfaçam os requisitos estruturais impostos pelo alinhamento e pelas coordenadas da estrutura-molde. Os modelos obtidos são posteriormente refinados através de vários passos de minimização de energia.

### 3.3.5.5. Avaliação do modelo

As estruturas finais obtidas após a etapa de minimização de energia foram posteriormente avaliadas quanto à sua qualidade estereoquímica através do programa PROCHECK (LASKOWSKI *et al.*, 1993). O RMSD entre a estrutura modelo de ClpS de *Xac* de menor energia e a estrutura de ClpS de *E.coli* (PDB 1MBX) foi calculado usando o programa PROFIT (MARTIN, 1996).

### **3.4. RESULTADOS**

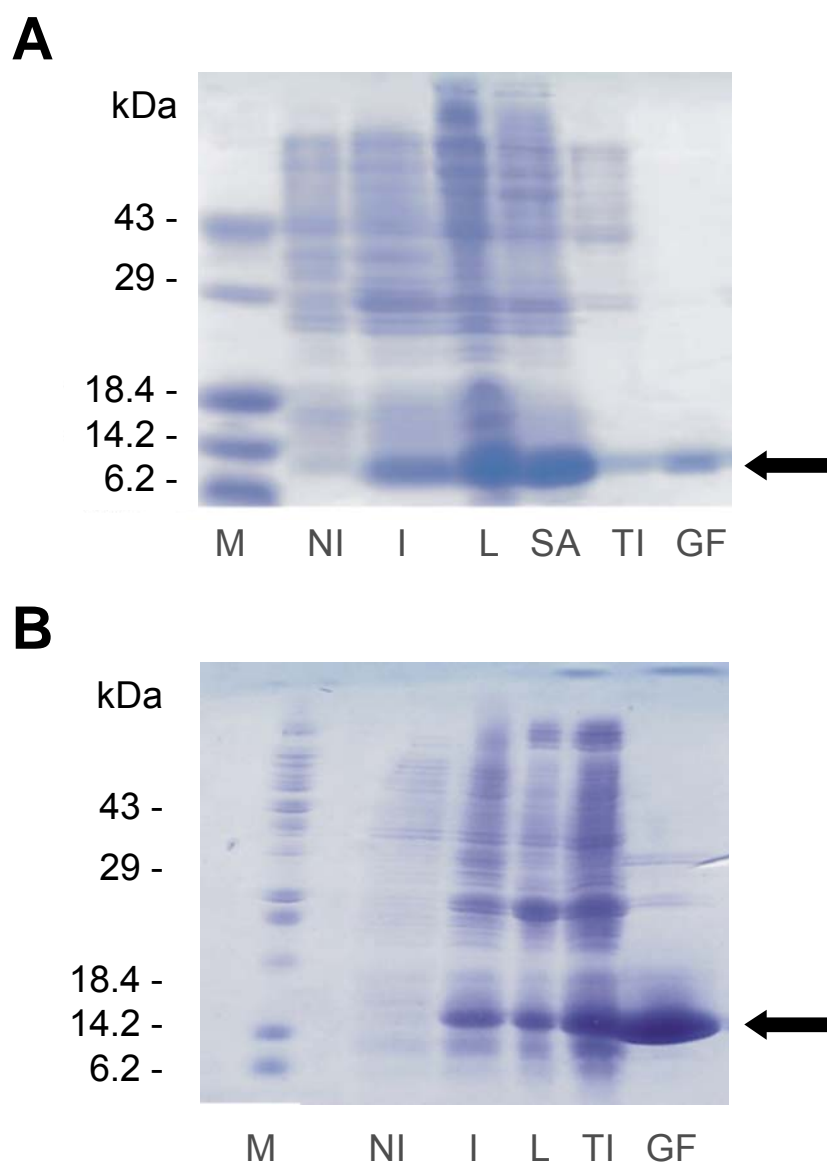
Considerando-se a complexidade e a variedade de resultados descritos neste capítulo, optou-se por discutir a relevância biológica dos resultados obtidos juntamente com os mesmos. Na seção intitulada DISCUSSÃO, será apresentada uma discussão detalhada dos resultados obtidos face à literatura atual.

#### **3.4.1. Expressão e purificação da proteína em estudo**

A proteína ClpS (XAC2000) e o N-terminal da proteína ClpA de *Xac* foram clonados, expressos e purificados conforme descrito na seção 3.3.1.

A Figura 25 A mostra as frações puras obtidas após a purificação da proteína ClpS. A proteína foi liofilizada e, após a sua ressuspensão em tampões de interesse, continuou apresentando boa solubilidade e estabilidade, permitindo prosseguir com os experimentos de estudos estruturais descritos a seguir. A Figura 25 B mostra as frações puras obtidas após a purificação do N-terminal da proteína ClpA. Esta proteína, apresentou-se agregada em concentrações superiores a 100  $\mu$ M. Além disso, o grau de pureza desta proteína é menor que o grau de pureza apresentado pela proteína ClpS. Uma vez que esta proteína seria usada apenas para estudos preliminares de interação, e considerando que estudos de interação normalmente não requerem um grau de pureza tão elevado quanto os estudos estruturais por RMN, optou-se por não otimizar o protocolo de purificação desta proteína.





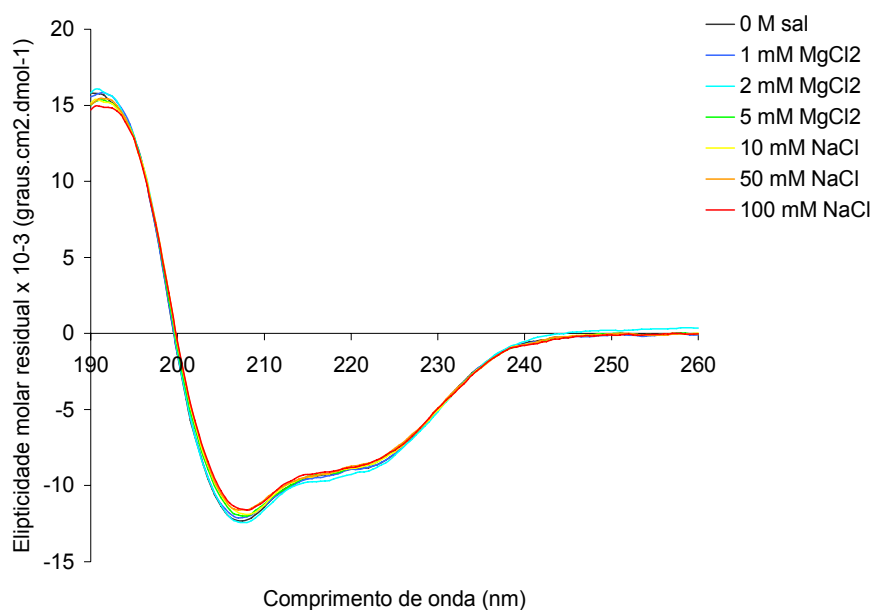
**Figura 25.** Expressão e purificação da proteína ClpS (A) e doN-terminal da proteína ClpA (B).

Gel de SDS-PAGE mostrando as várias etapas envolvidas. M= marcador de peso molecular; NI = amostra não induzida; I = amostra induzida; L = amostra lisada por French Press<sup>®</sup>; SA = amostra precipitada por sulfato de amônio, TI = amostra após a cromatografia de troca iônica; GF = amostra após a cromatografia de gel filtração.

### 3.4.2. Experimentos de Dicroísmo Circular (CD)

Visando uma caracterização estrutural preliminar, a proteína ClpS foi analisada por CD mediante várias condições de força iônica, pH e temperatura. A Figura 26 mostra espectros de CD da proteína ClpS obtidos durante a variação da força iônica. O perfil no UV distante da proteína ClpS apresenta duas bandas negativas, uma em ~ 222 nm e outra em ~208 nm, características de uma proteína com perfil principalmente helicoidal. Pode-se verificar que a variação da força iônica não teve efeito significativo nos espectros obtidos.

A Figura 27 mostra espectros de CD da proteína ClpS obtidos durante a variação do pH entre 3,0 e 9,0. Neste caso, percebe-se a ocorrência de uma desestruturação à medida em que se acidifica o meio, que é evidenciada pela diminuição e alteração de perfil das bandas positiva e negativas presentes no espectro da proteína em sua conformação inicial.



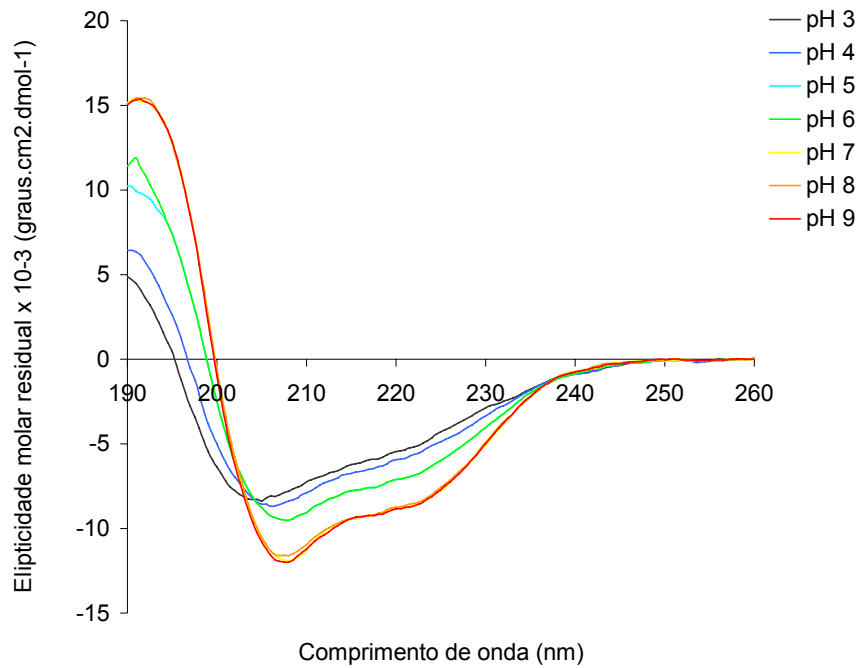
**Figura 26.** Espectros de CD da proteína ClpS obtidos em diferentes concentrações de sal. Condições da amostra: 10  $\mu$ M de proteína em 10 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8.

A ausência de um perfil típico de completo desenovelamento (“random coil”) a pH 3, indica que ClpS apresenta alguma resistência à desnaturação ácida. Em pH alto (pH < 9) não ocorre alteração conformacional de ClpS e os valores de pH nos quais a proteína se apresenta mais estruturada se encontram entre pH 7 e pH 9.

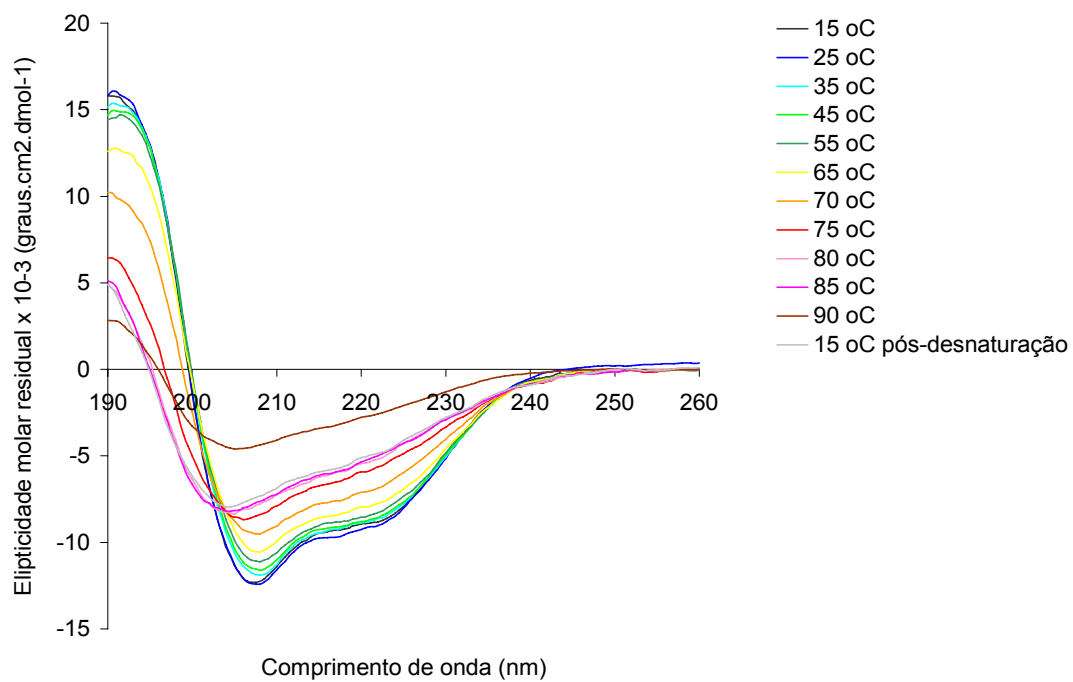
Na Figura 28 são mostrados os espectros da proteína ClpS a diferentes temperaturas, assim como seu perfil a 15 °C após a desnaturação térmica a 90 °C. Percebe-se que a proteína começa a mudar ligeiramente a sua conformação por volta de 70 °C, tornando-se apenas parcialmente desenovelada a 93 °C. Isto significa que a proteína ClpS é altamente termoestável tendo uma temperatura de desenovelamento ( $T_m$ ) bastante elevada. É também importante notar que o perfil a 15 °C após a desnaturação é diferente do obtido no início do experimento.

Neste sentido, podem ser levantadas algumas hipóteses: a) a proteína não consegue retornar à sua conformação original; b) as moléculas de proteína agregam-se após a desnaturação térmica e o sinal observado é menos intenso devido à presença de um número menor de moléculas enoveladas; c) um ou mais resíduos de prolina apresentam isomerização lenta, constituído assim um passo limitante do re-enovelamento resultando numa fração de proteínas na amostra que não conseguiria retornar ao estado nativo no período de tempo do experimento.

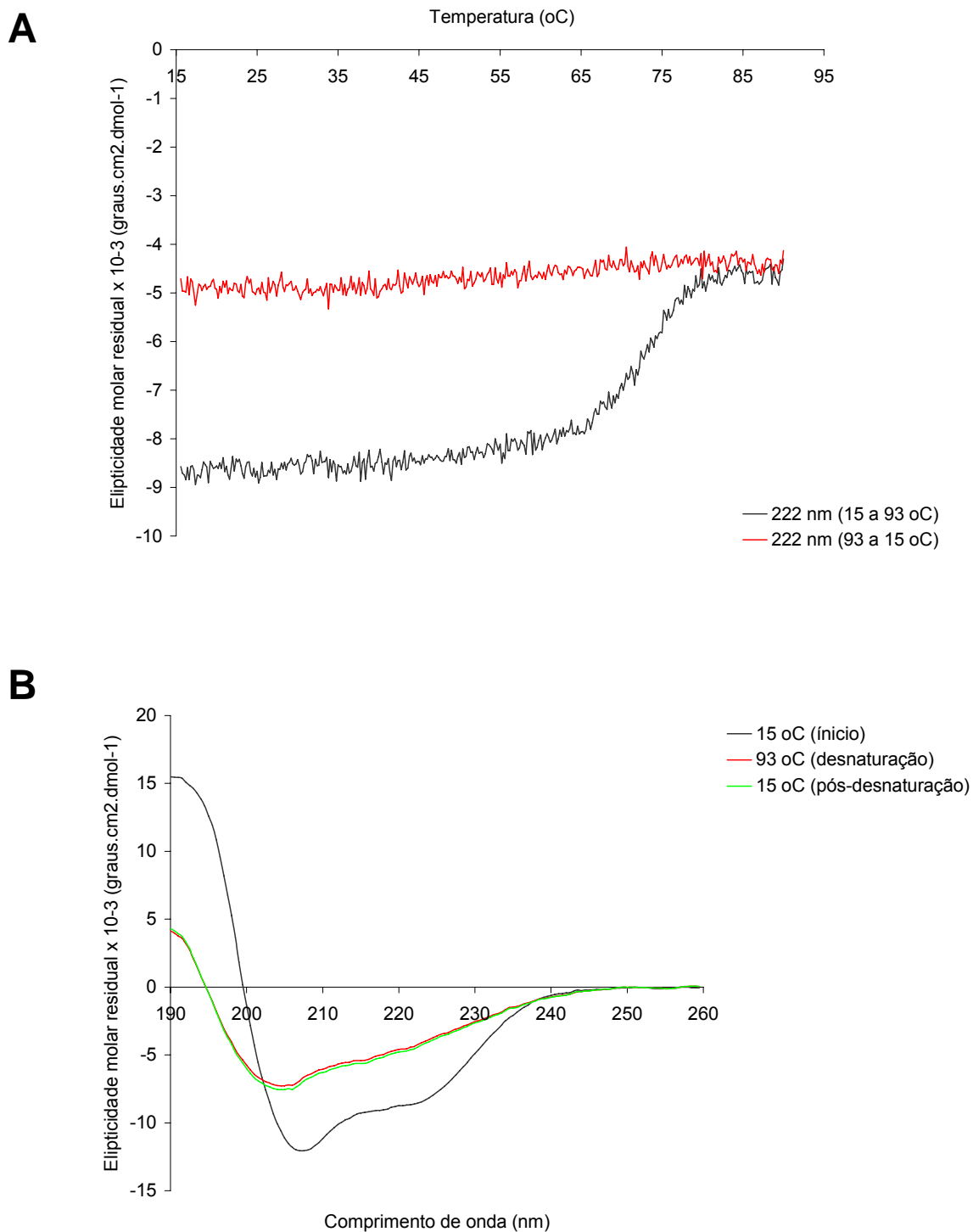
Na Figura 29, são mostrados os resultados do experimento de desnaturação térmica a 222 nm variando a temperatura em 1 °C/min (de 15 °C a 93 °C e vice-versa). A razão pela qual foi escolhido o comprimento de onda de 222 nm, está relacionada com o fato desta região do UV distante estar diretamente relacionada com a presença de estruturas protéicas helicoidais conforme descrito acima. Deste modo, podemos acompanhar a variação do perfil helicoidal da proteína em função da temperatura.



**Figura 27.** Espectros de CD da proteína ClpS mediante variação de pH entre 3,0 e 9,0. Condições da amostra: 20  $\mu$ M de proteína em 10 mM  $\text{NaH}_2\text{PO}_4$ ).



**Figura 28.** Espectros de CD de ClpS obtidos a diferentes temperaturas. Espectros de variação de 15 a 90 °C, e espectro a 15 °C após desnaturação térmica. Condições da amostra: 10  $\mu$ M de proteína em 10 mM  $\text{NaH}_2\text{PO}_4$ , pH 6,8.



**Figura 29.** Estudos de desnaturação térmica da proteína ClpS.

Condições da amostra: 10  $\mu$ M de proteína em 10 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 6,8.

A. Variação da elipticidade molar a um comprimento de onda fixo de 222 nm em função da temperatura. Desnaturação de 15 a 93 °C (linha preta) e retorno de 93 para 15 °C (linha vermelha).

B. Espectros de CD controle de 195 a 260 nm pré e pós desnaturação: pré-desnaturação a 15 °C (linha preta), pós-desnaturação a 93 °C (linha vermelha) e a 15 °C (linha verde).

A Figura 29 B mostra o espectro inteiro da proteína, de 195 a 260 nm, antes da desnaturação (15 °C), pós-desnaturação (93 °C) e no retorno pós-desnaturação para 15 °C, para controle dos valores obtidos nos espectros da Figura 29 A. Pela Figura 29 A, pode-se concluir que os valores de elipticidade molar obtidos durante a desnaturação até 93 °C são diferentes daqueles obtidos durante o retorno para 15 °C após a desnaturação, indicando que esta foi irreversível, o que leva às hipóteses já mencionadas sobre o comportamento da proteína ClpS sob estresse térmico. Os ensaios de CD descritos acima foram fundamentais para aferir rapidamente as condições (pH, temperatura, sal) em que a proteína ClpS se encontra estruturada e estável e portanto passível de ser testada por RMN.

### **3.4.3. Estudos estruturais da proteína ClpS de *Xac* por RMN**

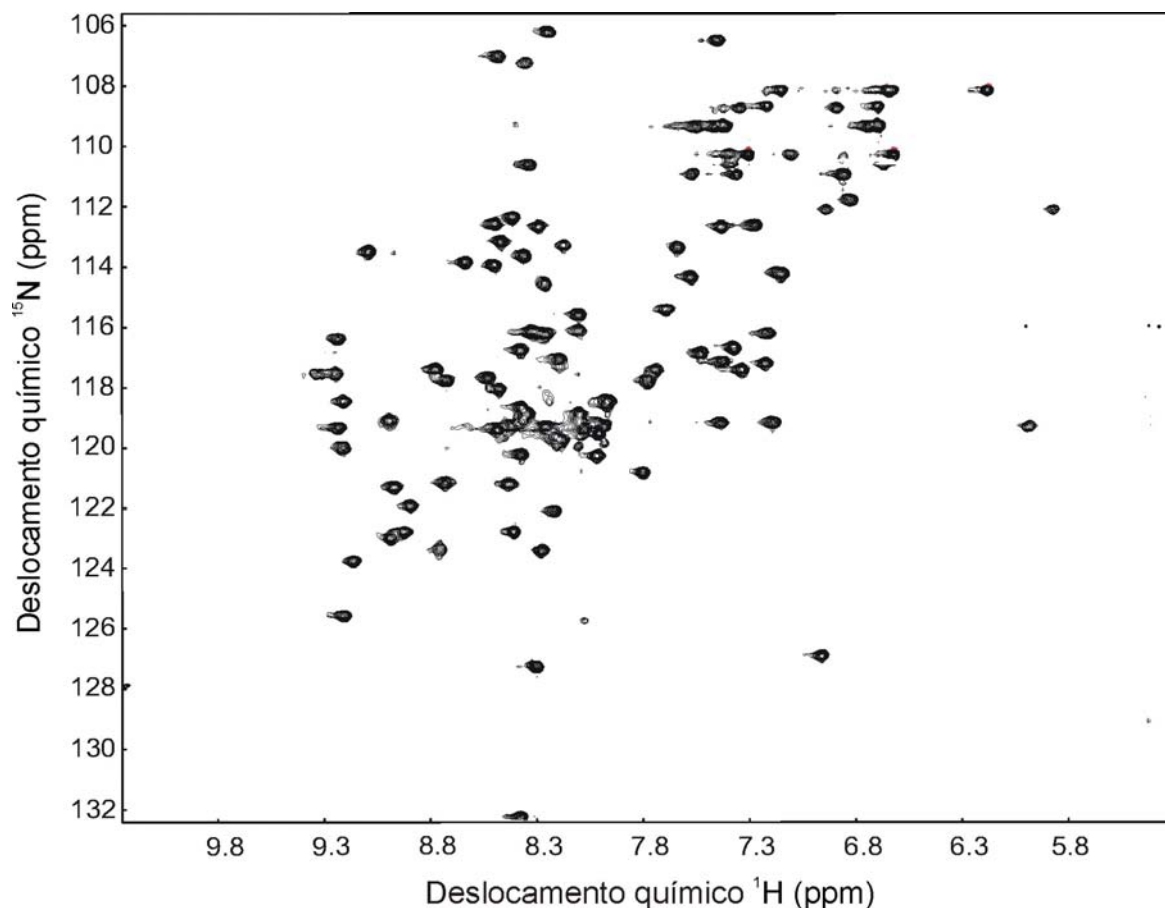
#### **3.4.3.1. Otimização das condições de RMN**

Antes de se iniciarem estudos estruturais de uma proteína por RMN é necessário determinar quais as melhores condições para a coleta dos vários experimentos. Conforme descrito anteriormente na secção 2.1.3, o espectro 2D <sup>15</sup>N-HSQC é muitas vezes denominado de “impressão digital” de uma proteína uma vez que nele podem ser observados um pico para cada próton amídico da cadeia principal de uma dada proteína e o padrão apresentado pelos picos (i.e. dispersão do deslocamento químico ( $\delta$ ), largura da linha, número e intensidade dos picos observados) revela informações únicas sobre a solubilidade, o grau de enovelamento e a conformação da proteína em estudo. Assim, ao adquirir vários espectros 2D <sup>1</sup>H-<sup>15</sup>N-HSQC de uma proteína pura em diferentes condições de concentração, pH, sal e temperatura pode-se determinar de forma simples e rápida quais as condições ideais para prosseguir com os demais estudos por RMN.

Desta forma, foram adquiridos espectros 2D  $^{15}\text{N}$ -HSQC em diferentes condições. Inicialmente, partiu-se de uma amostra marcada com  $^{15}\text{N}$  preparada a uma concentração de 500  $\mu\text{M}$  em 600  $\mu\text{L}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ . Esta amostra inicial foi diluída para 400  $\mu\text{M}$ , 300  $\mu\text{M}$ , e 100  $\mu\text{M}$ , utilizando o mesmo tampão e novos espectros foram adquiridos. A diluição da proteína provocou uma diminuição no sinal de RMN, mas não provocou qualquer alteração do posicionamento ou resolução dos picos indicando que não há alteração estrutural com a diluição e que a concentração de 500  $\mu\text{M}$  de ClpS é uma boa condição para a aquisição de dados experimentais por RMN. Além disso, foram também testadas diferentes concentrações de fosfato: 0, 10, 20 e 40 mM. Todas as concentrações apresentaram resultados semelhantes. Como se havia demonstrado anteriormente por CD que a proteína ClpS é altamente termoestável e bem estruturada em valores de pH próximos de 7,0, optou-se por adquirir os espectros a 303K e pH 6,8.

Assim, os vários espectros de RMN da proteína ClpS utilizados em estudos por RMN desta proteína foram adquiridos nas seguintes condições: 500  $\mu\text{M}$  de proteína ClpS em 600  $\mu\text{L}$  em tampão 40 mM  $\text{NaH}_2\text{PO}_4$ , pH 6,8, 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ .

Como pode ser visto na Figura 30, o espectro 2D  $^{15}\text{N}$ -HSQC adquirido nestas condições apresenta uma dispersão grande do deslocamento químico, linhas estreitas e picos intensos, indicativo de uma proteína estável e estruturada e com pouca variabilidade conformacional. Em especial, uma análise detalhada do espectro revelou que o número de picos observado é igual ao esperado a partir da seqüência de amino ácidos (94 picos de prótons amídicos da cadeia principal e 26 picos de prótons amídicos das cadeias laterais dos resíduos Asn e Gln).



**Figura 30.** Espectro 2D  $^{15}\text{N}$ -HSQC da proteína ClpS.

Amostra marcada com  $^{15}\text{N}$ , 500  $\mu\text{M}$  de proteína em tampão 40 mM  $\text{NaH}_2\text{PO}_4$  pH 6,8; 0,025% de  $\text{NaN}_3$  e 5% de  $\text{D}_2\text{O}$ .

### 3.4.3.2. Cálculo do raio hidrodinâmico

Conforme descrito na secção 3.3.3.2, o raio hidrodinâmico da proteína ClpS ( $R_H^{\text{proteína}}$ ) foi obtido utilizando a técnica de DOSY (“Diffusion Ordered Spectroscopy”) (SOUZA e LAVERDE JR, 2002).

Considerando que a proteína ClpS contém 106 amino ácidos e utilizando a equação 5 descrita anteriormente, o valor teórico (considerando-se uma



configuração esférica) do raio hidrodinâmico da proteína ClpS no estado monomérico é de 18,37 Å. Já no caso de um dímero de ClpS o valor teórico do raio hidrodinâmico da proteína é de 22,46 Å. O valor do raio hidrodinâmico da proteína ClpS obtido experimentalmente usando o método de DOSY foi de 18,87 Å. Uma vez que o método de DOSY assume a forma como sendo uma esfera e que valor do raio hidrodinâmico de ClpS obtido experimentalmente é muito semelhante ao valor teórico esperado para uma proteína monomérica, os resultados obtidos sugerem assim, que nas condições testadas, a proteína ClpS é monomérica e apresenta uma conformação globular e enovelada.

### **3.4.3.3. Assinalamento da proteína ClpS**

Uma vez determinadas as condições ideais para a aquisição de espectros de RMN, foi possível dar início aos estudos estruturais por RMN da proteína ClpS. Conforme descrito na secção 3.1.2.1.1, a primeira etapa de qualquer estudo estrutural de uma dada proteína por RMN envolve o assinalamento seqüencial de todos os amino ácidos presentes na molécula. Isto é, cada ressonância obtida nos espectros de RMN deve ser associada a um núcleo de um amino ácido específico da proteína.

Conforme visto anteriormente, a estratégia específica de assinalamento utilizada depende, entre outros fatores, do tamanho da proteína de interesse, da possibilidade de marcação isotópica da mesma e do equipamento de RMN disponível. A proteína ClpS de apresenta 106 resíduos de amino ácidos em sua seqüência primária o que já representa um elevado número núcleos a serem assinalados. Assim, de forma a minimizar possíveis ambigüidades, adotamos uma estratégia baseada em experimentos 3D de dupla e de tripla ressonância para o assinalamento da cadeia principal e das cadeias laterais da proteína ClpS.

### 3.4.3.3.1. Assinalamento da cadeia principal

Os núcleos da cadeia principal da proteína ClpS foram assinalados através da coleta e análise dos experimentos 3D de tripla ressonância HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH e HNCO. Os espectros foram adquiridos de acordo com os métodos descritos em 3.3.3.3 usando uma amostra da proteína ClpS duplamente marcada com  $^{13}\text{C}$  e  $^{15}\text{N}$ .

Conforme descrito anteriormente, os experimentos 3D de tripla ressonância usam acoplamentos escalares de uma ou duas ligações para correlacionar as frequências de ressonância de até três núcleos diferentes. A magnetização inicia-se no núcleo do hidrogênio amídico e é transferida sucessivamente através das ligações covalentes para outros núcleos fornecendo informações sobre os núcleos de hidrogênio, nitrogênio ( $^{15}\text{N}$ ) e carbono ( $^{13}\text{C}$ ) da cadeia principal. A Tabela 14 resume as correlações observadas para assinalar as frequências de ressonância dos núcleos da cadeia principal da proteína ClpS.

**Tabela 14.** Resumo das correlações observadas nos experimentos 3D de tripla ressonância utilizados para assinalar os núcleos da cadeia principal da proteína ClpS.  $i$  indica a posição de um resíduo de aminoácido ao longo da cadeia polipeptídica (adaptado de CLORE e GRONENBORN, 1998).

Experimento	Correlação observada
HNCA	$^{13}\text{C}^{\alpha}(i)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$ $^{13}\text{C}^{\alpha}(i-1)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$
HN(CO)CA	$^{13}\text{C}^{\alpha}(i-1)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$
HNCACB	$^{13}\text{C}^{\alpha}(i)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$ $^{13}\text{C}^{\beta}(i)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$
CBCA(CO)NH	$^{13}\text{C}^{\alpha}(i-1)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$ $^{13}\text{C}^{\beta}(i-1)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$
HNCO	$^{13}\text{CO}(i-1)-^{15}\text{N}(i)-^1\text{H}^{\text{N}}(i)$

Inicialmente foi analisado o espectro HNCA que fornece correlações entre as ressonâncias dos núcleos de  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  e o núcleo  $^{13}\text{C}^{\alpha}$  de um mesmo resíduo de amino ácido ( $i$ ) e com o núcleo de  $^{13}\text{C}^{\alpha}$  do resíduo anterior ( $i-1$ ). Os picos referentes às correlações intra-residuais no espectro são fortes, enquanto que aqueles referentes às correlações seqüenciais, são geralmente mais fracos, fornecendo assim um caminho para o assinalamento. Como exemplo deste procedimento, a Figura 31 mostra fatias selecionadas do espectro de HNCA de ClpS indicando o caminho percorrido no assinalamento seqüencial das ressonâncias dos resíduos 72 a 76.

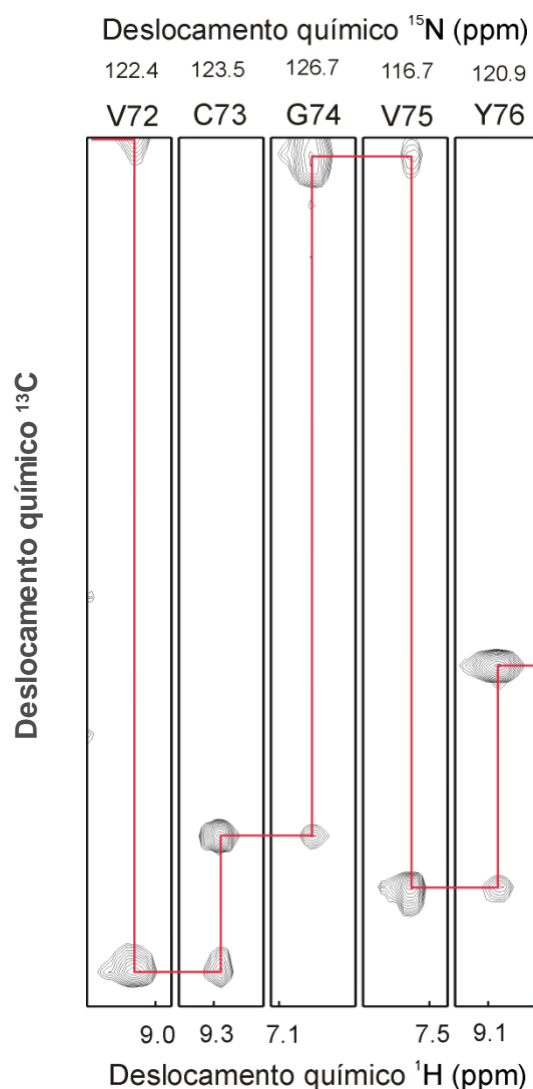
Para eliminar potenciais ambigüidades, este espectro foi analisado paralelamente com o espectro HN(CO)CA, que mostra apenas as correlações seqüenciais i.e., correlaciona as ressonâncias dos núcleos de  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  de um mesmo resíduo de amino ácido ( $i$ ) com o núcleo de  $^{13}\text{C}^{\alpha}$  do resíduo anterior ( $i-1$ ). Embora os espectros de HNCA e HN(CO)CA tenham possibilitado o assinalamento de grande parte dos núcleos de  $^{13}\text{C}^{\alpha}$  da cadeia principal da proteína ClpS, algumas ambigüidades resultantes da sobreposição de alguns picos permaneceram. Estas ambigüidades, foram resolvidas através da análise combinada dos espectros HNCACB e CBCA(CO)NH. O espectro HNCACB correlaciona as ressonâncias dos núcleos de  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  de um mesmo resíduo de amino ácido ( $i$ ) com as ressonâncias dos núcleos de  $^{13}\text{C}^{\alpha}$  e  $^{13}\text{C}^{\beta}$  do mesmo resíduo ( $i$ ) e do resíduo anterior ( $i-1$ ). Já o espectro CBCA(CO)NH correlaciona apenas as ressonâncias dos núcleos de  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  de um mesmo resíduo de amino ácido ( $i$ ) com o núcleos de  $^{13}\text{C}^{\alpha}$  e  $^{13}\text{C}^{\beta}$  do resíduo de amino ácido anterior ( $i-1$ ). Isto é, o espectro HNCACB fornece correlações intra-residuais entre todos os núcleos  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}^{\alpha}$  e  $^{13}\text{C}^{\beta}$  de um mesmo resíduo e correlações inter-residuais entre os mesmos núcleos, enquanto o

espectro CBCA(CO)NH nos fornece apenas correlações inter-residuais e seqüenciais entre os vários núcleos.

A Figura 32, mostra fatias selecionadas do espectro de HNCACB intercaladas com fatias selecionadas do espectro CACB(CO)NH, ilustrando dois caminhos para o assinalamento seqüencial dos resíduos 72 a 76, um baseado nos deslocamentos químicos de  $^{13}\text{C}^\alpha$  e outro, nos deslocamentos químicos  $^{13}\text{C}^\beta$ . Uma análise combinada destes espectros, associada aos resultados obtidos anteriormente através da análise dos espectros HNCA e HN(CO)CA, possibilitou a obtenção dos valores de deslocamentos químicos ( $\delta$ ) para cada um dos núcleos  $^1\text{H}^N$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}^\alpha$  e  $^{13}\text{C}^\beta$  da cadeia principal de cada amino ácido e da correlação destes com os núcleos da cadeia principal de amino ácidos vizinhos.

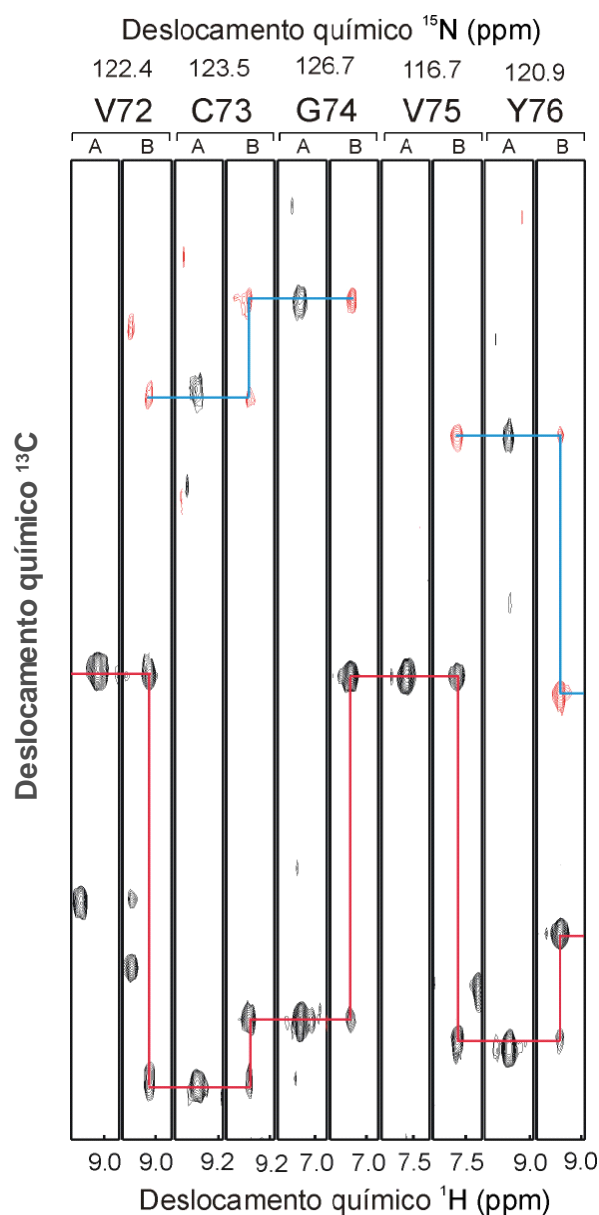
Paralelamente foi também analisado o espectro HNCO que correlaciona as ressonâncias dos núcleos de  $^1\text{H}^N$ ,  $^{15}\text{N}$  de um mesmo resíduo de amino ácido ( $i$ ) com a ressonância do núcleo  $^{13}\text{CO}$  do resíduo de amino ácido anterior ( $i-1$ ), fornecendo assim valores de deslocamento químico para cada um dos núcleos de  $^{13}\text{CO}$  da cadeia principal da proteína.

O processo de atribuição de cada pico do espectro 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC a um resíduo de amino ácido específico da proteína ClpS foi realizado selecionando-se inicialmente resíduos na seqüência primaria da proteína com deslocamento químicos ( $\delta$ ) característicos. Por exemplo, como pode ser visto na Tabela 8 (seção 3.1.2.1), apenas as glicinas possuem um valor médio de deslocamento químico para o núcleo  $^{13}\text{C}^\alpha$  em torno de 45 ppm e não apresentam valor de deslocamento químico para o núcleo  $^{13}\text{C}^\beta$  uma vez que este núcleo inexistente neste amino ácido.



**Figura 31.** Fatias selecionadas do espectro de HNCA da proteína ClpS.

A linha vermelha indica o caminho percorrido para o assinalamento seqüencial das ressonâncias dos núcleos de  $^{13}\text{C}^\alpha$  dos resíduos 72 a 76, baseado nas correlações existentes entre um amino ácido e o seu vizinho anterior. De um modo geral picos fortes correspondem às correlações entre o grupo amídico e o núcleo  $\text{C}^\alpha$  do mesmo resíduo (intra-residual) enquanto picos fracos referem-se às correlações seqüenciais entre o grupo amídico de um resíduo e o núcleo  $\text{C}^\alpha$  do resíduo anterior.



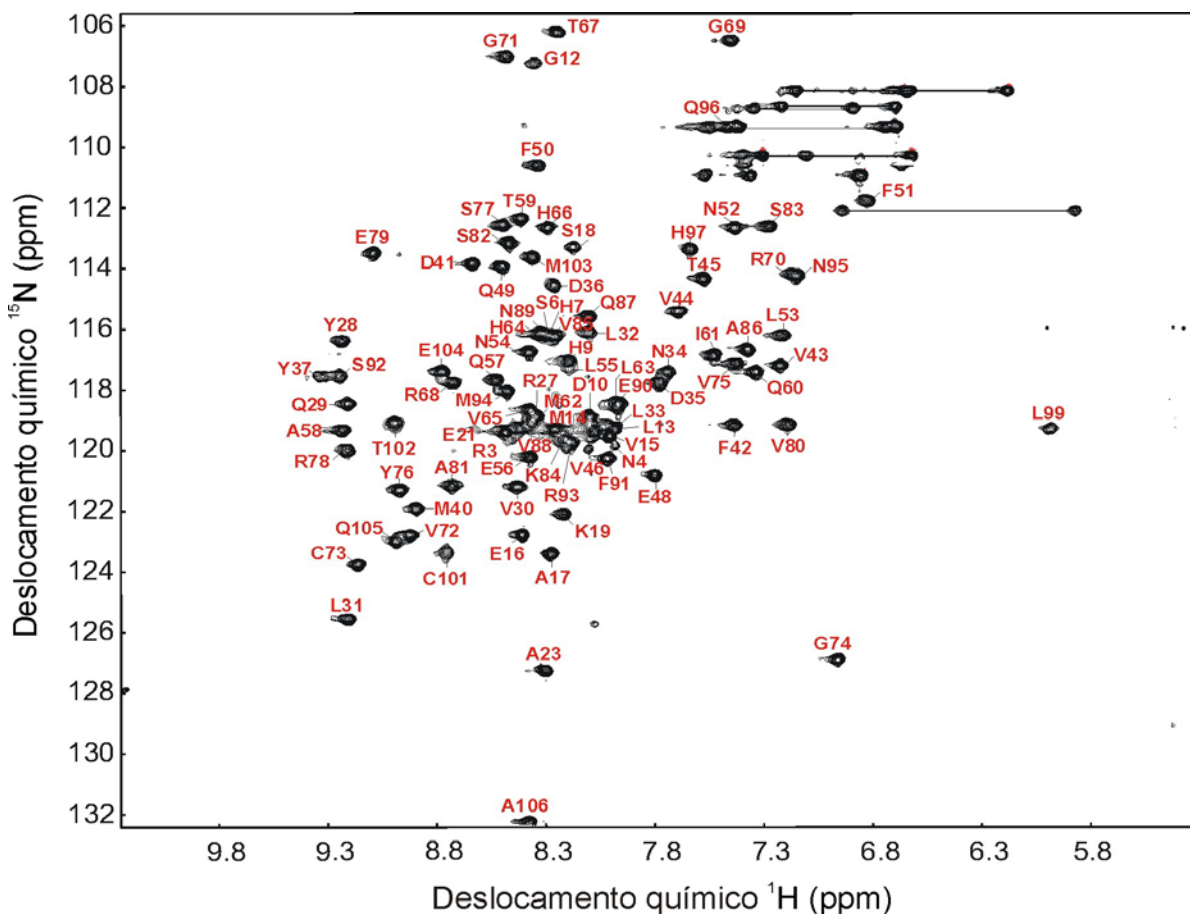
**Figura 32.** Fatias intercaladas selecionadas dos espectros de CACB(CO)NH (Fatias A) e HNCACB (Fatias B) da proteína ClpS.

As linhas indicam o caminho percorrido para o assinalamento seqüencial das ressonâncias dos núcleos de  $^{13}\text{C}^\alpha$  e  $^{13}\text{C}^\beta$  dos resíduos 72 a 76, baseado nas correlações existentes entre um amino ácido e o seu vizinho anterior. As linhas vermelhas referem-se ao assinalamento das ressonâncias  $^{13}\text{C}^\alpha$ . As linhas azuis referem-se ao assinalamento das ressonâncias  $^{13}\text{C}^\beta$ . Note-se que, conforme esperado, no caso das glicinas (G71 e G74) não se observam sinais de  $^{13}\text{C}^\beta$ . De um modo geral, picos fortes correspondem às correlações entre o grupo amídico e os núcleos  $^{13}\text{C}^\alpha$  e  $^{13}\text{C}^\beta$  do mesmo resíduo (intra-residual) enquanto os picos fracos referem-se às correlações seqüenciais entre o grupo amídico de um resíduo e os núcleos  $^{13}\text{C}^\alpha$  e  $^{13}\text{C}^\beta$  do resíduo anterior.

Assim, picos do espectro  $^1\text{H}$ - $^{15}\text{N}$ -HSQC que nos espectros de tripla ressonância apresentaram picos correspondentes com um deslocamento químico de  $^{13}\text{C}^\alpha$  em torno de 45 foram atribuídos aos quatro resíduos de glicina presentes na proteína ClpS.

Outros exemplos de amino ácidos facilmente identificáveis por terem deslocamentos químicos para os núcleos de  $^{13}\text{C}^\alpha$  e  $^{13}\text{C}^\beta$  característicos são as alaninas ( $^{13}\text{C}^\beta \sim 20$  ppm), as treoninas ( $^{13}\text{C}^\beta \sim 70$  ppm) e as serinas ( $^{13}\text{C}^\beta \sim 65$  ppm). Usando estes amino ácidos como pontos de referência na seqüência, e sabendo-se os  $\delta$  dos núcleos de  $^{13}\text{C}^\alpha$  e  $^{13}\text{C}^\beta$  do aminoácido anterior (i-1) é possível efetuar o encadeamento sucessivo dos vários amino ácidos. Este encadeamento é apenas interrompido pela presença de prolinas (que totalizam sete nesta proteína), por serem imino ácidos (i.e. o  $^{15}\text{N}$  da ligação peptídica faz parte de um anel imídico).

Usando a metodologia descrita acima foi possível atribuir as ressonâncias de aproximadamente 98% dos núcleos da cadeia principal da proteína ClpS. A Figura 33 mostra o espectro 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC da proteína ClpS contendo 96 picos amídicos assinalados de um total de 98 picos esperados para esta proteína (106 amino ácidos, menos 1 resíduo N-terminal, menos 7 resíduos de prolina). Os resíduos amídicos não assinalados pertencem aos amino ácidos T38 e L100, e não puderam ser atribuídos pois os seus picos correspondentes não foram observados no espectro  $^1\text{H}$ - $^{15}\text{N}$ -HSQC. Os restantes núcleos da cadeia principal também foram atribuídos com sucesso. A Tabela 15 apresenta a estatística da atribuição dos vários núcleos da cadeia principal de ClpS. O assinalamento completo dos vários núcleos pode ser visto no Apêndice 1 que contém as informações do arquivo "ppm.out" gerado pelo programa NMRView 5.0 (JOHNSON e BLAVINS, 1994).



**Figura 33.** Espectro 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC da proteína ClpS mostrando o assinalamento dos picos  $^1\text{H}$ - $^{15}\text{N}$  da cadeia principal.

As legendas acima dos picos indicam os resíduos de amino ácidos na seqüência da proteína. Os picos conectados por linhas e sem legenda são referentes a ressonâncias  $^1\text{H}$ - $^{15}\text{N}$  dos grupos  $\text{C}(=\text{O})\text{NH}_2$  das cadeias laterais de de resíduos de asparagina e glutamina.

**Tabela 15.** Estatística da atribuição dos núcleos da cadeia principal da proteína.

Núcleo	% de núcleos atribuídos	Resíduos não-atribuídos
HN	98	T38, L100
CO	93	M1, T5, K19, A23, P24, P25, Y37
C $\alpha$	97	M1, P24, P25
C $\beta$	96	M1, P24, P25, L100
H $\alpha$	94	M1, P2, P20, P24, P25, L100



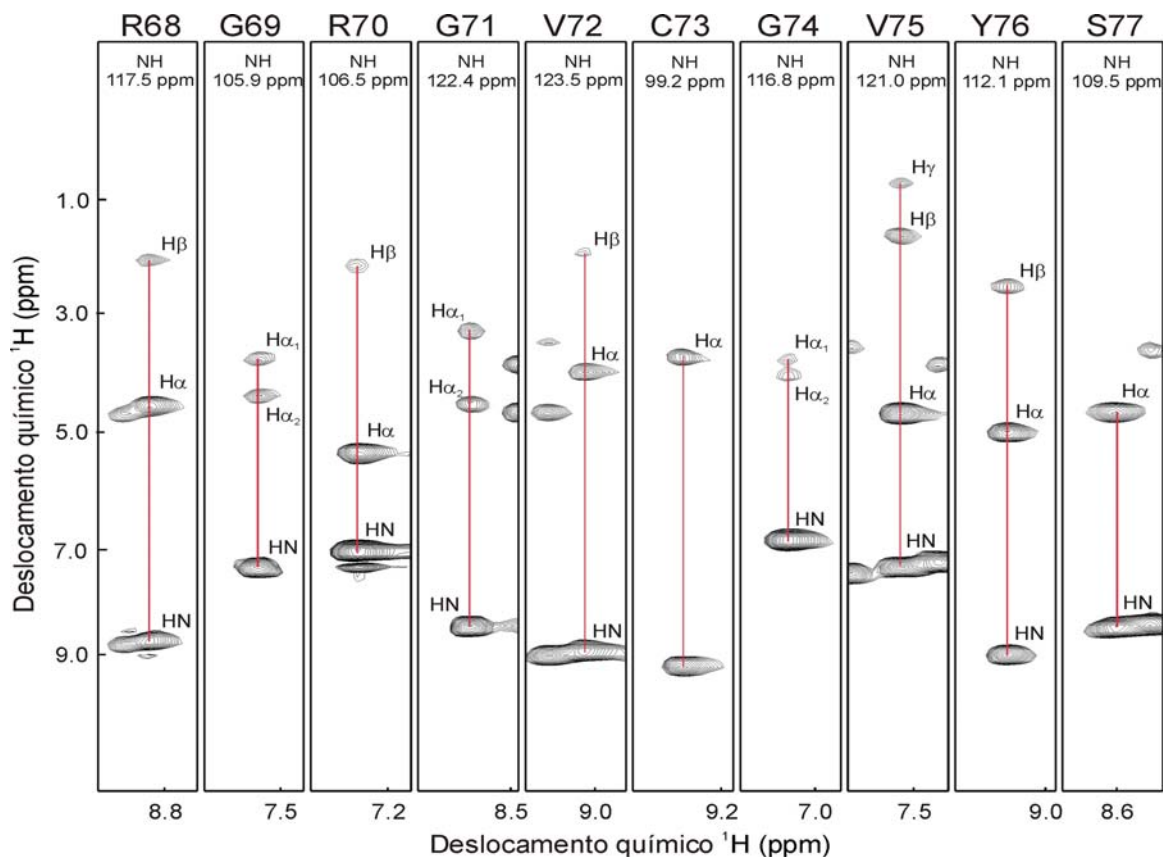
### 3.4.3.3.2. Assinalamento das cadeias laterais

Os núcleos das cadeias laterais alifáticas da proteína ClpS foram assinalados através da coleta e análise de quatro experimentos: 3D de  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TOCSY, 3D HCCH-COSY, 3D HCCH-TOCSY e 2D  $^1\text{H}$ - $^{13}\text{C}$ -HSQC. Os espectros foram adquiridos de acordo com os métodos descritos em 3.3.3.4 usando uma amostra da proteína ClpS duplamente marcada.

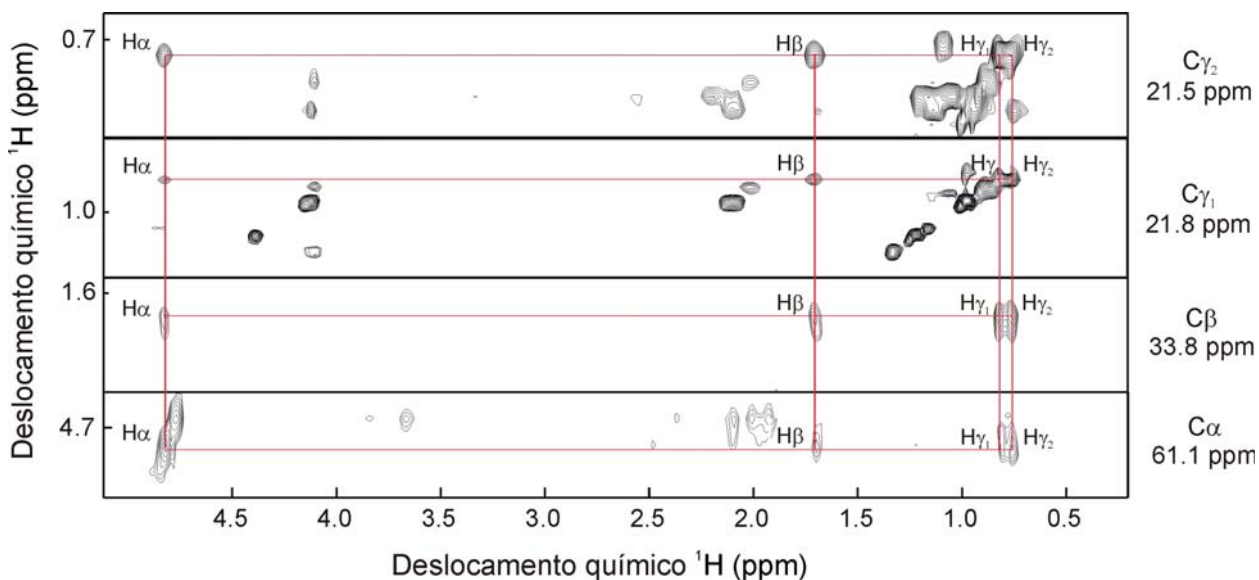
Inicialmente foi analisado o espectro 3D de  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TOCSY que fornece conectividades via ligação química entre o grupo amídico (i.e. núcleos  $^1\text{H}^{\text{N}}$  e  $^{15}\text{N}$ ), o hidrogênio alfa ( $^1\text{H}^{\alpha}$ ) da cadeia principal e os restantes núcleos de hidrogênio da cadeia lateral de um dado amino ácido possibilitando assim o assinalamento de todos os prótons alifáticos do amino ácido em estudo. A Figura 34 mostra fatias selecionadas do espectro  $^1\text{H}$ - $^{15}\text{N}$  TOCSY-HSQC de ClpS ilustrando o assinalamento dos núcleos de hidrogênio dos resíduos 68 a 77.

Apesar de na teoria ser possível assinalar todos os núcleos de hidrogênio alifáticos das cadeias laterais analisando apenas espectro  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TOCSY, em proteínas maiores nem sempre isso é possível. Assim, o assinalamento das ressonâncias de hidrogênio das cadeias laterais alifáticas da proteína ClpS foi complementado pela análise dos espectros 3D HCCH-COSY e 3D HCCH-TOCSY. Estes experimentos possibilitam a dispersão dos espectros 2D  $^1\text{H}$ - $^1\text{H}$  COSY ou TOCSY em uma terceira dimensão correspondente à frequência do  $^{13}\text{C}$  facilitando a atribuição das ressonâncias dos núcleos de carbono e de hidrogênio das cadeias laterais alifáticas.

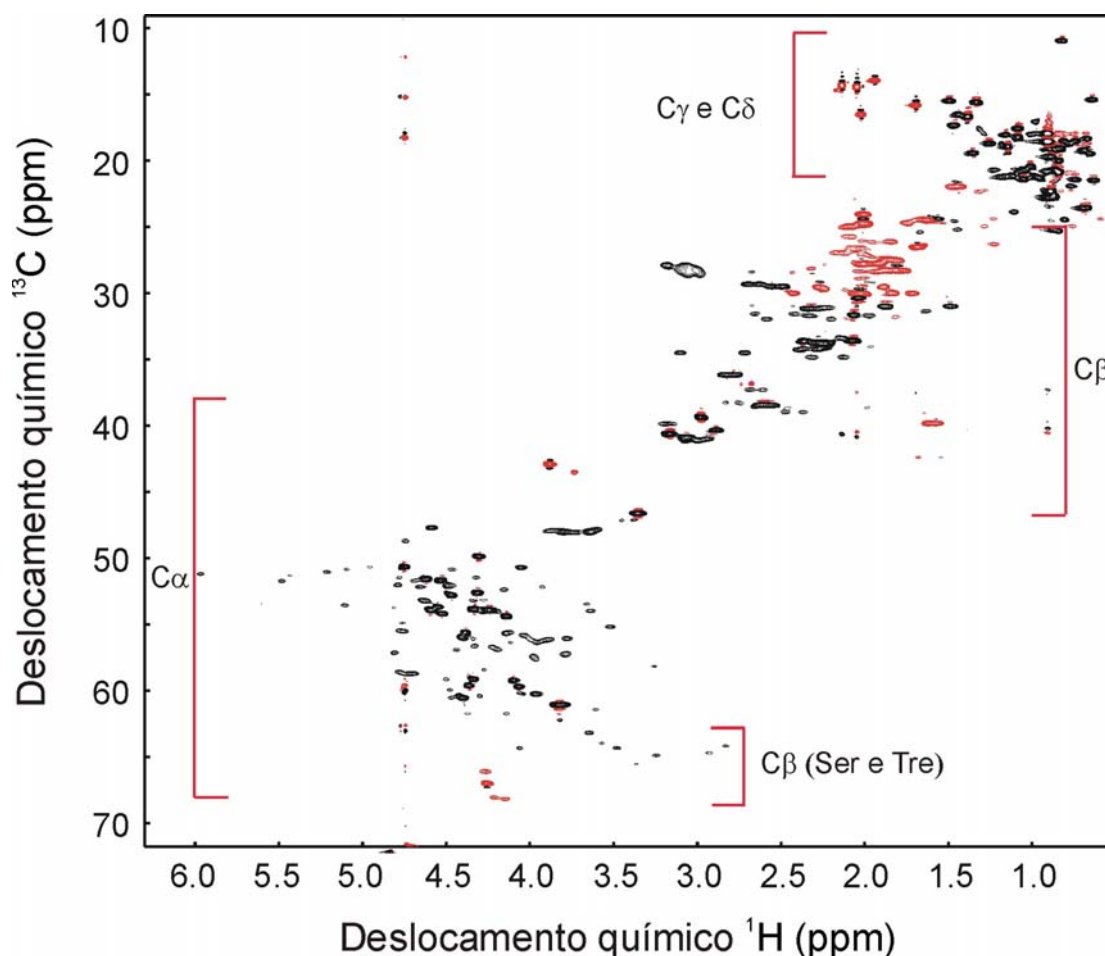
A Figura 35 mostra regiões selecionadas do espectro de HCCH-TOCSY da proteína ClpS, mostrando o assinalamento das ressonâncias da cadeia lateral do resíduo V75.



**Figura 34.** Fatias selecionadas do espectro  $^1\text{H}$ - $^{15}\text{N}$  TOCSY-HSQC de ClpS ilustrando o assinalamento dos prótons alifáticos dos resíduos 68 a 77.



**Figura 35.** Regiões selecionadas de fatias de  $^{13}\text{C}$  do espectro de HCCH-TOCSY da proteína ClpS. Em cada fatia são mostrados os deslocamentos químicos de  $^{13}\text{C}$  correspondentes. As linhas vermelhas indicam as correlações existentes entre os prótons da cadeia lateral do resíduo V75. O assinalamento de cada ressonância está indicado no próprio espectro.



**Figura 36.** Espectro de  $^1\text{H}$ - $^{13}\text{C}$  HSQC da proteína ClpS da região dos núcleos alifáticos. As regiões dos diferentes deslocamentos químicos dos picos de correlação dos núcleos  $\text{C}\alpha$ ,  $\text{C}\beta$ ,  $\text{C}\gamma$  e  $\text{C}\delta$  estão indicadas no espectro.

Uma vez assinaladas as ressonâncias dos núcleos de  $^1\text{H}$  e  $^{13}\text{C}$ , os sistemas de spins das cadeias laterais alifáticas foram conectados com as ressonâncias de  $^1\text{H}^{\text{N}}$  e  $^{15}\text{N}$  da cadeia principal da proteína via a correlação  $^1\text{HN}$ - $^{15}\text{N}$ - $^1\text{H}^{\alpha}$  observada no espectro 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TOCSY. Usando esta metodologia, aproximadamente 90% das cadeias laterais alifáticas de ClpS foram atribuídas com sucesso (vide Apêndice I para maiores detalhes).

O assinalamento não-ambíguo das ressonâncias das cadeias laterais alifáticas de ClpS foi confirmado através do assinalamento de cada um dos picos

observados no espectro  $^1\text{H}$ - $^{13}\text{C}$  HSQC, que ilustra as correlações entre todos os núcleos de hidrogênio e de carbono da proteína. A Figura 36 mostra o espectro  $^1\text{H}$ - $^{13}\text{C}$  HSQC da proteína ClpS das regiões dos núcleos alifáticos. É interessante notar que os deslocamentos químicos observados para cada par de núcleos CH depende estão relacionados com a posição em que estes núcleos se encontram nas cadeias laterais. Assim, picos cruzados de  $\text{C}\gamma\text{-H}\gamma$  e  $\text{C}\delta\text{-H}\delta$  geralmente apresentam deslocamentos químicos deslocados para regiões de campo baixo tanto na dimensão  $^1\text{H}$  como de  $^{13}\text{C}$ . Já picos cruzados de  $\text{C}\alpha\text{-H}\alpha$  e  $\text{C}\beta\text{-H}\beta$  aparecem no espectro em regiões de campo mais alto.

O assinalamento das ressonâncias de hidrogênio das cadeias laterais aromáticas de ClpS seria realizado através da análise dos espectros 3D  $^1\text{H}$ - $^{13}\text{C}$  NOESY-CHSQC e 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC. No entanto, conforme descrito anteriormente, durante o assinalamento desta proteína, dois grupos independentes determinaram a estrutura cristalográfica da proteína ClpS de *E. coli* (GUO *et al.*, 2002a; ZETH *et al.*, 2002), cuja seqüência é muito semelhante à seqüência da proteína ClpS de *Xac*.

Assim, estas publicações redirecionaram os objetivos deste trabalho, que originalmente visavam a determinação da estrutura tridimensional da proteína ClpS de *Xac* por RMN, para a obtenção de informações que complementassem os dados obtidos por GUO e colaboradores (2002a) e ZETH e colaboradores (2002) em seus estudos por cristalografia. Nestes novos estudos, o assinalamento das cadeias laterais aromáticas de ClpS de *Xac* não seria fundamental. Assim, foi decidido abandonar o assinalamento das cadeias laterais aromáticas de ClpS de *Xac* e focalizar em outros estudos por RMN desta proteína.

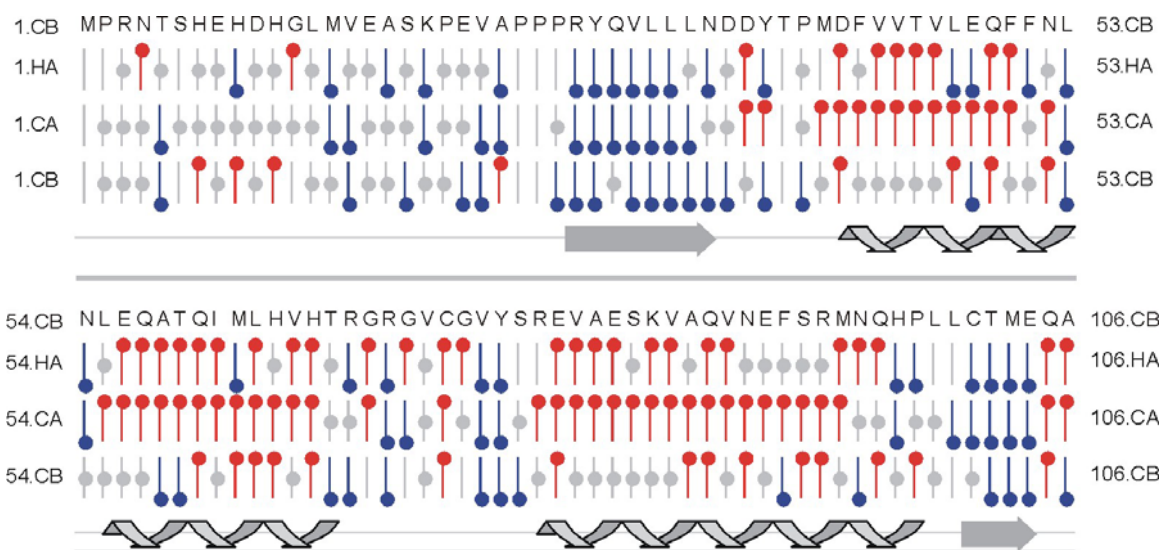
O assinalamento das ressonâncias dos núcleos da cadeia principal e das cadeias foi depositado no Biological Magnetic Resonance Bank (BMRB) sob o número de acesso BMRB ID 15448 (Fonte: [www.bmrb.wisc.edu/](http://www.bmrb.wisc.edu/), 2007).

#### 3.4.3.4. Identificação dos elementos de estrutura secundária

Uma vez obtido o assinalamento seqüencial de todos os amino ácidos presentes na molécula tem-se acesso imediato a informação estrutural que possibilita a identificação dos elementos de estrutura secundária presentes na proteína em estudo antes mesmo da determinação da estrutura tridimensional da mesma.

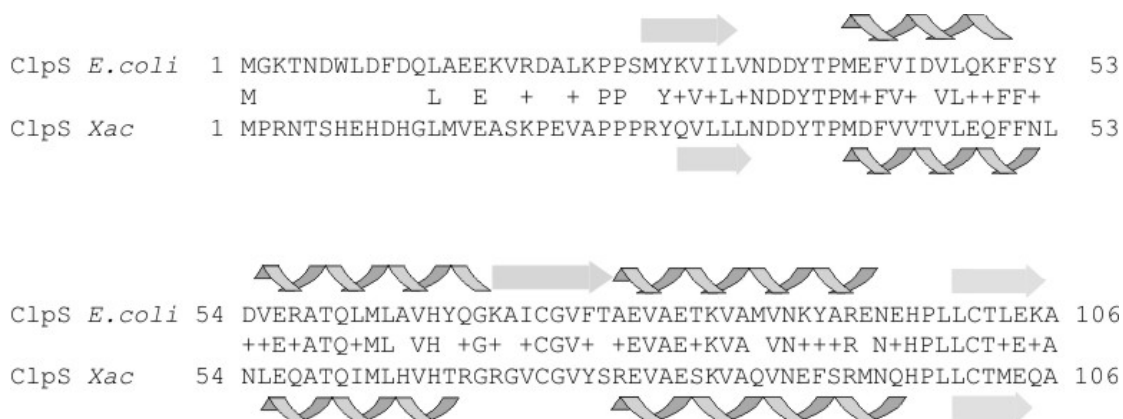
Inicialmente, o conteúdo de estrutura secundária da proteína ClpS foi predito usando o método de cálculo de “Chemical Shift Index (CSI)” descrito por WISHART, e colaboradores (1992) que faz uso dos valores de deslocamento químico obtidos experimentalmente para cada um dos núcleos da cadeia principal.

Conforme descrito anteriormente, o CSI ou índice de deslocamento químico, é a diferença entre os valores de deslocamento químico obtidos experimentalmente a partir do assinalamento seqüencial da proteína e os valores de deslocamento químico médios para uma conformação randômica. As regiões da proteína que apresentam diferenças negativas para  $^1\text{H}^\alpha$  e  $^{13}\text{C}^\beta$  e, positivas para  $^{13}\text{C}^\alpha$ , indicam a presença de uma conformação em  $\alpha$ -hélice. Já as regiões da proteína que apresentam diferenças negativas para  $^1\text{H}^\alpha$  e  $^{13}\text{C}^\beta$  e, positivas para  $^{13}\text{C}^\alpha$  indicam a presença de uma conformação de fita  $\beta$ . Nas regiões em que o índice de deslocamento químico está próximo de zero provavelmente não existe estrutura secundária regular.



**Figura 37.** “Chemical Shift Index” calculado para a proteína ClpS de *Xac*.

A seqüência primária da proteína está representada no topo da figura. Abaixo estão indicadas as diferenças de deslocamento químico dos núcleos  $^1\text{H}^\alpha$ ,  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ , respectivamente, para cada aminoácido de ClpS. Vermelho indica diferença positiva e azul, negativa. Em cinza encontram-se as diferenças não significativas (traço com bola) ou ressonâncias não assinaladas (traço sem bola) Os elementos de estrutura secundária estão representados por setas (fita- $\beta$ ) e hélices ( $\alpha$ -hélices).



**Figura 38.** Comparação da estrutura secundária da proteína ClpS de *E.coli* com a estrutura secundária predita para a proteína ClpS de *Xac*. Os elementos de estrutura secundária estão representados por setas (fita- $\beta$ ) e hélices ( $\alpha$ -hélices).

A Figura 37 apresenta o “Chemical Shift Index” (CSI) calculado e a estrutura secundária estimada para a seqüência da proteína ClpS de *Xac*. De acordo com os resultados obtidos, a proteína ClpS de *Xac* apresenta os seguintes elementos de estrutura secundária: 2 fitas  $\beta$  (resíduos 27 a 34 e 101 a 104) e 3  $\alpha$ -hélices (resíduos 41 a 53, 55 a 67 e 78 a 97).

É interessante notar que a região que compreende os resíduos 70 a 77 da proteína ClpS de *Xac* parece apresentar o que seria uma região de fita  $\beta$  apenas interrompida pelos valores discrepantes obtidos para o resíduo C73. De fato, analisando a estrutura secundária da proteína ClpS de *E.coli* (DOUGAN *et al.*, 2002a) observa-se a presença de uma fita  $\beta$  na região compreendida entre os resíduos 70 a 77 (Figura 38). Especificamente esta fita apresenta um “kink” ou pequena dobradura na posição do resíduo C73. Assim, dado o elevado grau de similaridade entre as seqüências das duas proteínas, é de esperar que a proteína ClpS de *Xac* também apresente estrutura secundária semelhante nesta região.

As estruturas secundárias regulares presentes na proteína ClpS de *Xac* foram também identificadas através das constantes de acoplamento de escalar  $^3J_{\text{HNH}\alpha}$  obtidas experimentalmente e dos respectivos ângulos diedros. As constantes de acoplamento escalar  $^3J_{\text{HNH}\alpha}$  foram obtidas através da análise do experimento 3D de dupla ressonância HNHA (VUISTER e BAX, 1993). Neste espectro, observam-se dois picos para cada resíduo de amino ácido. Na diagonal do espectro observa-se o pico de correlação dos núcleos HN e N equivalente ao pico de correlação observado no experimento 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC. Observa-se ainda um pico cruzado que corresponde à correlação dos núcleos  $\text{H}^\alpha$ , HN e N. Uma vez assinalados todos os picos do espectro, a constante  $^3J_{\text{HNH}\alpha}$  para cada resíduo de amino ácido foi calculada

usando a equação abaixo que relaciona esta constante às intensidades do pico cruzado ( $I_X$ ) e do pico da diagonal ( $I_D$ ) (KUBONIWA e col, 1994).

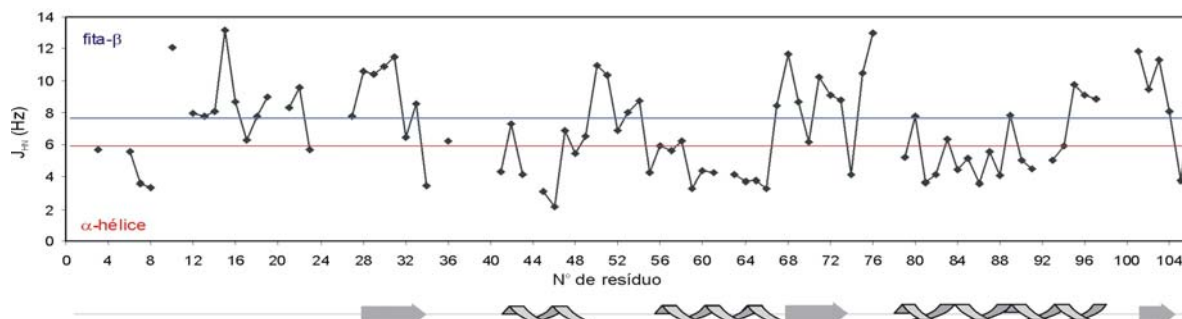
$$I_X/I_D = -\tan^2 (2\pi\xi J_{\text{HNH}\alpha}) \quad [\text{Equação 8}]$$

em que  $\xi$  se refere à soma dos dois períodos  $t_1$  e  $t_2$  em que  ${}^3J_{\text{HNH}\alpha}$  evolui, que neste caso foram  $t_1 = 0.0075$  e  $t_2 = 0.0125$  s respectivamente.

Desta forma, o valor da constante de acoplamento  ${}^3J_{\text{HNH}\alpha}$  foi calculado para 82 dos 106 resíduos de amino ácidos da proteína ClpS. Como seria de esperar, não foi possível obter valores de  ${}^3J_{\text{HNH}\alpha}$  para a metionina N-terminal, para os sete resíduos de prolina, nem para os resíduos T38 e L100 cujas ressonâncias dos núcleos HN não foram assinaladas. Além disso, também não foi possível quantificar esta constante para os resíduos N4, T5, H9, H11, D35, Y37, M40, V44, M62, S77, R78, S92 e L99 uma vez que os respectivos picos não puderam ser identificados no espectro HNHA devido à ocorrência de sobreposição espectral ou baixa intensidade do sinal.

Conforme descrito anteriormente, segmentos de três ou mais resíduos de amino ácidos consecutivos que apresentam valores de  ${}^3J_{\text{HNH}\alpha} < 6,0$  Hz indicam um região de  $\alpha$ -hélice. Os mesmos segmentos com valores de  ${}^3J_{\text{HNH}\alpha} > 8,0$  Hz indicam conformação em fita- $\beta$  (WÜTHRICH, 1986). Assim, como pode ser visto na Figura 39, os elementos de estrutura secundária da proteína ClpS identificados por este método correspondem essencialmente aqueles anteriormente preditos pelo método de CSI.





**Figura 39.** Constante de acoplamento  $^3J_{\text{HNH}\alpha}$  em função do número de resíduo de ClpS.

Valores de  $^3J_{\text{HNH}\alpha} < 6,0$  Hz indicam conformação em  $\alpha$ -hélice (abaixo da linha vermelha). Os mesmos segmentos com valores de  $^3J_{\text{HNH}\alpha} > 8,0$  Hz indicam conformação em fita- $\beta$  (acima da linha azul). Abaixo do gráfico estão representados por setas (fita- $\beta$ ) e hélices ( $\alpha$ -hélices) os elementos de estrutura secundária da proteína ClpS de Xac preditos pelo cálculo da constante de acoplamento  $^3J_{\text{HNH}\alpha}$ .

A estrutura secundária da proteína ClpS foi também predita usando o programa TALOS (CORNILESCU *et al.*, 1999). Inicialmente, os valores dos deslocamentos químicos obtidos experimentalmente para os núcleos da cadeia principal de ClpS foram comparados com a base de dados do programa composta por 20 proteínas para as quais os valores de deslocamentos químicos bem como as estruturas cristalinas de alta resolução são conhecidas. Para cada resíduo de ClpS, o programa procura na base de dados por um tripleto de amino ácidos cuja seqüência e deslocamentos químicos secundários sejam semelhantes aos do tripleto formado pelo resíduo de ClpS e pelos resíduos adjacentes. O programa seleciona os 10 tripletos que mais se assemelham ao tripleto de ClpS em estudo e a média dos ângulos  $\Phi$  e  $\Psi$  dos dez melhores “hits” é então usada como predição dos

valores dos ângulos diedros do resíduo analisado. Usando esta metodologia, foi possível prever os ângulos  $\Phi$  e  $\Psi$  para todos resíduos de ClpS cujos deslocamentos químicos dos núcleos da cadeia principal eram conhecidos.

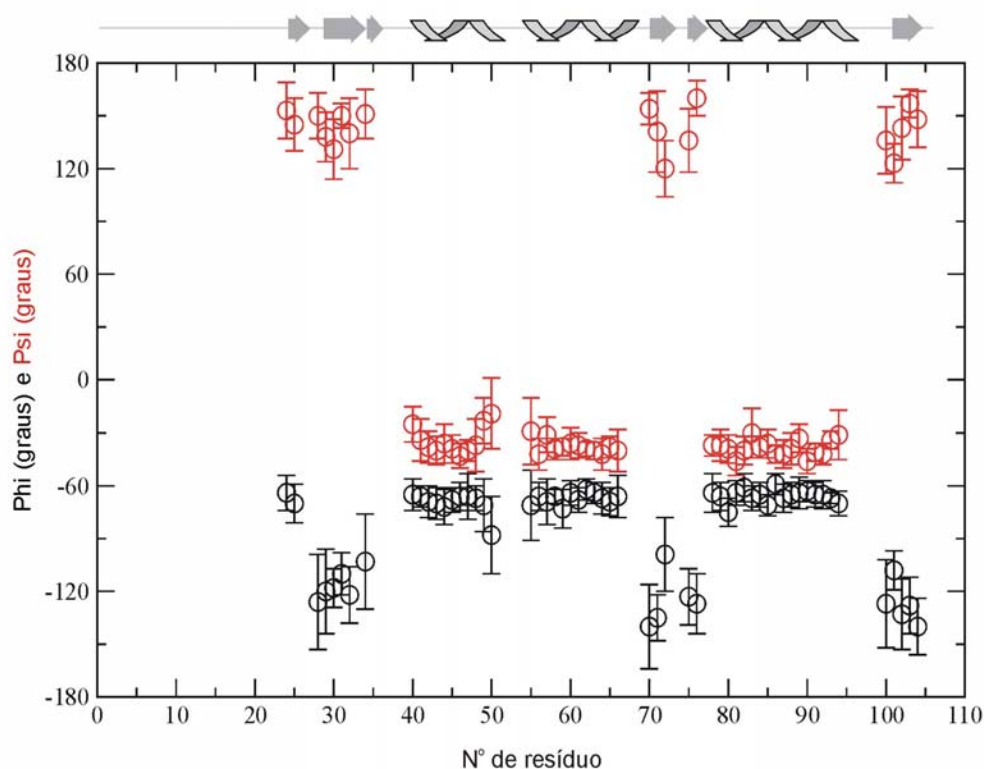
Posteriormente, usando um diagrama de Ramachandran, os ângulos diedros preditos para cada resíduo usando o programa TALOS foram correlacionados com os elementos de estrutura secundária correspondentes. Além disso, os valores de  $\Phi$  e  $\Psi$  preditos para proteína ClpS de *Xac* foram também comparados com os valores de  $\Phi$  e  $\Psi$  da proteína ClpS de *E.coli* observados na estrutura cristalina (GUO *et al.*, 2002a) (Figura 39). Finalmente, as predições de estrutura secundária foram classificadas como sendo “boas”, “ambíguas” ou “ruins”. Predições boas são aquelas para as quais os ângulos preditos para a proteína ClpS de *Xac* e os ângulos dos 10 melhores “hits” da base de dados TALOS bem como os ângulos da proteína ClpS de *E. coli* se encontram dentro da mesma região do diagrama de Ramachandran. Predições ambíguas, são aquelas para as quais a proteína em estudo e as várias proteínas analisadas possuem ângulos em diferentes regiões diagrama de Ramachandran porém os ângulos da proteína ClpS de *E. coli* encontra-se na mesma região. Uma predição foi considerada ruim quando os ângulos preditos para a proteína ClpS de *Xac* se encontram em regiões diagrama de Ramachandran distintas da proteína homóloga de *E. coli*.

A Figura 40 mostra um gráfico que apresenta as predições dos ângulos  $\Phi$  e  $\Psi$  consideradas “boas” e os elementos de estrutura secundária correspondentes para proteína ClpS de *Xac*. Uma tabela contendo todas as predições geradas pelo programa TALOS (“boas”, “ambíguas” e “ruins”) pode ser encontrada no Apêndice 2.

Conforme esperado, as predições de estrutura secundária usando o programa TALOS correspondem aos mesmos elementos de estrutura secundária

anteriormente preditos para a proteína ClpS de *Xac* pelos métodos de cálculo de CSI e  $^3J_{\text{HNH}\alpha}$ .

Pelos métodos descritos acima não foi possível identificar elementos de estrutura secundária na região N-terminal (resíduos 1 a 22) da proteína ClpS de *Xac*. A ausência de elementos de estrutura secundária nesta região proteínica foi confirmada pela análise do espectro 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY. Os resultados obtidos demonstraram que a região N-terminal da proteína ClpS de *Xac* apresenta apenas sinais de NOE intra-residuais. Sinais de NOE inter-residuais de média distância, típicos de estruturas secundárias regulares, não são observados indicando que a proteína se encontra em conformação randômica nesta região.



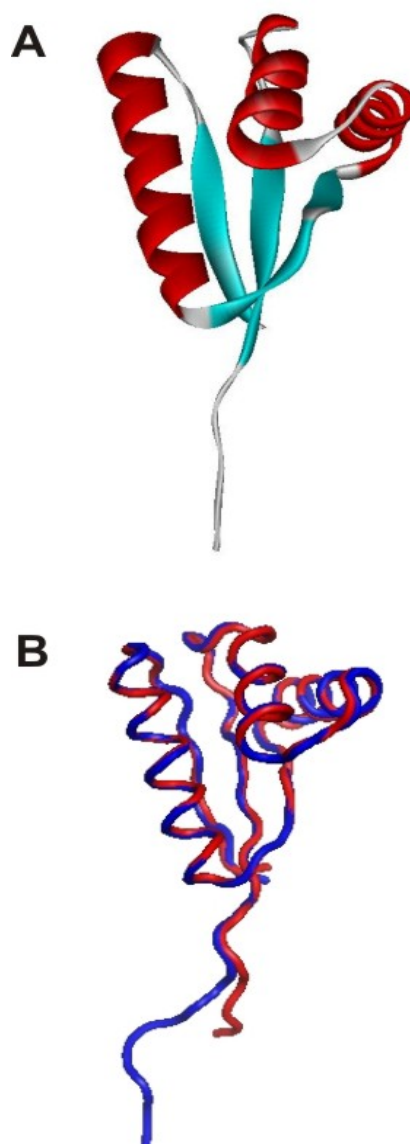
**Figura 40.** Predição dos ângulos diedros  $\Phi$  (preto) e  $\Psi$  (vermelho) da proteína ClpS de *Xac* usando o programa TALOS.

Apenas predições boas são mostradas. As predições dos elementos de estrutura secundária correspondente são representados acima do gráfico por setas (fita- $\beta$ ) e hélices ( $\alpha$ -hélices).

### 3.4.3.5. Validação do modelo estrutural da proteína ClpS de *Xac* por acoplamentos dipolares residuais (RDC)

#### 3.4.3.5.1. Cálculo do modelo estrutural da proteína ClpS de *Xac* usando modelagem por homologia

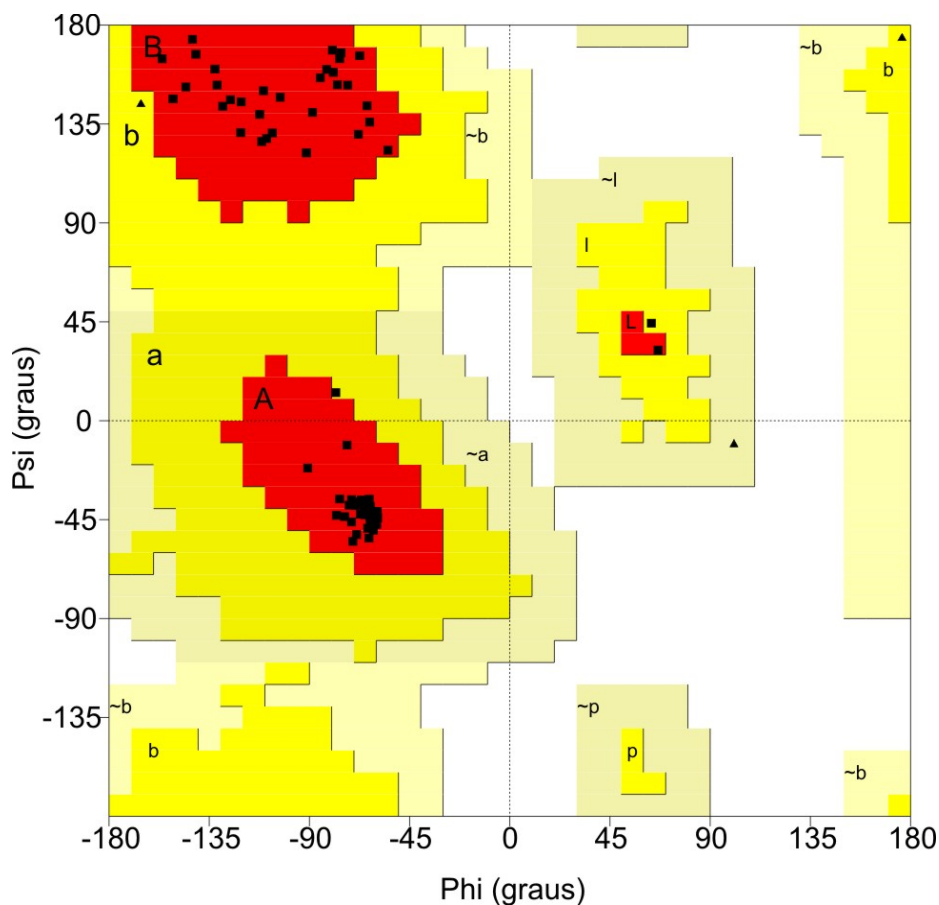
O modelo da estrutura tridimensional da proteína ClpS de *Xac* foi calculado usando modelagem por homologia a partir da estrutura cristalina de ClpS de *E.coli* (PDB 1 MBX) usando o programa MODELLER (SALI e BLUNDELL, 1993) conforme descrito na secção 3.3.5.4. O modelo gerado pode ser visto na Figura 41 A. De acordo com os resultados obtidos, a proteína ClpS de *Xac* apresenta 3 fitas  $\beta$  (resíduos 27 a 33, 70 a 77 e 100 a 105) e 3  $\alpha$ -hélices (resíduos 40 a 51, 55 a 68 e 78 a 95) e a sua estrutura tridimensional apresenta um formato cônico. É importante salientar que o modelo gerado não inclui a região N-terminal da proteína uma vez que a estrutura de ClpS de *E.coli* usada como molde não possui informações referentes a esta região. Assim, o modelo gerado refere-se apenas à região compreendida entre os resíduos P20 a A106 da proteína ClpS de *Xac*. A Figura 41 B que mostra uma sobreposição do modelo da proteína ClpS de *Xac* com a estrutura tridimensional da proteína ClpS de *E. coli*. O RMSD entre as duas estruturas é de 0.534 Å considerando-se os átomos N, C $\alpha$  e carbonila de toda região P20 a A106. Já se considerarmos os átomos N, C $\alpha$  e carbonila da região estruturada entre R27 e Q105 o RMSD é de apenas 0.147 Å.



**Figura 41.** Modelo tridimensional da proteína ClpS de *Xac* construído por modelagem por homologia com a proteína ClpS de *E. coli*.

A. Representação “cartoon” da estrutura tridimensional da proteína ClpS de *Xac* (resíduos P20-A106) gerada usando o programa WebLab Viewer Pro 3.7 (MOLECULAR SIMULATIONS, 2000). Hélices  $\alpha$  estão representadas em vermelho e fitas- $\beta$  estão representadas em turquesa.

B. Sobreposição da estrutura tridimensional da proteína ClpS de *E. coli* (resíduos E16-A106, em azul) com o modelo tridimensional da proteína ClpS de *Xac* (resíduos P20-A106, em vermelho). Representação “ribbon” gerada usando o programa PyMol (Fonte: <http://pymol.sourceforge.net/>, 2007). O RMSD entre as duas estruturas é de 0.534 Å considerando-se os átomos N,  $C_{\alpha}$  e carbonila de toda região P20 a A106 e de 0.147 Å considerando os átomos N,  $C_{\alpha}$  e carbonila da região estruturada entre R27 e Q105.



**Figura 42.** Diagrama de Ramachandran do modelo da proteína ClpS de *Xac* obtido usando o programa PROCHECK (LASKOWSKI *et al.*, 1993). O gráfico demonstra a inexistência de resíduos fora das regiões permitidas dentre os aminoácidos do modelo criado usando o programa MODELLER (SALI e BLUNDELL, 1993). 98.7% dos resíduos encontram-se nas regiões mais favorecidas (A,B,L) e 1.3% dos resíduos encontram-se em outras regiões permitidas (a,l). Os resíduos de glicina são mostrados como triângulos.

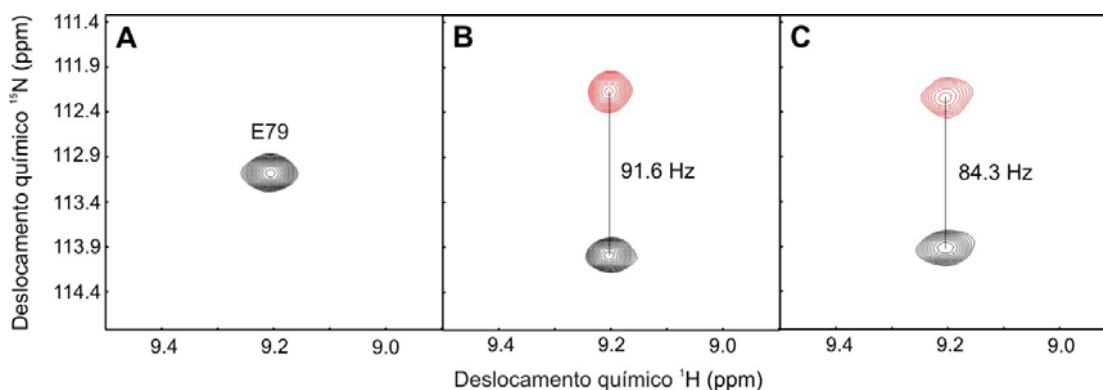
O modelo obtido foi também avaliada quanto à sua qualidade estereoquímica usando do programa PROCHECK (LASKOWSKI *et al.*, 1993). A Figura 42 apresenta o diagrama de Ramachandran do modelo da proteína ClpS de *Xac* mostrando que 98.7% dos resíduos encontram-se nas regiões mais favorecidas e 1.3% dos resíduos encontram-se em outras regiões permitidas estereoquimicamente.

### 3.4.3.5.2. Validação do modelo da proteína ClpS de *Xac* por acoplamentos dipolares residuais (RDC)

Conforme discutido anteriormente, os acoplamentos dipolares residuais, além de serem utilizados como restrições para o cálculo de estruturas por RMN, podem também ser usados para validar modelos estruturais e identificar enovelamentos de proteínas através da comparação dos acoplamentos dipolares residuais obtidos experimentalmente para a molécula em estudo com acoplamentos dipolares residuais teóricos calculados para uma estrutura conhecida (TJANDRA *et al.*, 1997; TOLMAN, 2001; PRESTGARD *et al.*, 2004).

Os acoplamentos dipolares residuais  $^1\text{H}$ - $^{15}\text{N}$  ( $D_{\text{HN}}$ ) da proteína ClpS de *Xac* foram medidos fazendo uso de espectros IPAP  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (OTTIGER *et al.*, 1998) adquiridos na ausência e na presença de fago Pf1 (HANSEN *et al.*, 1998). Ao não desacoplar os sinais de próton na dimensão indireta, o espectro  $^1\text{H}$ - $^{15}\text{N}$ -HSQC é desdobrado em dois sendo que para cada deslocamento químico de  $^1\text{H}$  se observam dois sinais de  $^{15}\text{N}$  correspondentes aos dois componentes do dubleto.

O método IPAP (“in-phase/anti-phase”) permite visualizar cada componente do dubleto como sinais com fases opostas em um mesmo espectro e a constante de acoplamento  $^1J_{\text{NH}}$  pode ser medida como a distância em Hz entre os dois sinais (PRESTGARD *et al.*, 2000). Desta forma, os acoplamentos dipolares residuais  $D_{\text{NH}}$  da proteína ClpS de *Xac* foram calculados a partir da diferença dos acoplamentos escalares  $^1J_{\text{NH}}$  medidos na ausência e presença de fago (Figura 43). É importante salientar que os deslocamentos químicos amídicos de ClpS não apresentaram variação na ausência e na presença de fago mostrando que a presença de fago não altera a conformação da proteína ClpS de *Xac*.



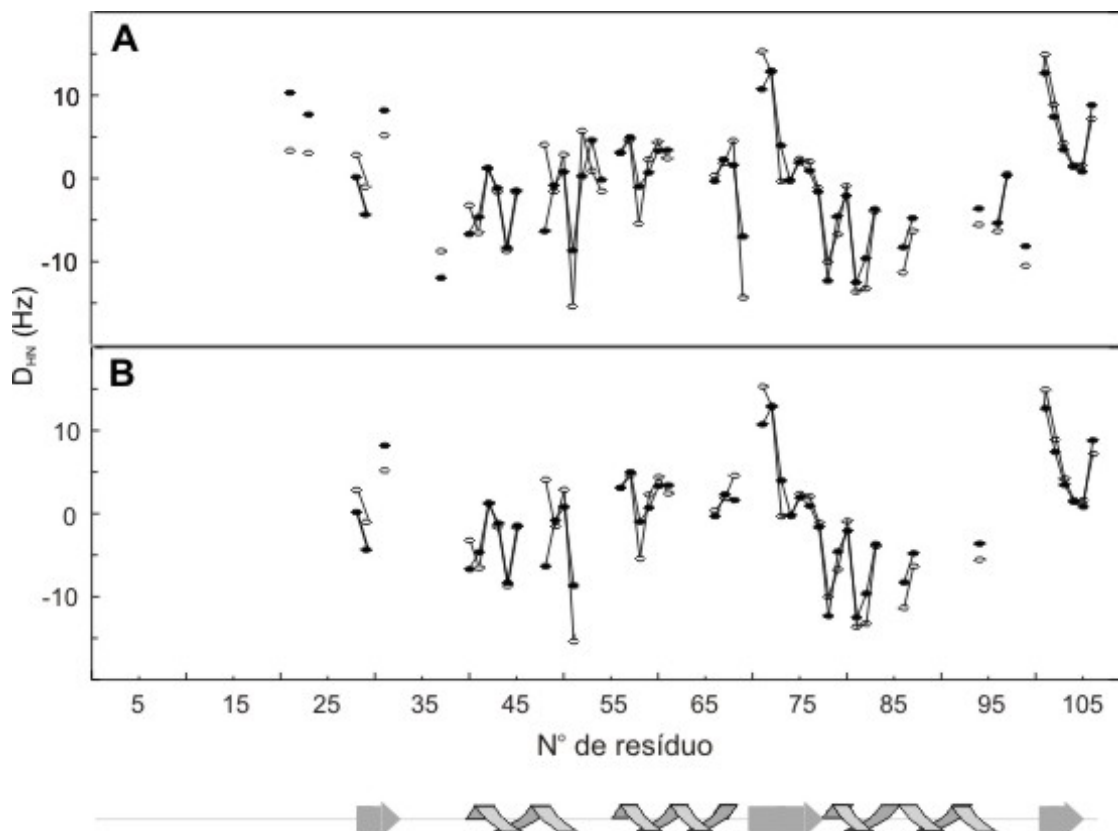
**Figura 43.** Experimentos utilizados para calcular os acoplamentos dipolares residuais  $D_{NH}$  da proteína ClpS de Xac.

A. Espectro  $^1\text{H}$ - $^{15}\text{N}$ -HSQC mostrando a região do pico HN do resíduo E79.

B e C. Espectros IPAP  $^1\text{H}$ - $^{15}\text{N}$ -HSQC na ausência (isotrópico) e na presença (anisotrópico) de 20mg/mL de fago Pf1. Picos vermelhos (negativos) representam os componentes de campo baixo do dubleto enquanto picos pretos (positivos) representam os componentes de campo alto do dubleto. O acoplamento  $^1J_{HN}$  medido na ausência e na presença de fago é mostrado em Hertz. Os desdobramentos de sinal medidos no painel B correspondem ao acoplamento escalar  $^1J_{NH}$ . Os desdobramentos de sinal medidos no painel C correspondem à soma do acoplamento escalar  $^1J_{NH}$  e do acoplamento dipolar  $D_{NH}$ .

Os acoplamentos dipolares residuais obtidos experimentalmente foram então analisados usando o módulo BestFit do programa PALES (ZWECKSTETTER e BAX, 2000). Este programa usa SVD (“Single Value Decomposition”) (LOSONCZI *et al.*, 1999) para prever os componentes de magnitude e orientação do tensor de alinhamento dos núcleos de uma determinada molécula a partir do modelo da sua estrutura tridimensional e a partir de acoplamentos dipolares residuais obtidos experimentalmente. Desta forma foi possível comparar os acoplamentos dipolares residuais teóricos calculados a partir do modelo tridimensional da proteína com os acoplamentos dipolares residuais obtidos experimentalmente. O resultado final é uma correlação quantitativa entre os acoplamentos dipolares residuais teóricos e experimentais, indicativa de quão bem o modelo teórico se ajusta aos dados experimentais.





**Figura 44.** Acoplamentos dipolares residuais  $D_{NH}$  da proteína ClpS de Xac em função da seqüência primária.

A. Acoplamentos dipolares observados para a proteína ClpS de Xac (círculos abertos) e os acoplamentos dipolares calculados (círculos fechados) a partir do modelo da proteína ClpS de Xac (total de 54 acoplamentos escalares).

B. Acoplamentos dipolares observados para a proteína ClpS de Xac (círculos abertos) e acoplamentos dipolares calculados (círculos fechados) a partir do modelo da proteína ClpS de Xac apenas para as regiões contendo elementos de estrutura secundária (total de 44 acoplamentos escalares). Os elementos de estrutura secundária do modelo tridimensional da proteína ClpS de Xac são mostrados abaixo dos gráficos.

A Figura 44 mostra um gráfico dos acoplamentos dipolares residuais obtidos experimentalmente e calculados a partir do modelo plotados em função da seqüência primária da proteína. Uma análise dos resultados obtidos mostra que os valores de  $D_{HN}$  medidos experimentalmente para a proteína ClpS de Xac e os valores de  $D_{HN}$  calculados a partir do modelo da proteína são bastante similares. Com exceção dos resíduos M40, L47 e G71, todos os resíduos da proteína ClpS

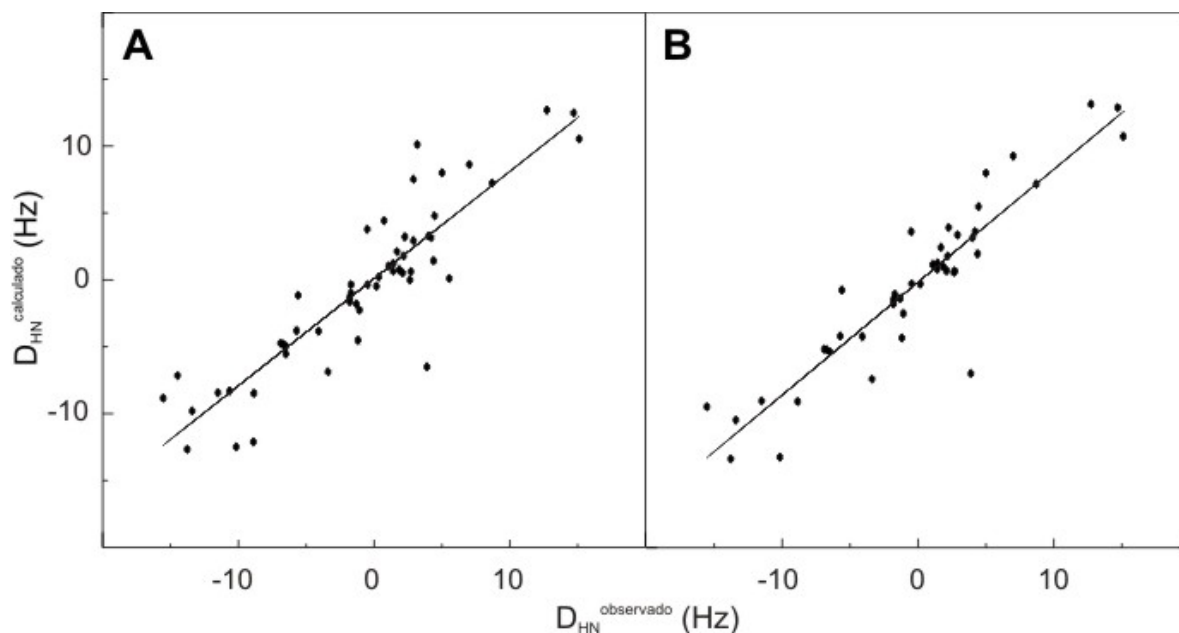
parecem encontrar-se na mesma orientação relativa ao tensor de alinhamento que os resíduos correspondentes no modelo tridimensional. Em particular, as fitas- $\beta$  apresentam uma correlação muito boa entre os valores de  $D_{HN}$  observados e calculados. Já na região N-terminal, embora tenham sido medidos valores de  $D_{HN}$ , não foi possível compará-los com o modelo estrutural uma vez que este compreende apenas os resíduos da região P20-A106.

A Figura 45 mostra uma correlação entre os acoplamentos dipolares residuais  $D_{HN}$  observados e calculados para a proteína ClpS de *Xac*. Para testar a concordância entre os dados experimentais e calculados foi utilizado o fator de qualidade Q calculado usando a seguinte equação (CORNILESCU *et al.*, 1998):

$$Q = \text{rms} (D_{HN}^{\text{obs}} - D_{HN}^{\text{calc}}) / \text{rms} (D_{HN}^{\text{obs}}) \quad [\text{Equação 9}]$$

O fator Q aproxima-se de zero quando existe uma concordância perfeita entre dados experimentais e calculados. Em estruturas de RMN refinadas usando acoplamentos dipolares residuais o valor de Q pode chegar próximo de 0.15 mas nunca se aproxima de zero pois o tensor de alinhamento não é conhecido com precisão e existem erros experimentais nas medidas dos acoplamentos dipolares (CORNILESCU e BAX, 2000).

Já no caso de estruturas cristalográficas de média resolução ( $>2\text{\AA}$ ) esse valor fica em torno de 0.3 para regiões estruturadas mas pode chegar até 0.7 em casos de regiões mais flexíveis (CORNILESCU *et al.*, 1998; BAX, 2003). No presente estudo foi obtido um fator Q de 0.448 quando todos os 54 os acoplamentos dipolares residuais  $D_{HN}$  de ClpS obtidos experimentalmente foram computados.



**Figura 45.** Correlação entre os acoplamentos dipolares residuais  $D_{\text{HN}}$  observados e calculados para a proteína ClpS de Xac.

A. Acoplamentos dipolares observados e calculados para a proteína ClpS de Xac (total 54 acoplamentos escalares). O fator de qualidade Q obtido foi de 0.448 e a inclinação da reta de 0.801.

B. Acoplamentos dipolares observados e calculados para as regiões da proteína ClpS de Xac contendo elementos de estrutura secundária (total de 44 acoplamentos escalares). O fator de qualidade Q obtido foi de 0.394 e a inclinação da reta de 0.843.

Quando apenas os 44 acoplamentos dipolares residuais  $D_{\text{HN}}$  obtidos experimentalmente presentes em regiões de estrutura secundária de ClpS foram computados o fator Q foi de 0.394.

Considerando-se os valores de fator Q obtidos e considerando-se que se trata de um modelo teórico e não de um modelo experimental, podemos afirmar que o modelo da proteína ClpS de Xac proposto por modelagem por homologia é representativo da estrutura tridimensional da proteína ClpS de Xac em solução.

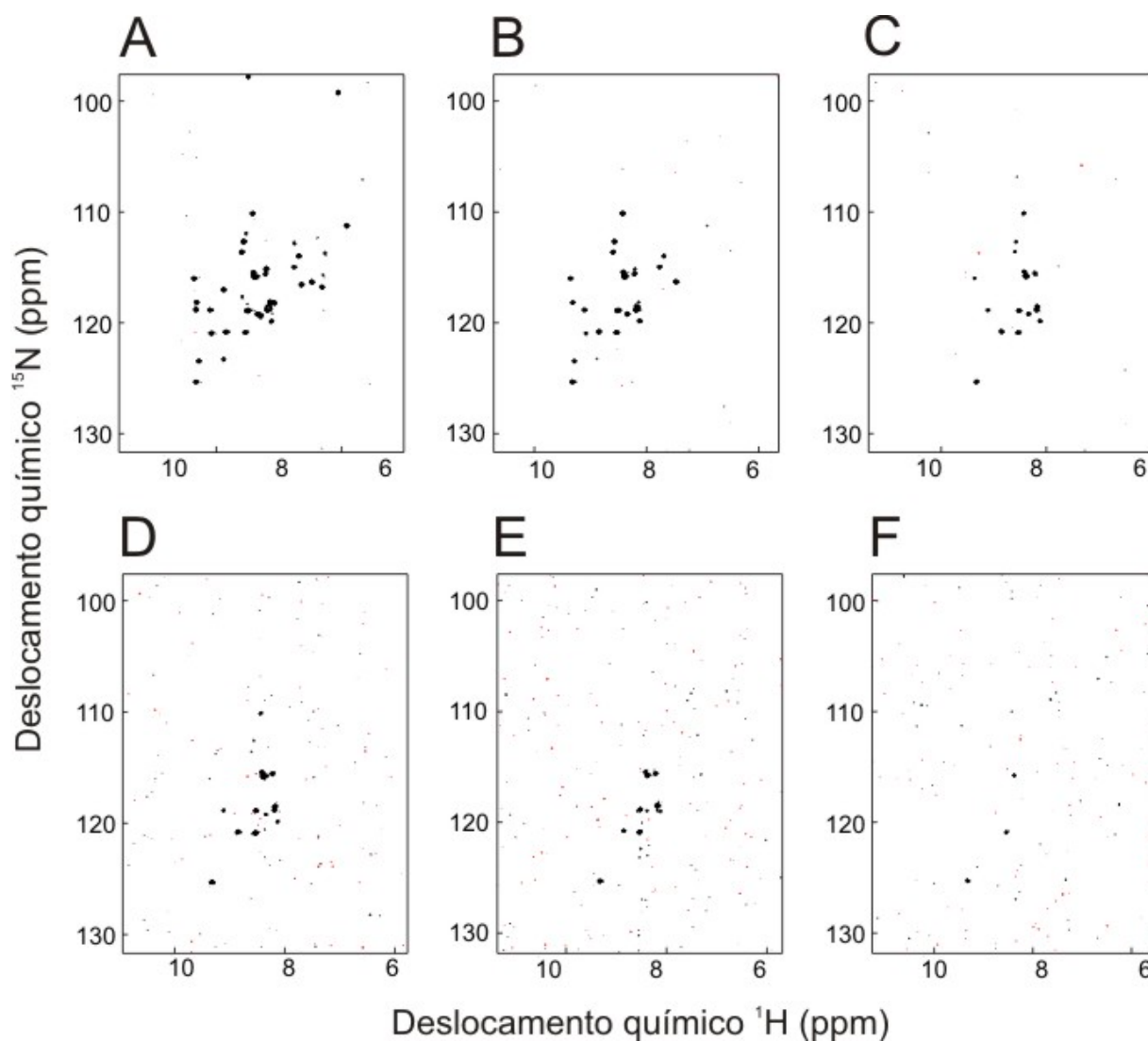
### 3.4.4. Estudos de dinâmica da proteína ClpS de Xac

#### 3.4.4.1. Experimento de troca-próton deutério

A proteção de prótons amídicos contra troca com o solvente está associada com a participação destes prótons em pontes de hidrogênio e com o seu grau de exposição ao solvente (ENGLANDER e KALLENBACH, 1983; SALINAS *et al.*, 2006). Embora não exista nenhum consenso em relação a qual grau de proteção contra troca com o solvente seria indicativo da presença de pontes de hidrogênio, a identificação de velocidades lentas de troca de prótons amídicos, é uma indicação qualitativa da presença de elementos de estrutura secundária (ENGLANDER e KALLENBACH, 1983).

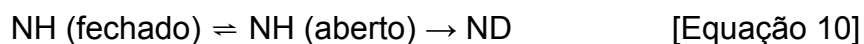
Assim, de forma a obter informações adicionais sobre a estrutura secundária da proteína ClpS em solução e sobre a exposição ao solvente dos seus grupos NH, foi realizado um experimento de troca de próton (H) por deutério (D) utilizando-se uma amostra de ClpS marcada com  $^{15}\text{N}$ . Na Figura 46 é possível observar vários espectros 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC adquiridos em tempos diferentes tomando como referência o momento da adição de  $\text{D}_2\text{O}$  a uma amostra de ClpS marcada com  $^{15}\text{N}$  e previamente liofilizada em tampão contendo água.

Entre o preparo da amostra e o início da coleta dos dados, demorou-se cerca de 20 minutos. Este tempo é razoavelmente curto e permitiu, ao longo da coleta dos vários espectros  $^1\text{H}$ - $^{15}\text{N}$  HSQC seqüenciais, notar gradualmente o desaparecimento dos picos nos espectros. Ao final de onze dias de coleta de dados, verifica-se que praticamente todos os sinais desapareceram, indicando a troca dos prótons amídicos da proteína por átomos de deutério presente em solução.



**Figura 46.** Espectros  $^1\text{H}$ - $^{15}\text{N}$  HSQC adquiridos após diferentes tempos de troca próton-deutério. A. 20 minutos. B. 100 minutos. C. 5 horas. D. 32 horas. E. 64 horas. F. 11 dias. À medida que os prótons amídicos trocam por átomos de deutério em solução o respectivo sinal no espectro  $^1\text{H}$ - $^{15}\text{N}$  HSQC desaparece.

O processo de troca próton-deutério em proteínas pode ser explicado por um modelo de dois-estados onde flutuações estruturais convertem o próton amídico de um estado protegido contra troca (fechado) para um estado favorável à troca (aberto) descrito pela seguinte equação (BAI, 1999; DEMPSEY, 2001):



Uma vez no estado aberto, a reação de troca por deutério é irreversível e ocorre com uma constante de velocidade intrínseca  $k_{\text{int}}$ , que é característica para cada próton amídico presente em uma dada seqüência peptídica em conformação randômica e pode ser facilmente calculada em função do pH, temperatura e amino ácidos vizinhos (BAI *et al.*, 1993).

Assim, o processo de troca próton-deutério em proteínas pode ser facilmente analisado em termos de fatores de proteção contra troca dos prótons amídicos calculados a partir da seguinte equação (BAI *et al.*, 1993):

$$P = (k_{\text{int}}/k_{\text{obs}}) \quad [\text{Equação 11}]$$

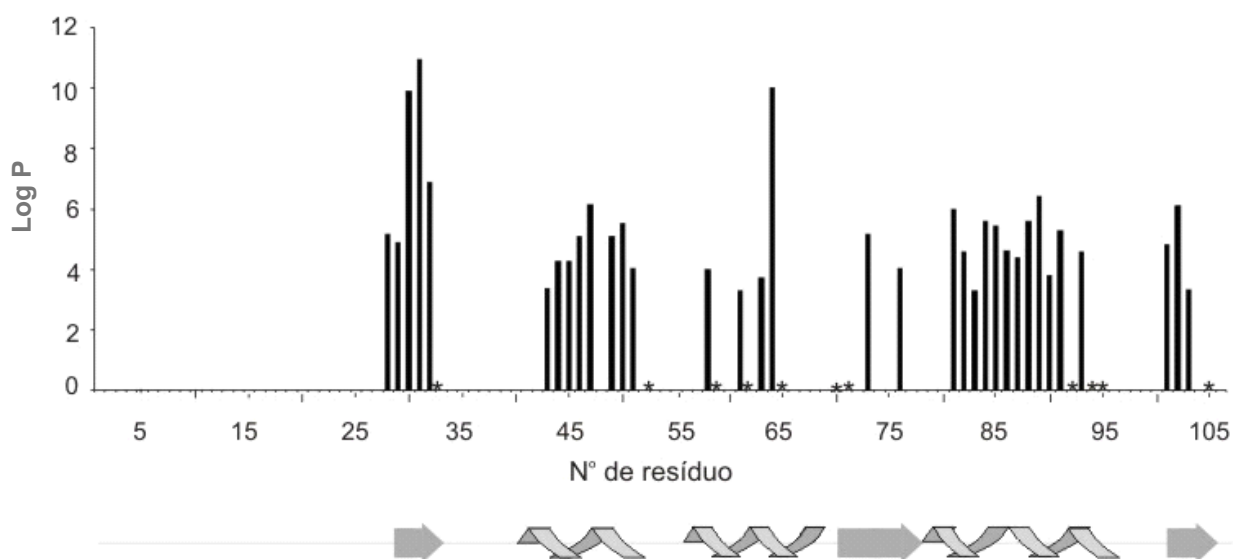
onde  $k_{\text{int}}$  é a constante de velocidade de troca intrínseca e  $k_{\text{obs}}$  é a constante de velocidade de troca observada experimentalmente.

A descrição do processo de troca em termos de fatores de proteção permite eliminar a influência das propriedades da solução (pH, força iônica, temperatura etc) e do ambiente químico dos grupos amídicos (i.e. do contexto da seqüência primária da proteína em estudo). Desta forma os fatores de proteção podem ser correlacionados diretamente com a atenuação relativa da velocidade de troca de um dado próton amídico da cadeia principal na estrutura nativa em relação a uma conformação randômica (BAI *et al.*, 1993; RABAH *et al.*, 2005).

A Figura 47 apresenta uma análise da cinética da troca próton-deutério dos amino ácidos da proteína ClpS expressa como o logaritmo do fator de proteção de troca (log P). Este fator pôde ser quantificado para 34 dos 98 prótons amídicos de ClpS. No caso de onze prótons amídicos, os picos eram visíveis no primeiro espectro  $^1\text{H}$ - $^{15}\text{N}$  HSQC mas a velocidade de troca foi muita rápida e portanto não foi

possível calcular a constante de velocidade de troca. Estes resíduos estão indicados na Figura 47 por asteriscos. Os restantes 53 prótons amídicos trocaram com o solvente muito rapidamente e não foi possível calcular a constante de velocidade de troca.

A Figura 47 mostra que existem regiões da molécula em que a maioria dos fatores de proteção apresenta valores altos (valor médio do  $\log P = 5.33$ ). Estes fatores de proteção observados estão associados com a presença de  $\alpha$ -hélices e fitas- $\beta$ . Três resíduos em particular (V30, L31 e H64) apresentam fatores de proteção extremamente elevados uma vez que mesmo após 11 dias de experimento os prótons amídicos destes resíduos não trocaram completamente com o solvente.



**Figura 47.** Análise da cinética da troca próton-deutério da proteína ClpS.

O logaritmo do fator de proteção de troca ( $\log P$ ) é mostrado em função da seqüência da proteína. Os asteriscos designam prótons amídicos cujos picos eram visíveis no primeiro espectro  $^1\text{H}$ - $^{15}\text{N}$  HSQC mas cuja velocidade de troca foi muito rápida e portanto não foi possível calcular a constante de velocidade de troca.

#### 3.4.4.2. Experimentos de relaxação

O fenômeno de relaxação é caracterizado pelo retorno do sistema ao equilíbrio após uma perturbação exterior. Na espectroscopia de RMN, este processo pode ser medido através dos parâmetros de relaxação e está relacionado com o movimento rotacional da molécula em solução. Além disso, a relaxação está relacionada com os processos de dinâmica da molécula em estudo tais como a mobilidade e flexibilidade estrutural intramolecular (WAGNER, 1995; KAY, 1998; PALMER, 2001; MITTERMAIER e KAY, 2006).

Os experimentos de relaxação envolvem a excitação da magnetização dos spins nucleares através da aplicação de um pulso de rádio frequência e o monitoramento da cinética de retorno ao equilíbrio usando experimentos de RMN multidimensional. Os parâmetros de relaxação medidos habitualmente são os tempos  $T_1$  e  $T_2$  de relaxação de  $^{15}\text{N}$  e os efeitos de NOE heteronuclear ( $^1\text{H}$ - $^{15}\text{N}$  NOE) (BRÜSCHWEILER, 2003).

Os tempos de relaxação  $T_1$  e  $T_2$  descrevem a relaxação do spin.  $T_1$  é chamado de tempo de relaxação longitudinal (ou spin-rede) e descreve o retorno da magnetização longitudinal ao estado de equilíbrio após um pulso de rádio frequência onde os spins nucleares excitados trocam de energia com o meio (rede), fazendo a magnetização voltar ao equilíbrio.  $T_2$  é chamado de tempo de relaxação transversa (ou spin-spin) e refere-se ao retorno da magnetização transversa ao estado de equilíbrio após um pulso de rádio frequência quando os spins nucleares excitados trocam de energia entre si.  $T_2$  é sempre menor ou igual a  $T_1$ . Em particular,  $T_2$  decresce com o aumento do tamanho molecular, o que o torna um fator limitante para a utilização do RMN para o estudo de macromoléculas (KAY, 1998; PALMER, 2001; BRÜSCHWEILER, 2003).



Os valores de  $T_1$  e  $T_2$  são usualmente obtidos coletando espectros de correlação  $^1\text{H}$ - $^{15}\text{N}$  usando tempos de espera de evolução diferentes. As intensidades dos picos são medidas para os diferentes tempos de espera e os tempos de relaxação são obtidos ajustando o decaimento do sinal a uma função exponencial simples de dois parâmetros:

$$I_i(\tau) = I_0 \exp(-\tau/T) \quad \text{[Equação 12]}$$

onde  $I_i(\tau)$  é a intensidade do sinal no tempo  $\tau$ , e  $I_0$  é a intensidade do sinal no tempo  $\tau$  e  $T$  é o tempo de relaxação  $T_1$  ou  $T_2$  (KAY, 1998).

Um terceiro parâmetro de relaxação é o efeito Overhauser nuclear (NOE) heteronuclear ( $^1\text{H}$ - $^{15}\text{N}$ ). Os efeitos de NOE heteronuclear são observados através da saturação do sinal de  $^1\text{H}$  que ocasiona uma alteração da intensidade do sinal de  $^{15}\text{N}$ . São adquiridos dois espectros que correlacionam os núcleos  $^1\text{H}$ - $^{15}\text{N}$ , com e sem saturação prévia do sinal de  $^1\text{H}$ . Considerando que a velocidade com que a relaxação cruzada entre os dois núcleos ocorre aproxima-se do estado estacionário quando a saturação do núcleo  $^1\text{H}$  é longa (BRÜSCHWEILER, 2003), o valor do NOE heteronuclear pode ser calculado a partir da seguinte equação:

$$^1\text{H}\text{-}^{15}\text{N NOE} = I_{\text{sat (H)}} \times I_0 \quad \text{[Equação 13]}$$

onde  $I_{\text{sat (H)}}$  é a intensidade do sinal de  $^{15}\text{N}$  com saturação prévia do sinal de  $^1\text{H}$  e  $I_0$  é a intensidade do sinal de  $^{15}\text{N}$  sem saturação prévia do sinal de  $^1\text{H}$  (KAY, 1998).

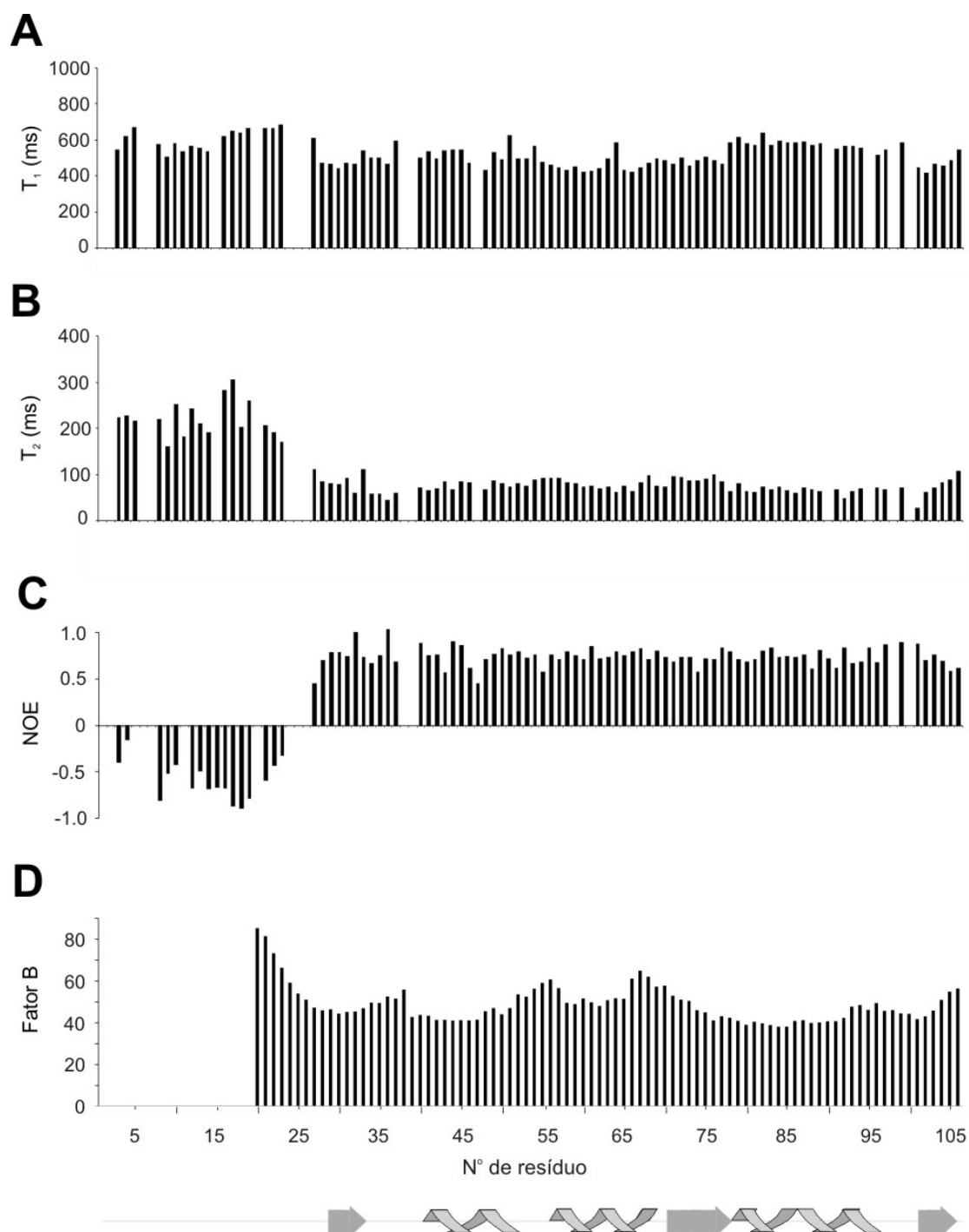
A Figura 48 apresenta os resultados dos experimentos de relaxação de  $^{15}\text{N}$  obtidos para proteína ClpS. Foram obtidos valores de  $T_1$ ,  $T_2$  e  $^1\text{H}$ - $^{15}\text{N}$  NOE para 90, 90 e 92 dos 98 prótons amídicos de ClpS, respectivamente.

Os valores de  $T_1$  apresentam uma variabilidade pequena ao longo da molécula. Porém, é notório um comportamento distinto na região N-terminal da

molécula. Na região compreendida entre os resíduos R27 e A106 (região da molécula que apresenta estrutura secundária definida), os valores de  $T_1$  variam entre 418 e 639 ms e o valor médio é de 510 ms. Já na região compreendida entre os resíduos R3 e A23 os valores de  $T_1$  variam entre 504 e 682 ms e o valor médio é de 569 ms.

Os valores de  $T_2$  apresentam uma variabilidade maior ao longo seqüência de ClpS. A região compreendida entre os resíduos R27 a A106 apresenta valores de  $T_2$  mínimo, máximo e médio de 27, 111 e 75 ms respectivamente. Estes valores são significativamente mais elevados na região compreendida entre os resíduos R3 e A23 que apresenta valores de  $T_2$  mínimo, máximo e médio de 160, 305 e 220 ms respectivamente. A diferença entre a região N-terminal e o restante da proteína é ainda maior quando os valores de  $^1\text{H}$ - $^{15}\text{N}$  NOE são analisados. A região compreendida entre os resíduos R27 e A106 apresenta valores de  $^1\text{H}$ - $^{15}\text{N}$  NOE positivos com valores mínimo, máximo e médio de 0.444, 0.999 e 0.740 respectivamente. A região compreendida entre os resíduos R3 e A23, por seu turno, apresenta valores de  $^1\text{H}$ - $^{15}\text{N}$  NOE negativos com valores mínimo, máximo e médio de -0.888, -0.151 e -0.585 respectivamente.

Os resultados obtidos mostram que os resíduos R27 a A106 apresentam valores de  $T_1$ ,  $T_2$  e  $^1\text{H}$ - $^{15}\text{N}$  NOE de estado estacionário indicativos de rigidez conformacional e pouca mobilidade interna, típicos para regiões estruturadas de proteínas (BRÜSCHWEILER, 2003). Já os resíduos R3 a A23 apresentam valores de  $T_1$ ,  $T_2$  e  $^1\text{H}$ - $^{15}\text{N}$  NOE de estado estacionário típicos de conformações randômicas altamente flexíveis e com elevada mobilidade interna (BRÜSCHWEILER, 2003).



**Figura 48.** Parâmetros de relaxação de  $^{15}\text{N}$  e fator B cristalográfico da proteína ClpS.

Valores de T<sub>1</sub> (A), T<sub>2</sub> (B) e  $^1\text{H}$ - $^{15}\text{N}$  NOE (C) obtidos experimentalmente para a proteína ClpS de *Xac* são plotados em função da seqüência da proteína. (D) Fator B dos átomos N<sup>H</sup> da estrutura da cristalográfica da proteína ClpS de *E. coli* descrita por GUO e colaboradores (2002a) plotados em função da seqüência da proteína.

Na Figura 48 também pode ser visto um gráfico ilustrando o fator B dos átomos  $N^H$  da estrutura da cristalográfica da proteína ClpS de *E. coli* descrita por GUO e colaboradores (1MBX, 2002a). O fator B está relacionado com a flexibilidade molecular a nível atômico e com desordens estáticas e dinâmicas no cristal (PETSKO e RINGE, 1984). A região compreendida entre os resíduos R27 e A106 apresenta uma variação pequena do fator B e o seu valor médio nesta região é de 47. Apenas algumas regiões compreendidas entre os vários elementos de estrutura secundária apresentam um fator B ligeiramente mais elevado. Por outro lado, a região compreendida entre os resíduos P20 e P26 apresenta valores de fator B mais elevado sendo que o seu valor médio para esta região é de 67, confirmando os resultados dos experimentos de relaxação de  $^{15}N$  que indicam que esta região da molécula apresenta mobilidade e flexibilidade elevada. A região compreendida entre os resíduos 1 a 19 não foi observada na estrutura cristalina, e portanto não existem valores de fator B para estes resíduos, o que indica uma flexibilidade ainda maior nesta região. Assim, os valores de fator B da estrutura cristalográfica da proteína ClpS de *E. coli* estão de acordo com os valores de  $T_1$ ,  $T_2$  e  $^1H$ - $^{15}N$  NOE heteronuclear obtidos para a proteína ClpS de *Xac* em solução.

#### **3.4.5. Estudos de interação da proteína ClpS de *Xac* com outras proteínas: resultados preliminares**

A perturbação de deslocamentos químicos é a técnica de RMN mais utilizada para mapear a interação entre duas proteínas. Resumidamente, o espectro  $^1H$ - $^{15}N$  HSQC de uma das proteínas (marcada com  $^{15}N$ ) é monitorado ao se adicionar a outra proteína de interesse (não-marcada) e as variações dos deslocamentos químicos da proteína marcada na ausência e na presença da proteína não-marcada

são analisadas. A interação entre as duas proteínas causa variações de ambiente na interface, afetando assim os deslocamentos químicos dos núcleos nesta área (ZUIDERWEG, 2002).

Inicialmente foi analisada a interação da proteína ClpS de *Xac* com o domínio N-terminal (resíduos 1-143) da proteína ClpA. A região de ClpA a ser estudada (resíduos 1 a 143) foi selecionada com base complexo cristalino da proteína ClpS e o N-terminal da proteína ClpA descrito por GUO e colaboradores (2002a).

Para analisar a interação entre as duas proteínas optou-se por preparar uma amostra contendo 100  $\mu\text{M}$  da proteína ClpS marcada com  $^{15}\text{N}$  e 100  $\mu\text{M}$  do N-terminal da proteína ClpA conforme descrito na secção 3.3.3.8. Esta proporção foi escolhida uma vez que a interação entre ClpS e ClpA foi descrita como ocorrendo na proporção 1:1 (DOUGAN *et al.*, 2002a). Assim, ao adicionarmos ClpA na mesma concentração que ClpS estaríamos analisando a interação de ClpS com ClpA em condições próximas das nativas e que haviam sido utilizadas anteriormente em estudos estruturais por cristalografia (GUO *et al.*, 2002a; ZETH *et al.*, 2002).

Porém, conforme descrito anteriormente ao prepararmos a amostra de ClpA, observou-se uma forte precipitação da amostra. É possível que a precipitação observada se deva ao fato de estarmos trabalhando muito próximo do PI do N-terminal da proteína ClpA, predito como 6.04 ou porque a proteína não permanece solúvel em concentrações elevadas.

Para eliminar esta última hipótese foram testadas outras concentrações do N-terminal da proteína ClpA (16.6, 20, 25, 33.3 e 50  $\mu\text{M}$ ) porém a proteína precipitou em todas as condições testadas. Uma vez que a proteína precipitava em todas as condições, optou-se por analisar os dados referentes à interação entre 100  $\mu\text{M}$  da proteína ClpS marcada com  $^{15}\text{N}$  e aproximadamente 100  $\mu\text{M}$  do N-terminal da

proteína ClpA agregada pois nestas condições haveria uma complexação maior com a proteína ClpS.

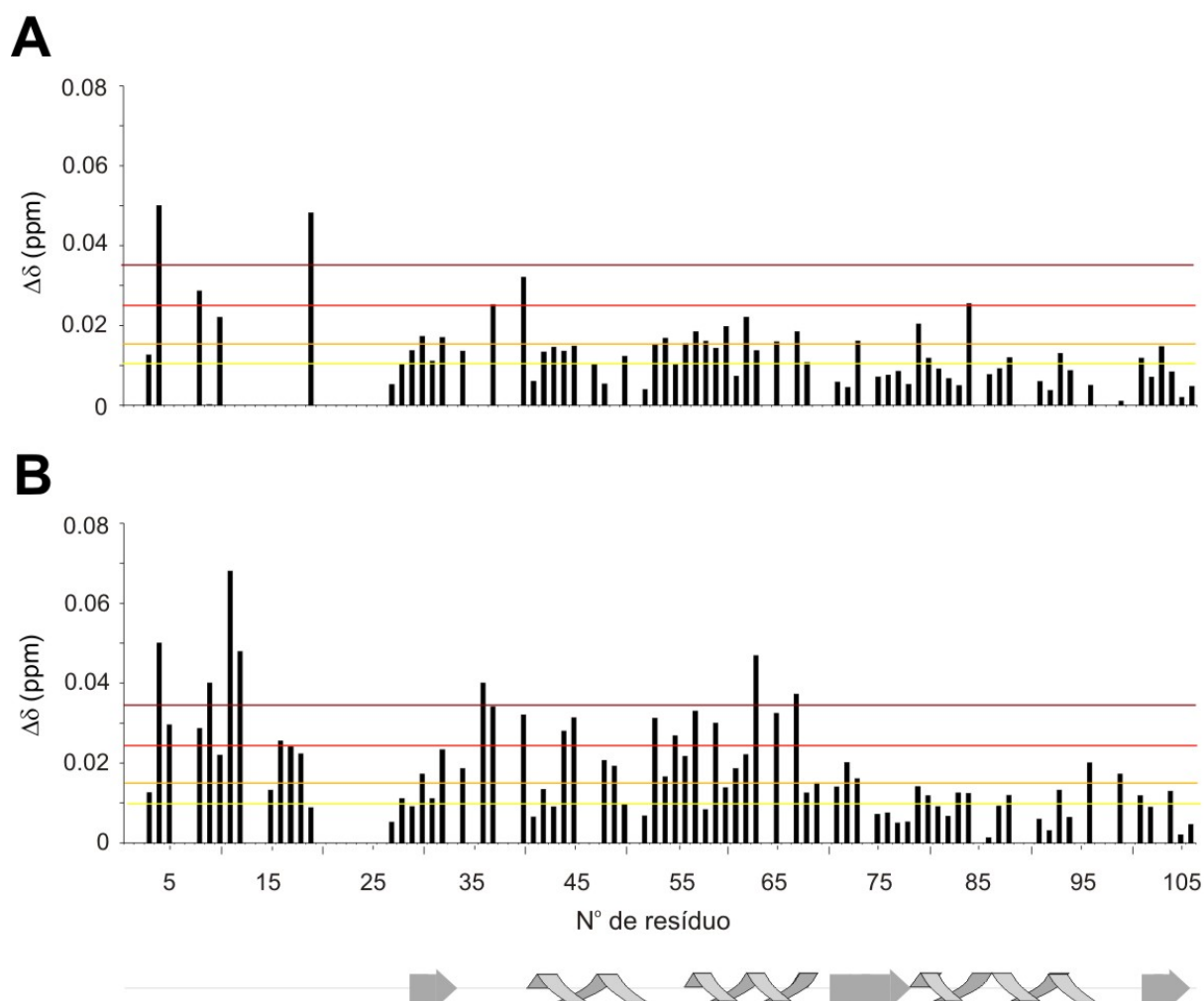
O fato do N-terminal da proteína ClpA se apresentar agregado levou-nos também a considerar a hipótese de que a proteína ClpS estaria interagindo com o N-terminal da proteína ClpA não como um parceiro do complexo ClpAPS, e sim como um substrato agregado. Para testar esta hipótese foi analisada a interação da proteína ClpS de *Xac* com a proteína apo Sm14-M20(C62V), um mutante da proteína Sm14 de *Schistosoma mansoni* (PERTINHEZ *et al.*, 2004), gentilmente cedida pelo Prof. Dr. Nilson Zanchin do Laboratório Nacional de Luz Síncrotron (Campinas, SP). Esta proteína tem um peso molecular similar ao N-terminal da ClpA (14 kDa) e foi analisada em concentração estimada de 100  $\mu$ M e apresentou-se agregada.

A Figura 49 mostra um gráfico ilustrando as variações dos deslocamentos químicos compostos dos grupos amídicos ( $\Delta_{1H+15N}$ ) da proteína ClpS em função da seqüência protéica após a adição do N-terminal da proteína ClpA agregada e da proteína apo Sm14-M20(C62V) agregada.

De um modo geral, os deslocamentos químicos dos grupos amídicos da proteína ClpS apresentam variações nas mesmas regiões na interação com ambas as proteínas. As maiores variações são observadas na região N-terminal da proteína e na região compreendida entre os resíduos D36 e T67, que inclui as hélices  $\alpha$  1 e 2, além de duas regiões sem estrutura secundária definida. Em particular, os resíduos D36 e Y37 apresentaram variações de deslocamento químico significativas. Na interação com a proteína apo Sm14-M20(C62V) agregada, os resíduos D36 e Y37 apresentaram variações de deslocamentos químicos de 0.040 e 0.034

respectivamente. No caso da interação com o domínio N-terminal da proteína ClpA, o sinal de D36 desapareceu enquanto Y37 apresentou uma variação de 0.025.

A fita  $\beta$  1, compreendida entre os resíduos 27 a 33, também apresenta variações de deslocamentos químicos dos grupos amídicos significativas, porém mais fracas.



**Figura 49.** Variações dos deslocamentos químicos dos grupos amídicos ( $\Delta_{1H+15N}$ ) da proteína ClpS em função da seqüência protéica após a adição do N-terminal da proteína ClpA agregado (A) e da proteína apo Sm14-M20(C62V) agregada (B).

As diferentes linhas coloridas ilustram a magnitude da variação do deslocamento químico dos grupos amídicos.  $\Delta_{1H+15N} > 0.01$  (amarelo);  $\Delta_{1H+15N} > 0.015$  (laranja);  $\Delta_{1H+15N} > 0.025$  (vermelho);  $\Delta_{1H+15N} > 0.035$  (vinho).

Apesar das semelhanças observadas na interação da proteína ClpS com o N-terminal da proteína ClpA agregada e com a proteína apo Sm14-M20(C62V) agregada, algumas diferenças importantes devem ser salientadas.

Em primeiro lugar, as variações dos deslocamentos químicos dos grupos amídicos observadas são maiores no caso da interação com a proteína apo Sm14-M20(C62V) sugerindo que a interação de ClpS com esta molécula é mais forte. Para testar esta hipótese seria interessante realizar titulações entre as várias proteínas de forma a estimar a afinidade, a estequiometria, e a cinética de ligação entre as proteínas em estudo. Porém, este tipo de experimento requer que se saiba a concentração protéica em solução com precisão o que não é possível com proteínas agregadas.

Em segundo lugar, no caso da interação com o N-terminal da proteína ClpA vários sinais dos grupos amídicos de ClpS desaparecem do espectro, em especial na região N-terminal onde as variações dos deslocamentos químicos no caso da interação com a proteína apo Sm14-M20(C62V) são maiores. O desaparecimento dos sinais de ClpS no caso da interação com o N-terminal da proteína ClpA pode ser possivelmente explicado pela formação de um complexo em troca intermediária entre os dois estados (ligado e não-ligado) na escala de tempo dos deslocamentos químicos, pela própria agregação da proteína que induz ao alargamento do sinal ou por uma combinação de ambos os fatores conforme sugerido por CICERO e colaboradores (2003).

É também interessante notar que no caso da interação com o N-terminal da proteína ClpA os resíduos E79 e K84 apresentam variações de deslocamento químico bastante significativas (0.021 e 0.026 respectivamente) enquanto que no



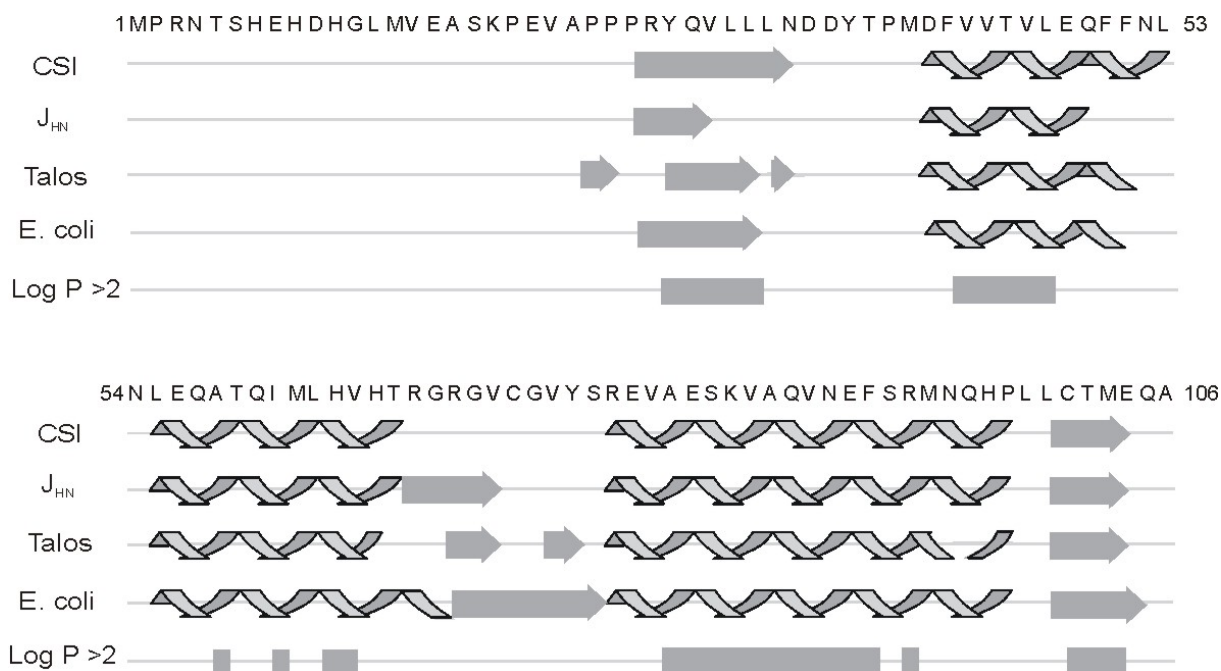
caso da proteína apo Sm14-M20(C62V) estas variações são consideravelmente menores (0.014 e 0.012 respectivamente).

### 3.5. DISCUSSÃO

#### 3.5.1. Estudos estruturais da proteína ClpS de *Xac* por RMN

Os estudos estruturais da proteína ClpS de *Xac* por RMN envolveram várias etapas. Primeiramente foi determinado pela técnica de DOSY que esta proteína se apresenta na forma monomérica nas condições estudadas. Posteriormente, foram utilizados os experimentos de tripla ressonância, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH e HNCO, e de dupla ressonância HCCH-COSY, HCCH-TOCSY,  $^1\text{H}$ - $^{15}\text{N}$ -HSQC-TOCSY,  $^1\text{H}$ - $^{13}\text{C}$  NOESY-HSQC e  $^1\text{H}$ - $^{13}\text{C}$ -HSQC, para assinalar as ressonâncias dos núcleos da cadeia principal e das cadeias laterais, respectivamente. Usando esta metodologia foi possível atribuir as ressonâncias de aproximadamente 98% dos núcleos da cadeia principal e 90% dos núcleos das cadeias laterais alifáticas de ClpS. É interessante notar, que os resíduos para os quais não foi possível atribuir algumas das ressonâncias da cadeia principal encontram-se, sem exceções, em regiões não estruturadas da proteína tais como a região N-terminal (resíduos M1, P2, T5, K19, P20, A23, P24 e P25) e regiões que não apresentam estrutura secundária definida (Y37, T38 e L100) (vide Tabela 15). Este tipo de observação é comum em estudos estruturais de proteínas por RMN uma vez que geralmente, quanto mais flexível a região em estudo, maior a sobreposição dos sinais de RMN e conseqüentemente mais difícil se torna o seu assinalamento. Além disso, é também comum que estas regiões apresentem troca conformacional rápida o que por vezes não permite a observação dos respectivos sinais de RMN nos espectros coletados.

O assinalamento das ressonâncias dos núcleos da cadeia principal da proteína ClpS permitiu, entre outras coisas, a identificação dos elementos de secundária desta proteína.



**Figura 50.** Elementos de estrutura secundária da proteína ClpS de *Xac*.

A seqüência primária da proteína está representada no topo da figura. Abaixo do gráfico estão representados por setas (fita- $\beta$ ) e hélices ( $\alpha$ -hélices) os elementos de estrutura secundária da proteína ClpS de *Xac* preditos pelo cálculo do CSI, pelo valor da constante de acoplamento  $^3J_{HNH\alpha}$  e pela predição dos ângulos diedros  $\Phi$  e  $\Psi$  usando o programa TALOS. São também apresentados os elementos de estrutura secundária presentes na proteína ClpS de *E. coli*. Na última linha estão representados por retângulos os resíduos cujo logaritmo do fator de proteção de troca do núcleo HN é superior a 2 ( $\log P > 2$ ).

A Figura 50 resume os elementos de estrutura secundária de ClpS de *Xac* preditos a partir dos deslocamentos químicos obtidos experimentalmente. Conforme pode ser observado, as predições obtidas usando três métodos distintos - CSI, constante de acoplamento  $^3J_{HNH\alpha}$  e predição dos ângulos diedros  $\Phi$  e  $\Psi$  usando o programa TALOS – são consistentes entre si. Além disso, os resultados obtidos coincidem também com os elementos de estrutura secundária descritos na literatura para a proteína ClpS de *E. coli* (GUO *et al.*, 2002a; ZETH *et al.*, 2002).

A principal variação entre as várias predições encontra-se na região de R70 a S77. Conforme mencionado anteriormente, na proteína ClpS de *E.coli*, esta região foi descrita como uma fita- $\beta$  que contém uma dobradura (“kink”) na região do resíduo C73. Por apresentar esta característica especial, os vários métodos usados para prever a estrutura secundária de ClpS de *Xac* nesta região apresentam resultados ligeiramente diferentes. Porém, em todos eles, fica evidente a predisposição para a formação de uma fita- $\beta$  nesta região, incluindo a respectiva dobradura na região do resíduo C73.

A outra discrepância observada entre os três métodos de identificação de estrutura secundária utilizados foi no caso da predição da região P24-N34 usando o programa TALOS. Conforme esperado, este método prediz a presença de estruturas de fita- $\beta$  nesta região. Porém, a posição exata destas estruturas não está bem definida. Esta ambigüidade na predição deve-se essencialmente a dois fatores. Em primeiro lugar, esta fita- $\beta$  encontra-se muito próxima da região N-terminal da proteína, região que apresenta alta flexibilidade (vide resultados de dinâmica na seção 3.4.4) e para a qual a estrutura cristalina não foi determinada na proteína homóloga de *E. coli*.

O fato de estarmos perante uma região mais flexível da molécula dificulta o uso do programa TALOS que se baseia justamente na predição de ângulos diedros baseado numa base de dados de estruturas de proteínas com estrutura tridimensional definida. Em segundo lugar, o método TALOS, apesar de rápido e eficaz, na média permite apresentar predições para apenas aproximadamente 65% dos resíduos de uma dada proteína. Além disso, o método apresenta cerca de 3% de erros em suas predições (CORNILESCU *et al.*, 1999).

Quando a estrutura cristalina de ClpS de *E.coli* foi publicada em 2002 por dois grupos independentes (GUO *et al.*, 2002a; ZETH *et al.*, 2002) sem conter informações estruturais sobre a região N-terminal da proteína, um dos principais objetivos deste trabalho passou a ser o de procurar obter informações estruturais por RMN desta região da proteína homóloga de *Xac*. Nenhum dos métodos utilizados conseguiu detectar evidências da presença de elementos de estrutura secundária na região N-terminal (resíduos M1 a P23), da proteína ClpS de *Xac*. A ausência de uma predição de estrutura nesta região, foi indicativo de que, conforme sugerido por GUO *et al.* e ZETH *et al.* em 2002, esta região da proteína provavelmente não apresentaria uma estrutura definida.

Esta suspeita foi confirmada pela análise do espectro 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY da proteína ClpS de *Xac*. A região N-terminal da proteína ClpS de *Xac* apresenta apenas sinais de NOE intra-residuais. Sinais de NOE inter-residuais de média distância, típicos de estruturas secundárias regulares, não são observados indicando que a proteína ClpS de *Xac*, tal como anteriormente descrito para a proteína ClpS de *E. coli* se apresenta uma conformação randômica nesta região.

Ao entrarmos na era da genômica e com o número crescente de projetos de proteômica estrutural é de se esperar que o uso de métodos que avaliem rapidamente modelos estruturais de proteínas que possuem homologia com proteínas cuja estrutura é conhecida aumente significativamente. A validação de modelos estruturais por acoplamentos dipolares determinados experimentalmente está de acordo com as mais recentes técnicas de biologia estrutural e promete revolucionar vários projetos de proteômica estrutural. Vários grupos têm demonstrado a utilidade de usar acoplamentos dipolares residuais para validação de estruturas (ANNILA *et al.*, 1999; MEILER *et al.*, 2000; CHOU *et al.*, 2000; BEWLEY e

CLORE, 2000). Em particular BEWLEY (2001) demonstrou que é possível validar modelos estruturais rapidamente e com precisão usando apenas um homólogo estrutural e valores de acoplamentos dipolares residuais  $D_{HN}$  para menos de um terço da molécula.

O modelo estrutural da proteína ClpS de *Xac* proposto neste trabalho, foi validado fazendo uso do homólogo ClpS de *E. coli* e de acoplamentos dipolares residuais  $D_{NH}$  para 56% da molécula obtidos experimentalmente. Esta abordagem possibilitou a obtenção e validação de um modelo estrutural de forma rápida e eficaz, imediatamente após o assinalamento das ressonâncias dos núcleos da cadeia principal da proteína ClpS.

Além disso, uma vez que o modelo estrutural de ClpS de *Xac* aqui proposto é muito semelhante à estrutura tridimensional de ClpS de *E. coli* determinada por cristalografia de raios-X, os resultados obtidos permitem responder alguns dos questionamentos que nos propusemos estudar neste trabalho:

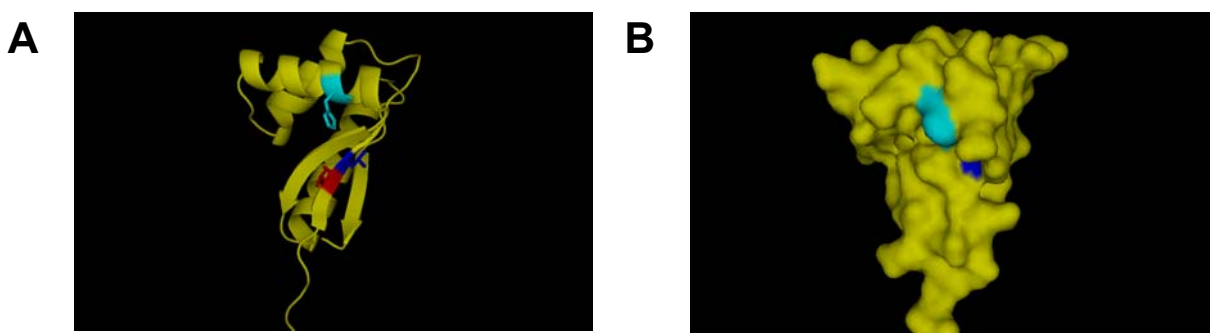
- a) Esta proteína apresenta uma elevada conservação estrutural o que evidencia a sua importância biológica em células procarióticas.
- b) A estrutura de ClpS isolada é semelhante à estrutura desta proteína em complexo com o domínio N-terminal da proteína ClpA, sugerindo que a modulação da atividade de ClpA por ClpS não envolve uma alteração estrutural significativa de ClpS.
- c) A estrutura de ClpS em solução é semelhante à estrutura cristalina sugerindo que a superfície de contato entre a proteína ClpS e o N-terminal da proteína ClpA observado nos cristais, também está provavelmente ocorrendo em solução.

- d) A região N-terminal de ClpS apresenta-se desestruturada em solução conforme proposto por GUO e colaboradores (2002a) e ZETH e colaboradores (2002).

### 3.5.2. Estudos de dinâmica da proteína ClpS de *Xac*

Estudos de troca-próton deutério da proteína ClpS de *Xac* permitiram avaliar a proteção dos seus prótons amídicos contra troca com o solvente. Uma vez que esta proteção está diretamente associada com a participação destes prótons em pontes de hidrogênio (WAGNER e WUTRICH, 1983; ENGLANDER e KALLENBACH, 1983) estes estudos permitiram uma avaliação qualitativa da presença de elementos de estrutura secundária na cadeia polipeptídica de ClpS. A proteção elevada dos prótons amídicos da região Y38 a M103 da proteína ClpS de *Xac* observada experimentalmente está de acordo com a estrutura homóloga de *E. coli* descrita na literatura (GUO *et al.*, 2002a; ZETH *et al.*, 2002) e sugere que esta proteína apresenta um núcleo estrutural de elevada estabilidade. Estas regiões de maior proteção correspondem também às regiões preditas como contendo elementos de estrutura secundária usando os métodos de CSI, cálculo da constante de acoplamento  $^3J_{\text{HNH}\alpha}$  e predição dos ângulos diedros  $\Phi$  e  $\Psi$  usando o programa TALOS (Figura 50). Em particular os resíduos D35, D36, M40, L63, V65, T67, E79 e K84 apresentam uma proteção elevada e todos estão localizados em regiões estruturadas da proteína. Dois destes resíduos, E79 e K84, foram descritos como estando envolvidos na interação com ClpA (GUO *et al.*, 2002a; ZETH *et al.*, 2002). Já os resíduos D35 e D36 foram descritos por ERBSE e colaboradores (2006) como necessários para a interação de ClpS com substratos da via de degradação da regra do N-terminal (vide discussão mais detalhada do papel destes resíduos na ligação

ao substrato na seção que se segue). Vale salientar que os três resíduos que apresentam fatores de proteção extremamente elevados e não trocaram completamente com o solvente (V30, L31 e H64) encontram-se quase que completamente enterrados no interior hidrofóbico da proteína (Figura 51). Conforme esperado, a região N-terminal da proteína não apresenta qualquer proteção à troca hidrogênio-deutério, característico de conformações randômicas.



**Figura 51.** Mapeamento dos resíduos V30 (vermelho), L31 (azul escuro) e H64 (azul claro) no modelo tridimensional da proteína ClpS.

Por outro lado, na espectroscopia de RMN, o movimento rotacional e os processos dinâmicos internos de uma molécula refletem-se no comportamento de relaxação da molécula (WAGNER, 1995; KAY, 1998; PALMER, 2001). Assim, ao estudar o fenômeno de relaxação de  $^{15}\text{N}$  da proteína ClpS de Xac foi possível analisar a dinâmica dos núcleos  $^{15}\text{N}$  da cadeia principal. Os dados obtidos revelaram que a região compreendida entre os resíduos R27 e A106 apresenta padrões de relaxação compatível com regiões estruturadas de proteínas. Já a região compreendida entre os resíduos R3 e A23 apresenta padrões de relaxação típicos de regiões bastante flexíveis.



Os resultados obtidos nos experimentos de troca-próton de deutério e nos experimentos de relaxação foram particularmente importantes para confirmar o teor altamente flexível da região N-terminal da proteína ClpS.

### 3.5.3. Estudos de interação da proteína ClpS de *Xac* com outras proteínas

Até recentemente a espectroscopia de RMN era talvez melhor conhecida como um método para determinar estruturas tridimensionais de macromoléculas biológicas em solução. Atualmente, no entanto, o estudo de interações proteína-proteína por RMN promete vir a ser a sua aplicação mais freqüente (ZUIDERWEG, 2002; BONVIN *et al.*, 2005).

Os resultados obtidos nos estudos de interação da proteína ClpS de *Xac* com outras proteínas são ainda preliminares e difíceis de interpretar detalhadamente. No entanto, é possível interpretar algumas das alterações observadas à luz da literatura atual. A Figura 52 mostra o mapeamento das variações dos deslocamentos químicos dos grupos amídicos ( $\Delta_{1H+15N}$ ) da proteína ClpS no modelo tridimensional da proteína após a adição do N-terminal da proteína ClpA agregada (Figura 52 A, C e E) e da proteína apo Sm14-M20(C62V) agregada (Figura 52 B, D e F). Em ambos os casos variações dos deslocamentos químicos podem ser mapeadas na lateral (Figura 52 A e B) e no topo (Figura 52 E e F) do cone formado pela proteína ClpS. Em particular, os resíduos localizados na primeira  $\alpha$ -hélice apresentam alterações significativas nos deslocamentos químicos.

É importante lembrar que o modelo da proteína ClpS inclui apenas a região compreendida entre os resíduos P20 a A106. Assim, não foi possível mapear no modelo tridimensional da proteína as fortes variações de deslocamento químico

observadas nos resíduos R3 a K19 na interação de ClpS com o N-terminal da proteína ClpA e com a proteína apo Sm14-M20(C62V).

É também interessante notar que as superfícies de interação da proteína ClpS com proteínas agregadas propostas acima estão distantes da superfície de interação com a proteína ClpA descrita na literatura compreendida pelos resíduos altamente conservados E79, E82 e K84 (ZETH *et al.*, 2002) (ilustrados em verde na Figura 51 C e D).

Os nossos resultados mostram que no caso da interação de ClpS com o N-terminal da proteína ClpA, os resíduos E79 e K84 apresentam variações de deslocamento químico bastante significativas (Figura 49). Já no caso da interação de ClpS com a proteína apo Sm14-M20(C62V) estas variações são consideravelmente menores. Estes resultados estão de acordo com o trabalho de ZETH e colaboradores (2002) que demonstraram através do estudo de mutantes de ClpS que os resíduos E79 e K84 são essenciais para a formação do complexo ClpA-ClpS em *E. coli*.

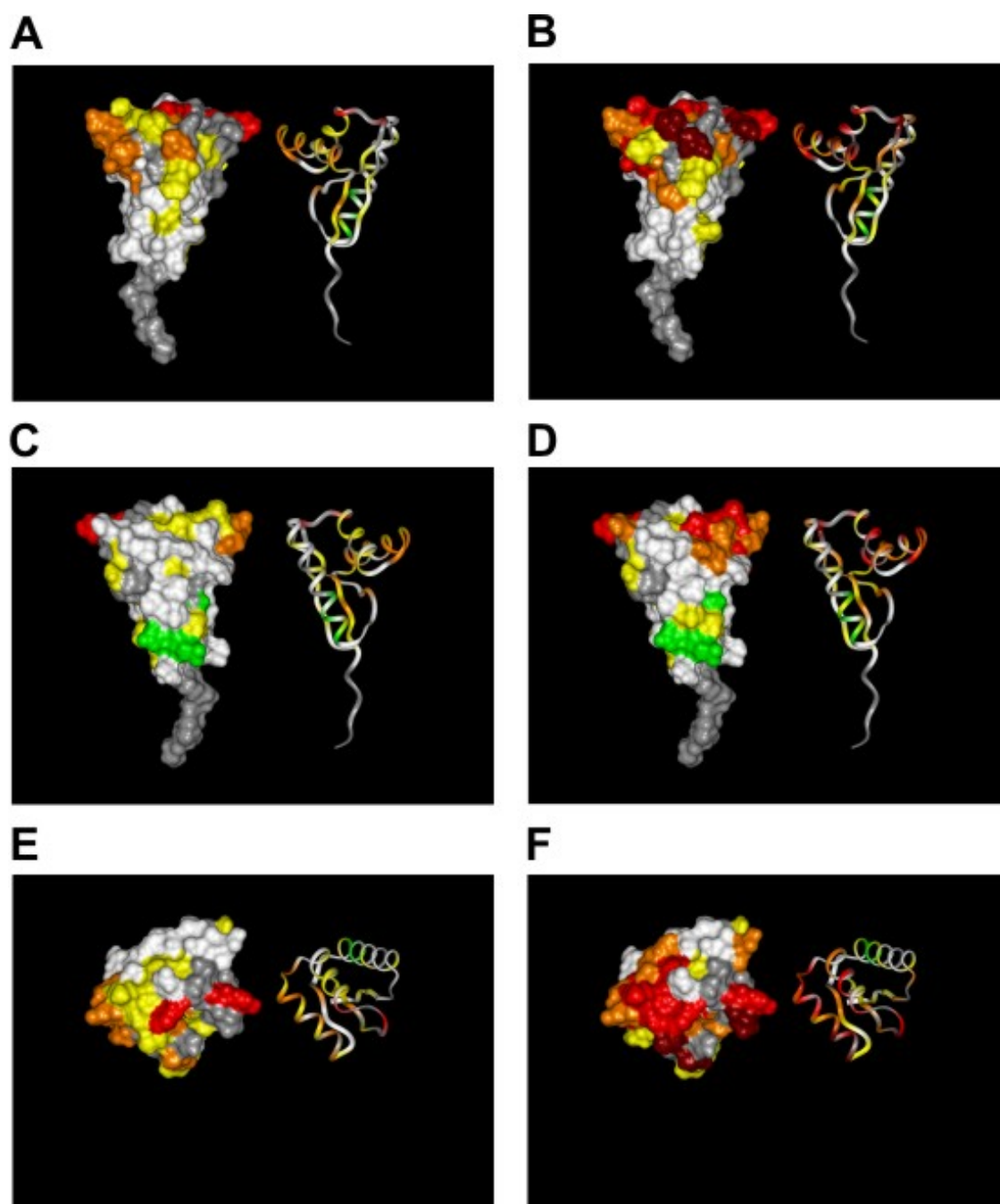
É importante lembrar no entanto, que a interação da proteína ClpS com o N-terminal de ClpA foi analisada em condições nas quais a proteína ClpA se encontrava agregada. Assim, é possível que parte das moléculas de ClpA presentes na amostra estivessem interagindo com ClpS como sua parceira na formação do complexo ClpS-ClpA enquanto as restantes moléculas de ClpA estariam interagindo com ClpS como um substrato agregado. De fato, na interação com o N-terminal de ClpA, e especialmente na interação com a proteína apo Sm14-M20(C62V), os resíduos D36 e Y37 de ClpS apresentaram alterações de deslocamento químico muito significativas. Além disso, na interação com ambas as proteínas os sinais do

resíduo D35 desaparecem sugerindo uma forte alteração do ambiente químico deste resíduo (Figura 49).

Estes resultados estão de acordo com um estudo recente publicado na revista *Nature* que sugere que os resíduos altamente conservados D35, D36 e Y37 e E41 de ClpS estão envolvidos na interação com substratos da via de degradação da regra do N-terminal (“N-end rule”) em *E. coli* (ERBSE *et al.*, 2006). Os resíduos D35, D36, Y37 e E41 estão localizados em uma região conservada da proteína ClpS, localizada na superfície da proteína numa posição distante da superfície de interação ClpS-ClpA compreendida pelos resíduos altamente conservados E79, E82 e K84.

Para investigar o papel destes resíduos na degradação de substratos via de degradação da regra do N-terminal, ERBSE e colaboradores (2006) construíram três mutantes de ClpS, Y37A, E41A e o mutante duplo D35A/D36A e testaram a sua capacidade de mediar a degradação de substratos da via N-terminal *in vitro*. Estes mutantes interagiram da forma esperada com ClpA porém a sua habilidade para mediar a degradação de substratos da regra do N-terminal foi afetada. A habilidade dos mutantes Y37A e E41A de mediar a degradação destes substratos foi parcialmente afetada enquanto o mutante duplo D35A/D36A não pode mediar a degradação destes substratos *in vitro* (ERBSE *et al.*, 2006).

Os autores deste trabalho especularam que a interação do grupo amino do resíduo N-terminal do substrato com as cargas negativas dos resíduos D35, D36 e E41, leva à ligação do substrato da regra do N-terminal com ClpS substrato enquanto a cadeia lateral hidrofóbica de Y37 será necessária para o posicionamento correto do N-terminal do substrato.



**Figura 52.** Mapeamento das variações dos deslocamentos químicos compostos dos grupos amídicos ( $\Delta\delta_{1H+15N}$ ) da proteína ClpS no modelo tridimensional da proteína após a adição do N-terminal da proteína ClpA agregado (A, C, E) e da proteína apo Sm14-M20(C62V) agregada (B, D, F). As diferentes cores ilustram a magnitude da variação do deslocamento químico dos grupos amídicos.  $\Delta_{1H+15N} = 0-0.01$  (branco);  $\Delta_{1H+15N} = 0.01-0.015$  (amarelo);  $\Delta_{1H+15N} = 0.015-0.025$  (laranja);  $\Delta_{1H+15N} = 0.015-0.025$  (vermelho);  $\Delta_{1H+15N} > 0.025$  (vinho). Resíduos não assinalados ou cujo sinal amídico não apareceu no espectro de interação estão representados em cinza. Em verde estão representados os resíduos E79, E82 e K84 descritos na literatura como estando envolvidos na interação com ClpA (GUO *et al.*, 2002a; ZETH *et al.*, 2002).

Os resultados obtidos em nosso estudo confirmam a suspeita de que os resíduos D36 e Y37 estão envolvidos na interação de ClpS com seus substratos uma vez que tanto na interação com o N-terminal de ClpA quanto na interação com a proteína apo Sm14-M20(C62V) estes resíduos apresentam variações de deslocamento químico muito significativas. Em particular, os resultados obtidos sugerem que estes resíduos podem estar possivelmente envolvidos na interação da proteína ClpS com substratos protéicos agregados independentemente do teor dos resíduos N-terminais dos mesmos e dos substratos serem ou não membros da via de degradação da regra do N-terminal.

Além disso, os nossos resultados sugerem ainda que toda a região N-terminal da proteína ClpS pode ter um papel importante na interação com substratos agregados. Na interação de ClpS com o N-terminal da proteína ClpA vários sinais dos grupos amídicos de resíduos N-terminais de ClpS desaparecem do espectro enquanto que no caso da interação com a proteína apo Sm14-M20(C62V) as variações dos deslocamentos químicos observadas para os resíduos N-terminais de ClpS são muito significativas. Estes resultados sugerem que a região N-terminal de ClpS pode estar envolvida no reconhecimento e ligação de substratos agregados e que a superfície de interação da proteína ClpS com proteínas agregadas estaria distante da superfície de interação de ClpS com a proteína ClpA compreendida pelos resíduos altamente conservados E79, E82 e K84 (ZETH *et al.*, 2002) (Figura 52).

Em estudos futuros será interessante analisar a interação da proteína ClpS com outras proteínas agregadas de forma a confirmar e mapear detalhadamente a superfície de interação de ClpS com substratos agregados. Além disso, será interessante analisar a interação de ClpS com ClpA e uma proteína agregada simultaneamente. É possível que numa análise deste tipo se observem interações

que não estão sendo observadas nos pares isolados ClpS-ClpA e ClpS-proteína agregada.

Caso seja possível encontrar condições em que N-terminal da proteína ClpA se mantém solúvel em altas concentrações, será também interessante realizar titulações onde as variações dos deslocamentos químicos são mapeadas em função da quantidade de proteína adicionada. Este tipo de estudo permitiria, além de mapear melhor a interface de interação entre ClpS e ClpA, estimar a afinidade, a estequiometria, a especificidade e a cinética de ligação entre as proteínas em estudo (ZUIDERWEG, 2002).

## **CAPÍTULO 4**

### **CONCLUSÕES**

Os resultados apresentados neste trabalho constituem exemplos claros da vasta aplicabilidade da técnica de RMN para o estudo de proteínas e de como esta técnica se pode tornar uma poderosa ferramenta em projetos de proteômica estrutural.

Vários grupos já haviam descrito que o principal valor da RMN em projetos de proteômica estrutural seria na caracterização de alvos para estudos estruturais por RMN e/ou cristalografia (MONTELLIONE *et al.*, 2000; PRESTGARD *et al.*, 2001; REHM *et al.*, 2002; AL-HASHIMI e PATEL, 2002.; YOKOYAMA, 2002). No Capítulo 2 foi apresentado um novo protocolo de triagem por RMN de proteínas alvo de projetos de proteômica estrutural. Este protocolo faz uso do antibiótico rifampicina e permite que as proteínas de interesse sejam analisadas sem a sua purificação prévia (GALVÃO-BOTTON *et al.*, 2003). O uso deste protocolo mostrou-se eficaz na seleção de alvos protéicos de *Xac* adequados para estudos estruturais futuros por RMN e/ou cristalografia de raios X (KATSUYAMA *et al.*, 2004; GUZZO *et al.*, 2005; GUZZO *et al.*, 2006; GUZZO *et al.*, 2007; CICERO *et al.*, 2007) e foi descrito por vários grupos como uma técnica promissora a ser utilizada em projetos futuros de proteômica estrutural de larga escala (MOREAU *et al.*, 2006; ETEZADY-ESFARJANI *et al.*, 2006).

No Capítulo 3 foram apresentados diversos estudos por RMN da proteína ClpS de *Xac*, uma das proteínas selecionadas para estudos estruturais pela nossa metodologia de triagem. O assinalamento das ressonâncias da cadeia principal e das cadeias laterais desta proteína usando experimentos de tripla ressonância e dados de dinâmica e de troca H/D forneceram informações sobre a sua estrutura secundária. Os elementos de estrutura secundária preditos estão de acordo com a estrutura descrita para o homólogo de *E. coli* (GUO *et al.*, 2002a; ZETH *et al.*, 2002).



Todos os dados RMN sugerem que a região N-terminal de ClpS de *Xac* se apresenta desestruturada. Um modelo tridimensional da proteína ClpS de *Xac* calculado por modelagem por homologia a partir da estrutura tridimensional do homólogo de *E. coli* (GUO *et al.*, 2002a) e validado a partir de acoplamentos dipolares residuais  $D_{NH}$  obtidos experimentalmente foi também proposto. Validações de modelos estruturais por acoplamentos dipolares residuais deste tipo terão um papel crescente em projetos de proteômica estrutural à medida em cada vez mais proteínas terão o seu modelo tridimensional determinado por modelagem por homologia. Neste trabalho foram também apresentados diversos estudos de dinâmica da proteína ClpS de *Xac* e estudos de interação da proteína ClpS de *Xac* com outras proteínas o que contribuiu para a elucidação da relação da estrutura tridimensional da proteína ClpS com a sua função biológica.

## REFERÊNCIAS

- AIZENMAN, E., ENGLEBERG-KULKA, H., GLASER, G. (1996). An *Escherichia coli* chromosomal "addiction module" regulated by 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc. Natl. Acad. Sci. U.S.A.*, v.93, p.6059-6063.
- AL-HASHIMI, H.M., PATEL, D.J. (2002). Residual dipolar couplings: Synergy between NMR and structural genomics. *J. Biomol. NMR*, v.22, p.1-8.
- ALMEIDA, F.C.L., AMORIM, G.C., MOREAU, V.H., SOU.S.A., V.O., CREAZOLA, A.T., AMÉRICO, T.A., PAIS, A.P.N., LEITE, A., NETTO, L.E.S., GIORDANO, R.J., VALENTE, A.P. (2001). Selectively labelling the heterologous protein in *Escherichia coli* for NMR studies: a strategy to speed up NMR spectroscopy. *J. Magn. Reson.*, v.148, p.142-146.
- ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W., LIPMAN, D.J. (1990). Basic local alignment search tool. *J. Molec. Biol.*, v.215, p.403-410.
- ALTSCHUL, S.F., MADDEN, T.L., SCHAFFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W.E., LIPMAN, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, v.25, p.3389-3402.
- ANNILA, A., AITIO, H., THULIN E., DRAKENBERG, T. (1999). Recognition of protein folds via dipolar couplings. *J. Biomol. NMR*, v.14, p.223-230.
- ARROWSMITH, C.H., WU, Y.S. (1998). NMR of large (>25 kDa) proteins and protein complexes. *Progress in NMR Spectroscopy*, v.32, p.277-286.
- BAI, Y. (1999). Equilibrium amide hydrogen exchange and protein folding kinetics. *J. Biomol. NMR*, v.15, p.65-70.
- BAI, Y., MILNE, J.S., MAYNE, L., ENGLANDER, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins*, v. 17, p.75-86.
- BAKER, D., SALI, A. (2001). Protein structure prediction and structural genomics. *Science*, v.294, p.93-96.
- BARBATO, G., IKURA, M., KAY, L.E., PASTOR, R.W., BAX, A. (1992). Backbone dynamics of calmodulin studied by <sup>15</sup>N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, v.31, p.5269-5278.
- BARNETT, M.E., ZOLKIEWSKA, A., ZOLKIEWSKI, M. (2000). Structure and Activity of ClpB from *Escherichia coli*. *J. Biol. Chem.*, v.275, p.37565-37571.
- BARTLAM, M., XU, Y., RAO, Z. (2007). Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases. *J. Struct. Funct. Genomics*. (Publicação eletrônica em 7 de agosto de 2007).
- BATEMAN, A., BIRNEY, E., DURBIN, R., EDDY, S.R., HOWE, K.L., SONNHAMMER E.L. (2000). The Pfam protein families database. *Nucleic Acids Res.*, v.28, p.263-266.
- BAX, A. (2003). Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.*, v.12, p.1-16.
- BEN-ZVI, A.P., GOLOUBINOL, P. (2001). Review: Mechanisms of Disaggregation and Refolding of Stable Protein Aggregates by Molecular Chaperones. *J. Struct. Biol.*, v.135, p.84-93.
- BERAUD, S., BERSCH, B., BRUTSCHER, B., GANS, P., BARRAS, F., BLACKLEDGE, M. (2002). Direct structure determination using residual dipolar couplings: reaction-site conformation of methionine sulfoxide reductase in solution. *J. Am. Chem. Soc.*, v.124, p.13709-13715.
- BERMAN, H.M., HENRICK, K., NAKAMURA, H. (2003). Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, v.10, p.980. *In* <http://www.pdb.org>.

- BEURON, F., MAURIZI, M.R., BELNAP, D.M., KOCSIS, E., BOOY, F.P., KESSEL, M., STEVEN, A.C. (1998). At Sixes and Sevens: Characterization of the Symmetry Mismatch of the ClpAP Chaperone-Assisted Protease. *J. Struct. Biol.*, v.123, p.248-259.
- BEVAN, M. (2000). The bugs from Brazil. *Nature*, v.406, p.140-141.
- BEWLEY, C.A. (2001). Rapid validation of the overall structure of an internal domain-swapped mutant of the anti-HIV protein cyanovirin-N using residual dipolar couplings. *J. Am. Chem. Soc.*, v.123, p.1014-1015.
- BEWLEY, C.A., CLORE, G.M. (2000). Determination of the Relative Orientation of the Two Halves of the Domain-Swapped Dimer of Cyanovirin-N in Solution Using Dipolar Couplings and Rigid Body Minimization. *J. Am. Chem. Soc.*, v.122, p.6009-6016.
- BEWLEY, M.C., GRAZIANO, V., GRIFFIN, K., FLANAGAN, J.M. (2006). The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. *J. Struct. Biol.*, v.153, p. 113-128.
- BHATTACHARYYA S., HABIBI-NAZHAD B., AMEGBEY G., SLUPSKY C.M., YEE A., ARROWSMITH C., WISHART D.S. (2002). Identification of a novel archaeobacterial thioredoxin: determination of function through structure. *Biochemistry*, v.41, p.4760-4770.
- BHAVESH, N.S., PANCHAL, S.C., & HOSUR, R.V. (2001). An efficient high-throughput resonance assignment procedure for structural genomics and protein folding research by NMR. *Biochemistry*, v.40, p.14727-14735.
- BIOLOGICAL MAGNETIC RESONANCE BANK (2001, 2007). *In* [www.bmrb.wisc.edu/](http://www.bmrb.wisc.edu/)
- BOETTNER, M., PRINZ, B., STAHL, U., LANG, C. (2002). High-throughput screening for expression of heterologous proteins in the yeast *Pichia pastoris*. *J. Biotechnol.*, v.99, p.51-62.
- BONVIN, A.M.J.J., BOELEN, R., KAPTEIN, R. (2005). NMR analysis of protein interactions. *Curr. Opin. Chem. Biol.*, v.9, p.501-508.
- BOVE, J.M., AYRES, A.J. (2007). Etiology of three recent diseases of citrus in Sao Paulo State: sudden death, variegated chlorosis and huanglongbing. *IUBMB. Life*, v.59, p.346-354.
- BRENNER, S.E. (2000). Target selection for structural genomics. *Nat. Struct. Biol. Supplement*, v.7, p.967-969.
- BRENNER, S.E. (2001). A tour of structural genomics. *Nat. Rev. Genet.*, v.10, p.801-809.
- BROWN, K. (2001). Florida fights to stop citrus canker. *Science*, v.292, p.2275-2276.
- BRUNGER, A.T., ADAMS P.D., CLORE, G.M., DELANO, W.L., GROS, P., GROSSE-KUNSTLEVE, R.W., JIANG, J.S., KUSZEWSKI, J., NILGES, M., PANNU, N.S., READ, R.J., RICE, L.M., SIMONSON, T., WARREN, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, v.54, p.905-921.
- BRUNINGS, A.M., GABRIEL, D.W. (2003). *Xanthomonas citri*: Breaking the Surface. *Molecular Plant Pathology*, v.4, p.141-157.
- BRÜSCHWEILER, R. (2003). New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.*, v.13, p.175-183.
- BUELL, C.R., JOARDAR, V., LINDEBERG, M., SELENGUT, J., PAULSEN, I.T., GWINN, M.L., DODSON, R.J., DEBOY, R.T., DURKIN, A.S., KOLONAY, J.F., MADUPU, R., DAUGHERTY, S., BRINKAC, L., BEANAN, M.J., HAFT, D.H., NELSON, W.C., DAVIDSEN, T., ZAFAR, N., ZHOU, L., LIU, J., YUAN, Q., KHOURI, H., FEDOROVA, N., TRAN, B., RUSSELL, D., BERRY, K., UTTERBACK, T., VAN AKEN, S.E., FELDBLYUM, T.V., D'ASCENZO, M., DENG, W.L., RAMOS, A.R., ALFANO, J.R., CARTINHO, S., CHATTERJEE, A.K., DELANEY, T.P., LAZAROWITZ, S.G., MARTIN, G.B., SCHNEIDER, D.J., TANG, X., BENDER, C.L., WHITE, O., FRASER, C.M., COLLMER, A. (2003). The

complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *Tomato* DC3000. Proc. Natl. Acad. Sci. U.S.A., v.100, p.10181-10186.

BÜRCKSTÜMMER, T., BENNETT, K.L., PRERADOVIC, A., SCHÜTZE, G., HANTSCH, O., SUPERTI-FURGA, G., BAUCH, A. (2006). An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. Nat. Methods, v.3, p. 1013-1019

BURLEY, S.K. (2000). An overview of structural genomics. Nat. Struct. Biol. Supplement, v.7, p.932-934.

CAI, M., HUANG Y., SAKAGUCHI, K., GLORE, G.M., GRONENBORN, A.M., CRAIGIE, R. (1998). An efficient and cost-effective isotope labeling protocol for proteins expressed in *Escherichia coli*. J. Biomol. NMR, v.11, p.97-102.

CARVALHO, M.L.V., ROSSETI, V., POMPEU-JUNIOR, J. (1995). Evolução da sintomatologia de clorose variegada dos citros em laranja doce (*C. sinensis*). Fitopatologia Brasileira, v.20, p.351.

CARVALHO, M.L.V., ROSSETI, V., POMPEU-JUNIOR, J. (1995). Evolução da sintomatologia de clorose variegada dos citros em laranja doce (*C. sinensis*). Fitopatologia Brasileira, v.20, p.351.

CHAMPIER, L., SIBILLE, N., BERSCH, B., BRUTSCHER, B., BLACKLEDGE, M., COVÈS, J. (2002). Reactivity, secondary structure, and molecular topology of the *Escherichia coli* sulfite reductase flavodoxin-like domain. Biochemistry, v.41, p.3770-3780.

CHEN, A.D., WU, D.H., JOHNSON JR., C.S. (1995). Determination of the binding isotherm and size of the bovine serum albumin-sodium dodecyl sulfate complex by diffusion-ordered 2D NMR. J. Phys. Chem., v.99, p.828-834.

CHOU, J.J., LI, S., BAX A. (2000). Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings. J. Biomol. NMR, v.18, p.217-227.

CHRISTENDAT, D., SARIDAKIS, V., KIM, Y., KUMAR, P.A., XU, X., SEMESI, A., JOACHIMIAK, A., ARROWSMITH, C.H., EDWARDS, A.M. (2002). The crystal structure of hypothetical protein MTH1491 from *Methanobacterium thermoautotrophicum*. Protein Sci., v.11, p.1409-1414.

CHRISTENDAT, D., YEE, A., DHARAMSI, A., KLUGER, Y., SAVCHENKO, A., CORT, J.R., BOOTH, V., MACKERETH, C.D., SARIDAKIS, V., EKIEL, I., KOZLOV, G., MAXWELL, K.L., WU, N., MCINTOSH, L.P., GEHRING, K., KENNEDY, M.A., DAVIDSON, A.R., PAI, E.F., GERSTEIN, M., EDWARDS, A.M., ARROWSMITH, C.H. (2000). Structural proteomics of an archaeon. Nat. Struct. Biol., v.7, p.903-909.

CICERO, D.O., CONTESSA, G.M., PERTINHEZ, T.A., GALLO, M., KATSUYAMA, A.M., PACI, M., FARAH, C.S., SPISNI, A. (2007). Solution structure of ApaG from *Xanthomonas axonopodis* pv. *citri* reveals a fibronectin-3 fold. Proteins, v.67, p.490-500.

CICERO, D.O., MELINO, S., ORSALE, M., BRANCATO, G., AMADEI, A., FORLANI, F., PAGANI, S., PACI, M. (2003). Structural rearrangements of the two domains of *Azotobacter vinelandii* rhodanese upon sulfane sulfur release: essential molecular dynamics, <sup>15</sup>N NMR relaxation and deuterium exchange on the uniformly labeled protein. Int. J. Biol. Macromol., v.33, p.193-201.

CLORE, G.M., GRONENBORN, A.M. (1998) NMR structure determination of proteins and protein complexes larger than 20 kDa. Curr. Opin. Chem. Biol., v.2, p.564-570.

CORNILESCU, G., BAX, A. (2000). Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. J. Am. Chem. Soc., v.122, p.10143-10154.

CORNILESCU, G., DELAGLIO, F., BAX, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR, v.13, p.289-302.

CORNILESCU, G., MAQUARDT, J., OTTIGER, M., BAX, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J. Am. Chem. Soc., v.120, p.6836-6837.

- DA SILVA, A.C. *et al.* (2002). Comparison of the genomes of two *Xanthomonas pathogens* with differing host specificities. *Nature*, v.417, p.459-463.
- DALVIT, C., HOMMEL, U. (1995). Sensitivity-improved detection of protein hydration and its extension to the assignment of fast-exchanging resonances. *J. Magn. Reson.*, v.109, p.334-338.
- DELAGLIO, F., GRZESIEK, S., VUISTER, G.W., ZHU, G., PFEIFER, J., BAX, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, v.6, p.277-293.
- DELAGLIO, F., KONTAXIS, G., BAX, A. (2000). Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.*, v.122, p.2142-2143.
- DEMPSEY, E. (2001). Hydrogen Exchange in peptides and proteins using NMR spectroscopy. *Prog. Nucl Magn. Res.*, v.39, p.135-170.
- DIECKMAN, L., GU, M., STOLS, L., DONNELLY, M.I., COLLART, F.R. (2002). High throughput methods for gene cloning and expression. *Protein Expr. Purif.*, v.25, p.1-7.
- DOUGAN D.A., REID B.G., HORWICH, A.L., BUKAU B. (2002a). ClpS, a substrate modulator of the ClpAP machine. *Mol. Cell*, v.9, p.673-683.
- DOUGAN, D.A., MOGK, A., ZETH, K., TURGAY, K., BUKAU, B. (2002b) AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett.*, v.529, p.6-10.
- DOW, J.M., DANIELS, M.J. (2000). *Xylella* genomics e bacterial pathogenicity to plants. *Yeast*, v.17, p.263-271.
- ENGLANDER, S.W., KALLENBACH, N.R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.*, v.16, p.521-655.
- ERBSE, A., SCHMIDT, R., BORNEMANN, T., SCHNEIDER-MERGENER, J., MOGK, A., ZAHN, R., DOUGAN, D.A., BUKAU, B. (2006). ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. *Nature*, v.439, p.753-756.
- ESTANISLAU *et al.* (2001). *Informativo agropecuário*, v.22, p.8-20.
- ETEZADY-ESFARJANI, T., HERRMANN, T., HORST, R., WÜTHRICH, K. (2006) Automated protein NMR structure determination in crude cell-extract. *J. Biomol. NMR*, v.34, p.3-11.
- EVANS, J.N.S. (1995). *In Biomolecular NMR Spectroscopy*. Oxford University Press.
- EXPASY (2003). *In* <http://expasy.org/>
- FARRELL, C.M., GROSSMAN, A.D., SAUER, R.T. (2005). Cytoplasmic degradation of ssrA-tagged proteins. *Mol. Microbiol.*, v.57, p. 1750-1761.
- FAWCET, H.H., JENKINS, A.E. (1993). Records of citrus canker from herbarium specimens of genus *Citrus* in England and the United States. *Phytopathology*, v.23, p.820-824.
- FEICHTENBERGER, E., MULLER, G.W., GUIRADO, N. (1997). Doença dos citrus (*Citrus spp*) Manual de Fitopatologia. Doenças das plantas cultivadas. São Paulo, Agronômica Ceres, p.261-296.
- FEIL, H., FEIL, W.S., CHAIN, P., LARIMER, F., DIBARTOLO, G., COPELAND, A., LYKIDIS, A., TRONG, S., NOLAN, M., GOLTSMAN, E., THIEL, J., MALFATTI, S., LOPER, J.E., LAPIDUS, A., DETTER, J.C., LAND, M., RICHARDSON, P.M., KYRPIDES, N.C., IVANOVA, N., LINDOW, S.E. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proc. Natl. Acad. Sci. U.S.A.*, v.102, p.11064-11069.
- FERENTZ, A.E., WAGNER, G. (2000). NMR Spectroscopy: a multifaceted approach to macromolecular structure. *Quarterly Reviews of Biophysics*, v.33, p.29-65.
- FISCHER, M.W., LOSONCZI, J.A., WEAVER, J.L., PRESTEGARD J.H. (1999) Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry*, v.38, p.9013-9022.

- FLANAGAN, J. M., WALL, J. S., CAPEL, M. S., SCHNEIDER, D. K., SHAKLIN, J. (1995). Scanning transmission electron microscopy and small-angle scattering provide evidence that native *Escherichia coli* ClpP is a tetradecamer with an axial pore. *Biochemistry*, v.34, p.10910-10917.
- FLEISCHMANN, R.D., ADAMS, M.D., WHITE, O., KIRKNESS, E.F., KERLAVAGE, A.R., BULT, C.J., TOMB, J.F., DOUGHERTY, B.A., MERRICK, J.M. e colaboradores (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, v.269, p.496-512.
- FLYNN, J.M., LEVCHENKO, I., SEIDEL, M., WICKNER, S.H., SAUER, R.T., BAKER, T.A. (2001). Overlapping recognition determinants within the *ssrA* degradation tag allow modulation of proteolysis. *Proc. Natl. Acad. Sci. U.S.A.*, v.98, p.10584-10589.
- FLYNN, J.M., LEVCHENKO, I., SEIDEL, M., WICKNER, S.H., SAUER, R.T., BAKER, T.A. (2001). Overlapping recognition determinants within the *ssrA* degradation tag allow modulation of proteolysis. *Proc. Natl. Acad. Sci. U.S.A.*, v.98, p.10584-10589.
- FOLKERS, G.E., BUUREN, B.N.M., KAPTEIN, R. (2004). Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. *J. Struct. Funct. Genomics*, v.5, p.119-131.
- FUNDECITRUS (2001, 2007). *In* <http://www.fundecitrus.com.br>
- GALVÃO-BOTTON, L.M.P., KATUSYAMA, A.M., GUZZO, C.R., FARAH, C.S., ALMEIDA F.C.L., VALENTE, A.P. (2003). High-throughput screening of structural proteomics targets using NMR. *FEBS Lett.*, v.552, p.207-213.
- GIAEVER, G. *et al.* (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, v.418, p.387-391.
- GIL, V.M., GERALDES, C.F.G.C. (1987). *In* Ressonância Magnética Nuclear: fundamentos, métodos e aplicações. Fundação Calouste Gulbenkian, Lisboa.
- GNU PLOT (2005). *In* [www.gnuplot.info](http://www.gnuplot.info)
- GOODNER, B. *et al.* (2001). Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science*, v.294, p.2323-2328.
- GOTTESMAN, S. (1996). Proteases and their targets in *Escherichia coli*. *Ann. Rev. Genet.*, v.30, p.465-506.
- GOTTESMAN, S. (1999). Regulation by proteolysis: developmental switches. *Curr. Opin. Microbiol.*, v.2, p.142-147.
- GOTTESMAN, S., CLARK, W.P., DE CRECY-LAGARD, V., MAURIZI, M.R. (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. Sequence and *in vivo* activities. *J. Biol. Chem.*, v.268, p.22618-22626.
- GOTTESMAN, S., MAURIZI, M.R., WICKNER, S. (1997). Regulatory Subunits of Energy-Dependent Proteases. *Cell*, v.91, p.435-438.
- GOTTESMAN, S., ROCHE, E., ZHOU, Y., SAUER, R.T. (1998). The ClpXP e ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & Development*, v.12, p.1338-1347.
- GRIGORIEV, I.V., CHOI, I.G. (2002). Target selection for structural genomics: a single genome approach. *Omics*, v.6, p.349-362.
- GUIGNARD, L., OZAWA, K., PURSGLOVE, S.E., OTTING, G., DIXON, N.E. (2002). NMR analysis of *in vitro*-synthesized proteins without purification: a high-throughput put approach. *FEBS Lett.*, v.254, p.159-162.
- GÜNTERT, P., (1998). Structure calculation of biological macromolecules from NMR data. *Q. Rev. of Biophysics*, v.31, p.145-237.

- GÜNTERT, P., MUMENTHALER, C., WUTHRICH, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, v.273, p.283-298.
- GUO, F., ESSER, L., SINGH, S., MAURIZI, M.R., XIA, D. (2002a). Crystal Structure of the Heterodimeric Complex of the Adaptor, ClpS, with the N-domain of the AAA+ Chaperone, ClpA. *J. Biol. Chem.*, v.227, p.46753-46762.
- GUO, F., MAURIZI, M.R., ESSER, L., XIA, D. (2002b). Crystal Structure of ClpA, an Hsp100 Chaperone and Regulator of ClpAP Protease. *J. Biol. Chem.*, v.227, p.46743-46752.
- GUZZO, C.R., NAGEM, R.A., BARBOSA, J.A., FARAH, C.S. (2007). Structure of *Xanthomonas axonopodis* pv. *citri* YaeQ reveals a new compact protein fold built around a variation of the PD-(D/E)XK nuclease motif. *Proteins* (Publicação eletrônica em 10 de julho de 2007).
- GUZZO, C.R., NAGEM, R.A.P., GALVÃO-BOTTON, L.M.P., GUIMARÃES, B.G., MEDRANO, F.J., BARBOSA, J.A.R.G., FARAH, C.S. (2005). Expression, purification, crystallization and preliminary X-ray analysis of YaeQ (XAC2396) from *Xanthomonas axonopodis* pv. *citri*. *Acta Cryst.*, F61, v.61, p.493-495.
- GUZZO, C.R., SILVA, L.R., GALVAO-BOTTON, L.M., BARBOSA, J.A., FARAH, C.S. (2006). Expression, crystallization and preliminary crystallographic analysis of SufE (XAC2355) from *Xanthomonas axonopodis* pv. *citri*. *Acta Crystallograph Sect F Struct Biol Cryst Commun.*, v.62, p.268-270.
- GYGI, S.P., AEBERSOLD, R. (2000). Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.*, v.4, p.489-494.
- HANASH, S. (2003). Disease proteomics. *Nature*, v.422, p.226-232.
- HANSEN, M.R., MUELLER, L., PARDI, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.*, v.5, p.1065-1074.
- HAYWARD, A.C. (1993). The hosts of *Xanthomonas*. In *Xanthomonas* (Swings, J.G. e Civerolo, E.L., Eds.), p.1-119. London: Chapman e Hall.
- HERMANN, J.C., MARTI-ARBONA, R., FEDOROV, A.A., FEDOROV, E., ALMO, S.C., SHOICHET, B.K., RAUSHEL, F.M. (2007). Structure-based activity prediction for an enzyme of unknown function. *Nature*, v.448, p.775-779.
- HERRMANN, T., GÜNTERT, P., WUTHRICH, K. (2002a). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, v.24, p.171-189.
- HERRMANN, T., GÜNTERT, P., WUTHRICH, K. (2002b). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, v.319, p.209-227.
- HIGGINS, D., THOMPSON, J., GIBSON, T., THOMPSON, J.D., HIGGINS, D.G., GIBSON, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, v.22, p. 4673-4680.
- HINNERWISCH, J., FENTON, W.A., FURTAK, K.J., FARR, G.W., HORWICH, A.L. (2005a). Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell*, v.121, p.1029-1041.
- HINNERWISCH, J., REID, B.G., FENTON, W.A., HORWICH, AL. (2005b). Roles of the N-domains of the ClpA unfoldase in binding substrate proteins and in stable complex formation with the ClpP protease. *J. Biol. Chem.*, v.280, p. 40838-40844.
- HOL, W.G.J. (2000) Structural genomics for science e society. *Nat. Struct. Biol. Supplement*, v.7, p.964-966.



- HORWICH, A.L., WEBER-BAN, E.U., FINLEY, D. (1999). Chaperone rings in protein folding and degradation. *Proc. Natl. Acad. Sci. U.S.A.*, v.96, p.11033-11040.
- HOSKINS, J.R., PAK, M., MAURIZI, M.R., WICKNER, S. (1998). The role of the ClpA chaperone in proteolysis by ClpAP. *Proc. Natl. Acad. Sci. U.S.A.*, v.95, p.12135-12140.
- HOSKINS, J.R., SINGH, S.K., MAURIZI, M.R., WICKNER, S. (2000). Protein binding and unfolding by the chaperone ClpA and degradation by the protease ClpAP. *Proc. Natl. Acad. Sci. U.S.A.*, v.97, p.8892-8897.
- HOSKINS, J.R., YANAGIHARA, K., MIZUUCHI, K., WICKNER, S. (2002). ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. *Proc. Natl. Acad. Sci. U.S.A.*, v.99, p.11037-11042.
- HUS, J.C., MARION, D., BLACKLEDGE, M. (2001). Determination of protein backbone structure using only residual dipolar couplings. *J. Am. Chem. Soc.*, v.123, p.1541-1542.
- HUTH, J.R., BEWLEY, C.A., JACKSON, B.M., HINNEBUSCH, A.G., CLORE, G.M., GRONENBORN, A.M. (1997). Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci.*, v.6, p.2359-2364.
- HUYNEN, M., DOERKS, T., EISENHABER, F., ORENGO, C., SUNYAEV, S., YUAN, Y., BORK, P. (1998). Homology-based fold prediction for *Mycoplasma genitalium* proteins. *J. Mol. Biol.*, v.280, p.323-326.
- HWANG, T.L., MORI, S., SHAKA, A.J., VAN ZIJL, P.C.M. (1997). Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water-protein proton exchange and intermolecular NOEs. *J. Am. Chem. Soc.*, v.119, p.6203-6204.
- HWANG, T.L., VAN ZIJL, P.C.M., MORI, S. (1998). Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR*, v.11, p.221-226.
- IKURA, M., KAY, L. E., BAX, A. (1990). A novel approach of sequential assignment of  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, v.29, p.4659-4667.
- INSTITUTO AGRONÔMICO DO PARANÁ (2004). *In* <http://www.iapar.br>
- ISHIKAWA, T., BEURON, F., KESSEL, M., WICKNER, S., MAURIZI, M.R., STEVEN, A.C. (2001). Translocation pathway of protein substrates in ClpAP protease. *Proc. Natl. Acad. Sci. U.S.A.*, v.98, p.4328-4333.
- JACKSON, R.M., RUSSELL, R.B. (2001). Predicting function from structure: examples of the serine protease inhibitor canonical loop conformation found in extracellular proteins. *Comput. Chem.*, v.26, p.31-39.
- JACOBS, D.M., SAXENA, K., VOGTHERR, M., BERNADO, P., PONS, M., FIEBIG, K.M. (2003). Peptide binding induces large scale changes in inter-domain mobility in human Pin1. *J. Biol. Chem.*, v.278, p.26174-26182.
- JAREN, O.R., KRANZ, J.K., SORENSEN, B.R., WAND, A.J., SHEA, M.A. (2002). Calcium-induced conformational switching of Paramecium calmodulin provides evidence for domain coupling. *Biochemistry*, v.41, p.14158-14166.
- JOARDAR, V., LINDEBERG, M., JACKSON, R., SELENGUT, J., DODSON, R., BRINKAC, L.M., DAUGHERTY, S.C., DEBOY, R.T., DURKIN, A.S., GIGLIO, M.G., MADUPU, R., NELSON, W.C., ROISOVITZ, M.J., SULLIVAN, S.A., CRABTREE, J., CREAMY, T., DAVIDSEN, T.M., HAFT, D.H., ZAFAR, N., ZHOU, L., HALPIN, R., HOLLEY, T., KHOURI, H.M., FELDBLYUM, T.V., WHITE, O., FRASER, C.M., CHATTERJEE, A.K., CARTINHO, S., SCHNEIDER, D., MANSFIELD, J., COLLMER, A., BUELL, R. (2005). Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J. Bacteriol.*, v.187, p. 6488-6498.

- JOHSON, B.A., BLAVINS, R.A. (1994). NMRView – a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, v.4, p.603-614.
- JONES, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, v.292, p.195-199.
- KARPLUS, M. (1959). Contact electron-spin coupling of nuclear magnetic moments. *J. Phys. Chem.*, v.30, p.11-15.
- KATAYAMA Y., GOTTESMAN, S., PUMPHREY, J., RUDIKOFF, S., CLARK, W.P., MAURIZI, M.R. (1988). The two-component, ATP-dependent Clp protease of *Escherichia coli*. Purification, cloning, and mutational analysis of the ATP-binding component. *J. Biol. Chem.*, v.263, p.15226-15236.
- KATSUYAMA, A.M., CICERO, D.O., SPISNI, A., PACI, M., FARAH, C.S., PERTINHEZ, T.A. (2004). <sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C resonance assignments of the ApaG protein of the phytopathogen *Xanthomonas axonopodis* pv. *citri*. *J. Biomol. NMR*, v.29, p.423-424.
- KAY, L.E. *et al.* (1990). Three-dimensional triple resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Res.*, v.89, p.495-514.
- KAY, L.E. (1998). Protein dynamics from NMR. *Nat. Struct. Biol.*, v.5, p.513-517.
- KAY, L.E., TORCHIA, D.A., BAX, A. (1989). Backbone dynamics of proteins as studied by <sup>15</sup>N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, v.28, p.8972-8979.
- KEDZIERSKA, S., CHESNOKOVA, L.S., WITT, S.N., ZOLKIEWSKI, M. (2005). Interactions within the ClpB/DnaK bi-chaperone system from *Escherichia coli*. *Arch. Biochem. Biophys.*, v.444, p. 61-65.
- KENNEDY M.A., MONTELLIONE, G.T., ARROWSMITH, C.H., MARKLEY, J. (2002). Role for NMR in structural genomics. *J. Struct. Funct. Genomics*, v.2, p.155-169.
- KESSEL, M., MAURIZI, M.R., KIM, B., KOCSIS, E., TRUS, B.L., SINGH, S.K., STEVEN, A.C. (1995). Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome. *J. Mol. Biol.*, v.250, p. 587-594.
- KIGAWA, T., YABUKI, T.; YOSHIDA, Y.; TSUTSUI, M.; ITO, Y.; SHIBATA, T., YOKOYAMA, S. (1999). Cell-free production and stable-isotope labelling of milligram quantities of proteins. *FEBS Lett.*, v.442, p.15-19.
- KIKUCHI, J., IWAHARA, J., KIGAWA, T., MURAKAMI, Y., OKAZAKI, T., YOKOYAMA, S. (2002). Solution structure determination of the two DNA-binding domains in the *Schizosaccharomyces pombe* Abp1 protein by a combination of dipolar coupling and diffusion anisotropy restraints. *J. Biomol. NMR*, v.22, p. 333-347.
- KIM, S.H., SHIN, D.H., CHOI, I.G., SCHULZE-GAHMEN U., CHEN, S., KIM, R. (2003). Structure-based functional inference in structural genomics. *Struct. Funct. Genomics.*, v.4, p.129-135.
- KIM, S.H., SHIN, D.H., LIU, J., OGANESVAN, V., CHEN, S., XU, Q.S., KIM, J.S., DAS, D., SCHULZE-GAHMEN, U., HOLBROOK, S.R., HOLBROOK, E.L., MARTINEZ, B.A., OGANESVAN, N., DEGIOVANNI, A., LOU, Y., HENRIQUEZ, M., HUANG, C., JANCARIK, J., PUFAN, R., CHOI, I.G., CHANDONIA, J.M., HOU, J., GOLD, B., YOKOTA, H., BRENNER, S.E., ADAMS, P.D., KIM, R. (2005). Structural genomics of minimal organisms and protein fold space. *J. Struct. Funct. Genomics*, v.6, p.63-70.
- KIM., K.I. , CHEONG., G.W., PARK., S.C., HA., J.S., WOO., K.M., CHOI, S.J., CHUNG., C.H. (2000). Heptameric ring structure of the heat-shock protein ClpB, a protein-activated ATPase in *Escherichia coli*. *J. Mol. Biol.*, v.303, p.655-666.
- KOPKE SALINAS, R., FOLKERS, G.E., BONVIN, A.M., DAS, D., BOELEN, R., KAPTEIN, R. (2005). Altered specificity in DNA binding by the lac repressor: a mutant lac headpiece that mimics the gal repressor. *ChemBiochem.*, v.6, p.1628-1637.

- KRESS, W., MUTSCHLER, H., WEBER-BAN, E. (2007). Assembly pathway of an AAA+ protein: tracking ClpA and ClpAP complex formation in real time. *Biochemistry*, v.46, p.6183-6193.
- KUBONIWA, H., GRZESIEK, S., DLAGLIO, F., BAX, A. (1994). Measurement of  $H^N$ - $H^\alpha$  J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR*, v.4, p.871-878.
- LAMZIN, V.S., PERRAKIS, A. (2000). Current state of automated crystallographic data analysis. *Nat. Struct. Biol.*, v.7, p.978-981.
- LARSEN, C.N., FINLEY, D. (1997). Protein translocation channels in the proteasome and other proteases. *Cell*, v.91, p.431-434.
- LASKOWSKI, R. A., RULLMANN, J.A.C., MACARTHUR, M.W., KAPTEIN, R., THORNTON, J.M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, v.8, p.477-496.
- LASKOWSKI, R.A., MACARTHUR, M.W.; MOSS, D.S., THORNTON, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Applied Crystallography*, v.26, p.283-285.
- LESLEY, S.A. *et al.* (2002). Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. U.S.A.*, v.99, p.11664-11669.
- LEVCHENKO, I., SEIDEL, M., SAUER, R.T., BAKER, T.A. (2000). A specificity-enhancing factor for the ClpXP degradation machine. *Science*, v.289, p.2354-2356.
- LIU, L., RODRIGUEZ-BELMONTE E.M., MAZLOUM, N., XIE, B., LEE, M.Y.W.T. (2003). Identification of a novel protein, PDIP38, that interacts with the p50 subunit of DNA polymerase  $\delta$  and proliferating cell nuclear antigen. *J. Biol. Chem.*, v.278, p.10041-10047.
- LO, J.H., BAKER, T.A., SAUER R.T. (2001). Characterization of the N-terminal repeat domain of *Escherichia coli* ClpA-A class I Clp/HSP100 ATPase. *Protein Sci.*, v.10, p.551-559.
- LOSONCZI, J.A., ANDREC, M., FISCHER, M.W., PRESTEGARD, J.H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.*, v.138, p.334-342.
- LUKIN, J.A., KONTAXIS, G., SIMPLACEANU, V., YUAN, Y., BAX, A., HO, C. (2003). Quaternary structure of hemoglobin in solution. *Proc. Natl. Acad. Sci. U.S.A.*, v.100, p.517-520.
- LUPAS, A., FLANAGAN, J.M., TAMURA, T., BAUMEISTER, W. (1997). Self-compartmentalizing proteases. *Trends Biochem. Sci.*, v.22, p.399-404
- LUPAS, A.N., KORETKE, K.K. (2003). Bioinformatic analysis of ClpS, a protein module involved in prokaryotic and eukaryotic protein degradation. *J. Struct. Biol.*, v.141, p.77-83.
- MARTIN, A.C.R. (1996). *In* PROFIT - <http://www.bioinf.org.uk/software/profit/>
- MATTE, A., JIA, Z., SUNITA, S., SIVARAMAN, J., CYGLER, M. (2007). Insights into the biology of *Escherichia coli* through structural proteomics. *J. Struct. Funct. Genomics*. (Publicação eletrônica em 1 de agosto de 2007).
- MAURIZI, M.R., CLARK, W.P., KATAYAMA, Y., RUDIKOFF, S., PUMPHREY, J., BOWERS, B., GOTTESMAN, S. (1990a) Sequence and Structure of ClpP, the proteolytic component of the ATP-dependent Clp Protease of *Escherichia coli*. *J. Biol. Chem.*, v.265, p.12536-12545.
- MAURIZI, M.R., CLARK, W.P., KIM S.H., GOTTESMAN, S. (1990b). ClpP represents a unique family of serine proteases. *J. Biol. Chem.*, v.265, p.12546-12552.
- MEILER, J., PETI, W., GRIESINGER, C. (2000). DipoCoup: A versatile program for 3D-structure homology comparison based on residual dipolar couplings and pseudocontact shifts. *J. Biomol. NMR*, v.17, p.283-294.

- MIKE, T., MANN, M. (2003). From genomics to proteomics. *Nature*, v.422, p.193-197.
- MITTERMAIER, A., KAY, L.E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science*, v.312, p.224-228.
- MITTL, P.R.E., GRÜTTER, M.G. (2001). Structural Genomics: opportunities and challenges. *Curr. Opin. Chem. Biol.*, v.5, p.402-408.
- MOGK, A., DOUGAN, D., WEIBEZAHN, J., SCHLIEKER, C., TURGAY, K., BUKAU, B. (2004). Broad yet high substrate specificity: the challenge of AAA<sup>+</sup> proteins. *J. Struct. Biol.*, v.146, p.90-98.
- MOLECULAR SIMULATIONS (2000). WebLab ViewerPro 3.7. Accelrys Software Inc., San Diego, California, USA.
- MONTELIONE, G.T. (2001). Structural genomics an approach to the protein folding problem. *Proc. Natl. Acad. Sci. U.S.A.*, v.98, p.13488-13489.
- MONTELIONE, G.T., ZHENG, D., HUANG Y.J., GUNSALUS, K.C., SZYPERKI, T. (2000). Protein NMR spectroscopy in structural genomics. *Nat. Struct. Biol.*, v.7, p.982-985.
- MUFFLER, D., FISCHER, S., ALTUVIA, G., STORZ AND R. HENGGE-ARONIS (1996). The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in *Escherichia coli*. *EMBO J.*, v.15, p.1333-1339.
- NACHIN, L., EL HASSOUNI, M., LOISEAU, L., EXPERT, D., BARRAS, F. (2001) SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. *Mol. Microbiol.*, v.39, p.960-972.
- NACHIN, L., LOISEAU, L., EXPERT, D., BARRAS, F. (2003). SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *EMBO J.*, v.22, p.427-437.
- NAKAI, K., HORTON, P. (1999). PSORT: a program for detecting sorting signals in proteins e predicting their subcellular localization. *Trends Biochem. Sci.*, v.24, p.34-36.
- NASCIMENTO, J.F., RODRIGUES NETO J., ALVES, J.M.A., REGO, M., ARAUJO, A.E.S. (2003). Ocorrência de cancro cítrico no estado de Rondônia. *Anais do Congresso Paulista de Fitopatologia*, p.26.
- NCBI (2007). In <http://www.ncbi.nlm.nih.gov/>
- NEHER, S.B., SAUER, R.T., BAKER, T.A. (2003). Distinct peptide signals in the UmuD and UmuD' subunits of UmuD/D' mediate tethering and substrate processing by the ClpXP protease. *Proc. Natl. Acad. Sci. U.S.A.*, v.100, p.13219-13224.
- NEUWALD, A.F., ARAVIND, L., SPOUGE, J.L., KOONIN, E.V. (1999). AAA<sup>+</sup>: A class of chaperonin-like ATPases associated with the assembly, operation, disassembly of protein complexes. *Genome Res.*, v.9, p.27-43.
- NORIN, M., SUNDSTROM, M. (2001). Protein models in drug discovery. *Curr. Op. in Drug Discovery & Development*, v.4, p.284-290.
- O'TOOLE, N., GRABOWSKI, M., OTWINOWSKI, Z., MINOR, W., CYGLER, M. (2004b). The structural genomics experimental pipeline: insights from global target lists. *Proteins*, v.56, p.201-210.
- O'TOOLE, N., RAYMOND, S., CYGLER, M. (2003). Coverage of protein sequence space by current structural genomics targets. *J. Struct. Funct. Genomics*, v.4, p.47-55.
- OCHIAI, H., INOUE, Y., TAKEYA, M., SASAKI, A., & KAKU, H. (2005). Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. *Jpn Agric Res Q.*, v.39, p.275-287.
- O'TOOLE, N., GRABOWSKI, M., OTWINOWSKI, Z., MINOR, W., CYGLER, M. (2004). The structural genomics experimental pipeline: insights from global target lists. *Proteins.*, v.56, p. 201-210.

- OTTIGER, M., DELAGLIO, F., BAX, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.*, v.131, p.373-378.
- PAGE, R., PETI, W., WILSON, I.A., STEVENS, R.C., WÜTHRICH, K. (2005). NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. *Proc. Natl. Acad. Sci. U.S.A.*, v.102, p.1901-1905.
- PAK, M., HOSKINS, J.R., SINGH, S.K., MAURIZI, M.R., WICKNER, S. (1999). Concurrent Chaperone and Protease Activities of ClpAP and the Requirement for the N-terminal ClpA ATP binding site for Chaperone Activity. *J. Biol. Chem.*, v.274, p.19316-19322.
- PALMER, A.G. (2001). NMR probes of molecular dynamics: overview and comparison with other techniques. *Ann. Rev. Biophys. Biomol. Struct.*, v.30, p.129-155.
- PARDI, A., BILLETER, M., WÜTHRICH, K. (1984). Calibration of the angular dependence of the amide proton-C alpha proton coupling constants,  $^3J_{HN\alpha}$ , in a globular protein. Use of  $^3J_{HN\alpha}$  for identification of helical secondary structure. *J. Mol. Biol.*, v.180, p.741-751.
- PERTINHEZ, T.A., SFORÇA, M.L., ALVES, A.C., RAMOS, C.R., HO, P.L., TENDLER, M., ZANCHIN, N.I., SPISNI, A. (2004).  $^1H$ ,  $^{15}N$  and  $^{13}C$  resonance assignments of the apo Sm14-M20(C62V) protein, a mutant of *Schistosoma mansoni* Sm14. *J. Biomol. NMR*, v.29 p.553-554.
- PETI, W., ETEZADY-ESFARJANI, T., HERRMANN, T., KLOCK H.E., LESLEY, S.A., WÜTHRICH, K. (2004). NMR for structural proteomics of *Thermotoga maritima*: screening and structure determination. *J. Struct. Funct. Genomics*, v.5, p.205-215.
- PETSKO, G.A., RINGE, D. (1984). Fluctuations in protein structure from X-ray diffraction. *Ann. Rev. Biophys. Bioeng.*, v.13, p.331-371.
- PHIZICKY, E., BASTIAENS, P.I., ZHU, H., SNYDER, M., FIELDS, S. (2003). Protein analysis on a proteomic scale. *Nature*, v.422, p.208-215.
- PICKART, C.M., COHEN, R.E. (2004). Proteasomes and their kin: proteases in the machine age. *Nat. Rev. Mol. Cell. Biol.*, v.5, p.177-187.
- PIOTTO, M., SAUDEK, V., SKLENÁO, V. (1992). Gradient-tailored Excitation for Single-quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR*, v.2, p.661-665.
- PISZCZEK, G., ROZYCKI, J., SINGH, S.K., GINSBURG, A., MAURIZI, M.R. (2005). The Molecular Chaperone, ClpA, has a Single High Affinity Peptide Binding Site per Hexamer. *J. Biol. Chem.*, v.280, p.12221-12230.
- PORANKIEWICZ, J., WANG, J., CLARKE, A.K. (1999). New Insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond. *Mol. Microbiol.*, v.32, p.449-458.
- PRESTEGARD, J.H., AL-HASHIMI, H.M., TOLMAN, J.R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q. Rev. Biophys.*, v.33, p.371-424.
- PRESTEGARD, J.H., BOUGAULT, C.M., KISHORE, A.I. (2004). Residual dipolar couplings in structure determination of biomolecules. *Chem Rev.*, v.104, p.3519-3540.
- PRESTEGARD, J.H., VALAFAR, H., TIAN, F. (2001). Nuclear magnetic resonance in the era of structural genomics. *Biochemistry*, v.40, p.8677-8685.
- PYMOL (2005, 2007). *In* <http://pymol.sourceforge.net/>
- RABAH, G., POPESCU, R., COX, J.A., ENGELBORGHES, Y., CRAESCU, C.T. (2005). Solution structure and internal dynamics of NSCP, a compact calcium-binding protein. *FEBS J.*, v.272, p.2022-2036.
- REHM, T., HUBER, R., HOLAK, T.A. (2002). Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure*, v.10, p.1613-1618.

- RIGAUT, G.; SHEVCHENKO, A., RUTZ, B., WILM, M., MANN, M., SÉRAPHIN, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotech.*, v.17, p.1030-1032.
- ROA, B.B., CONNOLLY, D.M., WINKLER, M.E. (1989). Overlap between pdxA and ksgA in the complex pdxA-ksgA-apaG-apaH operon of *Escherichia coli* K-12. *J. Bacteriol.*, v.171, p. 4767-4777.
- ROHRWILD, M., COUX, O., HUANG, H.C., MOERSCHELL, R.P., YOO, S.J., SEOL, J.H., CHUNG, C.H., GOLDBERG, A.L. (1996). HslV-HslU: A novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proc. Natl. Acad. Sci. U.S.A.*, v.93, p.5808-5813.
- ROMÁN, M.P., M. CAMBRA., J. JUAREZ., P. MORENO., N. DURAN-VILA, F.A.O., TANAKA, E., ALVES, E., W. KITAJIMA., P.T. YAMAMOTO., R.B. BASSNEZI., D.C. TEIXEIRA., W.C. JESUS, JR., A.J. AYRES., N. GIMENES-FERNANDES., F. RABENSTEIN., L.F. GIROTTO., AND J.M. BOVÉ. (2004). Sudden death of citrus in Brazil: a graft-transmissible bud union disease. *Plant. Dis.*, v.88, p.453-467.
- ROSSETTI, V. (1977). Citrus canker in Latin America: A review. *Proceedings of the International Society of Citriculture*, v.3, p.918-924.
- RUSSELL, R.B., EGGLESTON, D.S. (2000). New roles for structure in biology e drug discovery. *Nat. Struct. Biol.*, v.7, p.928-930.
- SALANOUBAT, M., GENIN, S., ARTIGUENAVE, F., GOUZY, J., MANGENOT, S., ARLAT, M., BILLAULT, A., BROTTIER, P., CAMUS, J.C., CATTOLICO, L., CHANDLER, M., CHOISNE, N., CLAUDEL-RENARD, C., CUNNAC, S., DEMANGE, N., GASPIN, C., LAVIE, M., MOISAN, A., ROBERT, C., SAURIN, W., SCHIEX, T., SIGUIER, P., THEBAULT, P., WHALEN, M., WINCKER, P., LEVY, M., WEISSENBACH, J., BOUCHER, C.A. (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*, v.415, p.497-502.
- SALI, A., BLUNDELL, T.L. (1993). Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, v.234, p.779-784.
- SALINAS, R.K., DIERCKS, T., KAPTEIN, R., BOELENS, R. (2006). Cooperative alpha-helix unfolding in a protein-DNA complex from hydrogen-deuterium exchange. *Protein. Sci.*, v.15, p.1752-1759.
- SAMBROOK, J., RUSSEL, D.W. (2001). *Molecular cloning. A laboratory protocol* (3<sup>rd</sup> Ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- SANDERS, C.R., SCHWONEK (1992). Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry*. v.31, p.8898-8905.
- SANDERS, C.R., HARE, B.J., HOWARD, K.P., PRESTGARD, J.H. (1994). Magnetically-oriented phospholipid micelles as a tool for the study of membrane-associated molecules. *Prog. Nucl. Magn. Reson. Spect.*, v.26, p.421-444.
- SANGER, F., NICKLEN, S.E., COULSON, A.R. (1977). DNA sequencing using chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, v.74, p.5463-5467.
- SASS, H.J., MUSCO, G., STAHL, S.J., WINGFIELD, P.T., GRZEIEK., S. (2001). An easy way to include weak alignment constraints into NMR structure calculations. *J. Biomol. NMR*, v.21, p.275-280.
- SAUER, R.T., BOLON, D.N., BURTON, B.M., BURTON, R.E., FLYNN, J.M., GRANT, R.A., HERSCH, G.L., JOSHI, S.A., KENNISTON, J.A., LEVCHENKO, I., NEHER, S.B., OAKES, E.S., SIDDIQUI, S.M., WAH, D.A., BAKER, T.A. (2004). Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell*, v.119, p.9-18
- SAUNDERS, M., WISHNIA, A., KIRKWOOD, J.G. (1957). The nuclear magnetic resonance of ribonuclease. *J. Am. Chem. Soc.*, v.79, p.3289-3290.
- SAUNDERS, M., WISHNIA, A., KIRKWOOD, J.G. (1957). The nuclear magnetic resonance of ribonuclease. *J. Am. Chem. Soc.*, v.79, p.3289.

- SAVCHENKO, A., YEE, A., KHACHATRYAN, A., SKARINA, T., EVDOKIMOVA, E., PAVLOVA, M., SEMESI, A., NORTHEY, J., BEASLEY, S., LAN, N., DAS, R., GERSTEIN, M., ARROWSMITH, C.H., EDWARDS, A.M. (2003). Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. *Proteins: Struct. Funct. Genetics.*, v.50, p.392-399.
- SAVEANU, C., MIRON, S., BORZA, T., CRAESCU, C.T., LABESSE, G., GAGYI, C., POPESCU, A., SCHAEFFER, F., NAMANE, A., LAURENT-WINTER C., BÂRZU, O., GILLES, A.M. (2002). Structural and nucleotide-binding properties of YajQ and YnaF, two *Escherichia coli* proteins of unknown function. *Protein Sci.*, v.11, p.2551-2560.
- SAVEANU, C., MIRON, S., BORZA, T., CRAESCU, C.T., LABESSE, G., GAGYI, C., POPESCU, A., SCHAEFFER, F., NAMANE, A., LAURENT-WINTER, C., BÂRZU, O., GILLES, A.M. (2002). Structural and nucleotide-binding properties of YajQ and YnaF, two *Escherichia coli* proteins of unknown function. *Protein Sci.*, v.11, p.2551-2560.
- SCHEICH, C., LEITNER, D. (2004). Fast identification of folded human protein domains expressed in *E. coli* suitable for structural analysis. *Struct. Biol.*, v.4, p.4.
- SCHIRMER, E.C., GLOVER, J.R., SINGER, M.A., LINDQUIST, S. (1996). HSP100/Cip proteins: a common mechanism explains diverse functions. *Trends Biochem. Sci.*, v.21, p.289-296.
- SCHIRRA, H.J. (1996). Determination of Protein Structure with NMR Spectroscopy. *In* <http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/home.htm>.
- SCHMIDT, M., LUPAS, A.N., FINLEY, D. (1999). Structure and mechanism of ATP-dependent proteases. *Curr. Opin. Chem. Biol.*, v.3, p.584-591
- SCHOOLNIK, G.K. (2002). Microarray analysis of bacterial pathogenicity. *Adv. Microb. Physiol.*, v.46, p.1-45.
- SELENKO, P., WAGNER, G. (2007). Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.*, v.158, p.244-253.
- SERBER, Z., KEATINGE-CLAY, A.T. LEDWIDGE, R., KELLY, A.E., MILLER, S.M., DOTSCH, V. (2001a). High-resolution macromolecular NMR spectroscopy inside living cells. *J. Am. Chem. Soc.*, v.123, p. 2446-2447.
- SERBER, Z., LEDWIDGE, R., MILLER, S.M., DÖTSCH, V. (2001b). Evaluation of parameters critical to observing proteins inside living *Escherichia coli* cells by in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, v.123, p.8895-8901.
- SERBER, Z., SELENKO, P., HANSEL, R., RECKEL, S., LOHR, F., FERRELL, J.E JR., WAGNER, G., DOTSCH, V. (2006). Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat. Protoc.*, v.1, p.2701-2709.
- SERVICE, R.F. (2002). Tapping DNA structures produces a trickle. *Science*, v.298, p.948-950.
- SHARMA, S., HOSKINS, J.R., WICKNER, S. (2005). Binding and Degradation of Heterodimeric Substrates by ClpAP and Clp XP. *J. Biol. Chem.*, v.280, p.5449-5455.
- SHIMBA, N., YAMADA, N., YOKOYAMA, K., SUZUKI, E. (2002). Enzymatic labelling of arbitrary proteins. *Anal. Biochem.*, v.301, p.123-127.
- SIMON, K., XU, J., KIM, C., SKRYNNIKOV, NR. (2005). Estimating the accuracy of protein structures using residual dipolar couplings. *J. Biomol. NMR*, v.33, p. 83-93.
- SIMPSON, A.J.G. *et al.* (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, v.406, p.151-157.
- SOUSA, A.A., LAVERDE JR., A. (2003). Aplicação da espectroscopia de ressonância magnética nuclear para estudos de difusão molecular em líquidos: a técnica DOSY. *Quím. Nova*, v.25, p.1022-1026.

- SPHERES (2005). In [www.fccc.edu/research/labs/roder/sphere](http://www.fccc.edu/research/labs/roder/sphere)
- STALL, R.E., CIVEROLO, E.L. (1993). *Xanthomonas campestris* pv *citri*: cause of citrus canker. In *Xanthomonas* (Swings, J.G. e Civerolo, E.L., Eds.), p.48-51. London: Chapman e Hall.
- STEVENS, R.C., YOKOYAMA, S., WILSON, I.A. (2001). Global Efforts in Structural Genomics. *Science*, v.294, p.89-92.
- STUDIER, F.W., ROSENBERG, A.H., DUNN, J.J., DUBENDORFF, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Method. Enzymol.*, v.185, p.60-89.
- SZYK, A., MAURIZI, M. (2006). Crystal structure at 1.9 Å of *E. coli* ClpP with a peptide covalently bound at the active site. *J. Struct. Biol.*, v.156, p.165-174.
- SZYPERSKI, T., YE, D.C., SUKUMARAN, D.K., MOSELEY, H.N. & MONTELEONE G.T. (2002). Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. *Proc. Natl. Acad. Sci. U.S.A.*, v.99, p.8009-8014.
- THIEME, F., KOEBNIK, R., BEKEL, T., BERGER, C., BOCH, J., BÜTTNER, D., CALDANA, C., GAIGALAT, L., GOESMANN, A., KAY, S., KIRCHNER, O., LANZ, C., LINKE, B., MCHARDY, A.C., MEYER, F., MITTENHUBER, G., NIES, D.H., NIESBACH-KLÖSGEN U., PATSCHKOWSKI, T., RÜCKERT, C., RUPP, O., SCHNEIKER, S., SCHUSTER, S.C., VORHÖLTER, F.J., WEBER, E., PÜHLER, A., BONAS, U., BARTELS, D., KAISER, O. (2005). Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.*, v.187, p. 7254-7266.
- THOMPSON, M.W., MAURIZI, M.E. (1994). Activity and specificity of *Escherichia coli* ClpAP protease in cleaving model peptide substrates. *J. Biol. Chem.*, v.269, p.18201-18208.
- TIAN, F., VALAFAR, H., PRESTEGARD, J.H. (2001). A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. *J. Am. Chem. Soc.*, v.123, p.11791-11796.
- TJANDRA, N. (1999). Establishing a degree of order: obtaining high-resolution NMR structures from molecular alignment. *Structure*, v.7, p.205-211.
- TJANDRA, N., BAX, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, v.278, p.1111-1114.
- TJANDRA, N., OMICHINSKI, J.G., GRONENBORN, A.M., CLORE, G.M., BAX, A. (1997). Use of dipolar  $^1\text{H}$ - $^{15}\text{N}$  and  $^1\text{H}$ - $^{13}\text{C}$  couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.*, v.4, p.732-738.
- TOBIAS, J. W., SHRADER, J., ROCAP, G., VARSHAVSKY, A. (1991). The N-end rule in bacteria. *Science*, v.254, p.1374-1377.
- TOLMAN, J.R. (2001). Dipolar couplings as a probe of molecular dynamics and structure in solution. *Curr. Opin. Struct. Biol.*, v.11, p.532-539.
- TYERS, M., MANN, M. (2003). From Genomics to proteomics. *Nature*, v.422, p.193-197.
- VARSHAVSKY, A. (1996). The N-end rule: functions, mysteries, uses. *Proc Nat. Acad. Sci. U.S.A.*, v.93, p.12142-12149.
- VICARI, D., ARTSIMOVITCH, I. (2004). Virulence regulators RfaH and YaeQ do not operate in the same pathway. *Mol. Genet Genomics*, v.272, p.489-496.
- VILLASENOR, A., SHA, M., THANA, P., BROWNER, M. (2002). Fast drops: a high-throughput approach for setting up protein crystal screens. *Biotechniques*, v.32, p.184, 186, 188-189.
- VITKUD, D., MELAMUD, E., MOULT, J., SANDER, C. (2001) Jun. Completeness in structural genomics. *Nat. Struct. Biol.*, v.8, p.482-484.



- VUISTER, G.W., BAX, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J( $H^N H^O$ ) coupling constants in  $^{15}N$ -enriched proteins. *J. Am. Chem. Soc.*, v.115, p.7772-7777.
- VUISTER, G.W. TESSARI, M., KARIMI-NEJAD, Y., WHITEHEAD, B. (2002). Pulse sequences for measuring coupling constants. In *Biological Magnetic Resonance*, v.16. *Modern Techniques in Protein NMR*, Ed. Rama N. Krishna and Lawrence J. Berliner. Springer, US.
- WAGNER, G. (1995). The importance of being floppy. *Nat. Struct. Biol.*, v.2, p.255-257.
- WAH, D.A., LEVCHENKO, I., BAKER, T.A., SAUER, R.T. (2002). Characterization of a specificity factor for an AAA+ ATPase: assembly of SspB dimers with ssrA-tagged proteins and the ClpX hexamer. *Chem. Biol.*, v.9, p.1237-1245.
- WAHL, M.C., BOURENKOV, G.P., BARTUNIK, H.D., HUBER, R. (2000). Flexibility, conformational diversity and two dimerization modes in complexes of ribosomal protein L12. *EMBO J.*, v.19, p.174-186.
- WALDO, G.S., STANDISH, B.M., BERENDZEN, J., TERWILLIGER, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.*, v.17, p.691-695.
- WANG, J., HARTLING J.A., FLANAGAN, J.M. (1998). Crystal structure determination of *Escherichia coli* ClpP starting from an EM-derived mask. *J. Struct. Biol.*, v.124, p.151-163.
- WANG, J., HARTLING, JAMES A., FLANAGAN, JOHN M. (1997). The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell*, v.91, p.447-456.
- WANG, J., LIU, H. (2007). A Bayesian-probability-based method for assigning protein backbone dihedral angles based on chemical shifts and local sequences. *J. Biomol. NMR*, v.37, p.31-41.
- WANG, K.H., SAUER, R.T., BAKER, T.A. (2007). ClpS modulates but is not essential for bacterial N-end rule degradation. *Genes Dev.*, v.21, p.403-408.
- WEBER-BAN, E.U., REID, B.G., MIRANKER, A.D., HORWICH, A.L. (1999). GLOBAL UNFOLDING OF A SUBSTRATE PROTEIN BY THE HSP100 CHAPERONE CLPA *Nature*, v.401, p.90-93.
- WEIBEZAHN, J., BUKAU, B., MOGK, A. (2004). Unscrambling an egg: protein disaggregation by AAA+ proteins. *Microb. Cell. Fact.*, v.3, p. 1-12.
- WEICHAERT, D., QUERFURTH, N., DREGER, M., HENGGE-ARONIS, R. (2003). Global Role for ClpP-Containing Proteases in Stationary-Phase Adaptation of *Escherichia coli*. *J. Bacteriol.*, v.185, p.115-125.
- WICKNER, S., MAURIZI, M.R., GOTTESMAN, S. (1999). Post-translational quality control: folding, refolding and Degrading Proteins. *Science*, v.286, p.1888-1893.
- WICKNER, S., MAURIZI, M.R., GOTTESMAN, S. (1999). Posttranslational quality control: folding, refolding, and degrading proteins. *Science*, v.286, p.1888-1893.
- WIERUSZESKI, J.M., BOHIN, A., BOHIN, J.P., LIPPENS, G. (2001). In vivo detection of the cyclic osmoregulated periplasmic glucan of *Ralstonia solanacearum* by high-resolution magic angle spinning NMR. *J. Magn. Reson.*, v.151, p.118-123.
- WILKINS, D.K., GRIMSHAW, S.B., RECEVEUR, V., DOBSON, C.M., JONES, J.A., SMITH, L.J (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, v.38, p.16424-16431.
- WILKINS, M.R. *et al.* (1996). From proteins to proteomics: large-scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology*, v.14, p.61-65.
- WILLIAMSON, M.P., HAVEL, T.F., WÜTRICH, K. (1985). Solution conformation and proteinase inhibitor IIA from bull seminal plasma by proton NMR and distance geometry. *J. Mol. Biol.*, v.182, p.295-315.

- WISHART, D.S., SYKES, B.D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.*, v.239, p.363-392.
- WISHART, D.S., SYKES, B.D., RICHARDS, F.M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.*, v.222, p.311-333.
- WISHART, D.S., SYKES, B.D., RICHARDS, F.M. (1992). The Chemical Shift Index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, v.3, p.1647-1651.
- WONG, K.R., HUGHES, C., KORONAKIS, V. (1998). A gene, *vaeQ*, that suppresses reduced operon expression caused by mutations in the transcription elongation gene *rfaH* in *Escherichia coli* and *Salmonella typhimurium*. *Mol. Gen. Genet.*, v.257, p.693-696.
- WONG, K.R., HUGHES, C., KORONAKIS, V. (1998). A gene, *yaeQ*, that suppresses reduced operon expression caused by mutations in the transcription elongation gene *rfaH* in *Escherichia coli* and *Salmonella typhimurium*. *Mol. Gen. Genet.*, v.257, p. 693-696.
- WOO, K. M., CHUNG, W. J., HA, D.B., GOLDBERG, A.L., CHUNG, C.H. (1989). Protease TI from *Escherichia coli* requires ATP hydrolysis for protein breakdown but not for hydrolysis of small peptides *J. Biol. Chem.*, v.264, p.2088-2091.
- WUTHRICH, K. (1986). *In* NMR of proteins and nucleic acids. Jon Wiley and Son, New York.
- WUTHRICH, K. (1990). Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.*, v.265, p.22059-22062
- XIA, D., ESSER, L., SINGH, S., GUO, F., MAURIZI, M.R. (2004). Crystallographic investigation of peptide binding sites in the N-domain of the ClpA chaperone. *J. Struct. Biol.*, v.146, p.166-179.
- YEE, A., CHANG, X., PINEDA-LUCENA, A., WU, B., SEMESI, A., LE, B., RAMELOT, T., LEE, G.M., BHATTACHARYYA, S., GUTIERREZ, P., DENISOV, A., LEE, C.H., CORT, J.R., KOZLOV, G., LIAO, J., FINAK, G., CHEN, L., WISHART, D., LEE, W., MCINTOSH, L.P., GEHRING, K., KENNEDY, M.A. EDWARDS, A.M., ARROWSMITH, C.H. (2002). An NMR approach to structural proteomics. *Proc. Natl. Acad. Sci. U.S.A.*, v.99, p.1825-1830.
- YOKOYAMA, S. (2003). Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.*, v.7, p.39-43.
- ZARTLER, E., JENNEY, F.E., TERRELL, M., EIDSNESS, M.K., ADAMS, M.W.W., PRESTEGART, J.H. (2001). Structural basis for thermostability in aporubredoxins from *Pyrococcus furiosus* and *Clostridium pasteurianum*. *Biochemistry*, v.40, p.7279-7290.
- ZETH, K., RAVELLI, R.B., PAAL, K. CUSAK, S., BUKAU, B., DOUGAN, D.A. (2002). Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat. Struct. Biol.*, v.9, p.906-911.
- ZHANG, C., KIM, S.H. (2003). Overview of structural genomics: from structure to function. *Curr. Opin. Chem. Biol.*, v.7, p. 28-32.
- ZHOU, S., GOTTESMAN, J.R., HOSKINS, M.R., MAURIZI AND S. WICKNER. (2001). The RssB response regulator directly targets sigma (S) for degradation by ClpXP. *Genes Dev.*, v.15, p.627-637.
- ZHU, H., SNYDE, M. (2001). Protein arrays and microarrays. *Curr. Opin. Chem. Biol.*, v.5, p.40-45.
- ZUIDERWEG, E.R.P. (2002). Mapping protein-protein interactions by NMR spectroscopy. *Biochemistry*, v.41, p.1-7.
- ZWECKSTETTER, M., BAX, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.*, v.122, p.3791-3792.

---

ZWECKSTETTER, M., BAX, A. (2001) Single-step determination of protein substructures using dipolar coupling: Aid to structural genomics. *J. Am. Chem. Soc.*, v.123, p.9490-9491.

## **APÊNDICES**

## APÊNDICE 1

**Tabela 1.** Atribuição dos núcleos  $^1\text{H}$ ,  $^{15}\text{N}$  e  $^{13}\text{C}$  da proteína ClpS de *Xac*.

<u>Núcleo</u>	<u>Deslocamento químico (ppm)</u>
2.CA	62.868
2.CB	31.921
2.C	175.253
3.N	119.320
3.HN	8.562
3.CA	56.273
3.HA	4.386
3.CB	30.679
3.HB2	1.800
3.HB1	1.800
3.CG	26.883
3.HG2	1.649
3.HG1	1.649
3.CD	43.026
3.HD2	3.184
3.HD1	3.184
3.C	171.811
4.N	119.532
4.HN	8.088
4.CA	53.279
4.HA	4.484
4.CB	38.645
4.C	174.000
5.N	112.065
5.HN	8.281
5.CA	61.551
5.HA	4.362
5.CB	69.030
5.HB	3.935
5.HG21	1.176
6.N	115.372
6.HN	8.397
6.CA	58.300
6.C	173.375
7.N	115.586
7.HN	8.407
7.CA	56.008
7.HA	4.570
7.CB	30.195
7.HB2	3.124
7.HB1	3.219
7.C	174.002
8.N	119.000
8.HN	8.372
8.CA	56.932
8.HA	4.161
8.CB	29.952
8.HB2	1.903
8.HB1	1.837
8.CG	36.003
8.HG2	2.095
8.HG1	2.095

---

8.C	174.947
9.N	116.653
9.HN	8.338
9.CA	56.151
9.HA	4.832
9.CB	30.392
9.HB2	3.101
9.HB1	3.101
9.C	173.685
10.N	118.573
10.HN	8.220
10.CA	54.087
10.HA	4.562
10.CB	40.861
10.HB2	2.587
10.HB1	2.634
10.C	174.935
11.N	117.613
11.HN	8.400
11.CA	56.328
11.HA	4.626
11.CB	30.275
11.HB2	3.023
11.HB1	3.098
11.C	174.313
12.N	106.623
12.HN	8.481
12.CA	45.325
12.HA2	3.903
12.HA1	3.903
12.C	172.745
13.N	119.000
13.HN	8.104
13.CA	55.014
13.HA	4.352
13.CB	42.232
13.HB2	1.630
13.HB1	1.630
13.CG	26.932
13.HG	1.640
13.CD1	24.694
13.HD11	0.909
13.CD2	24.695
13.HD21	0.909
13.C	175.876
14.N	118.893
14.HN	8.384
14.CA	55.202
14.HA	4.491
14.CB	32.424
14.HB2	2.035
14.HB1	2.068
14.CG	31.904
14.HG2	2.573
14.HG1	2.511
14.C	174.625
15.N	119.149
15.HN	8.130
15.CA	62.111
15.HA	4.087
15.CB	32.768

---

15.HB	2.051
15.CG2	20.379
15.HG21	0.922
15.CG1	20.986
15.HG11	0.933
15.C	174.625
16.N	122.414
16.HN	8.525
16.CA	56.326
16.HA	4.264
16.CB	30.206
16.HB2	1.944
16.HB1	2.028
16.CG	36.208
16.HG2	2.266
16.HG1	2.266
16.C	174.638
17.N	123.054
17.HN	8.383
17.CA	52.294
17.HA	4.325
17.CB	19.092
17.HB1	1.405
17.C	176.187
18.N	112.918
18.HN	8.284
18.CA	58.029
18.HA	4.408
18.CB	63.500
18.HB2	3.849
18.HB1	3.849
18.C	172.749
19.N	121.774
19.HN	8.342
19.CA	54.007
19.HA	4.643
19.CB	32.409
19.HB2	1.736
19.HB1	1.850
19.CG	24.346
19.HG2	1.481
19.HG1	1.481
19.CD	28.917
19.HD2	1.719
19.HD1	1.719
19.CE	41.769
19.HE2	3.019
19.HE1	3.019
20.CA	62.963
20.HA	4.420
20.CB	31.941
20.HB2	1.910
20.HB1	2.290
20.CG	27.156
20.HG2	2.023
20.HG1	2.023
20.CD	50.444
20.HD2	3.657
20.HD1	3.815
20.C	175.562
21.N	119.000

---

21.HN	8.593
21.CA	56.326
21.HA	4.264
21.CB	30.398
21.HB2	1.944
21.HB1	2.032
21.CG	36.208
21.HG2	2.320
21.HG1	2.320
21.C	172.745
22.N	119.000
22.HN	8.182
22.CA	61.630
22.HA	4.116
22.CB	32.768
22.HB	2.062
22.CG2	20.379
22.HG21	0.939
22.CG1	20.985
22.HG11	0.960
22.C	173.999
23.N	127.002
23.HN	8.410
23.CA	50.110
23.HA	4.606
23.CB	17.993
23.HB1	1.353
26.CA	62.882
26.CB	32.355
26.C	174.942
27.N	118.360
27.HN	8.455
27.CA	54.582
27.HA	4.680
27.CB	32.609
27.HB2	1.606
27.HB1	1.531
27.CG	27.128
27.HG2	1.764
27.HG1	1.764
27.CD	43.322
27.HD2	3.082
27.HD1	3.082
27.C	173.375
28.N	116.013
28.HN	9.365
28.CA	57.309
28.HA	4.781
28.CB	41.667
28.HB2	2.634
28.HB1	2.585
28.C	173.384
29.N	118.146
29.HN	9.318
29.CA	53.451
29.HA	5.233
29.CB	30.169
29.HB2	1.969
29.HB1	2.069
29.CG	33.521
29.HG2	2.265



---

29.HG1	2.265
29.C	173.695
30.N	120.814
30.HN	8.549
30.CA	61.040
30.HA	4.793
30.CB	33.408
30.HB	1.895
30.CG2	20.765
30.HG21	0.697
30.CG1	21.636
30.HG11	0.706
30.C	173.058
31.N	125.295
31.HN	9.318
31.CA	53.728
31.HA	5.454
31.CB	45.508
31.HB2	1.217
31.HB1	1.679
31.CG	27.280
31.HG	1.463
31.CD1	25.957
31.HD11	0.703
31.CD2	25.957
31.HD21	0.703
31.C	174.938
32.N	115.692
32.HN	8.208
32.CA	53.256
32.HA	5.121
32.CB	44.783
32.HB2	1.237
32.HB1	1.397
32.HG	1.593
32.CD1	23.230
32.HD11	0.928
32.CD2	23.230
32.HD21	0.928
32.C	174.311
33.N	118.679
33.HN	8.101
33.CA	53.302
33.HA	4.433
33.CB	42.722
33.C	175.878
34.N	117.079
34.HN	7.848
34.CA	53.068
34.HA	4.766
34.CB	40.669
34.HB2	2.776
34.HB1	2.842
34.C	173.057
35.N	117.400
35.HN	7.885
35.CA	54.073
35.HA	4.720
35.CB	42.762
35.HB2	2.863
35.HB1	2.863

---

35.C	173.373
36.N	114.199
36.HN	8.390
36.CA	55.176
36.HA	4.522
36.CB	41.058
36.HB2	2.005
36.HB1	2.487
36.C	174.001
37.N	117.186
37.HN	9.444
37.CA	59.218
37.HA	4.608
37.CB	41.689
37.HB2	3.117
37.HB1	3.117
39.CA	62.397
39.HA	4.386
39.CB	32.402
39.HB2	2.047
39.HB1	2.449
39.CG	26.472
39.HG2	2.033
39.HG1	2.033
39.CD	49.024
39.HD2	3.373
39.HD1	3.373
39.C	177.126
40.N	121.454
40.HN	9.005
40.CA	58.278
40.CB	33.324
40.CG	35.575
40.C	176.819
41.N	113.452
41.HN	8.755
41.CA	56.263
41.HA	4.350
41.CB	39.677
41.HB2	2.629
41.HB1	2.706
41.CG	175.685
41.C	177.132
42.N	118.787
42.HN	7.543
42.CA	61.269
42.HA	4.410
42.CB	39.112
42.HB2	3.113
42.HB1	2.908
42.C	174.947
43.N	116.759
43.HN	7.325
43.CA	67.285
43.HA	3.259
43.CB	31.065
43.HB	2.536
43.CG2	23.219
43.HG21	1.075
43.CG1	21.490
43.HG11	0.859

---

43.C	176.193
44.N	115.052
44.HN	7.771
44.CA	67.947
44.HA	3.385
44.CB	31.199
44.HB	2.157
44.CG2	22.117
44.HG21	0.952
44.CG1	21.141
44.HG11	0.796
44.C	175.883
45.N	113.879
45.HN	7.705
45.CA	66.759
45.HA	4.080
45.CB	68.538
45.HB	4.283
45.CG2	20.462
45.HG21	1.176
45.C	174.942
46.N	119.000
46.HN	8.184
46.CA	66.746
46.HA	3.502
46.CB	31.758
46.HB	2.289
46.CG2	23.981
46.HG21	0.772
46.CG1	22.853
46.HG11	1.053
46.C	176.814
47.N	118.573
47.HN	8.203
47.CA	58.011
47.HA	4.602
47.CB	40.981
47.C	179.939
48.N	120.494
48.HN	7.926
48.CA	59.211
48.HA	4.375
48.CB	30.486
48.HB2	2.058
48.C	178.064
49.N	113.559
49.HN	8.623
49.CA	58.498
49.HA	3.797
49.CB	28.719
49.HB2	1.245
49.HB1	1.655
49.CG	33.056
49.HG2	1.911
49.HG1	1.510
49.C	176.813
50.N	110.144
50.HN	8.457
50.CA	59.138
50.HA	4.247
50.CB	39.392

---

50.HB2	2.510
50.HB1	2.510
50.C	173.971
51.N	111.318
51.HN	6.922
51.CA	57.929
51.HA	4.781
51.CB	38.740
51.HB2	3.272
51.HB1	3.272
51.CG	138.372
51.C	173.062
52.N	112.172
52.HN	7.535
52.CA	54.284
52.HA	4.530
52.CB	36.923
52.HB2	2.739
52.HB1	3.121
52.CG	178.390
52.C	172.755
53.HN	7.304
53.CA	54.119
53.HA	4.710
53.CB	42.745
53.HB2	1.409
53.HB1	1.409
53.C	175.563
54.N	116.439
54.HN	8.482
54.CA	51.124
54.HA	4.766
54.CB	38.563
54.HB2	2.806
54.HB1	2.842
54.C	173.687
55.N	116.653
55.HN	8.306
55.CA	58.401
55.HA	4.424
55.CB	41.171
55.C	178.377
56.N	119.747
56.HN	8.496
56.CA	60.013
56.HA	3.994
56.CB	28.925
56.HB2	2.110
56.HB1	2.110
56.CG	36.417
56.HG2	2.264
56.HG1	2.264
56.C	178.072
57.N	117.293
57.HN	8.661
57.CA	58.247
57.HA	4.055
57.CB	29.726
57.HB2	1.982
57.HB1	2.291
57.CG	33.975

---

57.HG2	2.443
57.HG1	2.673
57.C	177.764
58.N	119.000
58.HN	9.353
58.CA	54.584
58.HA	3.945
58.CB	19.743
58.HB1	1.482
58.C	177.750
59.N	111.958
59.HN	8.551
59.CA	67.446
59.HA	3.689
59.CB	69.404
59.HB	4.277
59.CG2	21.345
59.HG21	1.164
59.C	174.000
60.N	117.079
60.HN	7.443
60.CA	58.624
60.HA	3.979
60.CB	28.155
60.HB2	2.107
60.HB1	2.107
60.CG	33.581
60.HG2	2.338
60.HG1	2.338
60.C	176.502
61.N	116.4391
61.HN	7.640
61.CA	63.835
61.HA	3.627
61.CB	37.019
61.HB	1.781
61.CG1	28.643
61.HG12	1.534
61.HG11	1.322
61.CD1	13.363
61.HD11	0.840
61.CG2	17.812
61.HG21	0.659
61.C	176.814
62.N	119.000
62.HN	8.472
62.CA	59.342
62.HA	4.460
62.CB	30.290
62.HB2	1.998
62.HB1	2.201
62.C	176.501
63.N	118.040
63.HN	8.112
63.CA	58.788
63.HA	4.137
63.CB	41.057
63.HD11	0.895
63.HD21	0.895
63.C	177.751
64.N	115.694

---

64.HN	8.420
64.CA	59.103
64.HA	4.582
64.CB	30.990
64.HB2	3.116
64.HB1	3.116
64.C	176.505
65.N	118.253
65.HN	8.493
65.CA	67.109
65.HA	2.950
65.CB	30.945
65.HB	2.340
65.CG2	22.103
65.HG21	0.907
65.CG1	20.734
65.HG11	0.667
65.C	176.195
66.N	112.065
66.HN	8.377
66.CA	58.500
66.HA	4.359
66.CB	30.267
66.HB2	3.429
66.HB1	3.244
66.C	175.243
67.N	105.557
67.HN	8.369
67.CA	63.521
67.HA	4.273
67.CB	70.475
67.HB	4.228
67.CG2	21.836
67.HG21	1.372
67.C	174.930
68.N	117.453
68.HN	8.842
68.CA	55.602
68.HA	4.652
68.CB	31.199
68.HB1	2.108
68.HG2	1.689
68.HG1	1.7392
68.HD1	3.151
68.C	175.876
69.N	105.877
69.HN	7.546
69.CA	46.016
69.HA2	4.460
69.HA1	3.861
69.C	172.122
70.N	113.772
70.HN	7.264
70.CA	54.136
70.HA	5.501
70.CB	32.669
70.HB2	1.544
70.HB1	1.528
70.HG2	2.048
70.HG1	2.099
70.HD2	3.020

---

70.HD1	3.088
70.C	172.124
71.N	106.517
71.HN	8.582
71.CA	44.038
71.HA2	4.659
71.HA1	3.331
71.C	169.303
72.N	122.414
72.HN	9.017
72.CA	62.622
72.HA	4.068
72.CB	31.523
72.HB	1.910
72.CG2	20.946
72.HG21	0.759
72.CG1	20.944
72.HG11	0.833
72.C	174.068
73.N	123.535
73.HN	9.289
73.CA	59.640
73.HA	3.811
73.CB	27.584
73.HB2	1.917
73.HB1	1.886
73.C	172.594
74.N	99.155
74.HN	7.063
74.CA	44.316
74.HA2	4.107
74.HA1	3.846
74.C	168.685
75.N	116.759
75.HN	7.535
75.CA	61.132
75.HA	4.720
75.CB	33.761
75.HB	1.653
75.CG2	21.485
75.HG21	0.750
75.CG1	21.841
75.HG11	0.674
75.C	173.376
76.N	121.027
76.HN	9.085
76.CA	55.965
76.HA	5.125
76.CB	45.332
76.HB2	3.586
76.HB1	2.598
76.C	173.999
77.N	112.065
77.HN	8.592
77.CA	58.069
77.CB	63.960
77.HB1	3.089
77.C	173.371
78.N	119.534
78.HN	9.311
78.CA	60.670

---

78.CB	29.855
78.C	175.875
79.N	113.025
79.HN	9.202
79.CA	59.997
79.HA	3.997
79.CB	28.845
79.HB2	2.083
79.HB1	2.083
79.CG	36.317
79.HG2	2.299
79.HG1	2.299
79.C	178.374
80.N	118.787
80.HN	7.292
80.CA	65.599
80.HA	3.665
80.CB	32.056
80.HB	2.193
80.CG2	22.109
80.HG21	0.993
80.CG1	21.372
80.HG11	0.686
80.C	176.504
81.N	120.707
81.HN	8.861
81.CA	55.867
81.HA	3.676
81.CB	18.941
81.HB1	1.457
81.C	176.819
82.N	112.705
82.HN	8.577
82.CA	59.432
82.HA	3.777
82.CB	29.676
82.HB2	1.884
82.HB1	1.884
82.C	178.378
83.N	112.172
83.HN	7.391
83.CA	60.843
83.HA	4.306
83.CB	62.751
83.HB2	3.985
83.HB1	3.982
83.C	175.247
84.N	119.320
84.HN	8.336
84.CA	58.587
84.HA	3.920
84.CB	32.017
84.CG	24.945
84.HG2	1.504
84.HG1	1.504
84.CD	26.899
84.HD2	1.725
84.HD1	1.725
84.CE	43.430
84.HE2	2.968
84.HE1	2.968



---

84.C	176.503
85.N	115.692
85.HN	8.366
85.CA	66.374
85.HA	3.591
85.CB	32.115
85.HB	2.059
85.CG2	23.340
85.HG21	1.030
85.C	176.192
86.N	116.333
86.HN	7.472
86.CA	54.790
86.HA	4.177
86.CB	17.889
86.HB1	1.514
86.C	179.643
87.N	115.159
87.HN	8.215
87.CA	58.885
87.HA	3.979
87.CB	29.042
87.HB2	2.057
87.HB1	2.057
87.CG	34.373
87.HG2	2.224
87.HG1	2.606
87.C	178.065
88.N	118.893
88.HN	8.538
88.CA	66.588
88.HA	2.854
88.CB	31.610
88.HB	1.971
88.CG2	21.556
88.HG21	0.879
88.CG1	23.862
88.HG11	0.649
88.C	176.812
89.N	115.793
89.HN	8.394
89.CA	56.601
89.HA	4.546
89.CB	36.510
89.C	177.124
90.N	118.040
90.HN	8.0831
90.CA	59.624
90.HA	4.250
90.CB	29.380
90.HB2	2.143
90.HB1	2.143
90.CG	36.633
90.HG2	2.390
90.HG1	2.390
90.CD	174.303
90.C	176.806
91.N	119.903
91.HN	8.122
91.CA	61.558
91.HA	4.520

---

91.CB	40.403
91.HB2	3.381
91.HB1	3.185
91.C	178.377
92.N	117.186
92.HN	9.376
92.CA	63.496
92.HA	4.410
92.C	174.942
93.N	119.534
93.HN	8.283
93.CA	59.041
93.HA	4.221
93.CB	28.906
93.HB2	2.183
93.HB1	2.138
93.HG2	2.390
93.HG1	2.390
93.C	180.257
94.N	117.613
94.HN	8.592
94.CA	58.076
94.HA	4.160
94.CB	32.078
94.HB2	2.277
94.HB1	2.139
94.CG	31.808
94.HG2	2.720
94.HG1	2.646
94.C	175.866
95.N	113.772
95.HN	7.261
95.HA	4.340
95.CB	40.395
95.C	169.629
96.N	108.864
96.HN	7.652
96.CA	56.384
96.HA	3.658
96.CB	25.326
96.HB2	1.983
96.HB1	2.083
96.CG	34.080
96.HG2	2.077
96.HG1	2.084
96.C	172.745
97.N	112.918
97.HN	7.745
97.CA	52.221
97.HA	5.479
97.CB	31.938
97.HB2	2.290
97.HB1	2.290
98.CA	62.919
98.HA	4.766
98.CB	30.802
98.HB2	1.950
98.HB1	2.450
98.CG	27.373
98.HG2	2.107
98.HG1	2.107

---

98.CD	50.416
98.HD2	3.897
98.HD1	3.657
98.C	173.062
99.N	118.893
99.HN	6.080
99.CA	55.626
99.CB	41.950
100.CA	54.089
100.C	172.120
101.N	123.054
101.HN	8.863
101.CA	55.855
101.HA	5.624
101.CB	28.640
101.HB2	2.707
101.HB1	2.438
101.C	173.061
102.N	118.787
102.HN	9.108
102.CA	59.545
102.HA	4.828
102.CB	70.588
102.HB	4.166
102.CG2	19.990
102.HG21	1.091
102.C	170.248
103.N	113.239
103.HN	8.473
103.CA	53.598
103.HA	5.992
103.CB	37.187
103.HB2	2.034
103.HB1	1.938
103.CG	35.820
103.HG2	2.145
103.HG1	2.052
103.C	173.062
104.N	116.973
104.HN	8.891
104.CA	54.429
104.HA	4.812
104.CB	34.343
104.HB2	2.133
104.HB1	1.835
104.CG	35.683
104.HG2	2.278
104.HG1	2.278
104.C	173.376
105.N	122.628
105.HN	9.098
105.CA	57.591
105.HA	3.540
105.CB	28.540
105.HB2	1.866
105.HB1	1.866
105.CG	34.138
105.HG2	1.996
105.HG1	2.344
105.C	173.376
106.N	132.043

106.HN	8.480
106.CA	53.117
106.HA	4.068
106.CB	19.990
106.HB1	1.115

## APÊNDICE 2

**Tabela 12** Predição dos ângulos diedros e da estrutura secundária da proteína ClpS de Xac usando o programa TALOS.

Resíduo	$\Phi$	$\Psi$	Predição	Estrutura 2°
M 1	9999.000	9999.000	Nenhuma	
P 2	-69.000	140.000	Ambígua	
R 3	-116.000	135.000	Ambígua	
N 4	-98.000	131.000	Ambígua	
T 5	-90.000	-16.000	Ambígua	
S 6	-82.000	148.000	Ambígua	
H 7	-94.000	151.000	Ambígua	
E 8	-92.000	152.000	Ambígua	
H 9	-97.000	149.000	Ambígua	
D 10	-88.000	130.000	Ambígua	
H 11	-84.000	-24.000	Ambígua	
G 12	89.000	9.000	Ambígua	
L 13	-95.000	95.000	Ambígua	
M 14	-96.000	144.000	Ambígua	
V 15	-102.000	141.000	Ambígua	
E 16	-102.000	139.000	Ambígua	
A 17	-100.000	145.000	Ambígua	
S 18	-99.000	127.000	Ambígua	
K 19	-110.000	122.000	Ambígua	
P 20	-69.000	143.000	Ambígua	
E 21	-105.000	140.000	Ambígua	
V 22	-110.000	114.000	Ambígua	
A 23	-102.000	123.000	Ambígua	
P 24	-64.000	153.000	Boa	Fita- $\beta$
P 25	-70.000	145.000	Boa	Fita- $\beta$
P 26	-84.000	138.000	Ambígua	
R 27	-112.000	133.000	Ambígua	
Y28	-126.000	150.000	Boa	Fita- $\beta$
Q 29	-120.000	138.000	Boa	Fita- $\beta$
V 30	-118.000	131.000	Boa	Fita- $\beta$
L 31	-110.000	150.000	Boa	Fita- $\beta$
L 32	-122.000	140.000	Boa	Fita- $\beta$
L 33	-95.000	140.000	Ambígua	
N 34	-103.000	151.000	Boa	Fita- $\beta$
D 35	-98.000	154.000	Ambígua	
D 36	-113.000	133.000	Ruim	
Y 37	-98.000	125.000	Ruim	
T 38	-110.000	117.000	Ambígua	
P 39	-60.000	144.000	Ambígua	
M 40	-65.000	-25.000	Boa	$\alpha$ -hélice
D 41	-66.000	-34.000	Boa	$\alpha$ -hélice
F 42	-69.000	-38.000	Boa	$\alpha$ -hélice
V 43	-70.000	-40.000	Boa	$\alpha$ -hélice
T 45	-68.000	-39.000	Boa	$\alpha$ -hélice
V 44	-72.000 -36.000	10.000	Boa	
V 46	-66.000	-43.000	Boa	$\alpha$ -hélice
L 47	-66.000	-40.000	Boa	$\alpha$ -hélice
E 48	-67.000	-37.000	Boa	$\alpha$ -hélice

Resíduo	$\Phi$	$\Psi$	Predição	Estrutura 2 <sup>o</sup>
Q 49	-71.000	-23.000	Boa	$\alpha$ -hélice
F 50	-88.000	-19.000	Boa	$\alpha$ -hélice
F 51	-105.000	-15.000	Ambígua	
N 52	64.000	21.000	Ruim	
L 53	-112.000	140.000	Ruim	
N 54	-97.000	159.000	Ruim	
L 55	-71.000	-29.000	Boa	$\alpha$ -hélice
E 56	-66.000	-42.000	Boa	$\alpha$ -hélice
Q 57	-69.000	-31.000	Boa	$\alpha$ -hélice
A 58	-66.000	-39.000	Boa	$\alpha$ -hélice
T 59	-73.000	-38.000	Boa	$\alpha$ -hélice
Q 60	-64.000	-36.000	Boa	$\alpha$ -hélice
I 61	-68.000	-37.000	Boa	$\alpha$ -hélice
M 62	-62.000	-39.000	Boa	$\alpha$ -hélice
L 63	-64.000	-40.000	Boa	$\alpha$ -hélice
H 64	-67.000	-42.000	Boa	$\alpha$ -hélice
V 65	-69.000	-38.000	Boa	$\alpha$ -hélice
H 66	-66.000	-40.000	Boa	$\alpha$ -hélice
T 67	-94.000	-19.000	Ambígua	
R 68	-102.000	-8.000	Ambígua	
G 69	76.000	20.000	Ruim	
R 70	-140.000	154.000	Boa	Fita- $\beta$
G 71	-135.000	141.000	Boa	Fita- $\beta$
V 72	-99.000	120.000	Boa	Fita- $\beta$
C 73	-75.000	-31.000	Ruim	
G 74	178.000	145.000	Ruim	
V 75	-123.000	136.000	Boa	Fita- $\beta$
Y 76	-127.000	160.000	Boa	Fita- $\beta$
S 77	-102.000	152.000	Ambígua	
R 78	-64.000	-37.000	Boa	$\alpha$ -hélice
E 79	-66.000	-37.000	Boa	$\alpha$ -hélice
V 80	-75.000	-39.000	Boa	$\alpha$ -hélice
A 81	-64.000	-46.000	Boa	$\alpha$ -hélice
E 82	-61.000	-40.000	Boa	$\alpha$ -hélice
S 83	-67.000	-30.000	Boa	$\alpha$ -hélice
K 84	-65.000	-38.000	Boa	$\alpha$ -hélice
V 85	-71.000	-36.000	Boa	$\alpha$ -hélice
A 86	-59.000	-42.000	Boa	$\alpha$ -hélice
Q 87	-67.000	-42.000	Boa	$\alpha$ -hélice
V 88	-65.000	-39.000	Boa	$\alpha$ -hélice
N 89	-64.000	-33.000	Boa	$\alpha$ -hélice
E 90	-63.000	-46.000	Boa	$\alpha$ -hélice
F 91	-65.000	-41.000	Boa	$\alpha$ -hélice
S 92	-65.000	-42.000	Boa	$\alpha$ -hélice
R 93	-67.000	-34.000	Boa	$\alpha$ -hélice
M 94	-70.000	-31.000	Boa	$\alpha$ -hélice
N 95	-85.000	-8.000	Ambígua	
Q 96	62.000	30.000	Ruim	
H 97	-122.000	151.000	Boa	$\alpha$ -hélice
P 98	97.000	150.000	Ruim	
L 99	-100.000	139.000	Ruim	
L 100	-127.000	136.000	Boa	Fita- $\beta$
C 101	-108.000	123.000	Boa	Fita- $\beta$
T 102	-133.000	143.000	Boa	Fita- $\beta$

<b>Resíduo</b>	<b><math>\Phi</math></b>	<b><math>\Psi</math></b>	<b>Predição</b>	<b>Estrutura 2°</b>
M 103	-128.000	157.000	Boa	Fita- $\beta$
E 104	-140.000	148.000	Boa	Fita- $\beta$
Q 105	-76.000	122.000	Ambígua	
A 106	9999.000	9999.000	Nenhuma	

## LISTA DE ANEXOS

### ANEXO A

- Súmula Curricular

### ANEXO B

- Artigo: High-throughput screening of structural proteomics targets using NMR.

Galvão-Botton, L.M.P., Katusyama, A.M., Guzzo, C.R., Farah, C.S., Almeida F.C.L., Valente, A.P. (2003) FEBS Lett., v. 552, p.207-213.

### ANEXO C

- Artigo: Expression, purification, crystallization and preliminary X-ray analysis of YaeQ (XAC2396) from *Xanthomonas axonopodis* pv. *Citri*.

Guzzo, C.R., Nagem, R.A.P., Galvão-Botton, L.M.P., Guimarães, B.G., Medrano, F.J., Barbosa, J.A.R.G., Farah, C.S. (2005). Acta Cryst., F61, p.493-495.

### ANEXO D

- Artigo: Expression, crystallization and preliminary crystallographic analysis of SufE (XAC2355) from *Xanthomonas axonopodis* pv. *citri*.

Guzzo C.R., Silva L.R., Galvao-Botton L.M., Barbosa J.A., Farah C.S. (2006). Acta Cryst. F62:268-270.



# SÚMULA CURRICULAR

## DADOS PESSOAIS

Nome: Leonor Magalhães Peres Galvão de Botton  
Local e data de nascimento: Lisboa, 20 de junho de 1978

## EDUCAÇÃO

**St. Dominic's International School, Lisboa, Portugal**  
International Baccalaureate Diploma (Ensino Médio), 1996.

**Imperial College of Science, Technology and Medicine, Universidade de Londres, Londres, Reino Unido**  
Bacharel em Bioquímica ("First Class Honours"), 2000.

**Instituto de Química, Universidade de São Paulo, São Paulo**  
Doutoranda em Bioquímica na área de biologia estrutural sob orientação do Prof. Dr. Chuck Farah (março de 2001 até ao presente).

## FORMAÇÃO COMPLEMENTAR

**Organização Mundial de Propriedade Intelectual**  
Curso de especialização em Propriedade Intelectual (2002).

## OCUPAÇÃO

**Imperial College of Science, Technology and Medicine, Universidade de Londres, Reino Unido**  
Bolsista de graduação (1996-2000).

**União Química/ Biolab Sanus Farmacêutica, São Paulo**  
Estágio junto ao departamento de Pesquisa e Desenvolvimento desenvolvendo novas formulações de fármacos (agosto a outubro de 2000).

**Fundação para a Ciência e Tecnologia (MCT - Portugal)**  
Bolsista de Doutorado (março de 2001 a junho de 2005).

**Momsen, Leonardos & Cia, São Paulo**  
Especialista de Biotecnologia e Farmácia junto ao Departamento de Patentes da empresa (agosto de 2005 até ao presente).

## **OUTRAS ATIVIDADES**

Membro da **Comissão de Bioética, Biodireito e Biotecnologia da Ordem dos Advogados do Brasil** - Secção de São Paulo (agosto de 2004 até ao presente).

Vice-coordenadora da **Comissão de Biotecnologia da Associação Brasileira de Propriedade Intelectual (ABPI)** (janeiro de 2004 até ao presente).

Membro da **Associação Nacional de Bioquímicos de Portugal (ANBIOQ)** (julho de 2002 até ao presente).

Membro da **Sociedade Brasileira de Bioquímica e Biologia Molecular (SBBq)** (maio de 2001 até ao presente).

**Adjunta de Agente Oficial de Propriedade Industrial** em Portugal (fevereiro de 2000 até Novembro de 2006).

## **PRÊMIOS**

Primeiro prêmio do concurso "**Biotechnology Young Entrepreneurs Scheme**" organizado pelo British Biotechnology and Biological Sciences Research Council (Londres, Reino Unido, 1998).

Prêmio "*David and Lucille Packard Fellow*" de melhor trabalho de doutorado apresentado no congresso Structural Genomics: From Gene Sequence to Function" da **Keystone Symposia Organization** (Colorado, janeiro de 2002).

## **IDIOMAS**

Inglês fluente, francês bom, espanhol razoável .

## **PUBLICAÇÕES EM PERIÓDICOS**

Galvão-Botton, L.M.P., Katsuyama, A.M., Almeida, F.C.L., Farah, C.S. and Valente, A.P. (2003) NMR High-throughput screening of structural proteomics targets using NMR. **FEBS Lett.** 552:207-13.

Galvão de Botton, Leonor (2004). Patenteamento de substâncias naturais. **Fármacos e Medicamentos.** 30:31-41.

Guzzo C.R., Nagem R.A., Galvão-Botton L.M., Guimaraes B.G., Medrano F.J., Barbosa J.A., Farah C.S. (2005). Expression, purification, crystallization and preliminary X-ray analysis of YaeQ (XAC2396) from *Xanthomonas axonopodis* pv. *citri*. **Acta Crystallograph.** Sect. F Struct. Biol. Cryst. Commun. 61:493-495.

Guzzo C.R., Silva L.R., Galvão-Botton L.M., Barbosa J.A., Farah C. (2006). Expression, crystallization and preliminary crystallographic analysis of SufE (XAC2355) from *Xanthomonas axonopodis* pv. *citri*. **Acta Crystallograph.** Sect F Struct. Biol. Cryst. Commun. 62:268-270.

Galvão de Botton, Leonor (2007). Licença compulsória de patentes: fatos e mitos. **Boletim da Associação Brasileira de Propriedade Intelectual.** 85:2.

Galvão-Botton, L.M.P., Salinas, R.K., Farah, C.S., Spisni, A., Pertinhez, T.A. (2007). NMR structural studies of the ClpS protein of the phytopathogen *Xanthomonas axonopodis* pv *citri*. **Manuscrito em preparação.**

#### **COMUNICAÇÕES ORAIS EM CONGRESSOS (2002-2004)**

“Biotecnologia: novas conquistas, novos desafios”, **2º Forum Iberoamericano sobre Inovação, Propriedade Industrial e Intelectual e Desenvolvimento**, 28-30 abril 2003, Lisboa, Portugal.

“Biotecnología: patentabilidad de sustancias naturales”, **XVII Jornadas Anuales de Propiedad Industrial – Asociación Argentina de Agentes de Propiedad Industrial**, 1-2 de setembro de 2003, Buenos Aires, Argentina.

“High-throughput Screening of *Xanthomonas axonopodis* pv *citri* targets for structural proteomics using NMR”. **2nd Portuguese-Brazilian NMR Conference**, 22-26 de setembro de 2003, Sintra, Portugal.

“Patentes de microorganismos”, **Congresso Internacional da Propriedade Intelectual, Associação Paulista de Propriedade Intelectual**, 18-20 de outubro de 2003, Águas de Lindóia, SP, Brasil.

#### **RESUMOS EM CONGRESSOS**

Galvão-Botton L. M. P., Katsuyama, A. M. e Farah, C. S. “Structural genomics of the phytopathogen *Xanthomonas axonopodis* pv *citri*”. **Structural Genomics: From Gene Sequence to Function, Keystone Symposia**, 5-11 de Janeiro de 2002, Breckenridge, Colorado, EUA.

Galvão de Botton L.M.P., Katsuyama, A.M. e Farah, C.S. “Structural genomics of the phytopathogen *Xanthomonas axonopodis* pv. *citri*”. **XXXI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular (SBBq)**, 18-21 de maio de 2002, Caxambu, MG, Brasil.

Katsuyama, A.M.; Farah, C.S.; Galvão-Botton, L. M.P., Almeida, F.C.L.; Valente, A.P.; Spisni, A. e Pertinhez, T.A. “Structural genomics of phytopatogen *Xanthomonas axonopodis* pv. *citri* – NMR characterization of ApaG/CorD protein” **Biophysical Society, 47<sup>th</sup> Annual Meeting**, 1-5 de março de 2003, San Antonio, Texas, EUA.

Guzzo, C.R., Galvão de Botton L.M.P., Medrano-Martin F.X. e Farah, C. S. “Purification and crystallization of three *Xanthomonas axonopodis* pv *citri* proteins for structural studies ”. **XXXII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular (SBBq)**, 17-20 de maio de 2003, Caxambu, MG, Brasil.

Galvão de Botton L. M. P., Katsuyama, Guzzo, C.R., Almeida, F. L., Farah, C. S. e Valente, A.P. “High-throughput Screening of *Xanthomonas axonopodis* pv *citri* targets for structural proteomics using NMR”. **XXXII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular (SBBq)**, 17-20 de maio de 2003, Caxambu, MG, Brasil.

Galvão de Botton, L. M. P., Katsuyama, A.M. Guzzo, C.R., Almeida, F. L., Farah, C. S. e Valente, A.P. “High-throughput Screening of *Xanthomonas axonopodis* pv *citri* targets for structural proteomics using NMR”. **2nd Portuguese-Brazilian NMR Conference**, 22-26 de setembro de 2003, Sintra, Portugal.

Galvão-Botton, L.M.P., Sforça M., Farah, C.S., Spisni, A. and Pertinhez, T.A. “NMR Structural studies of the ClpS protein of the phytopathogen *Xanthomonas axonopodis* pv *citri*.” **1st Latin American Protein Society Meeting**, 8-12 novembro de 2004, Angra dos Reis, RJ, Brasil.

Guzzo, C.R., Nagem, R.A.P., Galvão-Botton, L.M.P., Guimarães, B.G., Medrano, F.J., Barbosa, J.A.R.G., Farah, C.S. Crystallization and Structure Determination of YaeQ from *Xanthomonas axonopodis* pv *citri*. **XXXIV Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular (SBBq)**, 2-5 Julho de 2005, Águas de Lindoia, SP, Brasil.

T. A. Pertinhez, M. M. T. Vanini, G. M. Contessa, M. Gallo, L. M. Galvão-Botton, A. M. Katsuyama, M. Paci, D. O. Cicero, C. S. Farah, C. E. Benedetti and A. Spisni. “Using NMR for a structural proteomics study on the phytopathogen *Xanthomonas axonopodis* pv. *citri*. Structure-based investigation of the function of four unknown proteins” **30<sup>th</sup> FEBS Congress - 9<sup>th</sup> IUBMB Conference**, 2-7 de julho de 2005, Budapest, Hungria.

Guzzo, C.R. ; Nagem R.A.P; Galvão-Botton, L.M.P.; Guimarães, B.G; Medrano, F.J.; Barbosa, J.A.R.G.; Farah, C.S. New protein fold from *Xanthomonas axonopodis* pv. *citri*

YaeQ **Keystone Symposium: Frontiers in Structural Biology**. 29 janeiro a 3 fevereiro de 2006, Keystone, Colorado, EUA.

Guzzo, C. R., Ribeiro L., Zucker P. J., Galvão-Botton L.P.G., Barbosa, J.A.R.G., and Farah, C.S. Purificação, Cristalização e Análise Estrutural da Proteína SufE e do domínio Hpt da RpfC de *Xanthomonas axonopodis* pv. *citri* e da GAPDH de *Escherichia coli*. **XVI Reunião Anual de Utilitários do Laboratório Nacional de Luz Síncrotron**, 20-21 fevereiro de 2006, Campinas, SP.

Pertinhez, T. A.; Vanini, M.M.; Contessa, G. M.; Gallo, M.; Galvao-Botton, L.; Paci, M.; Cicero, D.O.; Farah, C.; Benedetti, C.; Spisni, A. Use of NMR for a Structural Proteomics Study on the Phytopathogen *Xanthomonas axonopodis* pv. *citri* Structure-based Investigation of the Function of Unknown Proteins. **XXXVI National Congress on Magnetic Resonance**, 2006, Vietri sul Mare, Itália.

Guzzo, C.R., Silva, L.R., Galvão de Botton, L.M.P., Barbosa, J.A.R.G., Farah, C.S. Crystal Structure of SufE from *Xanthomonas axonopodis* pv. *citri*. **XXXVI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular (SBBq) e 10th IUBMB Conference**, 21-25 de maio de 2007, Salvador, BA, Brasil.

# High-throughput screening of structural proteomics targets using NMR

Leonor M.P. Galvão-Botton<sup>a,1</sup>, Ângela M. Katsuyama<sup>a,1</sup>, Cristiane R. Guzzo<sup>a</sup>,  
Fábio C.L. Almeida<sup>b</sup>, Chuck S. Farah<sup>a,\*</sup>, Ana Paula Valente<sup>b,\*</sup>

<sup>a</sup>Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, CEP 05508-900, São Paulo, SP, Brazil

<sup>b</sup>Departamento de Bioquímica Médica, ICB, CCS, Universidade Federal do Rio de Janeiro, CEP 21941-590, Rio de Janeiro, RJ, Brazil

Received 3 July 2003; revised 12 August 2003; accepted 14 August 2003

First published online 28 August 2003

Edited by Robert B. Russell

**Abstract** We applied a high-throughput strategy for the screening of targets for structural proteomics of *Xanthomonas axonopodis* pv *citri*. This strategy is based on the rapid <sup>1</sup>H–<sup>15</sup>N HSQC NMR analysis of bacterial lysates containing selectively <sup>15</sup>N-labelled heterologous proteins. Our analysis permitted us to classify the 19 soluble candidates in terms of ‘foldedness’, that is, the extent to which they present a well-folded solution structure, as reflected by the quality of their NMR spectra. This classification allowed us to define a priority list to be used as a guide to select protein candidates for further structural studies.

© 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.

**Key words:** Structural proteomics; NMR; X-ray; Protein screening; Target ORF; *Xanthomonas citri*

## 1. Introduction

Structural proteomics is a rapidly developing field in biology [1–4]. It involves the determination of protein structures on a genome-wide scale and is expected to yield invaluable information about protein folding and protein function. Several recent examples have demonstrated the feasibility of obtaining functional information through structure [5–7].

Structural proteomics requires a large-scale approach to structural biology and, although rapidly evolving, still presents important bottlenecks that need to be overcome [8,9]. Methods have been developed for high-throughput cloning, expression and purification of target open reading frames (ORFs) [9–12]. Important developments have also been made in automated procedures for X-ray crystallography [12–14] and nuclear magnetic resonance (NMR) spectroscopy structure determination [15–19]. However, important issues, such as the acquisition of well-diffracting crystals and size limitation in NMR, decrease the speed and efficiency of the process. Furthermore, the study of uncharacterized proteins inevitably

begins with a large number of candidates that are progressively eliminated due to problems related to expression levels, solubility and stability.

A major challenge of structural proteomics is to select the target proteins in a realistic, efficient and cost-effective way. In order to optimize this process, it would be desirable to have a rapid and efficient method for screening target ORFs for their suitability for structural studies. Several methods have been developed to screen candidates for structural studies, such as cell free expression systems [20,21] and enzymatic labeling of glutamine in vitro [22] and selective isotopic labeling of the target proteins in vivo [23].

In this paper we describe the use of a high-throughput strategy for screening structural proteomics targets by applying it to a set of 35 previously uncharacterized ORFs of the recently sequenced *Xanthomonas axonopodis* pv *citri* (*X. a. pv citri*) phytopathogen [24]. This strategy is based on the NMR analysis of bacterial lysates containing selectively <sup>15</sup>N-labelled heterologous proteins as originally described by Almeida et al. [23]. By omitting the purification step, this protocol is a fast, efficient and inexpensive method to select proteins with conformational characteristics amenable to further structural analysis by NMR and/or X-ray crystallography.

## 2. Materials and methods

### 2.1. Cloning and expression

Targets were amplified by PCR from *X. a. pv citri* genomic DNA [24], subcloned into the pET-3a vector and expressed in *Escherichia coli* strain BL21(DE3)pLysS [25]. Cells were grown at 37°C in unlabelled M9 medium up to an OD<sub>600 nm</sub> of 0.8 and heterologous protein expression induced by 1 mM isopropyl-β-D-thiogalactoside (IPTG). After 15 min, rifampicin was added to a concentration of 200 μg ml<sup>-1</sup>. Upon 15 min of further incubation, the cells were pelleted and resuspended in M9 medium containing 1 g l<sup>-1</sup> <sup>15</sup>NH<sub>4</sub>Cl, 1 mM IPTG and 200 μg ml<sup>-1</sup> rifampicin. Cells were grown for 3.5 h before harvesting and storage at –70°C.

### 2.2. NMR screening

Frozen cell pellets were thawed by adding 500 μl of 20 mM sodium phosphate (pH 7.0), 10 mM β-mercaptoethanol, 40 μg ml<sup>-1</sup> PMSF and lysed by sonication. Lysates were clarified by centrifugation and the soluble fraction was concentrated using a 3 kDa cutoff Centricon filter. 10% D<sub>2</sub>O was added to the sample. Standard 1D <sup>1</sup>H spectra, 1D <sup>15</sup>N-edited <sup>1</sup>H spectra and 2D <sup>1</sup>H–<sup>15</sup>N heteronuclear single quantum coherence (HSQC) spectra were collected at 303 K in a Bruker Advance DRX 600 MHz. XACb0070 HSQC data were also acquired on a Varian INOVA 600 MHz instrument at the Laboratório Nacional de Luz Síncrotron in Campinas, Brazil. All NMR experiments took an average of 8 h for each sample.

\*Corresponding authors.

E-mail addresses: chsfarah@iq.usp.br (C.S. Farah), valente@cnrmn.bioqmed.ufrrj.br (A.P. Valente).

<sup>1</sup> These authors contributed equally to this work.

**Abbreviations:** ORF, open reading frame; NMR, nuclear magnetic resonance; HSQC, heteronuclear single quantum coherence

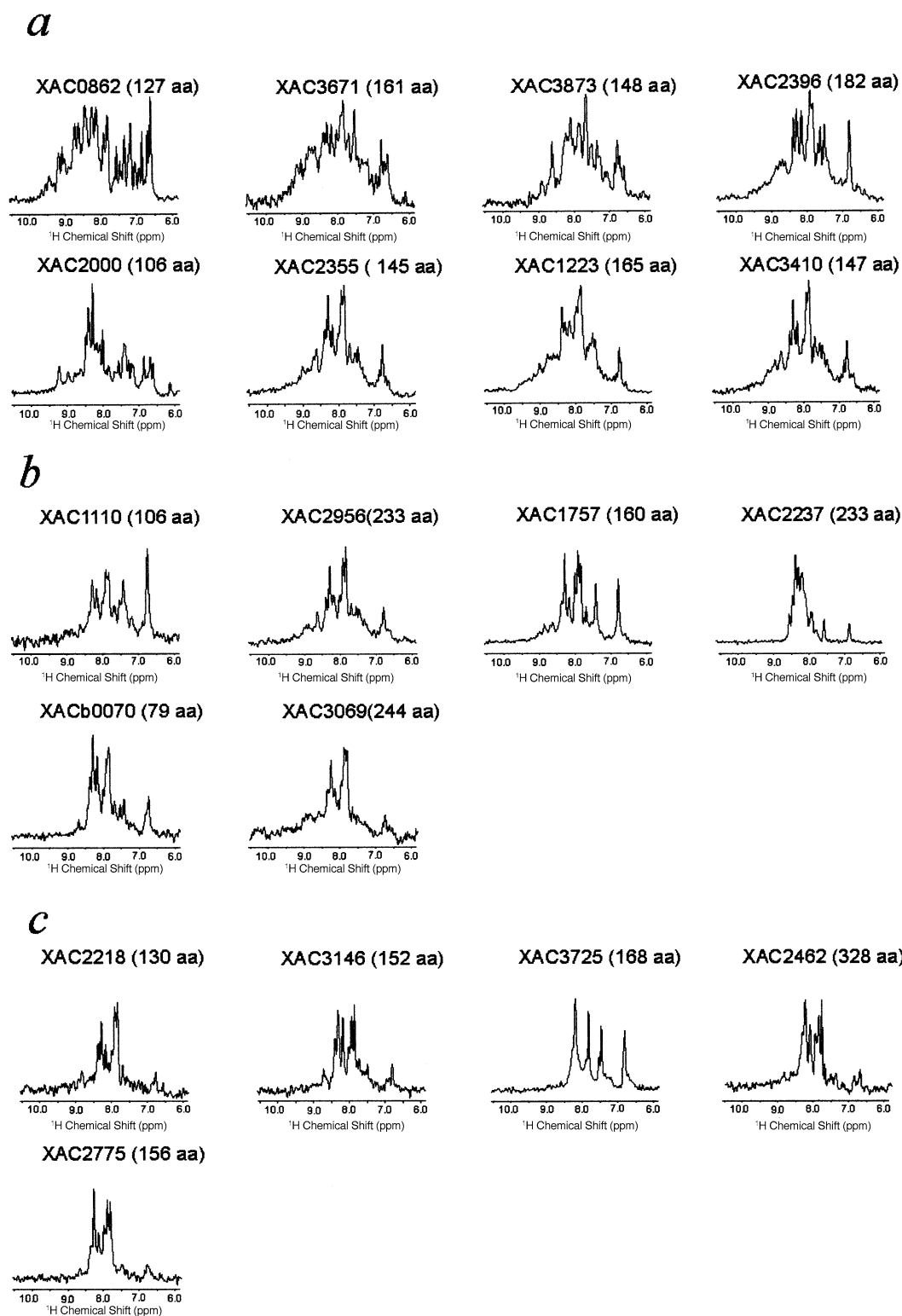


Fig. 1. 1D  $^{15}\text{N}$ -edited  $^1\text{H}$  HSQC spectra of the 19 soluble proteins obtained from soluble cell lysates after selective  $^{15}\text{N}$ -labelling. Spectra are grouped as *good* (a), *promising* (b) and *poor* (c). Classification is based on the analysis of chemical shift dispersion and line-width. The number above the spectra identifies the ORF encoding the expressed protein with the number of amino acid residues shown in parentheses.

### 3. Results and discussion

#### 3.1. Selection of target ORFs

Thirty-five *X. a. pv citri* gene products [24] were selected on the basis of size (79–330 amino acids), predicted cytosolic

localization using PSORT [27], methionine content (>1.1%), low sequence relatedness to proteins of known function and the presence of homologs in the genomes of other organisms (so-called ‘conserved hypothetical proteins’) [27]. Targets were also selected for absence of homologs in the

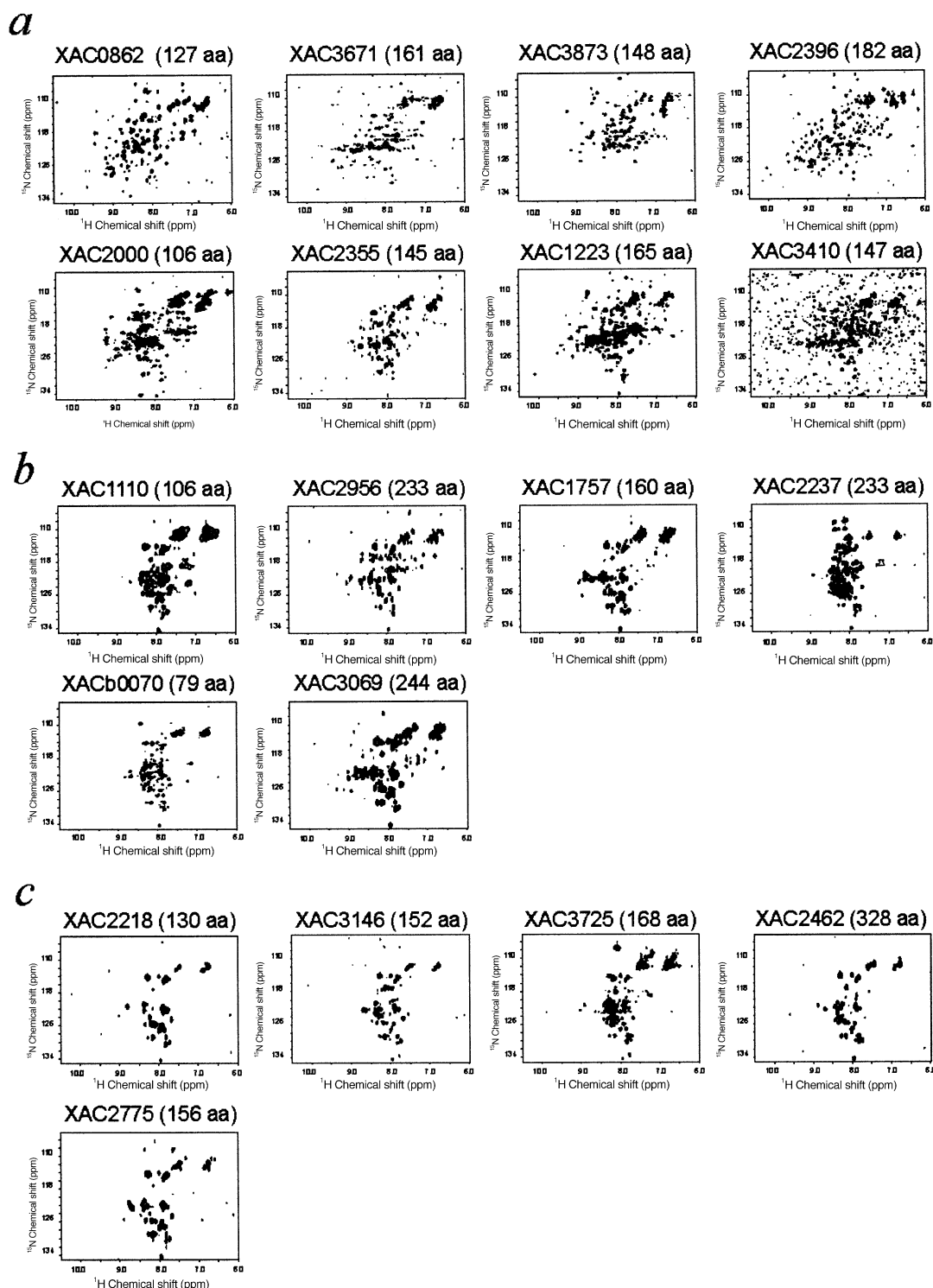


Fig. 2. 2D  $^{15}\text{N}$ -edited  $^1\text{H}$  HSQC spectra of the 19 soluble proteins obtained from soluble cell lysates after selective  $^{15}\text{N}$ -labelling. Spectra are grouped as *good* (a), *promising* (b) and *poor* (c). Classification is based on the analysis of chemical shift dispersion, line-width and number of peaks. Spectra are presented in the same order as shown in Fig. 1. The number above the spectra identifies the ORF encoding the expressed protein with the number of amino acid residues shown in parentheses.

PDB (Protein Data Bank, identified using PSI-BLAST [28] according to the method of Huynen et al. [29].

### 3.2. Expression and solubility

The 35 selected ORFs were amplified and expressed as described in Section 2 [25]. Expression levels and solubility were

monitored using SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) by the intensity of the induced band present in the cell lysate, pellet and supernatant. Most proteins (31/35) expressed well in *E. coli* grown in M9 minimum medium, and 19 of the expressed proteins remained in the soluble fraction of the cell lysates.



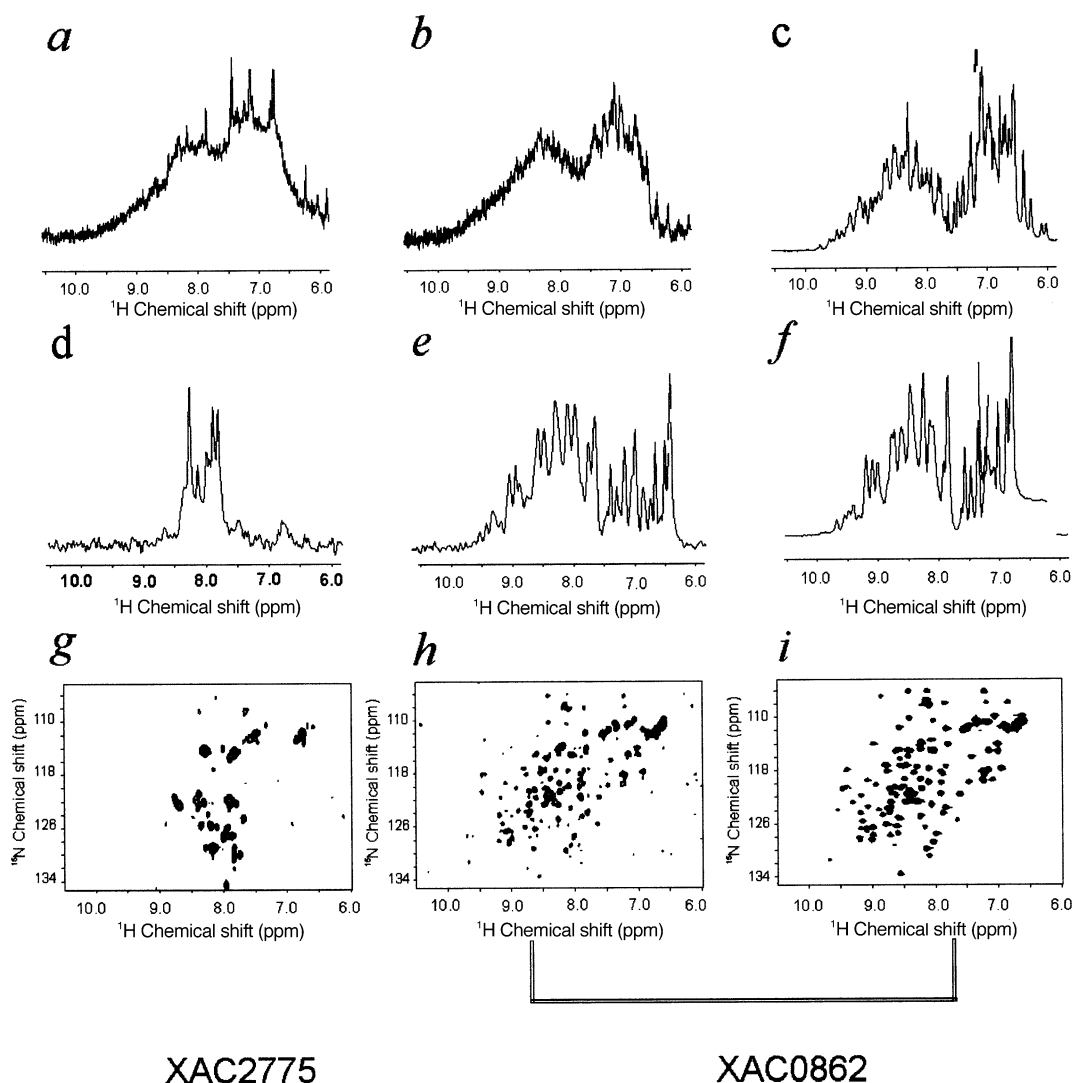


Fig. 3.  $^1\text{H}$  spectra (a,b,c),  $^{15}\text{N}$ -edited  $^1\text{H}$  spectra (d,e,f) and 2D  $^{15}\text{N}$ -edited  $^1\text{H}$  HSQC spectra (g,h,i) for the proteins encoded by ORFs XAC2775 and XAC0862: soluble bacterial lysates containing  $^{15}\text{N}$ -labelled proteins encoded by ORF XAC2775 (a,d,g) and ORF XAC0862 (b,e,h); purified XAC0862 protein product (c,f,i).

### 3.3. Selective $^{15}\text{N}$ -labeling

The 19 soluble proteins were selectively labelled with  $^{15}\text{N}$  as described by Almeida et al. [23]. This methodology uses the antibiotic rifampicin to selectively label heterologous proteins for NMR studies. By 'selective' we mean to indicate that only the heterologously expressed protein was labeled with  $^{15}\text{N}$ . Rifampicin is a bacterial RNA polymerase inhibitor that does not inhibit T7 polymerase used for expression of heterologous proteins cloned under the control of T7 promoters, as in the case of pET vectors [25]. Thus, heterologous protein expression induction following transfer of the bacteria into isotopically labelled minimal medium supplemented with rifampicin enables the selective labelling of heterologous proteins with an isotope of choice.

### 3.4. Screening by $^1\text{H}$ - $^{15}\text{N}$ HSQC

The soluble fractions of the 19 cell lysates containing the selectively  $^{15}\text{N}$ -labelled heterologous proteins were rapidly screened by 1D (Fig. 1) and 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC (Fig. 2). It has long been recognized that chemical shift contains struc-

tural information [30,31]. Furthermore, NMR spectra of well-folded proteins exhibit sharp lines and large chemical shift dispersion, while unfolded proteins present spectra with small chemical shift dispersion and often contain broad lines, due to conformational exchange. Thus, the combined analysis of chemical shift dispersion and line-width provides information about the conformational state of the protein in solution, or its 'foldedness', which could be used as criteria for selecting good protein candidates for further structural analysis [31,32,33].

Fig. 1 shows the 1D HSQC spectra obtained for all 19 *X. a. citri* protein targets screened in this study. According to the quality of the spectra proteins could be empirically grouped. The spectra obtained were classified as *good* when showing sharp, intense and well-dispersed peaks characteristic of well-structured, stable proteins (Fig. 1a). Spectra showing broader, less intense and less dispersed peaks than expected for a stable and well-structured protein were classified as *promising* (Fig. 1b). Such spectra often indicate conformational heterogeneity or dynamic processes on a slow time scale that

can broaden the NMR signal. *Poor* spectra were characterized by a cluster of broad peaks with low chemical shift dispersion (Fig. 1c). These spectra are likely to represent unfolded, conformationally unstable or aggregated proteins.

The 2D HSQC NMR spectra for all 19 soluble proteins are presented in Fig. 2 and again ordered according to the quality of the spectra. The conclusions derived from the 2D HSQC spectra are consistent with those from the 1D HSQC experiments. In addition, in the 2D spectra one can compare the observed number of peaks to that expected according to the primary sequence. As the 2D HSQC spectra are more informative than the 1D spectra, the final priority list as shown in Figs. 1 and 2 was built based mostly on the 2D spectra.

Fig. 3 compares the  $^1\text{H}$  spectra and the HSQC spectra of the cell lysates after expression of two proteins: XAC2775 (Fig. 3a,d,g) and XAC0862 (Fig. 3b,e,h), which presented *poor* and *good* spectra, respectively. Fig. 3 also shows the spectra of the purified XAC0862 protein (Fig. 3c,f,i). The  $^1\text{H}$  spectra of the two lysate samples (Fig. 3a,b) are very similar as resonances from all the soluble components in the cell lysate appear. In contrast, the HSQC spectra present fewer and more defined peaks that correspond to the amide protons of the  $^{15}\text{N}$ -labelled heterologous proteins (Fig. 3d,e,g,h). Note that the 1D  $^{15}\text{N}$ -edited  $^1\text{H}$  HSQC spectra (Fig. 3d,e) are clearly different from each other, reflecting distinct structural

characteristics. The spectrum for protein XAC2775 (Fig. 3d) shows low dispersion,  $\text{H}_\text{N}$  peaks falling between 7.5 and 8.5 ppm. On the other hand, the spectrum for protein XAC0862 (Fig. 3e) shows a wide dispersion of amide proton peaks (6.5–9.5 ppm), indicative of a well-folded protein. The 2D HSQC spectra (Fig. 3g,h) of the soluble bacterial lysates confirm these observations, but in more detail. Specifically, the spectrum for protein XAC2775 contains a small number of broad lines (Fig. 3g) while that of protein XAC0862 contains a large number of sharp well-resolved lines (Fig. 3h). Additionally, the number of peaks observed in the XAC0862 2D spectrum match with that expected from the primary sequence (143). Comparison of the spectra of the XAC0862 bacterial lysate (Fig. 3b,e,h) with the purified protein (Fig. 3c,f,i) confirms that rifampicin allows the selective labelling of the heterologous protein alone as previously shown [23], and that the lysate spectra of the  $^{15}\text{N}$ -labelled protein provides valuable information about protein conformation.

Besides XAC0862, three other *good* candidates were purified to homogeneity after labeling with  $^{15}\text{N}$ . The HSQC spectra of these purified proteins are presented in Fig. 4. The spectra for purified XAC2000, XAC2396 and XAC3873 (Fig. 4) are very similar to those observed for their corresponding bacterial lysates. For XAC2000 and XAC2396, the observed number of peaks in the 2D spectra are in good

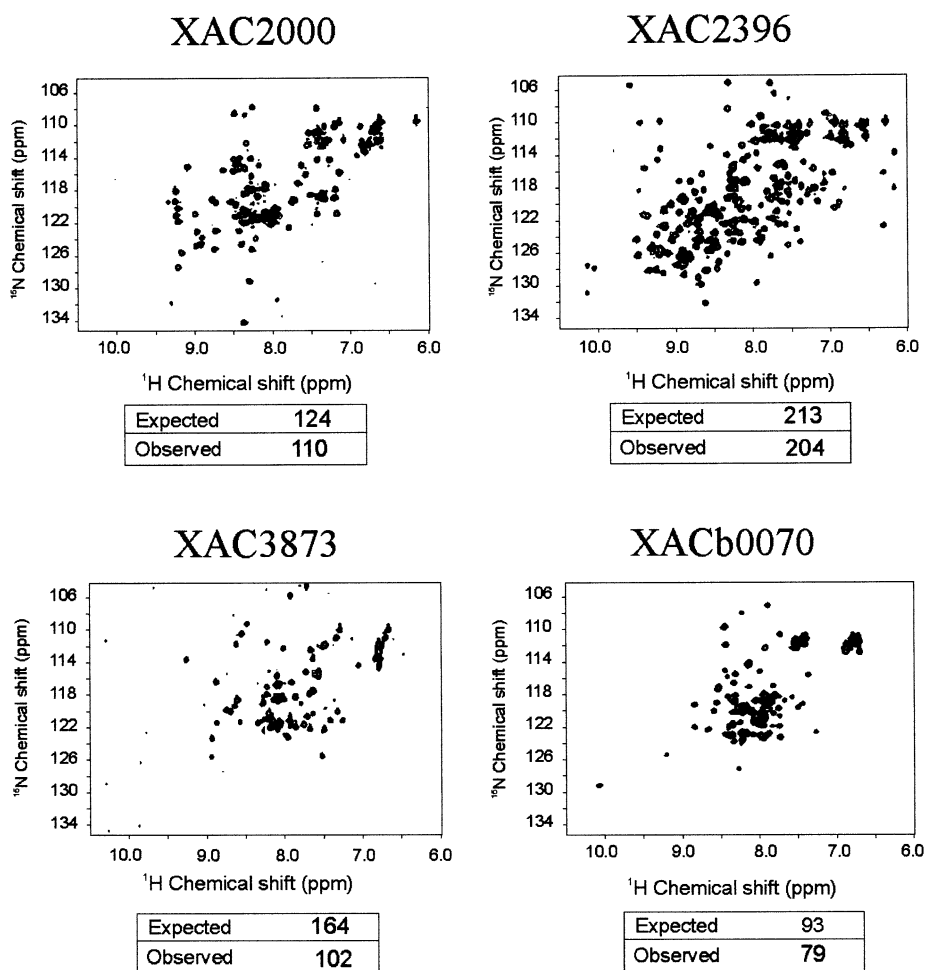


Fig. 4. 2D  $^{15}\text{N}$ -edited  $^1\text{H}$  HSQC spectra of purified proteins from ORFs XAC2000, XAC2396, XAC3873 and XACb0070. Spectra were acquired at 25°C and pH 7.0.

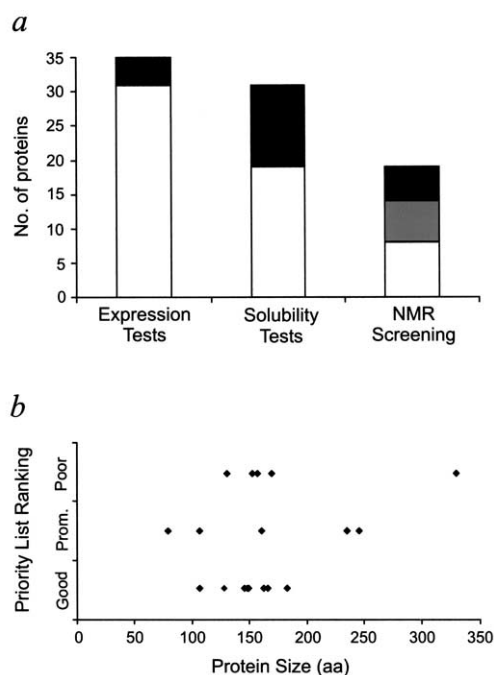


Fig. 5. a: Histogram of the number of proteins expressed (white), not expressed (black), soluble (white), insoluble (black) and with *good* (white), *promising* (gray) and *poor* (black) NMR spectra. b: Correlation between priority list ranking (good, promising and poor) and protein size (aa=number of amino acids and prom.=promising).

agreement with the expected number of peaks. These results are consistent with the hypothesis that most of the proteins presenting *good* spectra may in fact be well-folded proteins even in the bacterial lysates and that they maintain their folded state during the purification procedure. The 2D HSQC spectra of the unpurified and purified fractions of XAC3873 are also very similar and show good peak dispersion. However, in this case only 102 of the 164 expected peaks were observed in both. This may indicate overlap of resonances due to high helical content or to a well-folded domain consisting of a little more than half of the polypeptide chain. In the case of ambiguity, well-folded proteins may also be distinguished using pulse sequences that enable the identification of non-hydrogen bonded amide protons present in less structured regions of the protein and in fast exchange with water [26,36,37].

### 3.5. Prioritizing protein targets for further structural studies

The purpose of the screening protocol is to create a priority list to be used as a guide for choosing suitable protein candidates for the more expensive and time-consuming purification and structure determination phases of a medium- or large-scale structural biology project. The method described here allowed us to assess the conformational state of the protein in solution without the need of previous purification.

The spectra evaluated as 'promising' in Figs. 1 and 2 generally presented a lower degree of chemical shift dispersion than the *good* spectra but sharper and more intense lines than the *poor* spectra. In some cases, a narrow range of  $^1\text{H}$  chemical shifts may be indicative of high  $\alpha$ -helical content and not necessarily random coil conformations. One *promising* candidate, protein XACb0070, that showed sharp peaks but

a low degree of  $^1\text{H}$  chemical shift dispersion, was purified in a soluble and stable form. The 2D HSQC spectrum of the purified protein (Fig. 4) is similar to that observed in the bacterial lysate (Fig. 2b), although not the same. The number of amide peaks observed is close to that expected from its primary sequence (93). This protein aggregates at concentrations greater than 500  $\mu\text{M}$ , which may contribute, in part, to the observed differences. Furthermore, there are many examples in the literature of large-scale  $^1\text{H}$  and  $^{15}\text{N}$  chemical shift changes induced by the binding of inorganic and organic co-factors [33–35]. Such factors would be present in the bacterial lysate but probably absent in the purified protein. The CD spectrum of the purified spectrum obtained was typical for an  $\alpha$ -helical protein (data not shown).

It must be noted that during the screening procedure no protein-specific optimization whatsoever was carried out for the expression or lysis, prior to NMR analysis. Therefore proteins low in the priority list should not be discarded. Folded proteins that interact with cellular components would hamper the HSQC quality. In principle, it would be difficult to distinguish between natively unfolded proteins and misfolded proteins using our methodology [38].

Selective  $^{15}\text{N}$ -labelling using rifampicin for NMR screening was originally described and tested using proteins whose structure and stability were previously known [23]. It was not known whether this methodology would be effective when analyzing unknown structural proteomics targets. To determine the efficiency of this method for screening targets for structural proteomics, we screened 35 previously uncharacterized *X. a. pv citri* proteins. In total, 42% of the tested proteins showed good NMR spectra, 32% showed promising spectra and 26% showed poor spectra (Fig. 5a). One of the concerns of using HSQC spectra to screen for well-folded proteins was that, given the size limitation of NMR, the best spectra would correspond to the smaller proteins tested whereas poor spectra would be observed for larger proteins. We nevertheless found examples of small and large proteins producing all types of spectra (*good*, *promising* and *poor*) (Fig. 5b). This shows that this method is suitable for creating a priority list for structural proteomics candidates, regardless of their size, provided they are sufficiently small for NMR analysis. 'Foldedness' defined in this way may also be of value in the screening stages of large-scale protein crystallization projects [31,32].

There are several published methods to screen promising candidates for structural proteomics projects [12,31,32,33,39]. Yee et al. [33] recently described the screening of 513 structural proteomics candidates from several genomes using efficient large-scale affinity purification of His-tagged  $^{15}\text{N}$ -labelled proteins followed by HSQC analysis. The combination of selective  $^{15}\text{N}$ -labelling of heterologous proteins with the rapid analysis of soluble cell lysates by NMR spectroscopy described here provides a promising alternative method for rapidly and efficiently screening structural proteomics targets for future high-resolution studies by NMR or X-ray crystallography.

**Acknowledgements:** We thank A.S. Pinheiro for assistance. We also thank Dr. Thelma Pertinhez for help in the acquisition of the CD and HSQC spectra of XACb0070 at the Laboratório Nacional de Luz Síncrotron in Campinas, Brazil. This work was supported with grants from FAPESP, CNPq/PRONEX (Brazil) and The Third World Academy of Science (Trieste, Italy). L.M.P.G. is supported by a PhD

studentship from FCT/MCT (Portugal). A.M.K and C.R.G. received undergraduate research fellowships from FAPESP.

## References

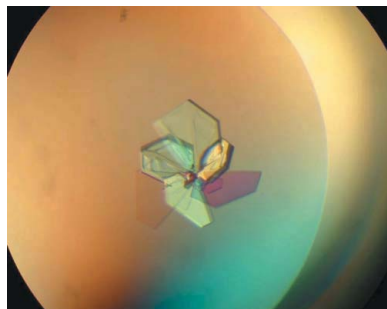
- [1] Burley, S.K. (2000) *Nat. Struct. Biol.* 7, 932–934.
- [2] Brenner, S.E. (2001) *Nat. Rev. Genet.* 10, 801–809.
- [3] Mittl, P.R.E. and Grütter, M.G. (2001) *Curr. Opin. Chem. Biol.* 5, 402–408.
- [4] Stevens, R.C., Yokoyama, S. and Wilson, I.A. (2001) *Science* 294, 89–92.
- [5] Bhattacharyya, S., Habibi-Nazhad, B., Amegbey, G., Slupsky, C.M., Yee, A., Arrowsmith, C. and Wishart, D.S. (2002) *Biochemistry* 41, 4760–4770.
- [6] Christendat, D., Saridakis, V., Kim, Y., Kumar, P.A., Xu, X., Semesi, A., Joachimiak, A., Arrowsmith, C.H. and Edwards, A.D. (2002) *Protein Sci.* 11, 1409–1414.
- [7] Jackson, R.M. and Russell, R.B. (2001) *Comput. Chem.* 26, 31–39.
- [8] Service, R.F. (2002) *Science* 298, 948–950.
- [9] Yokoyama, S. (2003) *Curr. Opin. Chem. Biol.* 7, 39–43.
- [10] Dieckman, L., Gu, M., Stols, L., Donnelly, M.I. and Collart, F.R. (2002) *Protein Exp. Purif.* 25, 1–7.
- [11] Boettner, M., Prinz, B., Stahl, U. and Lang, C. (2002) *J. Biotechnol.* 99, 51–62.
- [12] Lesley, S.A. et al. (2002) *Proc. Natl. Acad. Sci. USA* 99, 11664–11669.
- [13] Villasenor, A., Sha, M., Thana, P. and Browner, M. (2002) *Biotechniques* 32, 184, 186, 188–189.
- [14] Lamzin, V.S. and Perrakis, A. (2000) *Nat. Struct. Biol.* 7 (Suppl), 978–981.
- [15] Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N. and Montelione, G.T. (2002) *Proc. Natl. Acad. Sci. USA* 99, 8009–8014.
- [16] Bhavesh, N.S., Panchal, S.C. and Hosur, R.V. (2001) *Biochemistry* 40, 14727–14735.
- [17] Prestegard, J.H., Valafar, H. and Tian, F. (2001) *Biochemistry* 40, 8677–8685.
- [18] Zweckstetter, M. and Bax, A. (2001) *J. Am. Chem. Soc.* 123, 9490–9491.
- [19] Kim, S. and Szyperski, T. (2003) *J. Am. Chem. Soc.* 125, 1385–1393.
- [20] Guignard, L., Ozawa, K., Pursglove, S.E., Otting, G. and Dixon, N.E. (2002) *FEBS Lett.* 254, 159–162.
- [21] Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. and Yokoyama, S. (1999) *FEBS Lett.* 442, 15–19.
- [22] Shimba, N., Yamada, N., Yokoyama, K. and Suzuki, E. (2002) *Anal. Biochem.* 301, 123–127.
- [23] Almeida, F.C.L., Amorim, G.C., Moreau, V.H., Sousa, V.O., Creazola, A.T., Américo, T.A., Pais, A.P.N., Leite, A., Netto, L.E.S., Giordano, R.J. and Valente, A.P. (2001) *J. Magn. Reson.* 148, 142–146.
- [24] Da Silva, A.C. et al. (2002) *Nature* 417, 459–463.
- [25] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods Enzymol.* 185, 60–89.
- [26] Dalvit, C. and Hommel, U. (1995) *J. Magn. Reson.* 109, 334–338.
- [27] Nakai, K. and Horton, P. (1999) *Trends Biochem. Sci.* 24, 34–36.
- [28] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) *Nucleic Acids Res.* 25, 3389–3402.
- [29] Huynen, M., Doerks, T., Eisenhaber, F., Orengo, C., Sunyaev, S., Yuan, Y.P. and Bork, P. (1998) *J. Mol. Biol.* 280, 323–326.
- [30] Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.* 239, 363–392.
- [31] Rehm, T. and Huber, R. (2002) *Structure* 10, 1613–1618.
- [32] Montelione, G.T., Zheng, D., Huang, Y.J., Gunsalus, K.C. and Szyperski, T. (2000) *Nat. Struct. Biol.* 7 (Suppl), 982–985.
- [33] Yee, A. et al. (2002) *Proc. Natl. Acad. Sci. USA* 99, 1825–1830.
- [34] Zartler, E., Jenney, F.E., Terrell, M., Eidsness, M.K., Adams, M.W.W. and Prestegart, J.H. (2001) *Biochemistry* 40, 7279–7290.
- [35] Jaren, O.R., Kranz, J.K., Sorensen, B.R., Wand, A.J. and Shea, M.A. (2002) *Biochemistry* 41, 14158–14166.
- [36] Hwang, T.L., Mori, S., Shaka, A.J. and vanZijl, P.C.M. (1997) *J. Am. Chem. Soc.* 119, 6203–6204.
- [37] Hwang, T.L., van Zijl, P.C.M. and Mori, S. (1998) *J. Biomol. NMR* 11, 221–226.
- [38] Uversky, V.N. (2002) *Protein Sci.* 11, 739–756.
- [39] Christendat, D. et al. (2000) *Nat. Struct. Biol.* 7, 903–909.

**Cristiane R. Guzzo,<sup>a</sup>**  
**Ronaldo A. P. Nagem,<sup>b</sup>**  
**Leonor M. P. Galvão-Botton,<sup>a</sup>**  
**Beatriz G. Guimarães,<sup>c</sup>**  
**Francisco J. Medrano,<sup>c</sup>**  
**João A. R. G. Barbosa<sup>c\*</sup>** and  
**Chuck S. Farah<sup>a\*</sup>**

<sup>a</sup>Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, CEP 05508-000, São Paulo, SP, Brazil, <sup>b</sup>Instituto de Física de São Carlos, Universidade de São Paulo, CEP 13560-970, São Carlos, SP, Brazil, and <sup>c</sup>Centro de Biologia Molecular Estrutural, Laboratório Nacional de Luz Síncrotron, CEP 13084-971, Campinas, SP, Brazil

Correspondence e-mail: joao@lnls.br, chsfarah@iq.usp.br

Received 25 February 2005  
 Accepted 7 April 2005  
 Online 22 April 2005



© 2005 International Union of Crystallography  
 All rights reserved

## Expression, purification, crystallization and preliminary X-ray analysis of YaeQ (XAC2396) from *Xanthomonas axonopodis* pv. *citri*

*Xanthomonas axonopodis* pv. *citri* YaeQ (XAC2396) is a member of a family of bacterial proteins conserved in several Gram-negative pathogens. Here, the cloning, expression, purification and crystallization of the 182-residue (20.6 kDa) YaeQ protein are described. Recombinant YaeQ containing selenomethionine was crystallized in space group  $P2_1$  and crystals diffracted to 1.9 Å resolution at a synchrotron source. The unit-cell parameters are  $a = 39.75$ ,  $b = 91.88$ ,  $c = 48.03$  Å,  $\beta = 108.37^\circ$ . The calculated Matthews coefficient suggests the presence of two YaeQ molecules in the asymmetric unit. Initial experimental phases were calculated by the multiple-wavelength anomalous dispersion technique and an interpretable electron-density map was obtained.

### 1. Introduction

*Xanthomonas* is a genus of phytopathogenic bacteria that attack a variety of hosts with economic importance, including citrus fruits, rice, beans, grapes and cotton. Citrus canker is caused by the phytopathogen *Xanthomonas axonopodis* pv. *citri* (*Xac*), a rod-like Gram-negative motile bacterium. The *Xac* genome (Da Silva *et al.*, 2002) has a 5 175 554 bp chromosome and two megaplasmids: pXAC33 (33 699 bp) and pXAC64 (64 920 bp). Genome annotation identified over 4400 open reading frames, 2770 of which were assigned putative functions based on similarities with sequences in the public databases (Da Silva *et al.*, 2002).

*Xac* YaeQ (XAC2396) is a member of a family of bacterial proteins that are conserved in several Gram-negative pathogens, including *Salmonella* spp., *Escherichia coli*, *Pseudomonas* spp., *Shigella* spp., *Yersinia* spp., *Ralstonia* spp., *Erwinia* spp., *Bordetella* spp., *Burkholderia fungorum*, *Vibrio parahaemolyticus* and *Chromobacterium violaceum*, as well as *Azotobacter vinelandii*, *Dechloromonas aromatica*, *Shewanella oneidensis*, *Geobacter metallireducens* and *G. sulfurreducens*. This 182-amino-acid residue protein was selected for structural analysis owing to its well folded and non-aggregated state in solution, both in bacterial lysates and in the purified form (Galvão-Botton *et al.*, 2003).

It has been suggested that YaeQ is involved in the regulation of transcription of genes encoding virulence factors such as haemolysin and enzymes involved in the production of lipopolysaccharides. This hypothesis was presented as a YaeQ-encoding fragment from *Salmonella typhimurium* was shown to complement a defect in the RfaH-dependent expression of the *hlyCABD* operon (Wong *et al.*, 1998). However, Vicari & Artsimovitch (2004) have recently presented evidence that *E. coli* YaeQ does not complement the absence of RfaH in *hlyCABD* expression assays *in vitro* or *in vivo*. Thus, the precise function of YaeQ and its orthologues remains unknown. The YaeQ family does not show any sequence similarity with proteins of known structure or function. Determination of the three-dimensional structure of YaeQ may provide insights regarding its function. Several recent examples have demonstrated the feasibility of obtaining functional information from structure (Zarembinski *et al.*, 1998; Bhattacharyya *et al.*, 2002; Christendat *et al.*, 2002; Jackson & Russell, 2001). In this report, we describe the cloning and expression of recombinant *Xac* YaeQ containing selenomethionine and its crystallization and also the determination of an initial



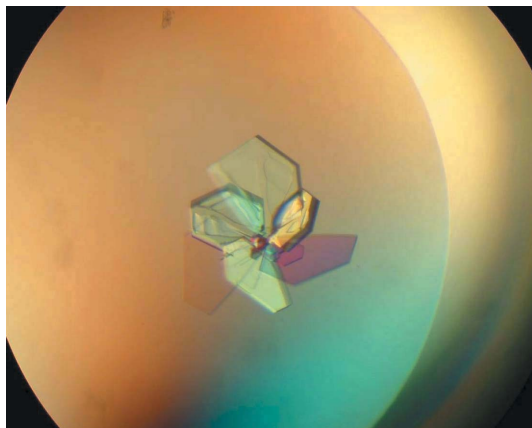
electron-density map of the *Xac* YaeQ crystal structure by multi-wavelength anomalous dispersion (MAD).

## 2. Cloning and expression of YaeQ

The YaeQ gene (XAC2396) was amplified by PCR from *Xac* genomic DNA using the following primers that were designed based on the published *Xac* genome sequence (Da Silva *et al.*, 2002): 5'-CATGC-CATGGCTCATATGGCCCTCACCGCCA-3' and 5'-GGAATT-CAAGCTTTCATTCGGCCGGGGCT-3'. The PCR product was digested with *Hind*III and *Nde*I and subcloned into the pET-3a vector (Studier *et al.*, 1990) previously digested with the same endonucleases. Selenomethionine-containing YaeQ was expressed in *E. coli* strain BL21(DE3)pLysS (Studier *et al.*, 1990) by growing a 500 ml culture in M9 medium to an optical density (600 nm) of 0.8, at which point 100 mg l<sup>-1</sup> lysine, 100 mg l<sup>-1</sup> phenylalanine, 100 mg l<sup>-1</sup> threonine, 50 mg l<sup>-1</sup> isoleucine, 50 mg l<sup>-1</sup> valine and 60 mg l<sup>-1</sup> selenomethionine were added (adapted from Berne *et al.*, 1999). After 15 min, heterologous protein expression was induced by adding 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) and the cells were grown for 4 h before harvesting and storage at 203 K.

## 3. Protein purification

Cells from 1 l culture were resuspended in 25 ml 50 mM Tris-HCl pH 8.0, 25% sucrose, 1 mM EDTA, 1 mM PMSF and lysed using a French press. 2 U ml<sup>-1</sup> DNase and 4 mM magnesium chloride were added to the soluble fraction, followed by incubation on ice for 30 min. This mixture was applied onto a Q-Sepharose Fast Flow (FF) Hiloadd 16/10 column (Amersham Pharmacia) previously equilibrated with 50 mM Tris-HCl pH 7.0, 1 mM EDTA and 14 mM  $\beta$ -mercaptoethanol. Bound proteins were eluted using a 0–300 mM NaCl gradient over 12 column volumes. Fractions containing YaeQ were concentrated using an Amicon system with a 10 kDa membrane and then further purified by gel filtration on a Superdex 75 prep-grade column (Amersham Pharmacia) equilibrated with 5 mM Tris-HCl pH 7.0. The molecular mass of the purified protein, as determined by MALDI-TOF mass spectrometry, was 20 982.749  $\pm$  5 for the selenomethionine protein and 20 668.582  $\pm$  5 for the methionine-containing protein. The difference in mass between the two preparations demonstrated that all seven methionines were exchanged for selenomethionines. The total protein yield was 46.5 mg l<sup>-1</sup>, with approximately 99% purity.



**Figure 1**  
Typical crystals of YaeQ. The approximate dimensions of individual plates are 0.40  $\times$  0.40  $\times$  0.03 mm.

**Table 1**

Recombinant SeMet-labelled *Xac* YaeQ crystal parameters and data-reduction statistics.

Values in parentheses refer to the highest resolution shell.

Space group	<i>P</i> 2 <sub>1</sub>		
Unit-cell parameters			
<i>a</i> (Å)	39.75		
<i>b</i> (Å)	91.88		
<i>c</i> (Å)	48.03		
$\beta$ (°)	108.37		
Data-set statistics	Peak	Inflection	Remote
Resolution range (Å)	37.80–1.90 (1.97–1.90)	46.13–2.40 (2.53–2.40)	45.64–2.40 (2.53–2.40)
No. of observed reflections	99056 (8105)	49405 (7232)	51465 (7478)
No. of unique reflections	24507 (2207)	12499 (1816)	12469 (1799)
$\langle I/\sigma(I) \rangle^\dagger$	25.1 (2.1)	11.1 (2.0)	13.0 (2.2)
Multiplicity	4.0 (3.7)	4.0 (4.0)	4.1 (4.2)
Multiplicity, anomalous <sup>†</sup>	2.1 (1.9)	2.1 (2.0)	2.1 (2.0)
Completeness (%)	95.0 (86.7)	97.5 (96.9)	97.5 (96.7)
Completeness, anomalous <sup>†</sup> (%)	93.5 (82.8)	94.0 (94.8)	94.4 (95.1)
<i>R</i> (%)	3.2 (34.9)	6.0 (35.2)	5.1 (31.7)
No. of images	200	195	199
$\lambda$ (Å)	0.97952	0.97962	1.00000

<sup>†</sup> Each member of a Friedel pair is counted as a separate reflection.

## 4. Crystallization

YaeQ crystals (Fig. 1) were grown using the hanging-drop vapour-diffusion technique at 291 K. Initial crystallization conditions of the methionine-containing protein were screened by the sparse-matrix sampling approach. Crystals were obtained in several conditions using the Crystal Screen (Hampton Research) and Clear Strategy (Molecular Dimensions Limited) screening kits. Optimization was then pursued by varying the precipitant concentration and the buffer pH. Suitable crystals for diffraction experiments were obtained by mixing equal volumes (1  $\mu$ l) of 10 mg ml<sup>-1</sup> protein solution (in 5 mM Tris-HCl pH 7.0) with reservoir solution, which consisted of either (i) 30–32% (w/v) PEG 4000, 0.2 M ammonium acetate and 0.1 M Tris-HCl pH 8.5–9.0 or (ii) 30% (w/v) PEG 8000, 0.1 M sodium cocadylate pH 6.7 and 0.2 M sodium acetate. Hanging drops were then equilibrated against 0.4 ml reservoir solutions. Crystals appeared within a few days and grew to mature size within a few weeks. Both methods gave the same crystal form.

## 5. Data collection and preliminary structure analysis

Initial X-ray diffraction data from the methionine-containing protein were collected at the D03B beamline of the Laboratório Nacional de Luz Síncrotron, Campinas, Brazil using a MAR CCD detector (data not shown). MAD X-ray diffraction data for selenomethionine-containing crystals were collected at the protein crystallography beamline 8.3.1 of the Advanced Light Source (ALS), Berkeley, California using a ADSC CCD detector. SeMet-YaeQ crystals equilibrated against 32% PEG 4000, 0.1 M Tris-HCl pH 9.0, 0.2 M ammonium acetate were flash-frozen in liquid nitrogen and maintained at 100 K in a nitrogen-gas stream during data acquisition (this crystallization solution was sufficient to cryoprotect the crystals). MAD data sets were collected using a single crystal at three wavelengths, 0.97952, 0.97962 and 1.00000 Å, corresponding to the peak, inflection (approximate) and remote points of the fluorescence spectrum, respectively. The peak data set was collected initially and the inflection and remote data sets were then both collected from another region of the same crystal. The exposure time for each image of the peak data set (5 s) was ten times that for the inflection and remote data sets (0.5 s). The oscillation range for each image was 1°. The diffraction patterns from these data sets extended to approxi-

mately 1.9 and 2.4 Å resolution, respectively (Table 1). The crystal belongs to space group  $P2_1$ , with unit-cell parameters  $a = 39.75$ ,  $b = 91.88$ ,  $c = 48.03$  Å,  $\beta = 108.37^\circ$ . The Matthews coefficient ( $V_M = 2.0$  Å<sup>3</sup> Da<sup>-1</sup> and 38.4% solvent content) suggests that there are two protein molecules per asymmetric unit.

Diffraction images data were indexed, integrated, scaled and merged using the *HKL2000* package (Otwinowski & Minor, 1997) for the peak data set and the programs *MOSFLM* (Leslie, 1992) and *SCALA* (Collaborative Computational Project, Number 4, 1994) for the inflection and remote data sets. The Friedel mates were scaled separately during data processing. Data sets were merged using *CAD* (Collaborative Computational Project, Number 4, 1994). Details of data-acquisition and data-processing statistics are shown in Table 1.

The program *SHELXD* (Sheldrick, 1998) was used to locate 14 selenium sites in the asymmetric unit of the crystal (seven selenomethionines per protein molecule are expected) using the anomalous differences from the peak data set. However, only the coordinates of 12 major selenium sites were used to calculate MAD phases using the program *SHARP* (de La Fortelle & Bricogne, 1997). Experimental phases obtained by *SHARP* were improved by density-modification protocols using the programs *SOLOMON* (Abrahams, 1997) and *DM* (Collaborative Computational Project, Number 4, 1994; Cowtan, 1994). The resulting electron-density map was used by the program *ARP/wARP* (Perrakis *et al.*, 1999) to produce a preliminary polyaniline model that included 285 of the 364 amino-acid residues found within the asymmetric unit of the crystal. On the basis of this preliminary model, the non-crystallographic symmetry relating the two monomers was determined and used for phase improvement and model building. Interpretation of electron-density maps and construction of missing residues is being performed using the program *O* (Jones *et al.*, 1991). Structural refinement of the XAC2396 atomic model is currently under way using *REFMAC* (Murshudov *et al.*, 1997) and *CNS* (Brünger *et al.*, 1998).

We would like to thank the ALS 8.3.1 beamline staff and Dr Mark Glover (University of Alberta) and members of his laboratory for the opportunity to collect MAD data at the ALS. We thank the laboratory of Dr Paolo di Mascio (Instituto de Química, Universidade de São Paulo) for MALDI-TOF MS analysis. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through grant 01/07534-3, Conselho Nacional de Desen-

volvimento Científico e Tecnológico (CNPq) (Brazil) and Associação Brasileira de Tecnologia de Luz Síncrotron (ABTLuS). CRG is a FAPESP fellow and LMPG was supported with a PhD studentship from FCT/MCT (Portugal).

## References

- Abrahams, J. P. (1997). *Acta Cryst.* **D53**, 371–376.
- Berne, P. F., Doublíć, S. & Carter, C. W. Jr (1999). *Crystallization of Nucleic Acids and Proteins: A Practical Approach*, 2nd ed., edited by A. Ducruix & R. Giegé, pp. 45–73. Oxford University Press.
- Bhattacharyya, S., Habibi-Nazhad, B., Amegbey, G., Slupsky, C. M., Yee, A., Arrowsmith, C. & Wishart, D. S. (2002). *Biochemistry*, **41**, 4760–4770.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Christendat, D., Saridakis, V., Kim, Y., Kumar, P. A., Xu, X., Semesi, A., Joachimiak, A., Arrowsmith, C. H. & Edwards, A. M. (2002). *Protein Sci.* **11**, 1409–1414.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cowtan (1994). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
- Da Silva, A. C. R. *et al.* (2002). *Nature (London)*, **417**, 459–463.
- Galvão-Botton, L. M. P., Katsuyama, A. M., Guzzo, C. R., Almeida, F. C. L., Farah, C. S. & Valente, A. P. (2003). *FEBS Lett.* **552**, 207–213.
- Jackson, R. M. & Russell, R. B. (2001). *Comput. Chem.* **26**, 31–39.
- Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **244**, 472–494.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- Sheldrick, G. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol.* **185**, 60–89.
- Vicari, D. & Artsimovitch, I. (2004). *Mol. Genet. Genomics*, **272**, 489–496.
- Wong, K. R., Hughes, C. & Koronakis, V. (1998). *Mol. Gen. Genet.* **257**, 693–696.
- Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R. & Kim, S.-H. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 15189–15193.

Cristiane R. Guzzo,<sup>a</sup>  
Lucicleide R. Silva,<sup>a</sup> Leonor M. P.  
Galvão-Botton,<sup>a</sup> João A. R. G.  
Barbosa<sup>b</sup> and Chuck S. Farah<sup>a\*</sup>

<sup>a</sup>Departamento de Bioquímica, Instituto de  
Química, Universidade de São Paulo,  
CEP 05508-000, São Paulo, SP, Brazil, and

<sup>b</sup>Centro de Biologia Molecular Estrutural, LNLS,  
CEP 13084-971, Campinas, SP, Brazil

Correspondence e-mail: chsfarah@iq.usp.br

Received 3 January 2006

Accepted 9 February 2006

## Expression, crystallization and preliminary crystallographic analysis of SufE (XAC2355) from *Xanthomonas axonopodis* pv. *citri*

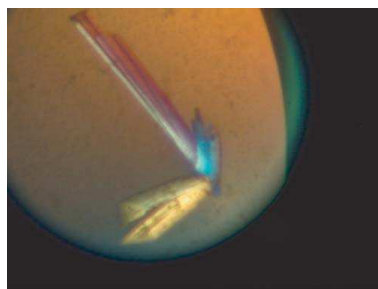
*Xanthomonas axonopodis* pv. *citri* (*Xac*) SufE (XAC2355) is a member of a family of bacterial proteins that are conserved in several pathogens and phytopathogens. The *Escherichia coli* *suf* operon is involved in iron–sulfur cluster biosynthesis under iron-limitation and stress conditions. It has recently been demonstrated that SufE and SufS form a novel two-component cysteine desulfurase in which SufS catalyses the conversion of L-cysteine to L-alanine, forming a protein-bound persulfide intermediate. The S atom is then transferred to SufE, from which it is subsequently transferred to target molecules or reduced to sulfide in solution. Here, the cloning, expression, crystallization and phase determination of *Xac* SufE crystals are described. Recombinant SufE was crystallized in space group  $P2_12_12_1$  and diffracted to 1.9 Å resolution at a synchrotron source. The unit-cell parameters are  $a = 45.837$ ,  $b = 58.507$ ,  $c = 98.951$  Å,  $\alpha = \beta = \gamma = 90^\circ$ . The calculated Matthews coefficient indicated the presence of two molecules in the asymmetric unit. Phasing was performed by molecular-replacement using *E. coli* SufE as a model (PDB code 1mzg) and an interpretable map was obtained.

### 1. Introduction

*Xanthomonas axonopodis* pv. *citri* (*Xac*) SufE (XAC2355) is a member of a family of proteins that are conserved in many prokaryotic pathogens and phytopathogens, including *Salmonella* spp., *Escherichia coli*, *Pseudomonas* spp., *Shigella* spp., *Yersinia* spp., *Leptospira interrogans*, *Erwinia carotovora*, *Vibrio* spp., *Azotobacter vinelandii*, *Shewanella baltica*, *Xanthomonas* spp. and *Xylella* spp. DNA sequences coding for SufE homologs have also been found in several eukaryotic genomes: *Plasmodium*, *Anopheles gambiae* and *Arabidopsis thaliana*. This protein was originally selected for structural analysis owing to its well folded and non-aggregated state in solution at a time when its function was unknown (Galvão Botton *et al.*, 2003).

In *E. coli*, the *sufABCDSE* cluster is expressed under conditions of iron limitation and oxidative stress, one of its functions being the restoration of damaged Fe–S clusters (Nachin *et al.*, 2001, 2003). In *E. coli*, SufA is a scaffold protein for the assembly of Fe–S clusters. SufB and SufD form a complex with SufC, a soluble cytoplasmic ABC-ATPase. SufBCD stimulates the cysteine desulfurase activity of the SufS–SufE complex (Outten *et al.*, 2003). In *Xac*, the cluster of genes *XAC2935–XAC2936–XAC2937–XAC2938* codes for homologues of SufB, SufC, SufD and SufS, respectively, although they were not annotated as such in the *Xac* genome (da Silva *et al.*, 2002). Genes coding for homologues of SufA (*XAC1619*) and SufE (*XAC2355*) are found elsewhere in the *Xac* genome.

Cysteine desulfurases are important pyridoxal 5'-phosphate (PLP) dependent enzymes that are involved in sulfur mobilization in many living organisms (Marquet, 2001; Mihara *et al.*, 2002; Fontecave *et al.*, 2003; Zheng *et al.*, 1994). The chemical steps involved in the cysteine desulfurase activity of the SufS–SufE complex consist of two stages. The first stage consists of the PLP-dependent transfer of sulfur from L-cysteine to a Cys residue on SufS, forming a persulfide linkage. The sulfur is then passed from SufS to a Cys residue on SufE, where it also forms a persulfide linkage (Loiseau *et al.*, 2003; Ollagnier-de-Choudens *et al.*, 2003). The cysteine desulfurase activity of SufS on its



© 2006 International Union of Crystallography  
All rights reserved



own is approximately 30-fold to 50-fold less than that of the SufS–SufE complex (Loiseau *et al.*, 2003).

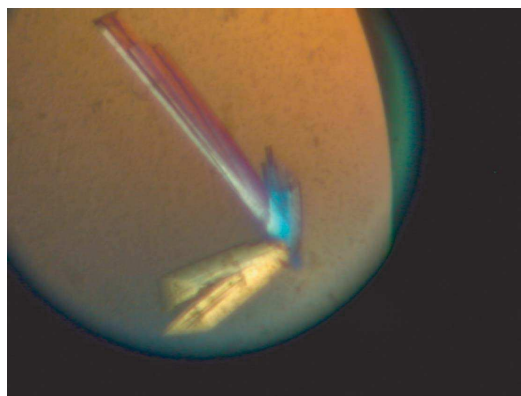
Recently, the structures of five SufE homologues have been determined: the NMR structures of *E. coli* YjdK (PDB code 1ni7; Liu *et al.*, 2005), *Mus musculus* IscU (PDB code 1wfz; unpublished data) and *Haemophilus influenzae* IscU (PDB codes 1r9p and 1q48; Ramelot *et al.*, 2004) and the crystal structures of *E. coli* SufE (PDB code 1mzg; Goldsmith-Fischman *et al.*, 2004) and *Bacillus subtilis* IscU (PDB code 1xjs; unpublished data). In this report, we describe the cloning, expression, purification and crystallization of recombinant *Xac* SufE. An initial electron-density map of the *Xac* SufE crystal structure has been obtained by molecular replacement using the *E. coli* SufE structure as a model (PDB code 1mzg; Goldsmith-Fischman *et al.*, 2004).

## 2. Cloning and expression of XAC2355

The gene coding for SufE (*XAC2355*) was amplified by PCR from *Xac* genomic DNA using the following primers designed based on the published *Xac* genome sequence (da Silva *et al.*, 2002): forward, 5'-CATGCCATGGCTCATATGACCACCTCCCCCTTCC-3'; reverse, 5'-GGAATTCAAGCTTTCCTGCTGCGCGCG-3'. The PCR product was digested with *Hind*III and *Nde*I and subcloned into the pET-3a vector (Studier *et al.*, 1990) previously digested with the same endonucleases. SufE was expressed in *E. coli* strain BL21(DE3)pLysS (Studier *et al.*, 1990). The protein was produced by growing a culture in 2×TY medium to an optical density (600 nm) of 0.8, at which point heterologous protein expression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside and the cells were harvested after growing for 4 h and stored at 203 K.

## 3. Protein purification

Cells from 1 l culture were resuspended in 25 ml 50 mM Tris–HCl pH 8.0, 25% sucrose, 1 mM EDTA, 1 mM PMSF and lysed using a French press. The soluble fraction was applied onto a Q-Sepharose Fast Flow (FF) Hi-Load 16/10 column (Amersham Pharmacia) previously equilibrated with 50 mM Tris–HCl pH 7.0, 1 mM EDTA and 14 mM β-mercaptoethanol. Bound proteins were eluted using a 12 column volume 0–300 mM NaCl gradient. Fractions containing SufE were concentrated using an Amicon system with a 10 kDa molecular-weight cutoff membrane and then purified further by gel filtration on a Superdex 75 prep-grade column (Amersham Pharmacia) equi-



**Figure 1**  
Crystals of SufE. The largest plates are approximately 400 μm in the longest dimension.

**Table 1**

Crystal parameters and data-reduction statistics.

The data sets were measured from a single crystal. Values in parentheses refer to the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters	
<i>a</i> (Å)	45.837
<i>b</i> (Å)	58.507
<i>c</i> (Å)	98.951
No. of images	212
Resolution range (Å)	50.00–1.90 (1.97–1.90)
Wavelength (Å)	1.438
No. of observed reflections	128189
No. of unique reflections	20431 (2002)
$\langle I/\sigma(I) \rangle$	25.6 (4.4)
Multiplicity	6.3 (5.8)
Completeness (%)	93.5 (93.3)
$R_{\text{int}}$ (%)	5.8 (37.3)

$$\dagger R = \sum |I(h) - \langle I(h) \rangle| / \sum I(h).$$

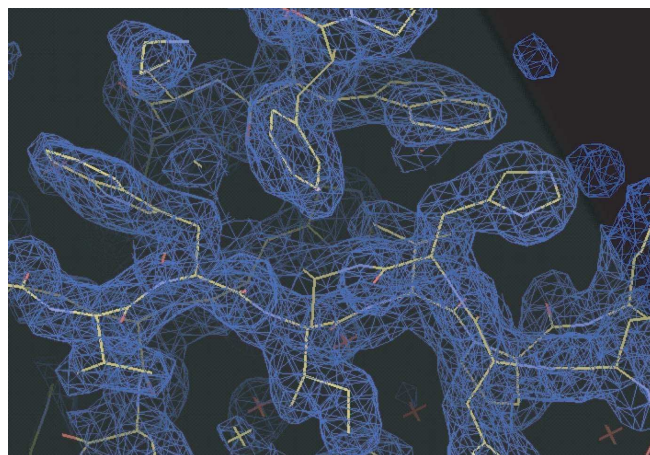
brated with 5 mM Tris–HCl pH 7.0. The purity of the protein was judged by visualization of Coomassie-stained SDS–PAGE.

## 4. Crystallization

*Xac* SufE crystals (Fig. 1) were grown using the hanging-drop vapour-diffusion technique at 291 K. Initial crystallization conditions that produced needle-like crystals were obtained by the sparse-matrix sampling approach using the Index crystallization screen kit from Hampton Research. Optimization was then pursued by varying the precipitant concentration and the buffer pH. Suitable crystals for diffraction experiments were obtained by mixing 1 μl of a 9.8 mg ml<sup>−1</sup> protein solution (5 mM Tris–HCl pH 7.0) with 1 μl reservoir solution consisting of 17–21% (w/v) PEG 5000 and 0.1 M Tris–HCl pH 7.0, followed by equilibration against 0.4 ml reservoir solution. Rectangular crystals appeared within a few days and grew to mature size within a few weeks.

## 5. Data collection and preliminary structure analysis

X-ray diffraction data were collected at the D03B beamline of the Laboratório Nacional de Luz Síncrotron, Campinas, Brazil using a MAR CCD detector. Crystals equilibrated against 19% PEG 5000 and 0.1 M Tris–HCl pH 7.0 were transferred to solution containing



**Figure 2**  
A section of the current SufE crystal  $2F_o - F_c$  electron-density map contoured at 1.0σ.

26% PEG 400 for cryoprotection. The crystal was then flash-frozen in liquid nitrogen and maintained at 100 K in a nitrogen-gas stream during data acquisition. A single data set was collected at 1.438 Å using 1 min exposure and 1° oscillation from 1 to 212°. The diffraction patterns from this data set extended to approximately 1.9 Å resolution (Table 1). The crystal belongs to space group  $P2_12_12_1$ , with unit-cell parameters  $a = 45.837$ ,  $b = 58.507$ ,  $c = 98.951$  Å,  $\alpha = \beta = \gamma = 90^\circ$ . There are two protein molecules per asymmetric unit, as suggested by the Matthews coefficient ( $V_M = 2.7$  Å<sup>3</sup> Da<sup>-1</sup>; solvent content 40.4%).

Diffraction data were indexed, integrated, scaled and merged using the *HKL2000* package (Otwinowski & Minor, 1997). Details of data acquisition and data-processing statistics are shown in Table 1. Initial phases of the *Xac* SufE crystal were calculated by molecular replacement using *Phaser* (Storoni *et al.*, 2004; Read, 2001) from the CCP4 suite (Collaborative Computational Project, Number 4, 1994) with the *E. coli* SufE structure 1mzg (Goldsmith-Fischman *et al.*, 2004) as the search model. The *E. coli* and *Xanthomonas* proteins share 34% sequence identity. The phases were improved by density-modification protocols using the programs *SOLOMON* (Abrahams, 1997) and *DM* (Collaborative Computational Project, Number 4, 1994; Cowtan, 1994). Interpretation of electron-density maps (Fig. 2) and model refinement are currently under way using the programs *COOT* (Emsley & Cowtan, 2004), *REFMAC* (Murshudov *et al.*, 1997) and *CNS* (Brünger *et al.*, 1998).

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Brazil), Associação Brasileira de Tecnologia de Luz Síncrotron (ABTLuS) and FCT/MCT (Portugal).

## References

Abrahams, J. P. (1997). *Acta Cryst.* **D53**, 371–376.

- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cowtan, K. (1994). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
- da Silva, A. C. *et al.* (2002). *Nature (London)*, **417**, 459–463.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Fontecave, M., Ollagnier-de-Choudens, S. & Mulliez, E. (2003). *Chem. Rev.* **103**, 2149–2166.
- Galvão Botton, L. M. P., Katsuyama, A. M., Guzzo, C. R., Almeida, F. C., Farah, C. S. & Valente, A. P. (2003). *FEBS Lett.* **552**, 207–213.
- Goldsmith-Fischman, S., Kuzin, A., Edstrom, W. C., Benach, J., Shastry, R., Xiao, R., Acton, T. B., Honig, B., Montelione, G. T. & Hunt, J. F. (2004). *J. Mol. Biol.* **344**, 549–565.
- Liu, G., Li, Z., Chiang, Y., Acton, T., Montelione, G. T., Murray, D. & Szyperki, T. (2005). *Protein Sci.* **14**, 1597–1608.
- Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. & Barras, F. (2003). *J. Biol. Chem.* **278**, 38352–38359.
- Marquet, A. (2001). *Curr. Opin. Chem. Biol.* **5**, 541–549.
- Mihara, H., Fujii, T., Kato, S., Kurihara, T., Hata, Y. & Esaki, N. (2002). *J. Biochem. (Tokyo)*, **131**, 679–685.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Nachin, L., El Hassouni, M., Loiseau, L., Expert, D. & Barras, F. (2001). *Mol. Microbiol.* **39**, 960–972.
- Nachin, L., Loiseau, L., Expert, D. & Barras, F. (2003). *EMBO J.* **22**, 427–437.
- Ollagnier-de-Choudens, S., Lascoux, D., Loiseau, L., Barras, F., Forest, E. & Fontecave, M. (2003). *FEBS Lett.* **555**, 263–267.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Otten, F. W., Wood, M. J., Munoz, F. M. & Storz, G. (2003). *J. Biol. Chem.* **278**, 45713–45719.
- Ramelot, T. A., Cort, J. R., Goldsmith-Fischman, S., Kornhaber, G. J., Xiao, R., Shastry, R., Acton, T. B., Honig, B., Montelione, G. T. & Kennedy, M. A. (2004). *J. Mol. Biol.* **344**, 567–583.
- Read, R. J. (2001). *Acta Cryst.* **D57**, 1373–1382.
- Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* **D60**, 432–438.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol.* **185**, 60–89.
- Zheng, L., White, R. H., Cash, V. L. & Dean, D. R. (1994). *Biochemistry*, **33**, 4714–4720.