



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**

ALESSANDRA PRANDO

**ESTUDOS BIOFÍSICOS DE CHAPERONAS DE SECREÇÃO E DE
INTERAÇÕES PROTEÍNA-LIGANTE**

TESE DE DOUTORADO APRESENTADA
AO INSTITUTO DE QUÍMICA DA
UNICAMP PARA A OBTENÇÃO DO
TÍTULO DE DOUTORA EM CIÊNCIAS.

ORIENTADORA: PROFA. DRA. LJUBICA TASIC

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA POR ALESSANDRA
PRANDO, E ORIENTADA PELA PROFA. DRA. LJUBICA TASIC.**

Assinatura da Orientadora

CAMPINAS, 2012

FICHA CATALOGRÁFICA ELABORADA POR SIMONE LUCAS - CRB8/8144 -
BIBLIOTECA DO INSTITUTO DE QUÍMICA DA UNICAMP

P886e Prando, Alessandra (1980-).
Estudos biofísicos de chaperonas de secreção e de interações proteína-ligante / Alessandra Prando. – Campinas, SP: [s.n.], 2012

Orientadora: Ljubica Tasic.
Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Química.

1. *Xanthomonas axonopodis* pv. *citri*. 2. Interação proteína-ligante. 3. Hsp90 da laranja. 4. STD.
I. Tasic, Ljubica. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em inglês: Biophysical studies on secretion chaperones and protein-ligand interactions

Palavras-chave em inglês:

Xanthomonas axonopodis pv. *citri*
protein-ligand interactions
Hsp90 from orange
STD NMR

Área de concentração: Química Orgânica

Titulação: Doutora em Ciências

Banca examinadora:

Ljubica Tasic [Orientadora]
Fabio Ceneviva Lacerda Almeida
Maurício Luís Sforça
Carlos Henrique Inácio Ramos
Fábio Cesar Gozzo

Data de defesa: 30/03/2012

Programa de pós-graduação: Química



Onde não falta vontade existe sempre um caminho.
(John Lennon)

Existem apenas duas maneiras de ver a vida. Uma é pensar que não existem milagres e a outra é
pensar que tudo é um milagre.
(Albert Einstein)

Quando somos bons para os outros, somos ainda melhores para nós.
(Benjamin Franklin)

Aos meus Pais Wilson e Maria,
com amor dedico.

Agradecimentos

À Profa. Dra. Ljubica Tasic pela amizade e dedicação durante todos esses anos de convivência, por confiar a mim este trabalho e acompanhá-lo sempre de tão perto. Também quero agradecer a Profa. Lúcia Baptistella pela orientação na Iniciação Científica e Mestrado, tudo que sei de Síntese Orgânica devo à vc!!!

À Universidade Estadual de Campinas e ao Instituto de Química, minha segunda casa há 10 anos, pela oportunidade de realizar este trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela concessão da bolsa de estudos e à Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) por todo o suporte financeiro.

Aos Profs. Carlos Ramos, Carlos Roque e Paulo Moran pelas valiosas sugestões feitas durante o exame de qualificação.

À Sonia, Paula e ao Anderson pela dedicação às análises de RMN e principalmente pelos encaixes concedidos quando mais precisamos.

Ao grupo do Prof. Fábio Gozzo, em especial a Alana e Alexandre por sempre estarem prontos para ajudar e realizar as análises de massas. Ao grupo do Prof. Carlos Ramos, em especial ao Daniel, Yuri, Letícia, Vivian Lee, Danieli, Bárbara e Melissa pelos ensinamentos e ajuda, sou muito grata à vcs, e principalmente pela amizade. Também não posso esquecer de agradecer aos técnicos, pessoal da limpeza (Dona Herotildes!!!) e da zeladoria.

Aos pesquisadores Ana Zeri e Mauricio Sforza por sempre cederem um espaço no LNBio no CNPEM sempre que precisei.

Agradeço aos professores Claudio Tormena, Francisco Reis, Paulo Miranda e Buba pela oportunidade de ser PED deles e aprender um pouco mais sobre a docência.

Aos amigos e colegas de laboratório de Química Biológica I-250, Dani loira, Dani potinho, Ana Paula, as codornas Jenifer e Ana Maria, Lucas, Elias, Fábio, Izabella, Juliana, Junko, Almas, Jéssica IC, Érica, Nelson Durán, Marcelita, Carla, Sandrinha, Rose, Thiago, Stephany, Bruna, Leonardo IC, Leandro IC, Diana IC, Eduardo IC, Willians IC, Larissa, Pryscila, Camila IC, Chico, Zaine, ufa acho que não esqueci ninguém!!! Quero agradecer em especial ao Fábio pela sua sabedoria e por sempre estar disponível para ajudar. E a Izita pelas tortas africanas...a minha nunca ficou igual a sua...ahahah.

Em especial quero agradecer as minhas amigas Vanessa e Deborah (corrigir vestibular sem vcs não seria a mesma coisa), Livon, Ana Paula Lemes, Danieli Ridolfi, Daniela Pott, Éricão e Letícia pelo carinho e amizade.

Aos meus eternos e valiosos amigos Juliana (e Bia), Fábio (Chico), Samanta (e Bia) e Mariana, eles sempre estiveram ao meu lado, sempre me dando apoio, mas também bronca quando merecia...

E agradeço a Deus a oportunidade me dada, Syngenta, que será agora a minha segunda casa por muitos anos.

Ao meu querido companheiro Guilherme, quem esteve sempre ao meu lado mesmo nas horas difíceis tendo muita paciência durante meus períodos de desespero...Vc Guigo é o amigo e companheiro que quero sempre ter por perto, aprendi muito com vc. EU TE AMO meu amor!!!!

Aos meus pais Wilson e Maria que sempre me motivaram a continuidade aos estudos, sem vocês não seria metade do que sou hoje, eu amo vocês. A você vovó

Theresinha *in memoriam*... sempre te amarei!!!! Aos meus irmãos Wilson, Juliana, Valéria e Luciana por tantos anos de convivência e amizade.

Curriculum Vitae

ALESSANDRA PRANDO

FORMAÇÃO ACADÊMICA

- 2008-2012 Instituto de Química – UNICAMP
Doutorado em Ciências
Projeto: Estudos Biofísicos de Chaperonas de Secreção e de Interações Proteína-Ligante.
Orientadora: Ljubica Tasic.
Agência Financiadora: Capes
- 2005-2007 Instituto de Química – UNICAMP
Mestrado em Química Orgânica
Projeto: Estudos sintéticos para a preparação de derivados lactâmicos: uso do ácido quínico.
Orientadora: Lúcia Helena Brito Baptistella.
Agência Financiadora: CNPq
- 2001-2004 Instituto de Química – UNICAMP
Bacharelado em Química Tecnológica
Projeto de Iniciação Científica: Estudos para transformações de dióis em hetero-amino derivados.
Orientadora: Lúcia Helena Brito Baptistella
Agência Financiadora: CNPq

PUBLICAÇÕES EM PERIÓDICOS

1. Cagliari, T.C.; da Silva, V.C.H.; Borges, J.C.; Prando, A.; Tasic, L.; Ramos, C.H.I.; Sugarcane Hsp101 is a hexameric chaperone that binds nucleotides. *International Journal of Biological of Macromolecules* **2011**, *42*, 1022-1030.

2. Fattori, J.; Prando, A.; Martins, A. M.; Rodrigues, F. H. S.; Tasic, L.. Bacterial Secretion Chaperones, *Protein and Peptide Letters* **2011**, *18*, 158-166.
3. Fattori, J. ; Prando, A. ; Assis, L. H. P. ; Aparicio, R. ; Tasic, L. . Structural Insights on Two Hypothetical Secretion Chaperones from *Xanthomonas axonopodis* pv. *citri*. *The Protein Journal* **2011**, *30*, 324-333.
4. Sussulini, A.; Prando, A.; Maretto, D.A.; Poppi, R.J.; Tasic, L.; Banzato, C.E.M.; Arruda, M.A.Z. Metabolic Profiling of Human Blood Serum from Treated Patients with Bipolar Disorder Employing 1H NMR Spectroscopy and Chemometrics, *Analytical Chemistry* **2009**, *81*, 9755-9763.

PATENTE

1. Baptistella, L.H.B.; Prando, A.; “Processo para a preparação de lactamas polissubstituídas” INOVA-Unicamp-aceita, 09.12.08, Prot. 018080075749.

APRESENTAÇÃO ORAL EM CONGRESSOS

1. Prando, A.; Tasic, T. *Aplicação de RMN em Estudos de Interações entre Proteínas de Choque Térmico e Ligantes*. XI Jornada Brasileira de Ressonância Magnética, Curitiba, PR, 2010.
2. Poster Award with a Mention of Honor, Hands-on Course on Protein and Proteomics, REQUINTE Universidade Nova de Lisboa, Lisboa, Portugal, 2008.

ESTÁGIO EXTERIOR

1. *Synthesis of Compounds with Biology Activity against Coxsackie Viruses* - Universität zu Lübeck, Biochemie Institute, Lübeck, Germany with scholarship DAAD (Deutscher Akademischer Austausch Dienst) - (08/08- 12/08).

TRABALHOS EXTRA-CURRICULARES

1. Programa Estágio Docente (PED C): QG 623 - Química Orgânica Experimental (Curso: Farmácia), 1º semestre de 2007, Supervisor: Prof. Dr. Cláudio Francisco Tormena. QO-422 - Química Orgânica Experimental (Curso: Engenharia Química),

2º semestre de 2009, Supervisor: Prof. Dr. Francisco de Assis Reis. QO-622 - Química Orgânica Experimental (Curso: Química), 2º semestre de 2010, Supervisor: Paulo Cesar Muniz de Lacerda Miranda. QG-651 - Bioquímica Experimental (Curso: Química), 1º semestre de 2011, Supervisor: Ljubica Tasic.

2. Comissão Organizadora do V-Fórum de Pós-Graduação do Instituto de Química da UNICAMP - 19 e 20 de outubro de 2011, "A trajetória da Química: História, Economia e Aplicações."

RESUMO

ESTUDOS BIOFÍSICOS DE CHAPERONAS DE SECREÇÃO E DE INTERAÇÕES PROTEÍNA-LIGANTE

Até o momento, pouco se sabe sobre os mecanismos de virulência da bactéria *Xanthomonas axonopodis* pv. *citri* (*Xac*), agente causador do cancro cítrico. Acredita-se que chaperonas de secreção (CS) estão envolvidas no processo de patogenicidade da *Xac* primeiramente formando complexos com fatores de virulência e auxiliando no encaminhamento desses para os sistemas de secreção utilizando o ATP como fonte de energia.

Neste trabalho foram adquiridos dados de fluorescência de emissão, dicroísmo circular, desenovelamento térmico e de ressonância magnética nuclear de ^1H NMR e de 2D $\{^{15}\text{N}, ^1\text{H}\}$ HSQC de duas proteínas da *Xac*, a XAC1990 (FlgN) e XACb0033. Para ambas proteínas foram propostas estruturas 3D usando a análise de *footprinting* com restrições SASA e RMSD. Para as estruturas propostas foi verificado que os dados de fluorescência corroboram com a estrutura 3D não ocorrendo o mesmo para os dados de CD e NMR que revelaram baixo conteúdo helicoidal além de ausência de estrutura 3D. A interação da proteína FlgN com a sua proteína parceira FlgK também foi sugerida através das análises de CD e fluorescência.

Na segunda parte do trabalho foram estudadas as interações entre a proteína Hsp90 da laranja com diferentes ligantes aplicando a técnica de **Saturation Transfer Difference** (STD-NMR) e espectroscopia de fluorescência. Estas análises revelaram dados que corroboraram com o modelo proposto e, além disso, indicaram que os hidrogênios H-8 e H-2 da adenina e H-1' da ribose estão localizados no sítio ligante da proteína com os fosfatos orientados para fora. Através da fluorescência foram calculados os valores de K_d e foi verificado que a geldanamicina é um potente inibidor de Hsp90 da laranja.

ABSTRACT

BIOPHYSICAL STUDIES ON SECRETION CHAPERONES AND PROTEIN-LIGAND INTERACTIONS

So far, the *Xanthomonas axonopodis* pv. *citri* (*Xac*) mechanisms of bacterial virulence is unknown. It is believed that secretion chaperones (CS) are involved in the *Xac*'s virulence process by first forming complexes with virulence factors, and assisting in their presentation to corresponding secretion systems using ATP as a source of energy.

Fluorescence emission, circular dichroism, thermal unfolding and nuclear magnetic resonance NMR ^1H and 2D $\{^{15}\text{N}, ^1\text{H}\}$ HSQC data from two proteins of *Xac*, XAC1990 (FlgN) and XACb0033 were collected. For both proteins, 3D structures were proposed using the *footprinting* analysis with RMSD and SASA restrictions. For the proposed structures were verified which the fluorescence data were consistent with the 3D structure. The CD and NMR data revealed low-helical content and absence of 3D structure. The interaction of the protein FlgN with its partner, FlgK, was suggested by CD and fluorescence analysis.

In the second part, the interactions between the orange's Hsp90 protein with different ligands using Saturation Transfer Difference (STD-NMR) and fluorescence spectroscopy techniques were studied. These analyzes revealed which the data were consistent with the proposed model and moreover showed that the adenine's hydrogens H-8 and H-2 and ribose's hydrogen H-1' are located in the protein binding site with the phosphate driven out. By fluorescence values were calculated K_d and it was verified that geldanamycin is a potent inhibitor of orange's Hsp90.

Sumário

LISTA DE ABREVIATURAS E SIGLAS	XIX
LISTA DE ILUSTRAÇÕES	XXI
LISTA DE TABELAS	XXVII
ANEXOS.....	XXIX
1.INTRODUÇÃO	3
1.1 Interesse.....	3
1.1.1. <i>Xanthomonas axonopodis</i> pathovar <i>citri</i>	5
1.1.2. Sistema de Secreção	7
1.1.3. Chaperonas de Secreção.....	13
1.1.4. Hsp90 da laranja	17
1.1.5. Proteínas FlgN e XACb0033.....	20
1.2 Estudos Estruturais	23
1.2.1. Estrutura e função de proteínas	25
1.2.2. Dicroísmo Circular	26
1.2.3. Fluorescência	29
1.2.4. Espectrometria de Massas em análise proteínas	32
1.2.5. Ressonância Magnética Nuclear em análises de interações proteínas-ligantes.....	37
2. OBJETIVOS	41
3. PARTE EXPERIMENTAL	45
3.1 Material: proteínas e ligantes.....	47
3.2 Extração do DNA plasmidial (lise alcalina)	48
3.3. Preparação de células competentes.....	50
3.4. Transformação pelo método de choque térmico.....	51
3.5. Expressão proteica em meio LB	52
3.6. Expressão proteica em Meio Mínimo	53
3.7. Eletroforese em gel	54
3.8. Lise celular	56
3.9. Tratamento com uréia (reenovelamento).....	56
3.10. Purificação das proteínas.....	57
3.11. Determinação da concentração de proteínas: Método de Edelhoch	58
3.12. Método de Bradford.....	59
3.13 Expressão e purificação da proteína FlgK	59
3.14 Expressão e purificação da proteína XACb0032	61
3.15. Espectrometria de Massas (MS) – Identificação das proteínas.....	62
3.15.1 Digestão e dessalinização.....	63
3.15.2 Identificação de peptídeos por cromatografia líquida acoplada a espectrometria de massas (LC-MS).....	63
3.15.3 Processamento dos dados de LC-MS e busca em banco de dados.....	65

3.16. Espectrometria de Massas: <i>Footprinting</i> e <i>Cross-Linking</i>	65
3.16.1 Reação com DEPC	65
3.16.2 Digestão proteolítica das proteínas	66
3.16.3 Análise por espectrometria de massas	66
3.16.4 Checagem manual dos espectros	66
3.17. Fluorescência	67
3.17.1 Interações proteína-ligante	68
3.18. Dicroísmo Circular (CD)	70
3.19. Ressonância Magnética Nuclear (NMR): ^1H e HSQC $\{^{15}\text{N}/^1\text{H}\}$	72
3.19.1 Ressonância Magnética Nuclear (NMR): STD e HETCOR	74
3.20. Bioinformática	75
4. RESULTADOS e DISCUSSÕES	77
4.1 A FlgN	81
4.2 XACb0033	102
4.3 Proposta de Estrutura para as proteínas FlgN e XACb0033	120
4.4 Interações Proteínas e Ligantes	135
5. CONCLUSÕES	159
ANEXOS: Espectros de NMR das proteínas XACb0033 e FlgN	163
ANEXOS: Artigos	191

LISTA DE ABREVIATURAS E SIGLAS

ϵ : coeficiente de absorvância molar ($M^{-1}.cm^{-1}$)

$[\theta]$: elipticidade molar residual

ALC: Agentes de ligação cruzada

ADP: adenosina 5'-difosfato

APS: persulfato de amônio

ATP: adenosina 5'-trifosfato

ATP γ S: adenosina 5'-[γ -tio]-trifosfato

BSA: albumina de soro bovino (do inglês *Bovine Serum Albumin*)

CD: Dicroísmo Circular (do inglês *Circular Dichroism*)

CNPEN: Centro Nacional de Pesquisa em Energia e Materias

CICS: deslocamento químico induzido por complexação (do inglês *Complexation-induced Chemical Shift*)

CS: Chaperona de secreção

Da: Dalton

DMSO: dimetilsulfóxido

DTT: ditioneitol

EDTA: ácido etilenodiamino tetra-acético

ESI: Ionização por Eletrospray

FPLC: cromatografia líquida (do inglês *Flow Protein Liquide Chromatography*)

Hetcor:experimento de correlação heteronuclear (do inglês Heteronuclear Correlation experiment)

Hsp90: proteína de choque térmico de ~90 kDa (do inglês *Heat shock protein 90 kDa*)

Hsp100: proteína de choque térmico de ~100 kDa

IPTG: isopropil β -D-1-tiogalactopiranosídeo

LB: Luria Bertrani

LC-MS: Cromatografia Líquida acoplada ao Espectrometria de Massas (Liquid Chromatography-Mass Spectrometry)

LNBio: Laboratório Nacional de Biociências que integra o CNPEM

MM: massa molecular

MALDI: Desorção/Ionização a laser assistida por matriz (do inglês *Matrix assisted laser desorption/ionization*)

Mpf: Formação do par conjugado (do inglês *Mating pair formation*)

MS: Espectrometria de Massas (do inglês *Mass Spectrometry*)

n: número de resíduos de aminoácidos da proteína

NMR: Ressonância Magnética Nuclear (do inglês *Nuclear Magnetic Resonance*)

pI: ponto isoelétrico

PMF: Força transiente de próton(saída/entrada) (do inglês *Proton Motive Force*)

RMSD: desvio da média quadrática (do inglês *Root Mean Square Deviation*)

rpm: rotações por minuto

SASA: área superficial acessível ao solvente (do inglês *solvent accessible surface area*)

SDS-PAGE: eletroforese em gel de poliacrilamida em presença de dodecil-sulfato de sódio (do inglês *sodium dodecyl-sulphate polyacrilamide gel electrophoresis*)

STD: Diferença de Saturação Transferida (do inglês *Saturation Transfer Difference*)

TE/RNase: solução tampão Tris-EDTA contendo a enzima ribonuclease

TEMED: N,N,N',N'-tetrametiletlenodiamina

Tris: 2-Amino-2-hidroximetil-propano-1,3-diol

ToF: Tempo de voo (do inglês *Time of Flight*)

T2SS: Sistema de secreção do tipo II (do inglês *Type II Secretion System*)

T3SS: Sistema de Secreção do tipo III (do inglês *Type III Secretion System*)

T4SS: Sistema de Secreção do tipo IV (do inglês *Type IV Secretion System*)

UV/VIS: Espectroscopia no ultravioleta/visível

Vis: visível

Xac: *Xanthomonas axonopodis* pv. *citri*

λ : comprimento de onda

θ : elipticidade

LISTA DE ILUSTRAÇÕES

Figura I-1: Lesões causadas pelo cancro cítrico em frutos, folhas e ramos.	4
Figura I-2: Sistemas de secreção das bactérias Gram-negativas (I-VI). ME= membrana externa e MI= membrana interna.	8
Figura I-3: Sistema de secreção flagelar mostrando as seis proteínas integrais da membrana (FlhA, FlhB, FliO, FliP, FliQ e FliR) e as três proteínas solúveis (FliH, FliI e FliJ).	12
Figura I-4: Ilustração dos sistemas de secreção bacteriano: T3SS (esquerda), T4SS (centro) e flagelar (direita). No sistema de secreção do tipo III a chaperona na forma de um dímero (1) encontra com a proteína parcialmente enovelada (2) forma um complexo proteína substrato/chaperona (3); então a chaperona mantém a proteína substrato parcialmente enovelada no seu domínio de ligação e esta é encaminhada para o T3SS para secreção. No sistema de secreção do tipo IV, as chaperonas agem do mesmo modo atuando como monômeros ou dímeros e exportando além de proteínas (I), fita simples de DNA (II) ou complexo proteína-DNA (III). No sistema flagelar as chaperonas agem como dímero (1) e ligam-se a proteína flagelar prevenindo agregação e precipitação além de participar na formação da maquinaria do flagelo.	14
Figura I-5: a) Estrutura da chaperona de secreção do T4SS, VirE1 (em vermelho) complexada com sua proteína parceira, a VirE2 (em azul; PDB:3BTP). b) Estrutura da chaperona de secreção flagelar, FliT de <i>Salmonella</i> sp. (PDB: 3A7M).	16
Figura I-6: Formas inativa (a) e nativa (b) da proteína Hsp90	20
Figura I-7: As três vistas do modelo gerado por SAXS da FlgN (em cinza) sobreposto com a estrutura calculada pelo I-TASSER (em vermelho).	22
Figura I-8: Estrutura dos reagentes de ligação cruzada.	34
Figura I-9: Esquema de reação entre o ALC reativo a grupos amino (~~~ cadeia espaçadora)	34
Figura I-10: Modificação dos resíduos de histidina pelo DEPC.	36
Figura I-11: Ilustração do mapeamento dos grupos do epítipo (Group epitope mapping - GEM) para dada interação proteína-ligante em uma troca rápida entre ligante livre e complexado. Os grupos representados por um átomo de hidrogênio grande (H) estão em contato mais próximo com a proteína, enquanto os H de tamanho mediano simbolizam grupos com menor interação. Os H menores representam um grupo com quase nenhum contato com a proteína, então recebendo saturação mínima.	39
Figura III-1: Curva de calibração de Bradford para quantificação da proteína FlgK (Espectrofotômetro HP 8453, cubeta de acrílico, 10 mm de caminho ótico).	61
Figura III-2: Curva de calibração de Bradford para quantificação da proteínas (Espectrofotômetro HP 8453, cubeta de acrílico, 10 mm de caminho ótico).	62
Figura IV-1: Fluxograma geral da expressão das proteínas em larga escala em meio LB e meio mínimo.	80
Figura IV-2: Sequencia primária da proteína FlgN. Resíduo de triptofano em destaque (vermelho).	81
Figura IV-3: Gel de eletroforese SDS-PAGE (15%) da expressão da FlgN, tendo na sequência, da esquerda para a direita: marcador de peso molecular de proteína; amostras não induzidas (NI) e 1h, e 2h após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.)	81

Figura IV-4: a) Cromatograma da purificação do sobrenadante do reenovelamento da FlgN, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequencia: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína FlgN pura mostrada em destaque (circulo vermelho).	82
Figura IV-5: a) Cromatograma total (LC-MS) dos peptídeos da digestão triptica da FlgN e b) Cromatogramas extraídos para íons dos peptídeos com maior score no MASCOT oriundos da FlgN.	83
Figura IV-6: Espectros de MS/MS gerados a partir dos cinco peptídeos com maior score da proteína FlgN. a) MS/MS do peptídeo TESAPSYDAK. b) MS/MS do peptídeo LSDALAGER. c) MS/MS do peptídeo H2N-MNVNEFLQR. d) MS/MS do peptídeo QALLENIDGLMR. e) MS/MS do peptídeo LSALQALEAAMPAGEEER.	84
Figura IV-7: Sequencia primária da proteína FlgN. Os peptídeos encontrados para a FlgN (MS/MS) estão mostrados em destaque (vermelho), sendo obtidos 68% de cobertura da sequencia.	85
Figura IV-8: Espectros de massas da proteína FlgN. a) Espectro de massa da proteína intacta e b) após processamento dos dados.	85
Figura IV-9: Espectros de fluorescência de excitação e de emissão da proteína FlgN com uma concentração de 78 $\mu\text{mol.L}^{-1}$ em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0).	86
Figura IV-10: Ilustração dos resultados obtidos em análise de CD para a proteína FlgN: a) Espectro de CD, b) Cálculo da estrutura secundária da proteína FlgN obtida pelo programa CDNN a partir dos dados de CD. Abaixo, gráfico gerado pelo CDNN.	87
Figura IV-11: a) Desenovelamento térmico 20 – 90°C acompanhado à 222 nm, b) Renovelamento térmico 90 – 20°C acompanhado à 222 nm, c) Espectros de desenovelamento e renovelamento térmico da FlgN e d) Desenovelamento térmico (5°C de incremento).	88
Figura IV-12: Espectro de ^1H NMR da FlgN (49 $\mu\text{mol L}^{-1}$) em tampão fosfato de sódio 100 mmol L^{-1} com adição de D_2O (5% v/v) obtido no equipamento INOVA 500 MHz (499,98 MHz). Expansão da região de 6,0 à 8,0 ppm mostrado em evidência.	89
Figura IV-13: Gel de eletroforese SDS-PAGE (15%) da expressão da FlgN, tendo na sequência: marcador de peso molecular de proteína; amostras não induzidas (NI), 30 minutos após a adição de IPTG e 1h, 2h e 3h após a adição de rifamicina, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.)	90
Figura IV-14: a) Cromatograma da purificação do sobrenadante do reenovelamento da FlgN, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequencia: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína FlgN pura mostrada em destaque (circulo vermelho).	91
Figura IV-15: Espectro de ^1H NMR da FlgN.	91
Figura IV-16: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN. A amostra proteica está em tampão fosfato de sódio 100 mmol.L^{-1} pH 8,0 com adição de (D_2O 5%, v/v) e os espectros foram obtidos em equipamento INOVA 600 MHz.	92
Figura IV-17: Sequencia primária da proteína FlgK. Resíduos de triptofano em destaque (vermelho).	93
Figura IV-18: Gel de eletroforese SDS-PAGE (15%) da expressão da FlgK, tendo na sequência: marcador de peso molecular de proteína; amostras não induzidas (NI) e 1h, 2h, 3h e 16h após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (S).	94
Figura IV-19: Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequencia: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína FlgK pura mostrada em destaque (circulo vermelho). Na ultima lacuna esta a proteína aplicada na coluna.	94

Figura IV-20: Espectros de fluorescência de excitação e emissão das proteínas a) FlgK e b) FlgN.	95
Figura IV-21: Espectros de fluorescência dos complexos de FlgN e FlgK na proporção a) 1:1 e b) 2:1 (FlgN/ FlgK).	95
Figura IV-22: Espectros de CD. a) proteína FlgK livre. b) proteína FlgN livre. c) Espectros de CD das proteínas e do complexo FlgN + FlgK (1:1). d) Espectros de CD das proteínas e do complexo FlgN + FlgK (2:1).	97
Figura IV-23: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN.	98
Figura IV-24: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo FlgN/FlgK (2:1).	99
Figura IV-25: Representação em cartoon da estrutura da FlgK com N- e C- terminais marcados (I-TASSER C-score = -2,20).	100
Figura IV-26: Representação em cartoon da estrutura da FlgN com N- e C- terminais marcados (I-TASSER C-score = -2,51) e o resíduo de Trp (W70) em verde (exposto ao solvente).	100
Figura IV-27: Sequencia primária da proteína XACb0033. Resíduos de triptofano em destaque (vermelho).	102
Figura IV-28: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0033, tendo na sequência: amostras não induzidas (NI) e 1h, 2h, 3h e 16hs após a adição de IPTG, marcador de peso molecular de proteína, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.). Os monômeros e dímeros estão indicados no gel.	102
Figura IV-29: a) Cromatograma da purificação do sobrenadante do reenovelamento da XACb0033, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequencia: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína XACb0033 pura mostrada em destaque (circulo vermelho).	103
Figura IV-30: a) Cromatograma total (LC-MS) dos peptídeos da digestão tríptica da XACb0033 e b) Cromatogramas extraídos para íons dos peptídeos com maior score no MASCOT oriundos da XACb0033.	104
Figura IV-31: Espectros de MS/MS gerados a partir dos três peptídeos com maior score da proteína XACb0033. a) MS/MS do peptídeo LAAQADMADTDMQR. b) MS/MS do peptídeo MAGLRPVQIWVPDTR. c) MS/MS do peptídeo FMDEALADMDGWTE.	105
Figura IV-32: Sequencia primária da proteína XACb0033. Os peptídeos encontrados para a XACb0033 (MS/MS) estão mostrados em destaque (vermelho), sendo obtidos 59% de cobertura da sequencia.	105
Figura IV-33: Espectros de massas da proteína XACb0033. a) Espectro de massa da proteína intacta e b) após processamento dos dados.	106
Figura IV-34: Espectros de fluorescência de excitação e de emissão da proteína XACb0033 com uma concentração de $24,8 \mu\text{mol.L}^{-1}$ em tampão fosfato de sódio 25 mmol.L^{-1} , pH 8,0.	107
Figura IV-35: Ilustração dos resultados obtidos em análise de CD para a proteína XACb0033: a) Espectro de CD, (b) Cálculo da estrutura secundária da proteína XACb0033 obtida pelo programa CDNN a partir dos dados de CD. Abaixo gráfico gerado pelo CDNN. Para os cálculos de CD o valor do branco (tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0) foi descontado.	108
Figura IV-36: Desenovelamento térmico $20 - 90^\circ\text{C}$ (5°C de incremento).	108
Figura IV-37: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0033, tendo na sequência: marcador de peso molecular de proteína; amostras não induzidas (NI), 30 minutos após a adição de IPTG e 1h, 2h e 16h após a adição de rifamicina, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.)	109

Figura IV-38: a) Cromatograma da purificação do sobrenadante do reenovelamento da XACb0033, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína XACb0033 pura mostrada em destaque (círculo vermelho).....	110
Figura IV-39: Espectro de ^1H NMR da XACb0033.....	110
Figura IV-40: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033. A amostra proteica está em tampão fosfato de sódio 100 mmol.L^{-1} com 50 mmol.L^{-1} de NaCl com adição de (D ₂ O 5%, v/v) e os espectros foram obtidos em equipamento INOVA 600.....	111
Figura IV-41: Sequência primária da proteína XACb0032.....	113
Figura IV-42: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0032, tendo na sequência: marcador de peso molecular de proteína, amostras não induzidas (NI) e 1h, 2h e 16hs após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.).....	113
Figura IV-43: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0032, tendo na sequência: marcador de peso molecular de proteína, amostras não induzidas (NI), 1h, 2h e 16hs após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.), sobrenadante após a mudança de tampão (S1) e pellet após a mudança de tampão (P1).	114
Figura IV-44: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0032, tendo na sequência: marcador de peso molecular de proteína, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.), pellet após a mudança de tampão (P1), sobrenadante após a mudança de tampão (S1), pellet com DNA-1 (P33), sobrenadante com DNA-1 (S33), pellet com DNA-2 (P306) e sobrenadante com DNA-2 (S306).....	115
Figura IV-45: a) Cromatograma da purificação do sobrenadante do complexo da XACb0032 + DNA-1, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína XACb0032 pura mostrada em destaque (círculo vermelho).....	116
Figura IV-46: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da proteína XACb0033.....	117
Figura IV-47: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo XACb0033/XACb0032 na razão molar 2:1.....	117
Figura IV-48: a) Representação em cartoon da estrutura da XACb0032 com N- e C- terminais marcados (I-TASSER C-score = -1,13 o único gerado pelo programa) e b) Representação em cartoon da estrutura da XACb0033 com N- e C- terminais marcados (I-TASSER C-score = -2,41, o valor de maior pontuação dentre os cinco gerados) e os resíduos de Trp (W70 e W26) em verde (exposto ao solvente).....	118
Figura IV-49: Espectro de MS do peptídeo LAAQADMADTDMQR da XACb0033 que reagiu com o DEPC.....	121
Figura IV-50: Espectro de MS do peptídeo MAGLRPVQIWVPDTR da XACb0033 que reagiu com o DEPC.....	121
Figura IV-51: Espectro de MS do peptídeo TESAPSYDAK da FlgN que reagiu com o DEPC.	122
Figura IV-52: Espectro de MS do peptídeo TESAPSYDAKGQSSVLR da FlgN que reagiu com o DEPC.....	122
Figura IV-53: Espectro de MS do peptídeo HTQDKLSALQALEAAMPAGEEER da FlgN que reagiu com o DEPC.....	123
Figura IV-54: Espectro de MS do peptídeo HTQDKLSALQALEAAMPAGEEER da FlgN que reagiu com o DEPC.....	123

Figura IV-55: Representação em cartoon da estrutura da XACb0033 com N- e C- terminais marcados (I-TASSER C-score = -2,41, o valor de maior pontuação dentre os cinco gerados) e os resíduos de Trp (W70 e W26) em verde (exposto ao solvente).	130
Figura IV-56: Representação em cartoon da estrutura da FlgN com N- e C- terminais marcados (I-TASSER C-score = -2,51) e o resíduo de Trp (W70) em verde (exposto ao solvente).	131
Figura IV-57: Os 15 modelos alinhados gerados pelos programas I-TASSER e QUARK para a proteína XACb0033 e ao lado valores de RMSD gerados no programa VMD usando-se como modelo a estrutura 1.	132
Figura IV-58: Os 15 modelos alinhados gerados pelos programas I-TASSER e QUARK para a proteína FlgN e ao lado valores de RMSD gerados no programa VMD usando-se como modelo a estrutura 1.	132
Figura IV-59: Representação em cartoon da estrutura da XACb0033 com N- e C- terminais marcados e os resíduos de Trp (W70 e W26) em verde. (a) Modelo 1 gerado pelo I-TASSER. (b) Modelo 6 gerado pelo Quark.....	133
Figura IV-60: Representação em cartoon da estrutura da FlgN com N- e C- terminais marcados e o resíduo de Trp (W78) em verde. (a) Modelo 1 gerado pelo I-TASSER. (b) Modelos 5 gerado pelo I-TASSER. (c) Modelo 13 gerado pelo Quark.	134
Figura IV-61: Espectros de ¹ H NMR (499,99 MHz) em D ₂ O. (a) ADP 1 mmol.L ⁻¹ e (b) ATP 1 mmol.L ⁻¹	136
Figura IV-62: Espectros de ¹ H NMR (499,99 MHz) em D ₂ O do ATP _γ S 1 mmol.L ⁻¹	137
Figura IV-63: Espectros de NMR da proteína Hsp100 com o nucleotídeo ATP na razão molar 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.	138
Figura IV-64: Espectros de NMR da proteína Hsp100 com o nucleotídeo ATP _γ S na razão 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.....	139
Figura IV-65: Espectros de NMR da proteína Hsp100 com o nucleotídeo ADP na razão 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.....	140
Figura IV-66: Mapas de contorno de experimentos HETCOR { ¹³ C- ¹ H}. (a) ATP _γ S (40 μL de D2O + 5,0 mg de ATP _γ S) e (b) ATP _γ S + Hsp100 (20μL de D2O + 2,5 mg de ATP _γ S + 20 μL de Hsp100, 120 μmol.L ⁻¹). Em vermelho estão destacadas as correlações ¹³ C- ¹ H em ATP _γ S que sofreram deslocamentos após complexação com a Hsp100.	142
Figura IV-67: Espectros de NMR da proteína Hsp90 da laranja com o nucleotídeo ATP na razão molar 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.	145
Figura IV-68: Espectros de NMR da proteína Hsp90 da laranja com o nucleotídeo ATP _γ S na razão molar 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.	146
Figura IV-69: Sequencia primária da proteína Hsp90 da laranja. Resíduos de triptofano em destaque (vermelho).	148
Figura IV-70: Espectro de fluorescência de excitação e emissão da Hsp90 da laranja 2,67 μmol.L ⁻¹	149
Figura IV-71: Espectros de fluorescência de excitação e emissão da geldanamicina 1,78 mmol L ⁻¹ e do ATP _γ S 7,30 mmol.L ⁻¹	149
Figura IV-72: Espectros de fluorescência de excitação e emissão do ADP 13,00 mmol L ⁻¹ e do ATP 13,00 mmol.L ⁻¹	150
Figura IV-73: Espectro de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com ADP 80 mmol.L ⁻¹	150
Figura IV-74: Espectro de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com ATP 80 mmol.L ⁻¹	151

Figura IV-75: Espectro de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com ATP γ S 80 mmol.L ⁻¹	151
Figura IV-76: Espectro de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com geldanamicina 1,78 μ mol.L ⁻¹	152
Figura IV-77: Gráficos de Stern-Volmer para os ligantes (a) geldanamicina; (b) ATP γ S frente a sua proteína alvo Hsp90 da laranja.....	153
Figura IV-78: Representação em cartoon da estrutura da Hsp90 da laranja em verde (I-TASSER C-score = -2,82) com seu sítio ligante em azul e em vermelho, o sítio ligante da 1yet (PDB)...	155
Figura IV-79: Representação em cartoon da estrutura da Hsp90 da laranja em verde com seu sítio ligante mostrado em azul com a estrutura do ATP.	156
Figura IV-80: Corte transversal da superfície do sítio ligante da Hsp90 da laranja com a estrutura do ATP.....	156

LISTA DE TABELAS

Tabela III-1. Soluções da lise alcalina	49
Tabela III-2. Soluções utilizadas na preparação de células competentes	50
Tabela III-3. Soluções utilizadas na transformação bacteriana	51
Tabela III-4. Solução estoque dos antibióticos e IPTG.....	52
Tabela III-5. Soluções utilizadas expressão proteica em meio mínimo	53
Tabela III-6. Soluções tampão utilizadas no preparo de amostras	55
Tabela III-7. Geis de acrilamida	55
Tabela III-8. Solução de acrilamida utilizada na preparação dos géis	55
Tabela III-9. Soluções corante, descorante e tampão de corrida.....	55
Tabela III-10. Soluções utilizadas na lise celular.....	56
Tabela III-11. Soluções utilizadas no processo de reatualização	57
Tabela III-12. Reagente de Edelhoch	59
Tabela III-13: Gradiente linear de eluição dos peptídeos	64
Tabela III-14: Concentrações utilizadas para os ensaios de interações proteína-proteína	68
Tabela III-15. Concentração e comprimento de onda de excitação e emissão máximos em ensaios de interação ATP, ADP, ATP γ S e geldanamicina	69
Tabela III-16. Condições testadas em ensaios de RMN para a FlgN	73
Tabela III-17. Condições testadas em ensaios de RMN para a XACb0033	73
Tabela IV-1: Absorbância da XACb0032 + DNA da XACb0033 à 280 nm e 260 nm.	115
Tabela IV-2: Valores de SASA (Å^2) para o peptídeo LAAQADMADTDMQR gerado a partir da análise de footprinting da proteína XACb0033. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de footprinting	125
Tabela IV-3: Valores de SASA (Å^2) para o peptídeo MAGLRPVQIWVDPDR gerado a partir da análise de footprinting da proteína XACb0033. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de footprinting	126
Tabela IV-4: Valores de SASA (Å^2) para o peptídeo TESAPSYDAK gerado a partir da análise de footprinting da proteína FlgN. Os aminoácidos em vermelho são os resíduos que reagem com o DEPC na análise de footprinting	126
Tabela IV-5: Valores de SASA (Å^2) para o peptídeo TESAPSYDAKGQSSVLR gerado a partir da análise de footprinting da proteína FlgN. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de footprinting	127
Tabela IV-6: Valores de SASA (Å^2) para o peptídeo HTQDKLSALQALEAAM PAGEEER gerado a partir da análise de footprinting da proteína FlgN. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de footprinting	127
Tabela IV-7: Valores de SASA (Å^2) para o peptídeo HTQDKLSALQALEA AMPAGEEER gerado a partir da análise de footprinting da proteína FlgN. Os aminoácidos em vermelho são os resíduos que reagem com o DEPC na análise de footprinting	128
Tabela IV-8: Mapa de epítomos obtido em interações da Hsp100 da cana de açúcar com ATP γ -S, ATP e ADP.....	141
Tabela IV-9: Atribuição dos sinais de HETCOR obtidos em interações da Hsp100 com ATP γ -S em comparação com ATP γ -S. A diferença entre os deslocamentos químicos (CICS-Complexation Induced Chemical Shift) foram calculados.....	143
Tabela IV-10: Mapa de epitopo obtido em STD-NMR para os hidrogênios H-8, H-2 e H-1' obtidos em interações da Hsp90 da laranja com ATP γ S e ATP.....	147

Tabela IV-11: Concentração e comprimento de onda de excitação e emissão máximos em ensaios de interação ATP, ADP, ATP γ S e geldanamicina. 148

ANEXOS

Anexo 1: Espectro de ^1H NMR da FlgN ($198 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	165
Anexo 2: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($198 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	166
Anexo 3: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($198 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz. Após adição de glicerol.....	167
Anexo 4: Espectro de ^1H NMR da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	168
Anexo 5: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	169
Anexo 6: Espectro de ^1H NMR da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5mmol.L^{-1} pH 8,0 a 35°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	170
Anexo 7: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5mmol.L^{-1} pH 8,0 a 35°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	171
Anexo 8: Espectro de ^1H NMR do complexo FlgN + FlgK na razão molar 2:1 em tampão fosfato de sódio (25mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	172
Anexo 9: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo FlgN + FlgK na razão molar 2:1 em tampão fosfato de sódio (25mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	173
Anexo 10: Espectro de ^1H NMR da XACb0033 ($45 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1}) com 50mmol.L^{-1} de NaCl, (pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	174
Anexo 11: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 ($45 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1}) com 50mmol.L^{-1} de NaCl, (pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	175
Anexo 12: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 ($45 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1}) com 50mmol.L^{-1} de NaCl, (pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz. Após adição de DTT.....	176
Anexo 13: Espectro de ^1H NMR da XACb0033 ($50 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (10mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	177
Anexo 14: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 ($50 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (10mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	178

Anexo 15: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 (50 $\mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (10 mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz. Após adição de glicerol.....	179
Anexo 16: Espectro de ^1H NMR da XACb0033 (50 $\mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (30 mmol.L^{-1} , pH 5,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	180
Anexo 17: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 (50 $\mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (30 mmol.L^{-1} , pH 5,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	181
Anexo 18: Espectro de ^1H NMR da XACb0033 (11 $\mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	182
Anexo 19: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 (11 $\mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	183
Anexo 20: Espectro de ^1H NMR da XACb0033 (11 $\mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0 a 40°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	184
Anexo 21: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 (11 $\mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0 a 40°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	185
Anexo 22: Espectro de ^1H NMR do complexo XACb0033 + XACb0032+DNA da XACb0033 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	186
Anexo 23: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo XACb0033 + XACb0032+DNA da XACb0033 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	187
Anexo 24: Espectro de ^1H NMR do complexo XACb0033 + XACb0032 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	188
Anexo 25: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo XACb0033 + XACb0032 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.....	189

Capítulo I - Introdução

1. Introdução

1.1 Interesse

O cancro cítrico foi introduzido no Brasil em 1957 e atinge todas as variedades de citrus desde então¹. Existem três variedades de cancro cítrico descritos na literatura: tipo A, B e C. O tipo A é oriundo da Ásia, provavelmente do sul da China, Indonésia e Índia, sendo o mais severo e o que gera maiores perdas econômicas; do tipo B encontrado na Argentina, Paraguai e Uruguai e do tipo C encontrado no Brasil². Esta doença é causada pela bactéria *Xanthomonas axonopodis* pv. *citri*, e provoca lesões nas folhas, frutos e ramos (Figura I-1). Os primeiros sintomas da contaminação surgem nas folhas como manchas amareladas salientes que aos poucos se tornam marrons no centro; nos ramos formam crostas pardas e nos frutos as lesões são salientes e, em estágios mais avançados, pode provocar o rompimento da casca e a sua queda.³

A bactéria do cancro é facilmente disseminada pela ação do homem, principalmente pelo uso de materiais de colheita contaminados, mas também, pela ação de chuvas e ventos fortes. Folhas, ramos e frutos jovens são preferidos pela bactéria, pois ela consegue penetrar através de aberturas naturais dos tecidos e, em plantas mais adultas, a disseminação é resultado de lesões causadas pelo material de colheita e insetos.³

¹ Belasque Jr., J.; Fernandes, N.; Massari, C. O Sucesso da Campanha de Erradicação do Cancro Cítrico no Estado de São Paulo, Brasil. *Summa Phytopathol* **2009**, *35*, 91-92.

² Moreira, L.; Almeida, N.; Potnis, N.; Digiampietri, L.; Adi, S.; Bortolossi, J.; da Silva, A.; da Silva, A.; Moraes, F.; Oliveira, J.; Souza, R.; Facincani, A.; Ferraz, A.; Ferro, M.; Furlan, L.; Gimenez, D.; Jones, J.; Kitajima, E.; Laia, M.; Leite Jr, R.; Nishiyama, M.; Neto, J.; Nociti, L.; Norman, D.; Ostroki, E.; Pereira Jr, H.; Staskawicz, B.; Tezza, R.; Ferro, J.; Vinatzer, B.; Setubal, J. Novel insights into genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifolli*. *BMC Genomics* **2010**, *11*, 238-262.

³ (a) http://www.agrobyte.com.br/cancro_citrico.htm (acessado em dezembro de 2011). (b) Lopes, T. Caracterização Estrutural do Complexo de Proteínas Hipotéticas – XACb0032/XACb0033 da Bactéria *Xanthomonas axonopodis* pv. *citri*. Dissertação de Mestrado, Instituto de Química, Universidade Quíivers Universidade Estadual de Campinas, SP, Novembro 2007.



Figura I-1: Lesões causadas pelo cancro cítrico em frutos, folhas e ramos.⁴

A *Xac* é uma bactéria altamente resistente, pois consegue sobreviver em diversos ambientes por vários meses ou até por anos e a única forma de eliminá-la é a erradicação das plantas contaminadas e a eliminação de todas as plantas dos talhões infectados. A área onde o foco da doença foi encontrado fica temporariamente interdita e não é permitido o replantio de citros por um período de dois anos.^{5,3}

De acordo com a Organização *FAO* (*Food and Agricultural Organization of United Nations*) das Nações Unidas, em 2007 um dos maiores produtores mundiais de citros foi o Brasil.⁶ Também, de acordo com a *UNCTAD* (*United Nations Conference on Trade and Development*) em 2004 havia 140 países produtores de citrus onde o Brasil foi o maior produtor mundial.^{6b}

Segundo um levantamento realizado pelo Fundo de Defesa da Citricultura (Fundecitrus) e divulgado em 2011, a incidência de cancro cítrico no Estado de São Paulo aumentou de 0,14% para 0,44% entre 2009 e 2010, apresentando um crescimento de 214,3%, decorrente da atenuação da legislação de controle da

⁴ (a) <http://www.madmoura.com.br/noticia/campanha-em-sp-orienta-citricultores-sobre-cancro-no-pomar.html> (acessado em Dezembro de 2011). (b) <http://www.ctexto.com.br/saladeimprensa-not.php?id=43> (acessado em Dezembro de 2011).

⁵ Brunings, A.; Gabriel, D. *Xanthomonas citri*: breaking the surface. *Molecular Plant Pathol* **2003**, *4*, 141-157.

⁶ (a) <http://www.xn--gew.net/articles/123535/Production> (acessado em Dezembro de 2011).

(b) http://www.ask.com/wiki/Citrus_production?qsrc=3044 (acessado em Dezembro de 2011).

doença pelo governo paulista.⁷ Em 1999, após o primeiro levantamento apontar incidência de 0,70% de cancro nos pomares, a legislação contra a doença passou a ser mais rígida tornando obrigatória a erradicação de todas as árvores com cancro em um raio de 30 metros de distância de uma contaminada e ainda de todo um talhão. Porém, por pressão dos citricultores, em 2009 a Secretaria da Agricultura de São Paulo revogou a determinação referente aos talhões e, seis meses após a mudança da lei, o número de casos novos de cancro cítrico cresceu quase 80% se comparado ao primeiro semestre de 2009. Por isso, agora o produtor precisa intensificar as medidas de controle do cancro, especialmente no período de chuvas, já que o clima quente e úmido favorece a proliferação da bactéria *Xanthomonas citri*, causadora da doença⁷

Desta forma, estudar o cancro cítrico e o seu agente causador é de extrema importância para que haja um melhor entendimento da doença e dos mecanismos de virulência do fitopatógeno, contribuindo para a melhoria da citricultura brasileira.

1.1.1. *Xanthomonas axonopodis* pathovar *citri*

A *Xanthomonas axonopodis* pv. *citri* (*Xac*) é uma bactéria Gram-negativa patogênica responsável pelo cancro cítrico em todas as variedades e espécies de citros e é uma das mais graves doenças da citricultura brasileira.⁵

A *Xac* se apresenta em formato de bastonetes e possui um flagelo polar responsável por sua mobilidade. Recentemente foi descoberto que este flagelo é requerido para a formação do biofilme maduro e conseqüentemente está envolvido no desenvolvimento do cancro cítrico.⁸

⁷ <http://www.agromundo.com.br/?p=19495> (acessado em Dezembro de 2011).

⁸ a) Malamud, F.; Torres, P.; Roeschlin, R.; Rigano, L.; Enrique, R.; Bonomi, H.; Castagnaro, A.; Marano, M.; Vojnov, A. The *Xanthomonas axonopodis* pv. *citri* flagellum is required for mature biofilm and canker development. *Microbiol* **2011**, *157*, 819-829. b) Martins, A. Análise de Chaperonas Hipotéticas da *Xanthomonas axonopodis* pv. *citri* por Espectrometria de Massas. Dissertação de Mestrado, Instituto de Química, Universidade Estadual de Campinas, SP, Outubro de 2010.

Por causa do seu grande impacto na citricultura brasileira, esta bactéria tornou-se alvo de estudos e teve o seu genoma sequenciado em 2002.⁹ A *Xac* possui um cromossomo circular de 5,1 Mb e dois plasmídeos pXAC64 (64 Kb) e pXAC33 (33 Kb).^{9,10}

Das proteínas anotadas em *Xac* muitas estão associadas à patogenicidade, virulência e adaptação (devido a sua similaridade com a proteínas com papéis patogênicos em outras bactérias) como as proteínas relacionadas ao sistema secretório do tipo III e IV, proteínas envolvidas em processos de adaptação e condições atípicas, proteínas de choque térmico¹⁰ e alguns dos genes codificam proteínas cujas características as classificam em possíveis chaperonas de secreção envolvidas no processo de virulência.¹¹

Durante o processo de infecção, as bactérias Gram-negativas patogênicas, como a *Xac*, secretam proteínas patógenas e outros fatores macromoleculares para o citoplasma da célula hospedeira.¹² Desse modo, a infecção começa com a atuação dos sistemas de secreção¹³, onde proteínas, ácidos nucleicos ou complexos de ácidos nucleicos e proteínas, são secretados para o citosol da célula hospedeira. Existem seis sistemas secretórios e, no caso da *Xac*, o seu genoma codifica os sistemas secretórios do tipo II, III e IV (T2SS, T3SS e T4SS).⁵

⁹ da Silva, A.; Ferro, J.; Reinach, F.; Farah, C.; Furlan, L.; Quaggio, R.; Monteiro-Vitorello, C.; Van Sluys, M.; Almeida, N.; Alves, L.; do Amaral, A.; Bertolini, M.; Camargo, L.; Camarotte, G.; Cannavan, F.; Cardozo, J.; Chambergo, F.; Ciapina, L.; Cicarelli, R.; Coutinho, L.; Cursino-Santos, J.; El-Dorry, H.; Faria, J.; Ferreira, A.; Ferreira, R.; Ferro, M.; Formighieri, E.; Franco, M.; Greggio, C.; Gruber, A.; Katsuyama, A.; Kishi, L.; Leite, R.; Lemos, E.; Lemos, M.; Locali, E.; Machado, M.; Madeira, A.; Martinez-Rossi, N.; Martins, E.; Meidanis, J.; Menck, F.; Miyaki, C.; Moon, D.; Moreira, L.; Novo, M.; Okura, V.; Oliveira, M.; Oliveira, V.; Pereira, H.; Rossi, A.; Sena, J.; Silva, C.; de Souza, R.; Spinola, L.; Takita, M.; Tamura, R.; Teixeira, E.; Tezza, R.; Trindade dos Santos, M.; Truffi, D.; Tsai, S.; White, F.; Setubal, J.; Kitajima, J. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **2002**, *417*, 459-463.

¹⁰ Docena, C. Identificação das interações das proteínas relacionadas aos Sistemas de Dois-Componentes e aos Sistemas Secretórios do Tipo III e IV do fitopatógeno *Xanthomonas axonopodis* pv. *Citri*. Tese de Doutorado, Instituto de Química, Universidade de São Paulo, SP, Outubro 2006.

¹¹ Khater, L.; Santos, T.; Alegria, M.; Docena, C.; Silva, A.; Ramos, C. *In silico* identification of potential chaperone genes that belong to type III and type IV secretion systems in *Xanthomonas axonopodis* pv. *citri*. *Genet Mol Biol* **2005** *28*, 321-327.

¹² Smith, C.; Hultgren, S. Bacteria thread the needle. *Nature* **2001**, *414*, 29-31.

¹³ (a) Papanikou, E.; Karamanou, S.; Economou, A. Bacterial protein secretion through the translocase nanomachine. *Nature Rev Microbiol* **2007**, *5*, 839-851. (b) Holland, I. Translocation of bacterial proteins—an overview. *Biochim Biophys Acta* **2004**, *1694*, 5-16.

No caso dos sistemas de secreção do tipo III e IV esses fatores de virulência são injetados diretamente no citosol da célula hospedeira, onde irão interferir nos processos celulares.¹⁴

1.1.2. Sistema de Secreção

Os sistemas de secreção são complexos multi-protéicos, e podem ser classificados em seis tipos diferentes, como ilustrado na Figura I-2, dependendo da forma como é realizada a translocação dos fatores de virulência para a célula hospedeira.^{14,15} Os sistemas de secreção III, IV e I utilizam um caminho de secreção *sec*-independente e, por isso, são capazes de exportar o substrato diretamente do citoplasma para fora da membrana externa da bactéria.^{14d,14b}

O sistema do tipo IV pode funcionar como *sec*-dependente e também como *sec*-independente¹⁶ e ele apresenta algumas similaridades com o sistema do tipo III, embora aparentemente eles não sejam relacionados. Estes dois sistemas necessitam de contato físico com a célula hospedeira, requerem que uma chaperona secretória (CS) se ligue à proteína a ser secretada e acredita-se que a exportação desta ocorra em somente uma etapa através do canal de secreção. Entretanto existe uma diferença fundamental entre eles: o sistema tipo IV pode

¹⁴ (a) Fattori, J.; Prando, A.; Martins, A.; Rodrigues, F.; Tasic, L. Bacterial Secretion Chaperones. *Protein Peptide Lett* **2011**, *18*, 158-166. (b) Gelvin, S. Finding a way to the nucleus. *Curr Opin Microbiol* **2010**, *13*, 53-58. (c) Cornelis, G.; Van Gijsegem, F. Assembly and function of Type III secretory systems. *Annu Rev Microbiol* **2000**, *54*, 735-774. (d) Thanassi, D.; Hultgren, S. Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr Opin Cell Biol* **2000**, *12*, 420-430. (e) Cheng, L.; Schneewind, O. Type III machines of gram-negative bacteria: delivering the goods. *Trends Microbiol* **2000**, *8*, 214-220. (f) Galán, J.; Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **1999**, *284*, 1322-1328. (g) Mota, L.; Sorg, I.; Cornelis, G. Type III secretion: The bacteria-eukaryotic cell express. *FEMS Microbiol Lett* **2005**, *252*, 1-10. (h) Aizawa, S. Bacterial flagella and type III secretion systems. *FEMS Microbiol Lett* **2001**, *202*, 157-164. (i) Saier Jr, M. Evolution of bacterial type III protein secretion systems. *TRENDS Microbiol* **2004**, *12*, 113-115.

¹⁵ (a) Economou, A. Following the leader: bacterial protein export through the *Sec* pathway. *Trends Microbiol* **1999**, *7*, 315-320. (b) Buttner, D.; Bonas U. Port of entry the type III secretion translocon. *Trends Microbiol* **2002**, *10*, 186-192. (c) Henderson, I. Navarro-Garcia F.; Desvaux, M.; Fernandez, R.; Aldeen, D. Type V proteins secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* **2004**, *68*, 692-744.

¹⁶ Burns, D. Biochemistry of type IV secretion. *Curr Opin Microbiol* **1999**, *2*, 25-29.

exportar DNA simples fita e complexo DNA-proteína até a célula e não há nenhuma evidência de que o sistema do tipo III possa fazer o mesmo.¹⁷

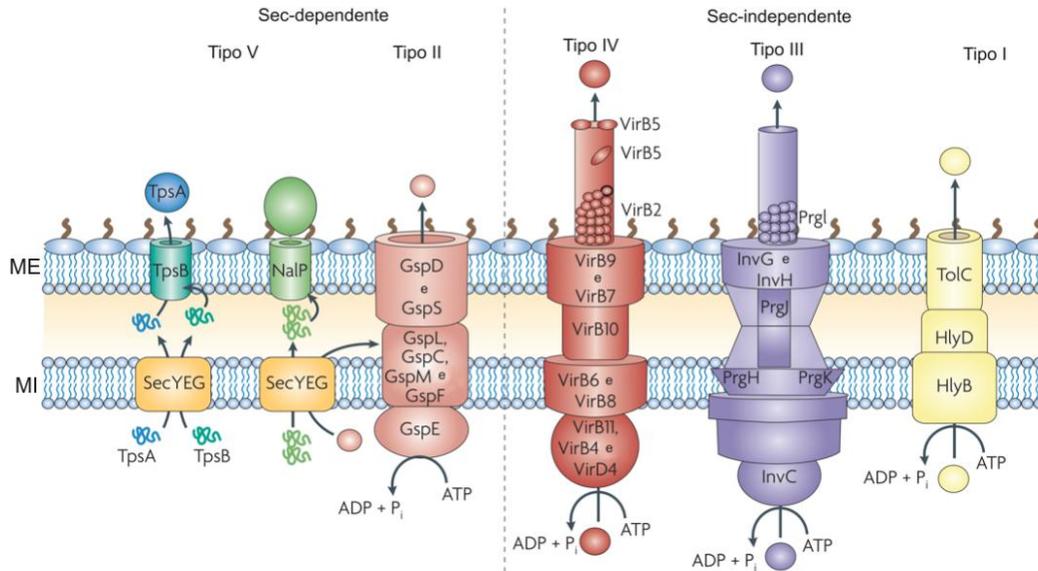


Figura I-2: Sistemas de secreção das bactérias Gram-negativas (I-VI).¹⁸ ME = membrana externa e MI = membrana interna.

1.1.2.1. Sistema de secreção do tipo III (T3SS)

O sistema de secreção do tipo III (T3SS – *Type Three Secretion System*) está relacionado com os componentes do sistema flagelar bacteriano¹⁹, sendo que um terço das proteínas que compõem a estrutura do T3SS são homólogas às proteínas constituintes da porção basal do flagelo.^{20,14h} Desta forma é possível que ele seja uma adaptação evolucionária do aparato flagelar, uma vez que na escala evolutiva a mobilidade tenha sido desenvolvida antes da virulência²¹. O aparato

¹⁷ Christie, P.; Vogel, J.; Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* **2000**, 8, 354-360.

¹⁸ Adaptado de Fronzes, R; Christie, P.; Waksman, G. The structural biology of type IV secretion systems. *Nat Rev Microbiol* **2009**, 7, 703-714.

¹⁹ Plano, G.; Day, J.; Ferracci, F. Type III export: new uses for an old pathway. *Mol Microbiol* **2001**, 40, 284-293.

²⁰ Hueck, C.; Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **1998**, 62, 379-433.

²¹ Bennett, J.; Hughes, C.; From flagellum assembly to virulence: the extended family of type III export chaperones. *Trends Microbiol* **2000**, 8, 202-204.

flagelar confere a bactéria mobilidade, o que é importante para o processo de infecção, entretanto a formação do sistema de secreção permite que proteínas possam ser secretadas.^{14f}

O formato cilíndrico do T3SS é conhecido por ser parecido com uma agulha de injeção. Ele é composto de aproximadamente 20 proteínas e é considerado o mais complexo dentre os sistemas de secreção, permitindo rápida e eficiente translocação das proteínas através do caminho: membrana interna, periplasma, membrana externa e finalmente membrana da célula eucariótica.²² A formação e ativação do complexo ocorrem quando a bactéria entra em contato com o hospedeiro^{14f} e então todas estas barreiras são atravessadas de uma só vez pelo estreito canal de aproximadamente 30 Å, do citoplasma bacteriano para o citosol da célula eucariótica e já dentro da célula eucariótica as proteínas chamadas de efetoras irão interferir nos processos celulares.^{14f,21}

Para que todo esse processo possa ocorrer, as proteínas que serão translocadas devem ser reconhecidas e engajadas na maquinaria de secreção. O estreito canal do sistema agulha mostra que para que o substrato possa ser secretado necessita estar em uma forma desenovelado e precisa atingir o sistema secretório correspondente, logo reconhecido e secretado. Para que este processo complexo ocorra, o mecanismo de reconhecimento do substrato deve envolver vários meios de sinalização e, também, o auxílio de proteínas acessórias.^{20,23}

Neste sistema, a maioria das proteínas secretadas possuem uma sequência sinal localizada dentro dos primeiros 20-30 aminoácidos^{24,25}, entretanto, diferentemente dos sistemas *sec*-dependentes, esta sequência sinal não é clivada durante o processo de secreção.¹⁰ No entanto, esta sequência sinal não assegura

²² Feldman, M.; Cornelis, G. The multitasking type III chaperones: all you can do with 15 kDa. *FEMS Microbiol Lett* **2003**, 219,151-158.

²³ Galán, J.; Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **2006**, 444, 567-573.

²⁴ Sory, M.; Boland, A.; Lambermont, I.; Cornelis, G. Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the *cyaA* gene fusion approach. *Proc Natl Acad Sci USA* **1995**, 92, 11998-12002.

²⁵ Schesser, K.; Frithz-Lindesten, E.; Wolf-Watz, H. Delineation and mutational analysis of the *Yersinia pseudotuberculosis* YopE domains which mediate translocation across bacterial and eukaryotic cellular membranes. *J Bacteriol* **1996**, 178, 7227-7233.

a especificidade do substrato, sendo sugerido que as chaperonas de secreção (CS) as quais se ligam especificamente ao substrato desempenhem este papel.^{26,21}

1.1.2.2. Sistema de secreção do tipo flagelar

O sistema de secreção das chaperonas flagelares²⁷ é similar e precedente ao sistema de secreção do tipo III e o corpo basal flagelar compartilha de características comuns com o aparato de secreção de fatores de virulência em forma de agulha, indicando que os dois sistemas são derivados de um ancestral comum.²⁸ No sistema flagelar, as proteínas flagelares são exportadas pelo T3SS para a formação da maquinaria flagelar. Estas polimerizam para formar a estrutura, incluindo a junção base do filamento, filamento e filamento final. As chaperonas flagelares auxiliam na transferência dos componentes estruturais flagelares através da ligação no domínio C-terminal anfipático de seus substratos.^{29,30,31}

O flagelo bacteriano é responsável pela mobilidade e está envolvido na adesão, formação do biofilme bacteriano e invasão das células hospedeiras.³² A bactéria pode se movimentar em líquidos através dos filamentos flagelares que são fixados e movidos pelo motor flagelar. Esses motores são nanomáquinas que

²⁶ Wattiau, P.; Cornelis, G. SycE, a chaperone-like protein of *Yersinia enterocolitica* involved in Ohe secretion of YopE. *Mol Microbiol* **1993**, *8*, 123-131.

²⁷ (a) Minamino, T.; Imada, K.; Namba, K. Mechanisms of type III protein export for bacterial flagellar assembly. *Mol BioSyst* **2008**, *4*, 1105-1115. (b) Buttner, D.; Bonas, U. Port of entry the type III secretion translocon. *Trends Microbiol* **2002**, *10*, 186-192.

²⁸ Iida, Y.; Hogley, L.; Lambert, C.; Fenton, A.; Sockett, R.; Aizawa, S. Roles of Multiple Flagellins in Flagellar Formation and Flagellar Growth Post Bdelloplast Lysis in *Bdellovibrio bacteriovorus*. *J Mol Biol* **2009**, *394*, 1011-1021.

²⁹ Cornelis, G. The type III secretion injectisome, a complex nanomachine for intracellular toxin delivery. *Biol Chem* **2010**, *391*, 745-751.

³⁰ Paul, K.; Erhardt, M.; Hirano, T.; Blair, D.; Hughes, K. Energy source of flagellar type III secretion. *Nature* **2008**, *451*, 489-492.

³¹ Khater, L.; Alegria, M.; Borin, P.; Santos, T.; Docena, C.; Tasic, L.; Farah, C.; Ramos, C. Identification of the flagellar chaperone FlgN in the phytopathogen *Xanthomonas axonopodis* pv. *citri* by its interaction with hook-associated FlgK. *Arch Microbiol* **2007**, *188*, 243-250.

³² Liu, R.; Ochman, H. Stepwise formation of the bacterial flagellar system. *PNAS*, **2007**, *104*, 7116-7121.

usam o potencial eletroquímico de gradiente de íons e assim, o motor pode girar em ambas direções, regulando a direção em resposta a estímulos quimioestáticos.³³ Um flagelo típico é constituído de cinco componentes: um corpo basal, um motor, um juntor, um filamento e um aparato de exportação (Figura I-3).²⁷ O aparato de exportação consiste de seis proteínas integrais da membrana (FlhA, FlhB, FliO, FliP, FliQ e FliR) e três proteínas solúveis (FliH, FliI e FliJ). São essas proteínas responsáveis pela formação do canal de exportação dirigido pela transferência de prótons. FliI é uma ATPase e forma um complexo heterotrimérico com um dímero do regulador dela, a FliH. O complexo FliH₂FliI se liga ao complexo chaperona-substrato e facilita a entrada inicial do complexo chaperona-substrato no canal de secreção. Já o desenovelamento e a translocação dos substratos através do canal é dirigida pelo gradiente de prótons (do inglês *Próton Motive Force* - PMF)²¹ cuja energia é gerada pela transferência de prótons ou elétrons através da membrana citoplasmática, enquanto a energia química obtida da hidrólise do ATP (FliI), é usada para a liberação do FliH e FliI do canal translocando o substrato.³⁴ Existem muitas outras proteínas que formam o sistema de secreção flagelar como as FlgK e FlgL formadoras da junção base dos filamentos, FliC formadora do filamento, FlgE formadora da junção, dentre outras.

³³ Kudryashev, M.; Cyrklaff, M.; Wallich, R.; Baumeister, W.; Frischknecht, F. Distinct *in situ* structures of the *Borrelia* flagellar motor. *J Struct Biol* **2010**, *169*, 54-61

³⁴ (a) Kazetani, K.; Minamino, T.; Miyata, T.; Kato, T.; Namba, K. ATP-induced FliI hexamerization facilitates bacterial flagellar protein export. *Biochem Biophys Res Comm* **2009**, *388*, 323-327. (b) Okabe, M.; Minamino, T.; Imada, K.; Namba, K.; Kihara, M. Role of the N-terminal domain of the FliI ATPase in bacterial flagellar protein export. *FEBS Lett* **2009**, *583*, 743-748. (c) Minamino, T.; Kazetani, K.; Tahara, A.; Suzuki, H.; Furukawa, Y.; Kihara, M.; Namba, K. Oligomerization of the bacterial flagellar ATPase FliI is controlled by its extreme N-terminal region. *J Mol Biol* **2006**, *360*, 510-519.

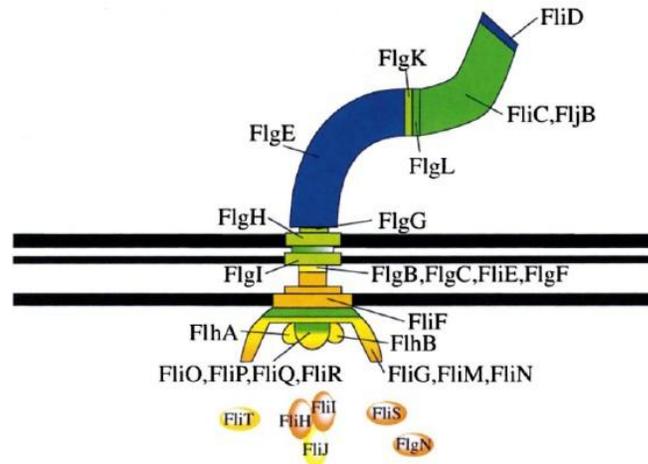


Figura I-3: Sistema de secreção flagelar mostrando as seis proteínas integrais da membrana (FlhA, FlhB, FliO, FliP, FliQ e FliR) e três proteínas solúveis (FliH, FliI e FliJ).³⁵

1.1.2.3. Sistema de secreção do tipo IV (T4SS)

O sistema de secreção do tipo IV (T4SS - *Type Four Secretion System*) é um sistema versátil encontrado em bactérias Gram-negativa e Gram-positiva e pode secretar uma ampla faixa de substratos desde proteínas, complexos proteína-proteína e complexos proteína-DNA.¹⁷ Aparentemente não está relacionado com o sistema do tipo III e sim relacionado evolutivamente com os componentes da maquinaria de conjugação bacteriana. A conjugação é o processo no qual o DNA é transferido de uma bactéria doadora para uma receptora por um mecanismo que envolve o contato celular e essa transferência ocorre através do envelope celular das bactérias Gram-negativas que é mediada por uma estrutura supramolecular denominada de formação do par conjugado (*Mating Pair Formation – Mpf*).³⁶ Este complexo é formado por um canal responsável por estabelecer o contato entre as células e há a formação de um

³⁵ Aizawa, S. Bacterial flagella and type III secretion systems. *FEMS Microbiol Lett* **2001**, *202*, 157-164.

³⁶ Willetts, N.; Wilkins, B. Processing of plasmid DNA during bacterial conjugation. *Microbiol Rev* **1984**, *48*, 24-41.

canal de conjugação por onde ocorre a translocação do DNA a ser transferido³⁷. Os aspectos clínicos deste tipo de transferência são extremamente problemáticos, pois o DNA plasmidial de uma bactéria é incorporado por outra, levando a rápida disseminação de genes resistentes a antibióticos e outros fatores de virulência entre as populações bacterianas.^{16,38}

1.1.3. Chaperonas de Secreção

As chaperonas de sistemas de secreção são pequenas proteínas, de até 20 kDa e pI baixo, podem interagir especificamente com uma ou mais proteínas de virulência e uma ou duas proteínas translocadoras e possuem hélice anfipática. Elas normalmente atuam como dímeros, se ligam à região N-terminal da proteína parceira e não possuem domínios de ligação de ATP, o que as diferencia das chaperonas moleculares mais conhecidas.^{33,26} Em geral, as CS são codificadas por genes localizados próximos aos genes que codificam as proteínas efetoras ou translocadoras as quais elas irão se associar, o que ajuda na sua identificação.^{22,14c,30} Estas chaperonas têm a função de proporcionar um estado transicional à proteína parceira, são responsáveis pelo encaminhamento do complexo chaperona/proteína parceira para a maquinaria de secreção³⁹ e previnem interações ou agregação prematuras da sua proteína parceira antes da secreção correspondente.^{14a,40} A Figura I-4 ilustra a ação das chaperonas secretórias no transporte da proteína até o sistema de secreção correspondente.

³⁷ Cascales, E.; Christie, P. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol* **2003**, *1*, 137-149.

³⁸ Christie, P. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol Microbiol* **2001**, *40*, 294-305.

³⁹ Lee, S.; Galán, J. Salmonella type III secretion-associated chaperones confer secretion-pathway specificity. *Mol Microbiol* **2004**, *51*, 483-495.

⁴⁰ Woestyn, S.; Sory, M.; Boland, A.; Leguenne, O.; Cornelis, G. The cytosolic SycE and SycH chaperones of *Yersinia* protect the region of YopE and YopH involved in translocation across eukaryotic cell membranes. *Mol Microbiol* **1996**, *20*, 1261-1271.

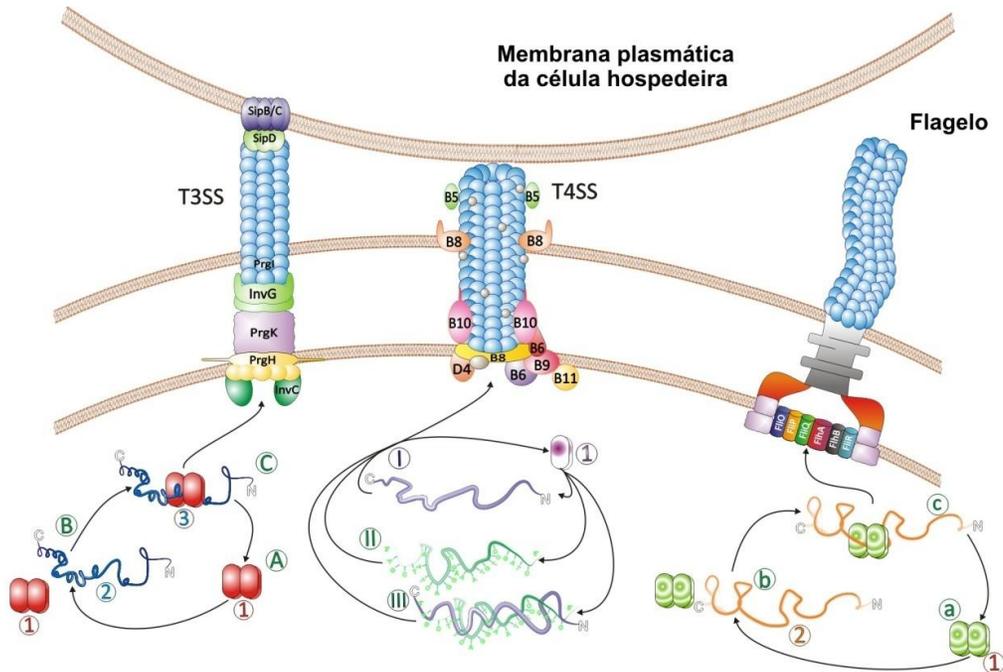


Figura I-4: Ilustração dos sistemas de secreção bacteriano: T3SS (esquerda), T4SS (centro) e flagelar (direita). No sistema de secreção do tipo III a chaperona na forma de um dímero (1) encontra com a proteína parcialmente enovelada (2) forma um complexo proteína substrato/chaperona (3); então a chaperona mantém a proteína substrato parcialmente enovelada no seu domínio de ligação e esta é encaminhada para o T3SS para secreção. No sistema de secreção do tipo IV, as chaperonas agem do mesmo modo atuando como monômeros ou dímeros e exportando além de proteínas (I), fita simples de DNA (II) ou complexo proteína-DNA (III). No sistema flagelar as chaperonas agem como dímero (1) e ligam-se a proteína flagelar prevenindo agregação e precipitação além de participar na formação da maquinaria do flagelo. (Figura adaptada).^{14a}

Estas chaperonas não partilham similaridades significantes entre suas estruturas primárias; entretanto a estrutura cristalográfica de algumas delas mostra que estas apresentam estrutura terciária muito similar, dependendo da sua subclasse.⁴¹ As chaperonas de secreção geralmente se ligam aos aminoácidos 50-100 da sua proteína parceira e, através das estruturas tridimensionais foi

⁴¹ (a) Stebbins, C.; Galán, J. Priming virulence factors for delivery into the host. *Nat Rev Mol Cell Biol* **2003**, 4, 738-743. (b) Page, A.; Parsot, C. Chaperones of the type III secretion pathway: jacks of all trades. *Mol Microbiol* **2002**, 46, 1-11.

observado que estas chaperonas mantêm o seu domínio de ligação na proteína parceira em um estado desenovelado.^{40,42}

O sistema flagelar é considerado o ancestral do T3SS,⁴³ e as chaperonas de secreção desta classe apresentam poucos dados estruturais. No entanto, cálculos de estrutura secundária sugerem que elas apresentam uma estrutura helicoidal estendida e, além disso, podem atuar como dímeros. Um exemplo de estrutura cristalográfica de uma chaperona flagelar pode ser representada pela FliT de *Salmonella* sp. (Figura I-5b), em que é possível verificar que esta chaperona consiste de quatro α -hélices e a região central da molécula é formada por hélices antiparalelas.

Já as chaperonas do sistema de secreção do tipo IV podem ser exemplificadas pela chaperona VirE1. No caso da bactéria *Agrobacterium tumefaciens*, a chaperone (VirE1) esta envolvida na transferência de proteínas independentemente de DNA, mediando a transferência da proteína VirE2. A VirE1 interage diretamente com o domínio interno da VirE2 estabilizando-a e prevenindo agregação da VirE2.⁴⁴ A estrutura cristalográfica da VirE1 consiste de uma única α -hélice entre os dois domínios estruturais da VirE2 e dados de microscopia eletrônica, SAXS e cristalografia de raios-X mostraram a formação de um heterodímero na razão 1:1 (Figura I-5a).

⁴² Stebbins, C.; Galán, J. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **2001**, *414*, 77-81.

⁴³ Macnab, R. The bacterial flagellum: reversible rotary propellor and type III export apparatus. *J Bacteriol* **1999**, *181*, 7149-7153.

⁴⁴ (a) Zhao, Z.; Sagulenko, E.; Ding, Z.; Christie, P. Activities of virE1 and the VirE1 secretion chaperone in export of the multifunctional VirE2 effector via an *Agrobacterium* type IV secretion pathway. *J Bacteriol* **2001**, *183*, 3855-3865. (b) Sundberg, C.; Ream, W. The *Agrobacterium tumefaciens* Chaperone-Like Protein, VirE1, Interacts with VirE2 at Domains Required for Single-Stranded DNA Binding and Cooperative Interaction. *J Bacteriol* **1999**, *181*, 6850-6855. (c) Duckely, M.; Oomen, C.; Axthelm, F.; Van Gelder, P.; Waksman, G.; Engel, A. The VirE1VirE2 complex of *Agrobacterium tumefaciens* interacts with single-stranded DNA and forms channels. *Mol Microbiol* **2005**, *58*, 1130-1142.

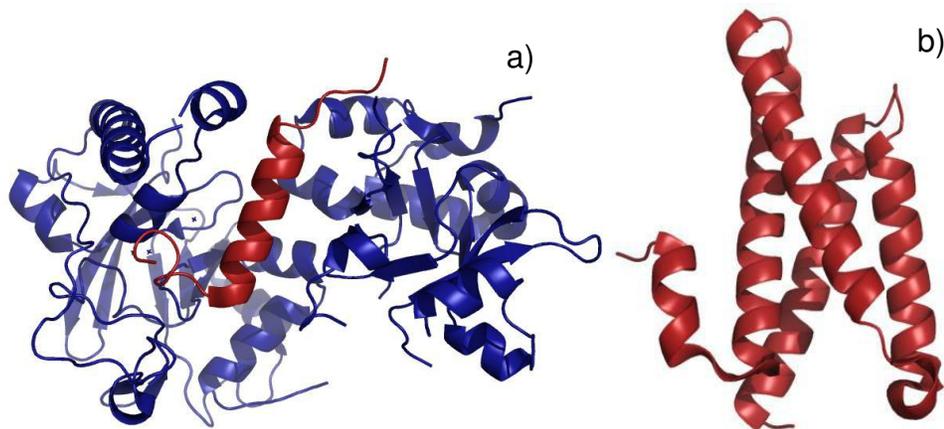


Figura I-5: a) Estrutura da chaperona de secreção do T4SS, VirE1 (em vermelho) complexada com sua proteína parceira, a VirE2 (em azul; PDB:3BTP).⁴⁵ b) Estrutura da chaperona de secreção flagelar, FliT de *Salmonella sp.* (PDB: 3A7M).⁴⁶

Durante a etapa de liberação e desenovelamento parcial do substrato a partir do complexo chaperona-substrato, a atividade ATPase pode estar associada fornecendo energia suficiente para que ocorra este processo. Outra fonte de energia associada ao T3SS e o sistema de secreção flagelar é obtida pelo gradiente de prótons (PMF) fenômeno pode ser provocado pela hidrólise do ATP. Por isso, tornou-se necessário também estudar as interações entre as chaperonas com o ligante ATP.

⁴⁵ Dym, O.; Albeck, S.; Unger, T.; Jacobovitch, J.; Branzburg, A.; Michael, Y.; Frenkiel-Krispin, D.; Wolf, S.; Elbaum, M. Crystal structure of the *Agrobacterium* virulence complex VirE1-VirE2 reveals a flexible protein that can accommodate different partners. *Proc Natl Acad Sci USA* **2008**, *105*, 11170-11175.

⁴⁶ Imada, K.; Minamino, T.; Kinoshita, M.; Furukawa, Y.; Namba, K. Structural insight into the regulatory mechanisms of interactions of the flagellar type III chaperone FliT with its binding partners. *Proc Natl Acad Sci USA* **2010**, *107*, 8812-8817.

1.1.4. Hsp90 da laranja

Acredita-se que durante o processo de infecção da *Xac*, as proteínas de virulência da *Xac* são secretadas para o citosol da célula hospedeira pela ação dos sistemas de secreção da bactéria. Então, após as proteínas efetoras serem secretadas para o citosol da célula hospedeira precisam enovelar-se novamente para adquirir a sua estrutura e conseqüentemente sua função.

Esse processo no caso da laranja (hospedeiro do cancro cítrico) e proteínas de virulência da *Xac* (hóspedes) é conduzido pelas chaperonas moleculares da planta e, entre elas, a Hsp90 pode ser a mais importante.⁴⁷

A Hsp90 compreende uma família de chaperonas moleculares de aproximadamente 90 kDa, altamente conservadas em procariotos e eucariotos e é uma das proteínas mais abundante em células.⁴⁸ Esta proteína pertence a família das *Heat shock proteins*, proteínas superexpressas sob condições de stress. Esta chaperona esta relacionada na manutenção da conformação, renovelamento de proteínas desnaturadas, estabilidade e função de uma enorme variedade de proteínas envolvidas na transdução de sinais, ciclo celular e apoptose.⁴⁹

Esta proteína funciona como um homodímero, ao qual se associam cochaperonas com a finalidade de catalisar a maturação e/ ou ativação de diversas proteínas, conhecidas como proteínas clientes, envolvidas em diferentes vias regulatórias da célula.⁵⁰ Dentre estas proteínas clientes, estão presentes proteínas cinases, receptores nucleares de hormônios, fatores de transcrição, entre outras.⁴⁸ Além disso, a Hsp90 esta envolvida na ativação da proteína p53,

⁴⁷ Mendonça, Y. Ramos, C. Cloning, purification and characterization of a 90 kDa heat shock protein from *Citrus sinensis* (sweet orange). *Plant Phys Biochem* **2012**, *50*, 87-94.

⁴⁸ Sun, X.; Kenney, S. Hsp90 inhibitors A potential treatment for latent EBV infection? *Cell Cycle* **2010**, *9*, 1665-1666.

⁴⁹ Chakraborty, A.; Koldobskiy, M.; Sixt, K.; Juluri, K.; Mustafa, A.; Snowman, A.; van Rossum, D.; Patterson, R.; Snyder, S. HSP90 regulates cell survival via inositol hexakisphosphate kinase-2. *PNAS* **2008**, *105*, 1134-1139.

⁵⁰ (a) Pearl, L.; Prodromou, C. Structure, function, and mechanism of the Hsp90 molecular chaperone. *Adv Protein Chem* **2001**, *59*, 157-186. (b) Brown, M.; Zhu, L.; Schmidt, C.; Tucker, P. Hsp90-From signal transduction to cell transformation. *Bioch Bioph Res Com* **2007**, *363*, 241-246.

uma proteína supressora de tumor⁵¹ e trabalhos recentes indicam que esta proteína contribui para a resistência de algumas drogas antifúngicas.⁵²

A Hsp90 se comporta essencialmente como um homodímero e esta dimerização é essencial para o desempenho de sua atividade biológica.⁵³ Cada monômero é dividido em três domínios estruturais: o domínio C-terminal que mantém a interface de dimerização molecular, o domínio central que é considerado a principal região de interação com suas proteínas clientes e o domínio N-terminal que possui um sítio de ligação a nucleotídeos responsável por uma fraca atividade da ATPase, o que provoca uma dimerização nesta região, ocasionando uma mudança de conformação fundamental para sua função.⁵⁰

Estas chaperonas usam ATP e sua hidrólise para promover desenovelamento e/ou enovelamento de proteínas. As interações entre proteínas estudadas com o ligante ATP γ -S também se faz necessária, pois o ATP é facilmente hidrolisado até o ADP pelas ATPases. Portanto, assim que a amostra de ATPase recebe a primeira quantidade de ATP, ela é convertida em ADP. Desse modo, ATP e ADP interagem com as proteínas alvos ao mesmo tempo, impossibilitando deste modo a observação de interações entre proteína e o ATP na ausência do ADP. O ATP γ -S apresenta ATP não hidrolisável, possibilitando monitorar as interações entre ATPases sem ocorrer a catálise (hidrólise).

Após a atuação dos sistemas de secreção e correto enovelamento dos fatores de virulência da *Xac* no citoplasma da célula hospedeira pela ação da Hsp90, acredita-se que as proteínas fitopatogênicas terão iniciado suas funções. Para tanto, inibindo a atividade da Hsp90 da laranja, o enovelamento dos fatores de fitopatogenicidade provenientes da *Xac* de natureza protéica serão bloqueados e, desta maneira, o cancro poderá ser prevenido, mesmo após a atuação dos

⁵¹ Hagn, F.; Lagleder, S.; Retzlaff, M.; Rohrberg, J.; Demmer, O.; Richter, K.; Buchner, J.; Kessler, H. Structural analysis of the interaction between hsp90 and the tumor suppressor protein p53. *Nature* **2011**, *18*, 1086-1093.

⁵² Cowen, L.; Singh, S.; Köhler, J.; Collins, C.; Zaas, A.; Schell, W.; Aziz, H.; Mylonakis, E.; Perfect, J.; Whitesell, L.; Lindquist, S. Harnessing Hsp90 function as a powerful, broadly effective therapeutic strategy for fungal infectious disease. *PNAS* **2009**, *106*, 2818-2823.

⁵³ Wayne, N.; Bolon, D. Dimerization of Hsp90 Is Required for *in Vivo* Function. *J Biol Chem* **2007**, *282*, 35386- 35395.

Sistemas de Secreção (SSs). Um inibidor conhecido de Hsp90 é a geldanamicina, que se liga competitivamente com o sítio de ligação do ATP/ADP desta ATPase.

O antibiótico, geldanamicina⁵⁴, é um produto natural isolado da *Streptomyces higroscopicus* com potente atividade antiproliferativa e efeito antitumoral.⁵⁵ Assim que ela se liga a proteína Hsp90, inibe a atividade chaperona desta proteína, conseqüentemente seus substratos (proteínas clientes) são degradados.^{56,57} A geldanamicina bloqueia o sítio ligante de ATP, resultando em complexos de Hsp90 e acúmulo de proteínas desenoveladas ou enoveladas incorretamente.⁵⁸

A estrutura cristalográfica do domínio ligante da Hsp90 revela que a geldanamicina se liga ao bolsão de 15 Å de profundidade formado pelos resíduos 9-232 da Hsp90 adotando uma estrutura compacta similar à da cadeia polipeptídica bloqueando a entrada do ATP.⁵⁶ A Hsp90 possui duas formas: uma ativa e uma inativa. A forma inativa é aberta e a ligação de ATP causa o seu fechamento tornando-a ativa (Figura I-6).⁵⁹

⁵⁴ (a) Fukumoto, R.; Kiang, J. Geldanamycin Analog 17-DMAG Limits Apoptosis in Human Peripheral Blood Cells by Inhibition of p53 Activation and its Interaction with Heat-Shock Protein 90 kDa after Exposure to Ionizing Radiation. *Radiat Res* **2011**, *176*, 333-345. (b) Mutsunguma, L.; Moethoa, B.; Edkins, A.; Luke, G.; Blatch, G.; Knox, C. Theiler's murine encephalomyelitis virus infection induces a redistribution of heat shock proteins 70 and 90 in BHK-21 cells, and is inhibited by novobiocin and geldanamycin. *Cell Stress Chaperon* **2011**, *16*, 505-515. (c) Dey, A.; Cederbaum, A. Geldanamycin, an inhibitor of Hsp90, potentiates cytochrome P4502E1-mediated toxicity in HepG2 cells. *J Pharmacol Exp Ther* **2006**, *317*, 1391-1399.

⁵⁵ (a) Schneider, T.; Somogyi, J.; Csermely, P.; Szamel, M. The Hsp90-specific inhibitor geldanamycin selectivity disrupts kinase-mediated signaling events of T-lymphocyte activation. *Cell Stress Chaperon* **2000**, *1*, 52-61. (b) Majeesh, N.; Post, D.; Willard, M.; Kaur, B.; Meir, E.; Simons, J. Geldanamycin induces degradation of hypoxia-inducible factor 1 α protein via the proteasome pathway in prostate cancer cells. *Cancer Res* **2002**, *62*, 2478-2482.

⁵⁶ Qin, H.; Panek, J. Total synthesis of the Hsp90 inhibitor geldanamycin. *Org Lett* **2008**, *10*, 2477-2479.

⁵⁷ Stebbins, C.; Russo, A.; Schneider, C.; Rosen, N.; Hartl, F.; Pavletich, N. Crystal Structure of an Hsp90-Geldanamycin Complex: Targeting of a Protein Chaperone by an Antitumor Agent. *Cell* **1997**, *89*, 239-250.

⁵⁸ Prodromou, C.; Roe, S.; O'Brien, R.; Ladbury, J.; Piper, P.; Pearl, L. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **1997**, *90*, 65-75.

⁵⁹ (a) Ali, M.; Roe, S.; Vaughan, C.; Meyer, P.; Panaretou, B.; Piper, P.; Prodromou, C.; Pearl, L. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **2006**, *440*, 1013-1017. (b) Shiao, A.; Harris, S.; Southworth, D.; Agard, D. Structural Analysis of *E.coli* Hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **2006**, *127*, 329-340.

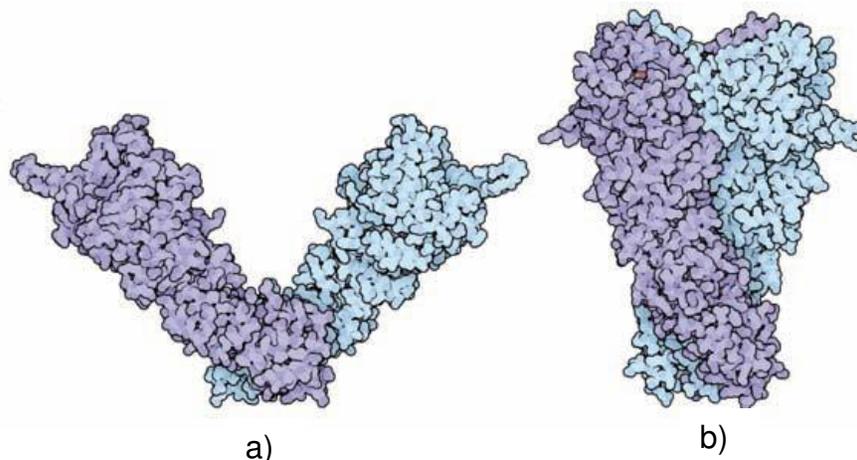


Figura I-6: Formas inativa (a) e nativa (b) da proteína Hsp90.⁵⁹

Por isso, o estudo das interações da Hsp90 da laranja com os ligantes geldanamicina, ATP, ADP e ATP γ S torna-se necessária e as técnicas propostas para este estudos foram Saturation Transfer Difference (STD-NMR) e fluorescência de emissão.

1.1.5. Proteínas FlgN e XACb0033

As proteínas FlgN (XAC1990) e XACb0033 já foram estudadas anteriormente e, neste capítulo, estão apresentados, de forma resumida, os principais resultados obtidos anteriormente.

Ambas proteínas FlgN e XACb0033 são consideradas chaperonas moléculares por possuírem características como baixo peso molecular, pI baixo e domínio anfipático. Além disso, estas proteínas são codificadas por genes localizados em cluster do sistema de secreção do tipo IV (XACb0033) e no cluster do sistema de secreção do tipo III (FlgN) e por isso, acredita-se que a XACb0033 seja uma chaperona de secreção do sistema de secreção do tipo IV (T4SS) a FlgN seja uma chaperona de secreção do sistema de secreção do tipo III (T3SS). Em 2007, através de ensaios de duplo-híbrido, a proteína FlgN foi identificada como uma chaperona flagelar, cuja proteína parceira é a FlgK.

Buscas por homologia e similaridade de sequência primária no BLAST⁶⁰ para ambas as proteínas não resultaram em alinhamentos significativos. Do mesmo modo, as buscas usando o PDB como banco de dados não geraram resultados significativos.

Em trabalhos anteriores do grupo, as massas molares destas proteínas foram preditas usando uma coluna de filtração em gel analítica calibrada. Os resultados mostraram que as proteínas XACb0033 e FlgN se encontraram monoméricas em solução, assim como para as medidas de DLS.⁶¹

Neste mesmo trabalho, amostras da proteína FlgN foram submetidas à análise de SAXS nas linhas de SAXS do LNLS e os resultados estão mostrados a seguir. Devido à ausência de modelos de alta resolução determinados por técnicas experimentais, a estrutura atômica da FlgN foi calculada pelo programa I-TASSER (que será discutido posteriormente) e está representada na Figura I-7, assim como a sobreposição deste modelo no envelope experimental obtido pelos dados de SAXS, publicados recentemente.⁶²

⁶⁰ <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

⁶¹ Fattori, J. Resolução Estrutural de Proteína Hipotéticas, Chaperonas de Secreção, da bactéria *Xanthomonas axonopodis* pv. *citri*. Tese de Doutorado, Instituto de Química, Universidade Estadual de Campinas, SP, setembro 2011.

⁶² Fattori, J.; Prando, A.; Assis, L.; Aparicio, R.; Tasic, L. Structural Insights on Two Hypothetical Secretion Chaperones from *Xanthomonas axonopodis* pv. *citri*. *Protein J* **2011**, *30*, 324-333.



Figura I-7: As três vistas do modelo gerado por SAXS da FlgN (em cinza) sobreposto com a estrutura calculada pelo I-TASSER (em vermelho).⁶²

ESTUDOS ESTRUTURAIS

1.2 Estudos Estruturais

1.2.1. Estrutura e função de proteínas

Uma peça chave para decifrar a função de uma proteína é o entendimento da sua estrutura. As proteínas são poliamidas cujas unidades monoméricas são L- α -aminoácidos e podem ser descritos em níveis de organização da sua estrutura: primária, secundária, terciária e quaternária. Os aminoácidos polimerizados formam cadeias polipeptídicas através das ligações peptídicas e as variações na sequência de aminoácidos contribuem para a diversidade na forma e nas funções biológicas das proteínas⁶³

A estrutura primária de uma proteína, caso ela seja composta de somente uma cadeia peptídica, é definida pela sequência de α -aminoácidos na cadeia polipeptídica unida por ligações covalentes ou ligações peptídicas a partir do N- até o C- terminal. A estrutura secundária representa o arranjo espacial dos átomos da cadeia principal $[-N(H)C\alpha C(O)-]_n$, sem levar em consideração a conformação das cadeias laterais e inclui estruturas regulares, tais como α -hélices, folhas β , voltas e regiões randômicas. Estes elementos de estrutura secundária, por sua vez, interagem entre si formando a estrutura terciária da proteína. A estrutura terciária especifica as posições de cada átomo na proteína, incluindo as cadeias laterais e, muitas vezes, para desempenharem suas funções algumas proteínas atuam como oligômeros ou requerem a ligação de um cofator gerando a estrutura quaternária.⁶³

A complexidade estrutural e a variedade das proteínas permitem que elas realizem várias tarefas biológicas como, por exemplo, agirem como catalisadores enzimáticos, como transportadoras e reguladoras, participar da defesa e na decodificação e transmissão da informação genética, entre outras. Para entendimento detalhado da função de uma macromolécula é necessário o conhecimento da sua estrutura tridimensional. Atualmente, duas principais

⁶³ Nelson, D.; Cox, M. Lehninger Principles of Biochemistry, 4th Ed.; W.H. Freeman & Company: New York, 2004.

técnicas são utilizadas para a determinação de estruturas de macromoléculas em nível atômico: a difração de raios-X e a ressonância magnética nuclear (NMR).⁶⁴ A difração de raios-X é sem dúvida, o principal método para a determinação da estrutura de proteínas uma vez que a maioria das proteínas foram analisadas por esta técnica (PDB), e para estudos de proteína em solução, a NMR é necessária. Para o melhor entendimento dos modelos obtidos através das técnicas supracitadas informações estruturais complementares e de grande importância podem ser fornecidas por diferentes técnicas como: Dicroísmo Circular (CD), Fluorescência de emissão, Espectrometria de Massas (MS).

As técnicas de caracterização estrutural utilizadas neste trabalho estão descritas a seguir.

1.2.2. Dicroísmo Circular

Uma das técnicas mais importantes na determinação de estrutura secundária das proteínas é a espectroscopia de dicroísmo circular (CD) na região do UV-distante (180-240 nm).^{65,66} O sinal de CD é resultante de uma absorção diferencial da luz circularmente polarizada a esquerda e a direita após sua passagem pela amostra, sendo nas proteínas, a ligação peptídica a responsável por esta diferença na absorção. As amidas, os cromóforos da ligação peptídica, têm duas transições eletrônicas: $n-\pi^*$ que é responsável pelos sinais negativos em 222 nm (característica de α -hélice) e 216-218 nm (característica de folhas- β) e $\pi_0-\pi^*$ que é responsável pelo sinal positivo em 190 nm e negativo em 208 nm (característica de α -hélice) e um sinal positivo em 198 nm (característica de folhas- β).⁶⁵

⁶⁴ Krishna, N.; Berliner, L. *Biological Magnetic Resonance: Modern Techniques in Protein NMR*, Kluwer Academic Publishers, New York, **2002**.

⁶⁵ Corrêa, D.; Ramos, C. The use of circular dichroism spectroscopy to study protein folding, form and function. *African J Biochem Res* **2009**, *3*, 164-173.

⁶⁶ Edelhoch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **1967**, *6*, 1948-1954.

Uma proteína com estrutura majoritária do tipo α -hélice apresenta mínimos locais em 222 e 208 nm e máximo em 190 nm. Este sinal em 222 nm está relacionado com as ligações de hidrogênio que estabilizam esta estrutura. Já uma proteína essencialmente do tipo folhas β tem um sinal negativo entre 210 - 220 nm e um positivo entre 195 - 200 nm. Os espectros deste tipo são mais dispersos devida a variedade de conformações que ela pode apresentar como folhas β em paralelo, antiparalelo, uma mistura de ambas e ainda pode estar torcida. Além disso, a proteína também pode ter uma estrutura desordenada conhecida como randômica apresentando uma banda negativa em 200 nm.⁶⁵

Sendo assim, os arranjos regulares das proteínas resultam na obtenção de espectros de CD como função de elipticidade (θ em graus) em função do comprimento de onda (λ em nm). O CD pode ser usado também em estudos conformacionais e estruturais de proteína em diferentes condições como temperatura, força iônica e presença de solutos e pequenas moléculas.⁶⁵

Os espectros de CD das proteínas podem ser usados para monitorar mudanças conformacionais devido a temperatura, mutações, aquecimento, desnaturantes e interações com ligantes. Quando ligantes com cromóforos se ligam as proteínas, as proteínas desenvolvem fortes sinais de CD extrínseco que podem ser usadas para monitorar o ligante.⁶⁷ Além disso, a titulação de proteínas com ligantes pode ser usada para calcular a constante de ligação das interações proteína-proteína, proteína-DNA e proteína-ligante.⁶⁸

Há vários métodos para analisar espectros de CD e estimar a estrutura secundária e, todos assumem que o espectro da proteína é uma combinação linear dos espectros dos elementos estruturais secundários e de um termo ruído, que inclui a contribuição dos cromóforos aromáticos. Então, conformação da proteína desconhecida pode ser avaliada usando programas como o CDNN que compara o espectro da proteína desconhecida com os espectros de proteínas que

⁶⁷ Greenfield, N. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **2006**, *6*, 2876-2890.

⁶⁸ Greenfield, N. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat Protoc* **2006**, *6*, 2733-2741.

já tiveram a sua estrutura determinada por cristalografia de raios-X como padrão. O programa CDNN analisa dados por comparação e determina conteúdo (%) de α -hélice, folhas- β paralelas e antiparalelas e as voltas.⁶⁷

O CD pode ser usado também para monitorar o enovelamento e desenovelamento proteico. Geralmente proteínas com elevado grau de ordem, como aquelas com alto conteúdo de α -hélice, tem sinais de CD com um mínimo local e absoluto maiores que não estão presentes nas proteínas desenoveladas.⁶⁹ Quando a proteína esta desenovelada ela perde elementos estruturais secundários e o sinal de CD muda.⁷⁰

A estabilidade térmica de uma proteína pode ser monitorada analisando as mudanças causadas no espectro de CD de uma proteína com o aumento da temperatura. Normalmente um comprimento de onda é escolhido (222 nm, por exemplo) e o sinal é gravado continuamente conforme a temperatura é aumentada, obtendo-se um conjunto de dados de elipticidade molar *versus* temperatura. Mudanças de pH, tampão, e aditivos como açúcar, sal e ou aminoácidos alteram a estabilidade térmica da proteína. Muitas proteínas agregam e precipitam rapidamente depois que elas são desenoveladas tornando o desenovelamento irreversível. Quando a curva de desnaturação é reversível, a temperatura é relacionada com a estabilidade conformacional e indica a temperatura que a proteína desnatura.⁷¹ As mudanças no CD como função da temperatura, em comprimentos de onda característicos, podem ser usados para determinar os parâmetros termodinâmicos de desenovelamento (mudança de estado enovelado - desenovelado) que é a entalpia de van't Hoff (ΔH) e entropia (ΔS) do desenovelamento, o ponto médio da curva de desnaturação térmica (T_m) e a energia livre (ΔG) do enovelamento.^{69,72}

⁶⁹ Greenfield, N. Analysis of the kinetics of folding of proteins and peptides using circular dichroism. *Nat Protoc* **2006**, *6*, 2891-2899.

⁷⁰ Greenfield, N. Using circular dichroism collected as a function of temperatura to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* **2006**, *6*, 2527-2535.

⁷¹ http://www.ap-lab.com/circular_dichroism.htm (acessado em fevereiro de 2012).

⁷² a) Consalvi, V.; Chiaraluce, R.; Giangiacomo, L.; Scandurra, R.; Christova, P.; Karshikoff, A.; Knapp, S.; Landenstein, R. Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile *Thermotoga maritima*. *Protein*

Quando duas proteínas interagem, mudanças substanciais não ocorrem nos componentes da estrutura secundária com a formação do complexo e estas mudanças no espectro de CD na região do UV-distante são muito pequenas, porém, podem ser um indicativo da interação. No entanto, quando elas interagem há mudanças na estrutura terciária de uma ou de ambas as proteínas no complexo e esta interação é melhor visualizada no espectro de CD na região do UV-próximo (260-320 nm). Por exemplo, a associação de monômeros da insulina para formar um dímero e então hexamero é acompanhado por mudanças no meio da superfície do Trp (W) das cadeias laterais, registrada no UV-próximo apesar de não demonstrar mudanças significativas na região do UV-distante. Uma mudança significativa no espectro de CD na região do UV-distante pode ser exemplificada pela adição de íons Ca^{+2} à troponina C resultando em um considerável aumento do conteúdo de α -hélice.⁷³

1.2.3. Fluorescência

A fluorescência é um fenômeno que ocorre com as substâncias fluorescentes quando um elétron após excitado retorna ao seu estado fundamental (S_0), em que, nem toda a energia da excitação é convertida em fluorescência, podendo uma fração da mesma ser dispersa por processos não radioativos. Este fenômeno é denominado deslocamento de Stokes, sendo amplamente observado em ensaios de fluorescência. Devido a essa perda parcial de energia, a energia da luz emitida é sempre menor que a energia da luz utilizada para a excitação^{74,75}

Eng **2000**, *13*, 501-507. b) Benjwal, S.; Verma, S.; Röhm, K.; Gursky, O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Science* **2006**, *15*, 635-639.

⁷³ Kelly, S.; Price, N.; Circular Dichroism to study protein interactions. *Curr Protoc Protein Sc* **2006**, John Wiley & Sons, Inc.

⁷⁴ Lackovicz, J. *Principles of Fluorescence Spectroscopy* **1983**, New York: Kluwer Academy/Plenum Publisher.

⁷⁵ Rodrigues, F. Derivados de Quinazolininas na inibição da adenosina quinase. Dissertação de Mestrado, Instituto de Química, Universidade Estadual de Campinas, SP, Outubro de 2011.

As proteínas possuem certos aminoácidos que são fluoróforos e que podem contribuir para a fluorescência intrínseca como: triptofano (Trp), tirosina (Tyr) e fenilalanina (Phe). Dos três aminoácidos, o triptofano é o mais utilizado nos estudos de fluorescência de proteínas devido ao seu rendimento quântico ser alto o suficiente para a obtenção de um sinal visível de emissão. Esse aminoácido é extremamente sensível ao ambiente no qual se encontra, e, portanto, apresenta características de fluorescência diferenciadas em cada grau de exposição aos diferentes meios, os quais podem ser hidrofílicos ou hidrofóbicos.⁷⁴

Em ensaios conduzidos com as proteínas, os deslocamentos dos máximos de fluorescência podem ser indicativos de uma mudança conformacional decorrente de algum processo com variações no pH, ou através da adição de ligantes, os quais podem perturbar a estrutura proteica.⁷⁴ Durante o processo de supressão de fluorescência, os máximos de emissão podem ser deslocados para comprimentos de onda menores (deslocamento para o azul) ou maiores (deslocamento para o vermelho) ou ainda ficarem inalterados. A supressão de fluorescência envolve a ocorrência de um fenômeno denominado Transferência de Energia de Ressonância (*Ret; Resonance Energy Transfer*). Este fenômeno ocorre quando um fluoróforo, denominado doador, transfere energia para outra espécie, denominada acceptor (no caso os ligantes); resultando em uma diminuição na intensidade de fluorescência emitida (supressão) do doador. Neste fenômeno não há emissão de fluorescência, uma vez que o processo não envolve a absorção de um fóton emitido pelo doador por parte do acceptor.⁷⁵

A fluorescência também pode ser usada para acompanhar interações proteína-proteína. Quatro parâmetros fotofísicos de fluorescência de emissão podem ser usados para quantificar interações *in vivo* e *in vitro* como polarização, tempo de vida, energia média e rendimento quântico.⁷⁶ As interações entre proteínas podem provocar as mudanças conformacionais na estrutura das proteínas levando uma maior exposição dos resíduos de Trp. Estas mudanças conformacionais podem ser verificadas também quando, por exemplo, uma

⁷⁶ Yan, Y.; Marriott, G. Analysis of protein interactions using fluorescence technologies. *Curr Opin Chem Biol* **2003**, 7, 635-640.

proteína é levada a condições de desnaturação. Além da supressão da fluorescência é observado um deslocamento químico para comprimentos de onda maiores (vermelho).⁷⁷

Como exposto, a fluorescência permite verificar a existência de mudanças conformacionais em proteínas, monitorar interações proteínas-proteínas, proteínas e ligantes e, também, determinar constante de dissociação (K_d) através de gráficos de Stern-Volmer e de Regressão Não Linear de Mínimos Quadrados (rnlmq). Para proteínas com apenas um triptofano em sua estrutura primária, os Gráficos de Stern-Volmer apresentam um comportamento linear e o valor de K_d pode ser obtido através do coeficiente angular da equação proveniente de uma regressão linear. Mas quando mais fluoróforos estão presentes, desvios da linearidade são comumente observados.⁷⁵

A fim de calcular o K_d , uma regressão não linear de mínimos quadrados se faz necessária, e é executada através do ajuste (regressão) da seguinte equação:

$$Q = \frac{Q_{\max} k_a C}{1 + k_a C}$$

Onde Q é a supressão de fluorescência observada (“*quenching dinâmico*”)⁷⁸, Q_{\max} é a máxima supressão de fluorescência observada, K_a a constante de associação, e C é a concentração do ligante. O valor de K_d é o inverso do valor de K_a e é então obtido como $1/K_a$. O valor de K_d descreve a tendência de um ligante dissociar-se de seu hospedeiro, sendo que quanto menor o valor da constante de dissociação, menor é a tendência em desligar-se de seu sítio de ligação, indicando uma maior afinidade.⁷⁵

⁷⁷ <http://www.chembio.uoguelph.ca/merrill/research/optical.html> (acessado em Fevereiro de 2012).

⁷⁸ a) Gratton, E.; Jameson, D.; Weber, G. A Modelo f Dynamic Quenching of Fluorescence. *Biophys. J.* **1984**, *45*, 789-794. b) Htun, T. A Negative Deviation from Stern–Volmer Equation in Fluorescence Quenching. *J. Fluorescence* **2004**, *14*, 217-222. c) Webber, S. The Role of Time-Dependent Measurements in Elucidating Static Versus Dynamic Quenching Processes. *Photoch. Photobiol.* **1997**, *65*, 33-38.

1.2.4. Espectrometria de Massas em análise proteínas

A espectrometria de massas é uma técnica capaz de determinar massas molares de forma muito precisa em experimentos rápidos sendo possível averiguar a sequência de aminoácidos de uma proteína, identificar proteínas, sequenciar peptídeos e proteínas, dentre outros. Através da espectrometria de massas é possível obter a massa molar proteica exata com alta precisão (até 6 algarismos significativos). A MS também pode fornecer informações importantes em relação à pureza de proteínas e peptídeos e formação de complexos proteicos.⁷⁹

Em uma das duas técnicas usadas para ionização das proteínas e posterior análise por espectrometria de massas (MS), proteínas são ionizadas aplicando laser e analisadas no espectrômetro de massas que poderia ser do tipo time-of-flight (MALDI-TOF/MS). Neste caso, devido a tempos diferentes de voo por causa da diferenças em massa/carga (m/z) de proteínas ocorre uma separação. Quanto mais pesada a proteína for, maior será seu tempo de voo. O sequenciamento de proteínas pode ser feito utilizando os métodos sequenciais, sendo que antes de serem levadas ao espectrômetro de massas, as amostras de proteínas são digeridas pela aplicação de proteases tais como tripsina e quimiotripsina.⁸⁰

No caso de ESI a amostra é submetida a um fluxo em uma pequena agulha em um compartimento contendo o gás de nebulização nitrogênio. Entre a saída da agulha e o compartimento há um campo elétrico que gera uma diferença de potencial fazendo com que a amostra saia como um spray de microgotas carregadas, que, direcionadas pelo campo, seguem para o analisador (por exemplo, um TOF). O campo elétrico é forte o suficiente para extrair íons isolados das microgotas. O tempo que cada íon leva para atingir o detector depende de sua

⁷⁹ Cunha, R.; Castro, M.; Fontes, W. Espectrometria de massa de proteínas. *Biociência* 2006, 36, 40-46.

⁸⁰ Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* 2006, 312, 212-217.

razão massa carga (m/z) sendo assim possível determinar a massa molecular do analito de interesse.⁸¹

Os sequenciamentos por MS são realizados em equipamentos que possuem pelo menos dois analisadores de massa. Estes analisadores são dispostos em série e, por isso, esta técnica é conhecida por MS/MS. Alguns exemplos de analisadores de série incluem aprisionamento de íons (*ion trap*) e equipamentos com ionização do tipo MALDI com dois analisadores TOF (MALDI-TOF-TOF). Nestas análises uma amostra de peptídeo puro ou uma mistura obtida através da digestão enzimática é injetada no espectrômetro de massas e os íons de interesse são selecionados e acelerados para uma câmara de colisão. Nesta câmara os íons sofrem colisão com o gás inerte (nitrogênio ultrapuro) resultando na fragmentação da cadeia polipeptídica, fenômeno conhecido como CID (*collision induced dissociation*). Como as ligações peptídicas são as mais propensas ao rompimento, obtém-se um espectro contendo todos os aminoácidos e peptídeos gerados a partir do lado C- e N- terminais.^{79,82}

Atualmente a MS teve grandes avanços para estudos estruturais de proteínas (terciária e quaternária) principalmente na área denominada proteômica estrutural. Dentre elas a técnica de ligação cruzada (*Cross-Linking*) tem fornecido resultados muito promissores.⁸³ Nos ensaios de ligação cruzada, proteínas são modificadas quimicamente com um escolhido agente de ligação cruzada (ALC) que interage covalente com aminoácidos próximos espacialmente e os conecta (grampos). Os ALC (Figura I-8) contêm pelo menos dois grupos reativos unidos por uma cadeia conhecida como cadeia espaçadora. Estes compostos, na

⁸¹ Colnago, L.; Almeida, F.; Valente, A. Espectrometria de massa e RMN multidimensional e multinuclear: revolução no estudo de macromoléculas biológicas. *Química Nova na Escola* **2002**, *16*, 9-14.

⁸² a) Divito, E.; Davic, A.; Johnson, M.; Cascio, M. Electrospray Ionization and Collision Induced Dissociation Mass Spectrometry of Primary Fatty Acid Amides. *Anal Chem* **2012**, *84*, 2388-2394. b) Nili, M.; Mukherjee, A.; Shinde, U.; David, L.; Rotwein, P. Defining the Disulfide Bonds of Insulin-like Growth Factor-binding Protein-5 by Tandem Mass Spectrometry with Electron Transfer Dissociation and Collision-induced Dissociation. *J Biol Chem* **2012**, *287*, 1510-1519.

⁸³ Chen, Z.; Jawari, A.; Ficher, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, M.; Wills, J.; Nilges, M.; Cramer, P.; Rappsilber, J. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectroscopy. *EMBO J* **2010**, *29*, 717-726.

presença de proteínas, reagem com as cadeias laterais dos aminoácidos de acordo com suas especificidades (Figura I-9).⁸⁴

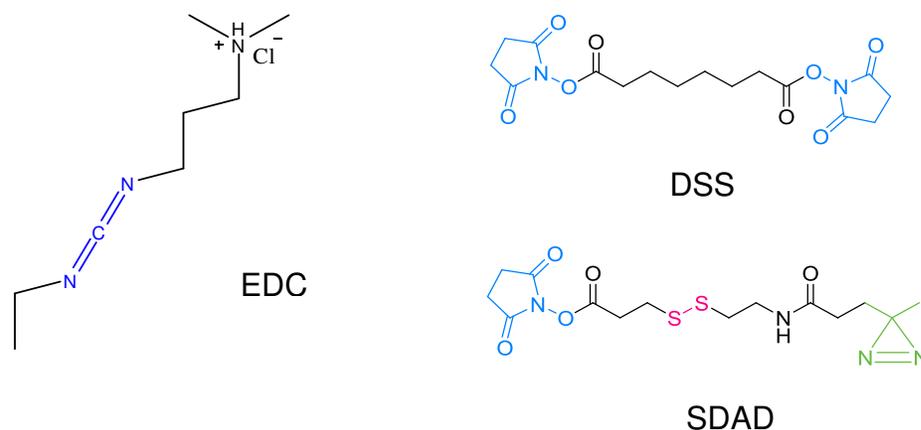


Figura I-8: Estrutura dos reagentes de ligação cruzada. EDC = (1-ethyl-3-(3-dimethylaminoisopropyl)carbodiimide). DSS = (disuccinimidyl suberate). SDAD = (succinimidyl-2-([4,4'-azipentanamido]ethyl)-1,3'-dithiopropionate).

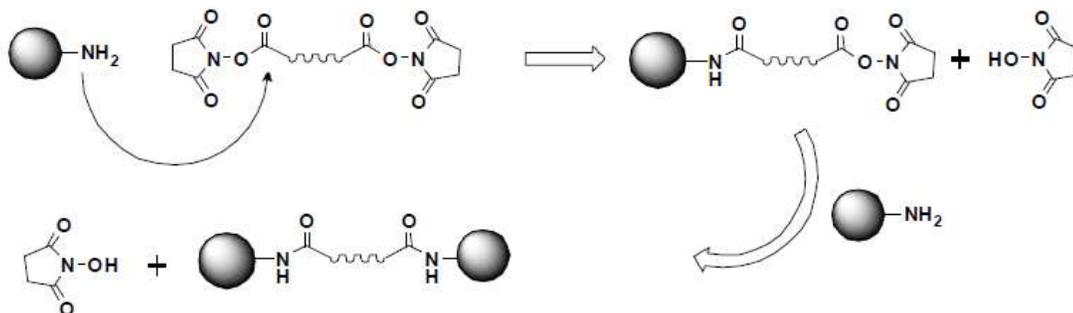


Figura I-9: Esquema de reação entre o ALC reativo a grupos amino (~~~ cadeia espaçadora).⁸⁴

Através dos experimentos de ligação cruzada é possível determinar pares ligados pelo agente de ligação cruzada e por meio de modelagem molecular obter

⁸⁴ Figueredo, A. Estudo Teórico e Experimental de Proteômica Estrutural por Espectrometria de Massas Acoplada à Ligação Cruzada. Dissertação de Mestrado, Instituto de Química, Universidade Estadual de Campinas, SP, Fevereiro de 2010.

estruturas tridimensionais que satisfaçam as distâncias definidas experimentalmente (restrições espaciais).⁸⁵

Uma outra técnica usada para proteínas em MS é conhecida como *Footprinting*. Esta técnica baseia-se na determinação do grau de exposição ao solvente de porções de uma macromolécula (proteínas ou ácidos nucleicos) por meio de reações de modificação química ou enzimática.⁸⁶ Tais modificações químicas são de natureza covalente, e possibilitam avaliar mudanças conformacionais devido à interação da macromolécula com um determinado ligante. Esta abordagem consiste na reação das proteínas isoladas e do complexo proteína-ligante com um reagente químico, como, por exemplo, o dietil pirocarbonato (DEPC). Este reagente apresenta alta reatividade à resíduos de histidina, levando a uma modificação covalente no resíduo de histidina, como mostrado na Figura I-10. Ainda que menos pronunciado, os resíduos de tirosina, serina, lisina e arginina também podem ser modificados por ele além de outros nucleófilos.⁸⁷

⁸⁵ a) Rappsilber, J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* **2011**, *173*, 530-540. b) Back, J.; Jong, L.; Muijsers, A.; Koster, C. Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J Mol Biol* **2003**, *331*, 303-313. c) Young, M.; Tang, N.; Hempel, J.; Oshiro, C.; Taylor, E.; Kuntz, I.; Gibson, B.; Dollinger, G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *PNAS* **2000**, *97*, 5802-5806.

⁸⁶ (a) Maleknia, S.; Ralston, C.; Brenowitz, M.; Downard, K.; Chance, M. Determination of Macromolecular folding and structure by Synchrotron X-Ray Radiolysis Techniques. *Anal Biochemistry* **2001**, *289*, 103-115. (b) Maleknia, S.; Downard, K. Radical approaches to probe protein structure, folding and interactions by mass spectrometry. *Mass Spectrom. Rev* **2001**, *20*, 388-401. (c) Guan, J.; Almo, S.; Chance M. Synchrotron Radiolysis and mass spectrometry: A new Approach to Research on the Actin Cytoskeleton. *Acc Chem Res* **2004**, *37*, 221-229.

⁸⁷ Adaptado de Dage, J., Sun, H.; Halsall, H. Determination of Diethylpyrocarbonate-Modified Amino Acid Residues in alpha 1-Acid Glycoprotein by High-Performance Liquid Chromatography Electro-spray Ionization-Mass Spectrometry and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight-Mass Spectrometry *Anal Biochem* **1998**, *257*, 176-185.

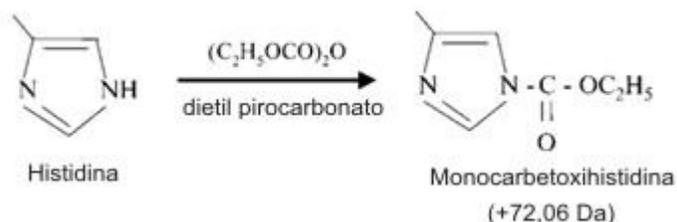


Figura I-10: Modificação dos resíduos de histidina pelo DEPC.⁸⁷

Após isso, as amostras são digeridas e analisadas por MS. Análises do tipo LC-MS/MS permitem identificar os sítios das modificações pelo reagente, enquanto análises de LC/MS são utilizadas para quantificá-las. Nas análises de LC-MS as relações m/z dos peptídeos inteiros são medidas. Comparando-se as intensidades dessas relações m/z entre um experimento controle (só a proteína ou o complexo proteico) e um experimento de *footprinting*, aqueles peptídeos que reagiram com o DEPC apresentarão diferença de intensidade (pois o peptídeo foi "consumido" para gerar o produto de reação com DEPC, então nesse espera-se que a intensidade do sinal dele diminua no experimento de *footprinting*, às custas do surgimento de um sinal do peptídeo modificado pelo DEPC). Por essa diferença de intensidade dos sinais, é possível estimar quantitativamente o quanto cada aminoácido (His) reagiu com o DEPC. Os resíduos (peptídeos) que não reagiram com o DEPC, por exemplo, são aqueles que não são acessíveis ao solvente. O LC-MS/MS, no caso, é usado para a fragmentação dos peptídeos no massas, *i.e.*, sequenciamento. Nesse caso, os resíduos de aminoácido que sofreram reação com o DEPC apresentarão um ganho de massa. As regiões que estiverem em contato com o solvente serão mais acessíveis ao DEPC. Métodos computacionais são utilizados para a construção de modelos tridimensionais.⁸⁸

⁸⁸ Kamal, J.; Chance, M. Modeling of protein binary complexes using structural mass spectrometry data. *Protein Sci* **2008**, 17, 79-94.

Os métodos baseados em *footprinting* têm possibilitado a caracterização da estrutura de complexos proteínas-ligantes em solução⁸⁹ e estudos de dinâmica de formação/enovelamento resolvidos no tempo. Além disso, também é possível a determinação de sítios de interação proteína-ligante, para diversos tipos de ligantes, tais como íons metálicos,⁹⁰ peptídeos,⁹¹ proteínas⁹² e ácidos nucleicos.⁹³

1.2.5. Ressonância Magnética Nuclear em análises de interações proteínas-ligantes.

As interações intermoleculares entre ligantes e receptores macromoleculares são responsáveis pelos vários processos biológicos, tais como a transdução de sinal, vias metabólicas, regulação de expressão de genes entre outras. No contexto de entender melhor interações intermoleculares, a ressonância magnética nuclear (NMR) tornou-se uma técnica importante e indispensável sendo capaz de detectar e quantificar as interações com alta sensibilidade e sem o conhecimento prévio da estrutura da macromolécula.⁹⁴

⁸⁹ Zheng, X.; Wintrode, P.; Chance, M. Complementary structural mass spectrometry techniques reveal local dynamics in functionally important regions of a metastable serpin. *Structure* **2008**, *16*, 38-51.

⁹⁰ (a) Guan, J.; Almo, S.; Reisler, E.; Chance, M. Structural Reorganization of Proteins Revealed by Radiolysis and Mass Spectrometry: G-Actin Solution Structure Is Divalent Cation Dependent. *Biochem* **2003** *42*, 11992-12000. (b) Kiselar, J.; Janmey, P.; Almo, S.; Chance, M. Visualizing the Ca²⁺-dependent activation of gelsolin by using synchrotron footprinting. *Proc Natl Acad Sci USA* **2003**, *100*, 3942-3947.

⁹¹ Gupta, S.; Mangels, W.; McGrath, W.; Perek, J.; Lee, D.; Takamoto, K.; Chance, M. DNA binding provides a molecular strap activating the adenovirus proteinase. *Mol Cell Proteomics* **2004**, *3*, 950-959.

⁹² (a) Xu, G.; Liu, R.; Zak, O.; Aisen, P.; Chance, M. Structural allostery and binding of the transferrin-receptor complex. *Mol Cell Proteomics* **2005**, *4*, 1959-1967. (b) Liu, R.; Guan, J.; Zak, O.; Aisen, P.; Chance, M. Structural Reorganization of the Transferrin C-Lobe and Transferrin Receptor upon Complex Formation: The C-Lobe Binds to the Receptor Helical Domain. *Biochem* **2003**, *42*, 12447-12454. (c) Guan, J.; Vorobiev, S.; Almo, S.; Chance, M. Mapping the G-Actin Binding Surface of Cofilin Using Synchrotron Protein Footprinting. *Biochem* **2002**, *41*, 5765-5775.

⁹³ (a) Gupta, S.; Cheng, H.; Mollah, A.; Jamison, E.; Morris, S.; Chance, M.; Khrapunov, S.; Brenowitz, M. DNA and Protein Footprinting Analysis of the Modulation of DNA Binding by the N-Terminal Domain of the *Saccharomyces cerevisiae* TATA Binding Protein. *Biochem* **2007**, *46*, 9886-9898. (b) Rashidzadeh, H.; Khrapunov, S.; Chance, M.; Brenowitz, M. Solution Structure and Interdomain Interactions of the *Saccharomyces cerevisiae* "TATA Binding Protein" (TBP) Probed by Radiolytic Protein Footprinting. *Biochem* **2003**, *42*, 3655-3665.

⁹⁴ Meyer, B.; Peters T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed* **2003**, *42*, 864-890.

A técnica de NMR nomeada STD (**S**aturation **T**ransfer **D**ifference) foi inicialmente proposta para investigar a ligação de pequenos ligantes à macromoléculas biológicas, incluindo proteínas e ácidos nucleicos e uma de suas vantagens é atribuída a habilidade de detectar fracas ligações com grande sensibilidade.⁹⁵ A STD em sistemas proteína-ligante é baseada na transferência de magnetização de proteínas para ligantes através da aplicação de um pulso de saturação seletivo no sinal de hidrogênio da proteína. A saturação propaga-se através dos hidrogênios da proteína via rede de interações dipolares intramoleculares H-H (difusão de spin). A saturação é transferida aos compostos ligados via relaxação cruzada intermolecular para a interface proteína-ligante. As pequenas moléculas dissociam-se do receptor, mas permanecem em um estado “saturado” devido aos seus longos tempos de T1 (Figura I-11). Também, o grau de saturação de hidrogênios de moléculas de ligantes menores reflete a proximidade destes com a superfície da proteína. Informação pode ser obtida rapidamente e facilmente com esta técnica e apenas uma pequena quantidade de proteína não marcada isotopicamente é necessária.^{96,97,98}

⁹⁵ Huang, H.; Milojevik, J.; Melacini, G. Analysis and Optimization of Saturation Transfer Difference NMR Experiments Designed to Map Early Self-Association Events in Amyloidogenic Peptides. *J Phys Chem* **2008**, *112*, 5795-5802.

⁹⁶ Ji, Z.; Yao, Z.; Liu, M. Saturation transfer difference nuclear magnetic resonance study on the specific binding of ligand to protein. *Anal Biochem* **2009**, *385*, 380-382.

⁹⁷ Streiff, J.; Juranic, N.; Macura, S.; Warner, D.; Jones, K.; Perkins, W. Saturation Transfer Difference Nuclear Magnetic Resonance Spectroscopy as a Method for Screening Proteins for Anesthetic Binding. *Mol Pharmacol* **2004**, *66*, 929-935.

⁹⁸ Figueiredo, I.; Marsaioli, A. Mapeamento das interações proteína-ligante através de técnicas de RMN de ¹H utilizando detecção do ligante. *Quim Nova*, **2007**, *30*, 1597-1605.

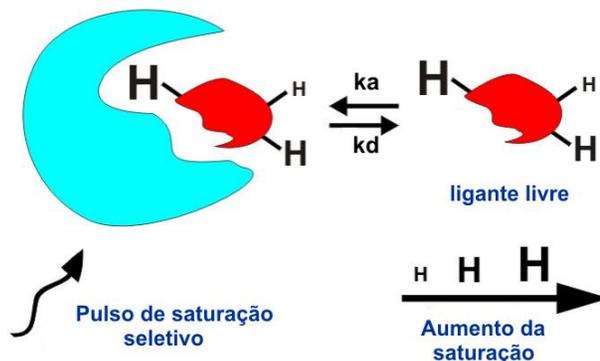


Figura I-11: Ilustração do mapeamento dos grupos do epítipo (*Group Epitope Mapping* - GEM) para dada interação proteína-ligante em uma troca rápida entre ligante livre e complexado. Os grupos representados por um átomo de hidrogênio grande (H) estão em contato mais próximo com a proteína, enquanto os H de tamanho mediano simbolizam grupos com menor interação. Os H menores representam um grupo com quase nenhum contato com a proteína, então recebendo saturação mínima (Figura adaptada)⁹⁹

Existem outros métodos para estudar as interações entre proteínas e ligantes, em que a obtenção de deslocamento químico induzido por complexação (CICS do inglês *Complexation-Induced Chemical Shift*) é um deles. Pode ser usado para mapear as interfaces do complexo e localizar sítios ligantes nas proteínas explorando a influência do meio circundante ao deslocamento químico dos hidrogênios. Por exemplo, quando não há uma mudança conformacional significativa induzida pelo ligante na proteína, as mudanças podem ser atribuídas ao contato direto do ligante com a proteína através de ligações de hidrogênio, ou por efeitos indiretos como produzidos por grupos aromáticos. Resíduos da proteína que sofrem perturbação são considerados próximos dos sítios de interação com o ligante. Na descoberta de novos fármacos baseado na estrutura, o mapeamento CICS é muito importante em NMR para estudar a relação

⁹⁹ Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J Am Chem Soc* **2001**, *123*, 6108-6117.

estrutura-atividade (SAR).¹⁰⁰ Com o CICS é possível obter a localização aproximada do sítio ligante através da comparação dos deslocamentos químicos da proteína livre e da proteína complexada. O mesmo pode ser realizado para o ligante livre e complexado com a proteína para descobrir quais hidrogênios do espectro de HETCOR {¹³C, ¹H} do ligante livre e depois complexado, por exemplo, pode fornecer dados sobre quais hidrogênios e carbonos do ligante estão interagindo com a proteína, através de mudanças no deslocamento químico observadas.¹⁰¹ Do mesmo modo, o espectro de HSQC {¹⁵N, ¹H} da proteína livre e depois complexada com o ligante pode fornecer a localização provável do sítio ligante da proteína e determinar a orientação provável do ligante dentro do sítio ligante.¹⁰²

¹⁰⁰ Cioffi, M.; Hunter, C.; Packer, M.; Spitaleri, A. Determination of protein-ligand binding modes using complexation-induced changes in ¹H NMR chemical shift. *J Med Chem* **2008**, *51*, 2512-2517.

¹⁰¹ Cagliari, T.; da Silva, V.; Borges, J.; Prando, A.; Tasic, L.; Ramos, C.; Sugarcane Hsp101 is a hexameric chaperone that binds nucleotides. *Int J Biol Macromol* **2011**, *42*, 1022-1030.

¹⁰² Cioffi, M.; Hunter, C.; Packer, M.; Pandya, M.; Williamson, M. Use the quantitative ¹H NMR chemical shift changes for ligand docking into barnase. *J Biomol NMR* **2009**, *43*, 11-19.

Capítulo II - Objetivos

2. Objetivos

Este trabalho tem como objetivos: (1) determinar características estruturais de duas chaperonas de secreção da bactéria *Xanthomonas axonopodis* pv. *citri* (FlgN e XACb0033) via estudos biofísicos; e (2) elucidar as interações intramoleculares entre a Hsp90 da laranja e alguns dos ligantes, para melhor compreensão da estrutura e função desta proteínas.

Para tal finalidade foram propostos os seguintes objetivos específicos:

1. Expressar e purificar as proteínas alvos em altas concentrações;
2. Determinar a concentração das proteínas em solução;
3. Caracterizar as proteínas alvos XACb0033 e FlgN por métodos biofísicos (Dicroísmo Circular, Fluorescência de emissão, Ressonância Magnética Nuclear, Espectrometria de Massas) e realizar experimentos de Dicroísmo Circular, Fluorescência de emissão e Ressonância Magnética Nuclear para a FlgN na presença de sua proteína parceira, FlgK, em razão molar de 1:1 e 1:2.
4. Propor as estruturas para as XACb0033 e FlgN de acordo com dados de *Footprinting*, SASA e RMSD.
5. Obter os dados estruturais através de ferramentas de bioinformática e compara-los com os dados obtidos experimentalmente;
6. Estudar as interações da chaperona Hsp90 da laranja com os ligantes ATP, ADP, ATP γ S e geldanamicina (possível inibidor) usando as técnicas de Saturation Transfer Difference (STD-NMR) e Fluorescência de emissão.

Capítulo III - Parte Experimental

3. Parte Experimental

3.1. Material: proteínas e ligantes

As ORFs referentes às proteínas de interesse XACb0033, XACb0032, FlgN e FlgK foram anteriormente clonadas^{103,31} no vetor de expressão pET23a (Novagen - XACb0032, XACb0033 e FlgN) e pET28a (Novagen – FlgK) e armazenados à -80°C. Após extração do DNA plasmidial por lise alcalina, estes foram transformadas em cepas de *Escherichia coli* (*E. coli*) BL21 (DE3) pLysS através do método de choque térmico. As proteínas foram expressas em meio LB contendo os antibióticos necessários (ampicilina, cloranfenicol ou canamicina) à 37°C, 200 rpm por tempo de 16hs. Aliquotas de 5 mL da cultura das linhagens de *E.coli* BL21(DE3)pLysS contendo o plasmideo recombinante foram inoculados para o mesmo meio (500mL). As culturas cresceram em meio LB à 37°C, 200 rpm até atingirem A_{600} de 0,8 e então, a expressão foi induzida pela adição de 1 mM de IPTG. O crescimento bacteriano foi procedido de 3 ou 16hs dependendo da proteína de interesse e então o meio foi centrifugado e o pellet foi armazenado à -80°C. Em seguida os pellets bacterianos foram ressuspensos no tampão de lise e as células foram lisadas em sonicador. Após centrifugação a proteína foi para corpo de inclusão e sua solubilização foi realizada com uréia. Nesta etapa, o renovelamento do pellet bacteriano pode levar a um renovelamento parcial ou errôneo. Já as proteínas solúveis não apresentam este tipo de problema. Depois da purificação, as proteínas foram usadas nos ensaios biofísicos (CD, fluorescência de emissão e MS). Para os ensaios de NMR a expressão das proteínas foi realizada da mesma maneira que para a expressão em meio LB mas, a indução (500mL) foi realizada em meio mínimo. Após 30 min. de indução com IPTG, rifamicina foi adicionada. O procedimento posterior é igual ao supracitado.

¹⁰³ Tasic, L.; Borin, P.; Khater, L.; Ramos, C. Cloning and characterization of three hypothetical secretion chaperones proteins from *Xanthomonas axonopodis* pv. *citri*. *Protein Expres Purif* **2007**, *53*, 363-369.

A proteína Hsp90 da laranja e Hsp100 da cana-de-açúcar¹⁰⁴ foram clonadas e purificadas pelo grupo do Prof. Carlos Ramos.

A Hsp100 da cana-de-açúcar é uma proteína com atividade ATPase e tem como principal função auxiliar na solubilização de proteínas formadores de agregados. Além disso, esta proteína é considerada uma chaperona molecular apresentando funções características deste grupo de proteínas como: enovelamento proteico, transporte através de membranas de polipeptídios recém sintetizados para as organelas; prevenção da agregação proteica e interações não desejadas/produzidas durante o enovelamento e re-solubilização/desagregação de agregados proteicos, bem como o encaminhamento dos mesmos para vias de degradação.¹⁰⁴

Os ligantes usados em ensaios de monitoramento de interações proteína-ligante foram: ATP, ADP, ATP_γS e geldanamicina (Sigma-Aldrich).

3.2. Extração do DNA plasmidial (lise alcalina)

Foi preparado um inóculo de uma colônia de *E. coli*, cepa DH5 α transformada com o DNA referente à proteína de interesse (pET23a-DNA para XACb0033, XACb0032, FlgN e pET28a-DNA para FlgK) em 2,0 mL de meio LB (Tabela III-1). Ao meio LB foram adicionados 50 $\mu\text{g.mL}^{-1}$ de ampicilina, ou 30 mg.mL^{-1} de canamicina (dependendo do sistema pET-DNA) e então foi incubado por 16 horas a 37°C e 200 rpm (Incubadora Tecnal TE-420). No caso da linhagem DH5 α nenhum antibiótico é adicionado, exceto do sistema pET pois, ela não apresenta resistência a cloranfenicol. Transcorrido este período, a suspensão foi transferida para um microtubo e centrifugada (Centrífuga Eppendorf 5804R) a 14000 rpm durante 1 minuto à temperatura ambiente. O sobrenadante foi descartado e o pellet foi ressuspensionado em 100 μL da solução I (Tabela III-1), agitando-a em vortex. Esta solução foi encubada durante 10 minutos no gelo e,

¹⁰⁴ Cagliari, T. Análise da expressão de chaperonas moleculares em plantas e clonagem, purificação e caracterização inicial das proteínas Hsp100 e Hsp90 de cana-de-açúcar. Tese de Doutorado, Instituto de Biologia, Universidade de São Paulo, SP, Maio 2009.

então, 200 μL da solução II recém-preparada (Tabela III-1) foram adicionadas. O tubo foi mantido no gelo por 5 minutos. A esta suspensão foram adicionados 150 μL da solução III (Tabela III-1) e homogeneizou-se a mistura por agitação em vortex. O microtubo foi incubado em banho de gelo por mais 5 minutos e a mistura foi centrifugada a 14000 rpm por 5 minutos. O sobrenadante foi transferido para um novo microtubo e ao sobrenadante foram adicionados 225 μL de fenol e 225 μL de clorofórmio, agitando-se em vortex. Em seguida, foi feita uma centrifugação a 12000 rpm por 5 minutos e o sobrenadante foi transferido novamente para um novo microtubo.

A este foram adicionados 1000 μL de etanol absoluto gelado para precipitação do DNA e foi incubado por 30 minutos em gelo. Então a solução foi centrifugada a 12000 rpm por 20 minutos. Desprezou-se o sobrenadante e o pellet foi lavado com 1000 μL de etanol 70% gelado. O sobrenadante foi centrifugado por 5 minutos a 12000 rpm e após este período, o sobrenadante foi então removido e o pellet foi seco a temperatura ambiente. O DNA foi então ressuspendido em 37 μL de TE pH 8,0 e 3 μL de RNase (Tabela III-1) e armazenado a -80°C .

Tabela III-1. Soluções da lise alcalina

Meio de cultura LB	Solução I	Solução II
Peptona 1,0% (m/v)	Glicose 50 mmol.L^{-1}	NaOH 0,2 mol.L^{-1}
Extrato de levedura 0,5% (m/v)	Tris 25 mmol.L^{-1}	SDS 1% (m/v)
NaCl 1,0% (m/v)	EDTA 10 mmol.L^{-1}	Preparada no dia do uso
pH 7,0; autoclavado	pH 8,0; autoclavada e armazenado a 4°C	
Tampão TE/RNase		Solução III
Tris 10 mmol.L^{-1}		CH_3COOH 2 mol.L^{-1}
EDTA 1 mmol.L^{-1}		CH_3COONa 3 mol.L^{-1}
RNase 10 mg m.L^{-1}		Armazenada 4°C
pH 8,0; armazenado a -4°C		

3.3. Preparação de células competentes

O pré-inóculo foi preparado a partir de uma placa de cultura fresca da cepa BL21(DE3)pLysS de *E. coli* (Invitrogen), em 20 mL de meio de cultura PSI (Tabela III-2) contendo 20 μ L de cloranfenicol e deixado por 16 horas sob agitação (200 rpm – Incubadora Tecnal TE-420) a 37°C. Transcorrido este período, 5 mL do pré-inóculo foi diluído em 200 mL de meio PSI contendo 200 μ L de cloranfenicol a 37°C e 200 rpm. A absorbância foi monitorada em 600 nm (Espectrofotômetro Bel Photonics SP1105) e, quando se atingiu a absorbância de 0,4, o crescimento foi interrompido colocando a cultura em banho de gelo. A suspensão foi então centrifugada (Centrífuga Beckman Coulter Allegra X-22R) por 10 minutos a 4000 rpm e 4°C. Durante todos os passos banho de gelo foi utilizado para estacionar o crescimento. O pellet foi então ressuscitado em 25 mL do tampão de transformação (Tampão I – Tabela III-2), deixado 15 minutos em repouso no gelo e posteriormente centrifugado por 15 minutos a 4000 rpm, 4°C. O sobrenadante foi descartado e o pellet foi novamente ressuscitado em 8 mL Tampão II (Tabela III-2). Foram feitas alíquotas de 100 μ L em microtubos resfriados e estas células foram congeladas em nitrogênio líquido e estocadas à -80°C.

Tabela III-2. Soluções utilizadas na preparação de células competentes

Meio de cultura PSI	Tampão I	Tampão II
Peptona 1,6% (m/v)	NaOH 30 mmol.L ⁻¹	MOPS 10mmol.L ⁻¹
Extrato de levedura 1,0% (m/v)	CH ₃ COOH 30 mmol.L ⁻¹	RbCl 7 mmol.L ⁻¹
NaCl 0,5 (m/v)	MgCl ₂ 50 mmol.L ⁻¹	CaCl ₂ 10 mmol.L ⁻¹
KH ₂ PO ₄ 2 mmol.L ⁻¹	RbCl 77 mmol.L ⁻¹	Glicerol 13,5%
Na ₂ PO ₄ 2 mmol.L ⁻¹	CaCl ₂ 100 mmol.L ⁻¹	pH 7,0; autoclavado
pH 7,2; autoclavado	Glicerol 13,5%	
	pH 6,0; autoclavado	

Todas as soluções foram armazenadas à 4°C

3.4. Transformação pelo método de choque térmico

Adicionou-se 1 μL do DNA da proteína de interesse (pET-proteína alvo) aos 100 μL das células competentes BL21(DE3)pLysS, com leve agitação. Esta suspensão foi mantida por 30 minutos em banho de gelo, depois 1,5 minutos a 42°C e, em seguida 1 minuto no gelo novamente. Em seguida, 1 mL de meio SOC (Tabela III-3) foi adicionado e incubou-se por uma hora a 37°C, 200 rpm (Incubadora Tecnal TE-420). Transcorrido este período, 200 μL desta suspensão foram plaqueadas em placas de Petri contendo meio LB sólido (Tabela III-3) e os antibióticos necessários. As placas foram deixadas por 16 horas em estufa (Nova Ética – 410/1 ND) a 37°C para crescimento bacteriano. A partir das colônias que cresceram, foi selecionada uma colônia e uma mini-indução foi realizada. As colônias que apresentaram o inserto, observado por eletroforese em gel das mini-induções, foram replaqueadas e estocadas a 4°C, sendo a manutenção das células feita a cada 15 dias.

Tabela III-3. Soluções utilizadas na transformação bacteriana

Meio de cultura LB ágar	Meio de cultura SOC
Peptona 1,0% (m/v)	Peptona 2,0%
Extrato de levedura 0,5% (m/v)	Extrato de levedura 0,5%
NaCl 1,0% (m/v)	NaCl 10 mmol.L ⁻¹
Ágar 1,5% (m/v)	KCl 1 mol.L ⁻¹
pH 7.0; autoclavado	Glicose 10 mmol.L ⁻¹
	MgCl ₂ 20 mmol.L ⁻¹
	pH 7.0; armazenado a 4°C
	Tudo autoclavado, exceto a glicose adicionada em fluxo laminar

3.5. Expressão proteica em meio LB

O pré-inóculo foi preparado a partir de uma placa de cultura fresca da cepa BL21(DE3)plysS de *E. coli* transformada com o DNA de interesse, em 10 mL de meio LB (Tabela III-1) contendo os antibióticos apropriados (Tabela III-4). A concentração final dos antibióticos no meio é de $\mu\text{g.mL}^{-1}$. O meio foi incubado por 16 horas a 37°C e 200 rpm (Incubadora Tecnal TE-420). Em seguida estes 10 mL foram diluídos em 1000 mL de meio LB contendo os respectivos antibióticos nas mesmas concentrações e esta nova suspensão foi mantida a 37°C, sob agitação (200 rpm), monitorando a absorbância em 600 nm (Espectrofotômetro Bel Photonics SP1105). Atingida a absorbância de aproximadamente 0,8, a expressão protéica foi induzida adicionando-se IPTG (isopropil β -D-1-tiogalactopiranosídeo) na concentração final de 1 mmol.L⁻¹ (Tabela III-4) e a suspensão foi deixada sob agitação (200 rpm), a 37°C por 3 ou 16 horas – dependendo da proteína em questão. Transcorrido este período, a suspensão foi centrifugada a 3500 rpm (Centrífuga Cientc CT-6000R), 4°C por 15 minutos e o pellet foi armazenado a -20°C.

Tabela III-4. Solução estoque dos antibióticos e IPTG

Cloranfenicol	Ampicilina	Canamicina
Cloranfenicol 35 mg.mL ⁻¹ em etanol absoluto	Ampicilina sódica 50 mg.mL ⁻¹ em água/ etanol (1:1)	Sulfato de canamicina 30 mg.mL ⁻¹ em água
Rifamicina	IPTG	
Rifamicina sódica 20 mg.mL ⁻¹ em etanol absoluto	IPTG 1 mol.L ⁻¹ em água MilliQ	
As soluções foram filtradas em filtros de 0,22mm e armazenadas a - 4°C		

3.6. Expressão proteica em Meio Mínimo

O pré-inóculo foi preparado a partir de uma placa de cultura fresca da cepa BL21(DE3)plysS de *E. coli* transformada com o DNA referente à proteína de interesse, em 10 mL de meio LB (Tabela III-1) contendo os antibióticos apropriados (Tabela III-4), incubando por 16 horas a 37°C e 200 rpm (Incubadora Tecnal TE-420). Em seguida estes 10 mL foram diluídos em 1000 mL de meio mínimo (Tabela III-5) contendo os antibióticos apropriados. O meio foi mantido sob agitação (200 rpm) a 37°C, monitorando a absorbância em 600 nm (Espectrofotômetro Bel Photonics SP1105). Atingida a absorbância de aproximadamente 0,8, a expressão proteica foi induzida adicionando-se IPTG na concentração final de 1 mmol.L⁻¹ (Tabela III-4) e a suspensão foi deixada sob agitação (200 rpm), a 37°C por 30 minutos. Em seguida, rifamicina (Tabela III-4) foi adicionada ao meio e este foi deixado sob agitação (200 rpm), a 37°C por 3 horas ou 16 horas – dependendo da proteína em questão. Transcorrido este tempo, a suspensão foi centrifugada a 3500 rpm (Centrífuga Cientific CT-6000R), 4°C por 15 minutos e o pellet foi acondicionado a -20°C.

Tabela III-5. Soluções utilizadas expressão proteica em meio mínimo

Meio M9	Meio Mínimo (1000 mL)
Na ₂ HPO ₄ ·7H ₂ O 239 mmol.L ⁻¹	200 mL de meio M9
KH ₂ PO ₄ 110 mmol.L ⁻¹	2 mL de MgSO ₄ 2 mol.L ⁻¹
NaCl 43 mmol.L ⁻¹	100 mL de CaCl ₂ 1 mol.L ⁻¹
¹⁵ NH ₄ Cl 94 mmol.L ⁻¹	20 mL de glicose 20% (m/m)

3.7. Eletroforese em gel

Todas as amostras coletadas durante a expressão proteica foram analisadas através de eletroforese em gel de acrilamida (SDS-PAGE 15% - Tabela III-7). As amostras de 5-20 μL (5 μL para amostras de indução e 20 μL para amostras de purificação) foram aplicadas no gel de empacotamento e separadas no gel de separação pela aplicação de voltagem em tampão de corrida (Tabela III-9). A eletroforese foi conduzida utilizando o aparato MiniVE Vertical Electrophoresis System (Amershan Biosciences), utilizando 180 V, 50 mA durante aproximadamente 80 minutos. Transcorrido este tempo o gel foi corado com uma solução corante (Tabela III-9) por cerca de 20 minutos sob agitação em mesa agitadora (Tecnal TE 241). Em seguida, os géis foram lavados abundantemente com água corrente e descorado com solução descorante (Tabela III-9) sob agitação em mesa agitadora (Tecnal TE 241) por 20 minutos, trocando a solução descorante de 3 a 4 vezes. Como padrão molecular foi utilizado o kit LMW (Low Molecular Weight Calibration Kit for SDS Electrophoresis – GE Healthcare), contendo proteínas de 14,4 a 97 kDa, diluído em tampão para padrão (Tabela III-6).

As amostras de indução foram preparadas do seguinte modo: alíquotas de 1 mL da indução foram centrifugadas a 14000 rpm (centrífuga Eppendorf 5424) por um minuto e o sobrenadante foi descartado. Ao pellet adicionou-se 50 μL de água e 50 μL de tampão de amostra. As amostras foram misturadas em vortex, sonicadas (sonicador Sonic VCX-750) utilizando amplitude de pulso de 20% por 5 segundos e aquecidas em banho seco (Nova Ética) por 10 minutos a 80°C. As amostras de purificação foram apenas diluídas pela metade com tampão de amostra antes da aplicação no gel.

Tabela III-6. Soluções tampão utilizadas no preparo de amostras

Tampão de amostra	Tampão padrão molecular
Tris 50 mmol.L ⁻¹	Tris 50 mmol.L ⁻¹
DTT 100 mmol.L ⁻¹	DTT 100 mmol.L ⁻¹
SDS 2% (m/v)	SDS 2% (m/v)
Glicerol 10% (m/v)	Glicerol 10% (m/v)
Azul de bromofenol 0,1% (m/v)	Azul de bromofenol 0,1% (m/v)
pH 6,8	pH 6,8

Tabela III-7. Geis de acrilamida

Gel de empacotamento	Gel de separação
Mix 30% 5% (v/v)	Mix 30% 15% (v/v)
Tris pH 6,8 0,19 mol.L ⁻¹	Tris pH 8,8 0,375 mol.L ⁻¹
SDS 0,1% (m/v)	SDS 0,1% (m/v)
Persulfato de amônio (APS) 0,1% (m/v)	Persulfato de amônio (APS) 0,1% (m/v)
Tetrametilenodiamina (TEMED) 0,04% (v/v)	Tetrametilenodiamina (TEMED) 0,04% (v/v)

Tabela III-8. Solução de acrilamida utilizada na preparação dos géis

MIX 30%	
acrilamida	29% (m/v)
bis-acrilamida	1% (m/v)

Tabela III-9. Soluções corante, descorante e tampão de corrida

Tampão de Corrida	Corante	Descorante
Tris-HCl 125mmol.L ⁻¹	Comassie brilliant blue R-250 0,25% (m/v)	CH ₂ COOH 10% (v/v)
SDS 0,5% (m/v)	CH ₂ COOH 10% (v/v)	CH ₃ CH ₂ OH 10% (v/v)
Glicina 1,25 mol.L ⁻¹	CH ₃ CH ₂ OH 30% (v/v)	
pH 8,0		

3.8. Lise celular

O pellet referente a 1 L da expressão proteica foi ressuspendido em 15 mL de tampão de lise (Tabela III-10) e as células foram lisadas por sonicação (sonicador de ponta Sonic VCX-750) utilizando amplitude de pulso de 30% por 100 segundos (pulsando 10 s e descansando 50 s) em banho de gelo. Esta suspensão foi centrifugada a 15000 rpm (Centrífuga Beckman Coulter Allegra X-22R), 4°C por 20 minutos. Após análise de alíquotas do lisado, sobrenadante e do pellet por eletroforese, foi verificado que as proteínas FlgN e XACb0033 encontravam-se no pellet. O pellet foi armazenado a -20°C.

Tabela III-10. Soluções utilizadas na lise celular

Tampão de lise
Tris 100 mmol.L ⁻¹
EDTA 1 mmol.L ⁻¹
NaCl 100 mmol.L ⁻¹
pH 8,0; autoclavado

3.9. Tratamento com uréia (reenovelamento)¹⁰⁵

Ao pellet lisado (1 L de indução) foram adicionados 10 mL de tampão de solubilização (Tabela III-11) e esta suspensão foi mantida sob agitação contínua por 30 minutos em banho de gelo. Transcorrido este tempo, a suspensão foi submetida à diálise em tampão de diálise (Tabela III-11), por 30 minutos a 4°C sob agitação, com cerca de 6 trocas do tampão para posterior purificação. A diálise foi feita utilizando membranas de celulose regenerada com porosidade de 3,5 kDa (Dialysis Tubing – Fisher Scientific) fervida por 15 minutos antes do uso.

¹⁰⁵ Ribeiro Jr, E.; Regis, W.; Tasic, L.; Ramos, C. Fast purification of the Apo form and of a non-binding heme mutant of recombinant sperm whale myoglobin. *Protein Expr Purif* **2003**, *28*, 202-208.

Após a diálise, a suspensão foi centrifugada (Centrífuga Beckman Coulter Allegra X-22R) a 15000 rpm, 4°C por 20 minutos e as amostras do sobrenadante e do pellet foram analisadas por eletroforese em gel.

Tabela III-11. Soluções utilizadas no processo de reenovelamento

Tampão de solubilização	Tampão de diálise pH 8.0
Tampão fosfato de sódio 25mM, pH 8,0 Ureia 8 mol.L ⁻¹	Tampão fosfato de sódio 25mM, pH 8,0

3.10. Purificação das proteínas

As proteínas da fração sobrenadante foram filtradas em filtro de 0,22 µm e purificadas por cromatografia (FPLC ÄKTA Purifier UPC900 – GE Healthcare).

O sobrenadante foi então submetido a cromatografia por filtração em gel utilizando uma coluna SuperdexTM 200 Prep Grade (17 1043-01) pré-equilibrada com o mesmo tampão da diálise. A pureza das proteínas foi verificada por eletroforese em gel e as amostras contendo a proteína de interesse pura foram concentradas. As amostras foram concentradas em filtro AMICON de 3 kDa (Millipore), centrifugando a 4000 rpm, 4°C (Centrífuga Eppendorf 5804R), por 15 minutos, até obter um volume de 500 µL.

Os experimentos foram realizados no FPLC UPC900, utilizando um fluxo de 1,0 mL.min, com pressão máxima de 0,3 MPa na coluna e um detector de UV, monitorando a absorbância em 280 nm.

3.11. Determinação da concentração de proteínas: Método de Edelhoç.^{106,107}

A concentração de proteínas foi determinada por espectroscopia UV-VIS (Espectrofotômetro HP 8453 com cubeta de quartzo de 10 mm de caminho ótico) verificando a absorbância em 280 nm, utilizando o método de Edelhoç com modificações. O coeficiente de absorbância molar dessa proteína foi calculado por:

$$\epsilon_{280} = n_{\text{Trp}} \times \epsilon_{\text{Trp}} + n_{\text{Tyr}} \times \epsilon_{\text{Tyr}} + n_{\text{S-S}} \times \epsilon_{\text{S-S}} \quad (\text{Equação 3.1})$$

Onde ϵ_{280} é o coeficiente de absorbância molar ($\text{M}^{-1}\text{cm}^{-1}$) da proteína de interesse em 280 nm; n_{Trp} – o número de triptofanos da proteína; n_{Tyr} – o número de tirosinas e $n_{\text{S-S}}$ o número de ligações de dissulfeto (cistinas); ϵ_{Trp} , ϵ_{Tyr} e $\epsilon_{\text{S-S}}$ são os coeficientes de absorbância molar do triptofano, da tirosina e das cistinas, respectivamente. Os valores de ϵ_{Trp} , ϵ_{Tyr} e $\epsilon_{\text{S-S}}$ a 280 nm correspondem a $5690 \text{ M}^{-1}\text{cm}^{-1}$, $1280 \text{ M}^{-1}\text{cm}^{-1}$ e $125 \text{ M}^{-1}\text{cm}^{-1}$, respectivamente. Estes valores foram estimado pelo programa ProtParam tool disponível no Expasy.¹⁰⁸ Para as proteínas tem-se

$$\epsilon_{280} (\text{FlgN}) = 6970 \text{ M}^{-1}\text{cm}^{-1}$$

$$\epsilon_{280} (\text{XACb0033}) = 11630 \text{ M}^{-1}\text{cm}^{-1}$$

A concentração foi então calculada através da lei de Beer:

$$A = \epsilon \times l \times C \quad (\text{Equação 3.2})$$

onde A é a absorbância; l o comprimento do caminho ótico (cm), ϵ o coeficiente de absorbância molar ($\text{M}^{-1}.\text{cm}^{-1}$) e C a concentração molar (mol.L).

Para a leitura foi usado como branco tampão diluído na proporção 1:4 (branco/ reagente) com reagente de Edelhoç (Tabela III-12). A proteína também foi diluída com reagente de Edelhoç na proporção 1:4.

¹⁰⁶ Edelhoç, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **1967**, *6*, 1948-1954.

¹⁰⁷ Ramos, C. A spectroscopic-based laboratory experiment for protein conformational studies. *Biochem Mol Biol Edu* **2004**, *32*, 31-34.

¹⁰⁸ Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.; Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **2003**, *31*, 3784-3788.

Tabela III-12. Reagente de Edelhoch

Reagente de Edelhoch
Fosfato de sódio 25 mmol.L ⁻¹
Cloreto de guanidina 7,2 mol.L ⁻¹
pH 6,5

3.12. Método de Bradford¹⁰⁹

Para as proteínas que não apresentam resíduos aromáticos ou cistinas em sua estrutura primária, a quantificação foi feita pelo método de Bradford. Para isso foi construída uma curva de calibração utilizando 10 soluções com diferentes concentrações de BSA (Sigma-Aldrich) de 1 a 10 $\mu\text{g.mL}^{-1}$ no mesmo tampão que a proteína de interesse. Para cada amostra de BSA (800 μL) adicionou-se 200 μL do reagente de Bradford comercial (B6916–Sigma), misturou e aguardou 5 minutos. Então foi efetuada a leitura da absorbância em 595 nm no espectrofotômetro Agilent 8453 em cubetas de plástico. No caso da proteína de interesse foi preparada uma solução com as diluições necessárias e a partir dela foi preparada a amostra para leitura (800 μL da solução + 200 μL de reagente de Bradford).

3.13. Expressão e purificação da proteína FlgK

Iniciou-se com a transformação da cepa da *Escherichia coli* BL21(DE3)pLysS com vetor pET28a pelo método de choque térmico. A expressão desta proteína foi conduzida em meio LB do mesmo modo que para a expressão da FlgN em meio LB, mas, usando-se como agentes seletores os antibióticos cloranfenicol e canamicina, em concentração 35 e 20 $\mu\text{g.mL}$, respectivamente.

As condições para a expressão da FlgK foram iguais da FlgN, porém, esta proteína expressa melhor depois de 16h. Transcorrido o período de indução, todo

¹⁰⁹ Bradford, M. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal Biochem* **1976**, 72, 248-254.

o meio foi centrifugado (4°C, 15000 rpm e 20 minutos) e o material bacteriano separado do meio líquido.

O pellet bacteriano foi ressuspendido em 15 mL de tampão de lise¹¹⁰ (15 mL/ litro de indução) e lisado por sonicação. Esta suspensão foi centrifugada a 15000 rpm, 4°C por 20 minutos.

Como esta proteína tem 5 His no seu N-terminal ela foi purificada utilizando cromatográfica de afinidade (Hi-Trap Chelating GE) e, para isso, foi necessário trocar o tampão da proteína (tampão de lise) porque o EDTA por ser um quelante, retira o Ni²⁺ da coluna. Então a proteína foi dialisada contra tampão fosfato de sódio 100 mmol.L⁻¹ pH 7,4 + NaCl 500 mmol.L⁻¹ (o mesmo tampão usado para equilibrar a coluna) durante um período de 24 horas, com consecutivas trocas do meio da diálise.

As proteínas foram eluídas com um gradiente de tampão fosfato de sódio 100 mmol.L⁻¹ pH 7,4 + NaCl 500 mmol.L⁻¹ + imidazol 500 mmol.L⁻¹ (2% - 100%). O imidazol compete com as histidinas pela ligação ao metal e assim, a proteína se desliga da resina e é eluída. Esta purificação foi realizada novamente com a mesma coluna, obtendo-se a proteína FlgK.

Em seguida a proteína foi dialisada com fosfato de sódio 25 mmol.L⁻¹ pH 8,0 para a retirada de cloreto de sódio e imidazol e assim, estar no mesmo tampão da proteína FlgN para os estudos estruturais. Após a diálise, a proteína FlgK foi concentrada em filtro AMICON (Millipore). A sua concentração foi determinada como sendo 127 μmol.L⁻¹ pelo método de Bradford.¹⁰⁵ A curva de calibração foi construída utilizando como padrão BSA (Figura III-1). O branco e as soluções de BSA foram preparadas em tampão fosfato de sódio 25 mmol.L⁻¹, pH 8.0.

¹¹⁰ Tris-HCl 50 mmol.L⁻¹ + KCl 500 mmol.L⁻¹ + EDTA 10 mmol.L⁻¹, pH 8,0.

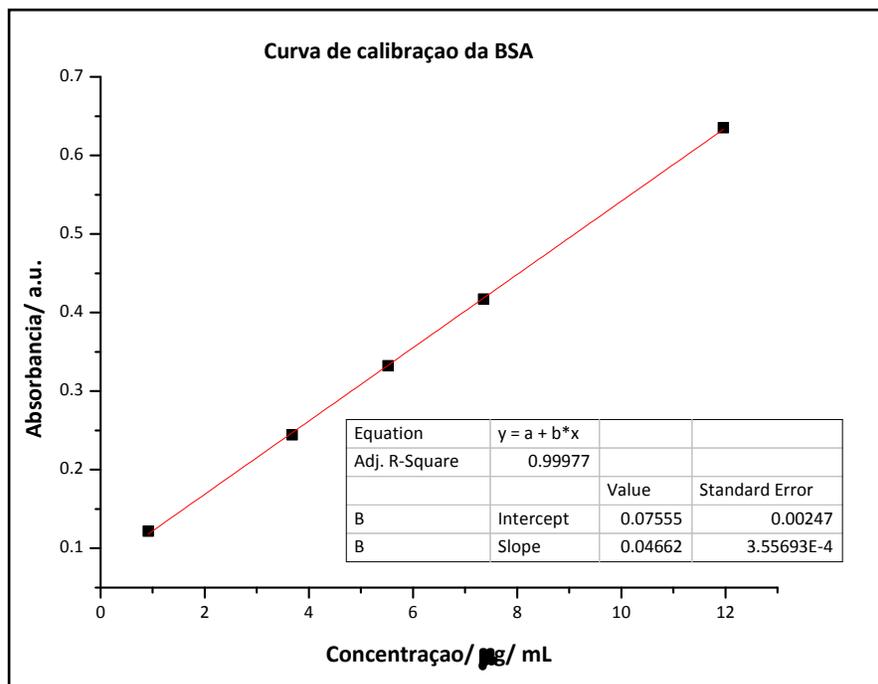


Figura III-1: Curva de calibração de Bradford para quantificação da proteína FlgK (Espectrofotômetro HP 8453, cubeta de acrílico, 10 mm de caminho ótico).

3.14. Expressão e purificação da proteína XACb0032

Iniciou-se com a transformação da cepa da *Escherichia coli* BL21(DE3)pLysS com vetor pET23a pelo método de choque térmico. A expressão desta proteína foi conduzida em meio LB do mesmo modo que para a expressão da XACb0033 em meio LB.

As condições de desenvolvimento bacteriano foram as mesmas utilizadas para a XACb0033, e esta proteína também expressa melhor depois de 16h. Transcorrido o período de indução, todo o meio foi centrifugado (4°C, 15000 rpm e 20 minutos) e o material bacteriano separado do meio líquido.

O pellet (precipitado) armazenado foi lisado com tampão de lise e ultrassom e, em seguida centrifugado (4°C, 15000 rpm e 20 minutos). Ao pellet foram adicionados 4 mL de uréia 8 mol.L⁻¹, 6 mL de tampão fosfato de sódio 25 mmol.L⁻¹ e 20 µL de DNA da XACb0033. Em seguida a proteína foi dialisada com fosfato de

sódio 25 mmol.L⁻¹ pH 8,0 para os estudos estruturais. Após a diálise, a proteína XAb0032 + DNA da XACb0033 foi concentrada em filtro AMICON (Millipore).

Como a XACb0032 não apresenta resíduos de triptofano que são responsáveis pela absorção de radiação em 280 nm, para a quantificação desta proteína utilizou-se o método de Bradford.¹⁰⁵ A concentração da proteína foi determinada como sendo 423 μmol.L⁻¹ pelo método de Bradford. A curva de calibração foi construída utilizando como padrão BSA (Figura III-2). O branco e as soluções de BSA foram preparadas em tampão fosfato de sódio 25 mmol.L⁻¹, pH 8.0.

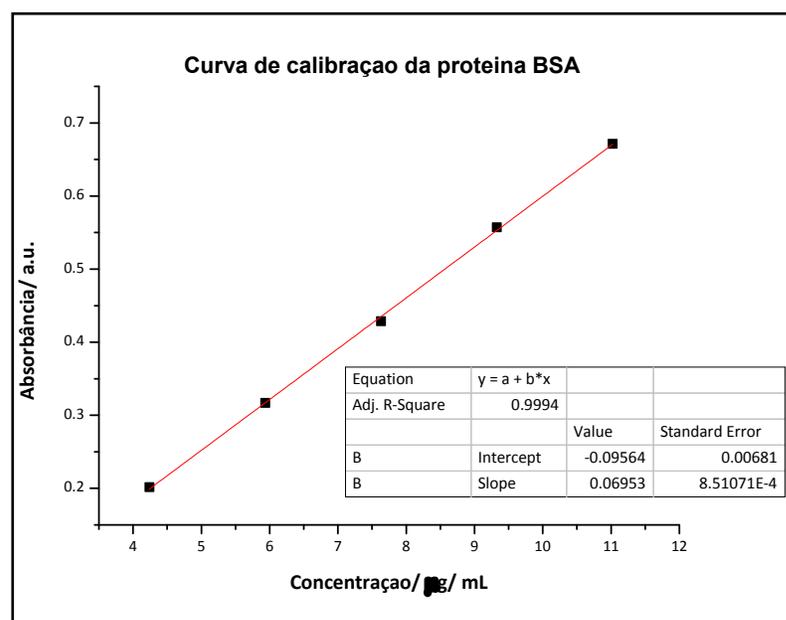


Figura III-2: Curva de calibração de Bradford para quantificação da proteínas (Espectrofotômetro HP 8453, cubeta de acrílico, 10 mm de caminho ótico).

3.15. Espectrometria de Massas (MS) – Identificação das proteínas

Os experimentos de Espectrometria de Massas foram realizados com auxílio do grupo do Prof. Dr. Fabio C. Gozzo (IQ-UNICAMP), utilizando o equipamento Waters nanoAcquity UPLC acoplado a um espectrômetro Waters Synapt HDMS, equipado com fonte de nanoESI.

3.15.1 Digestão e Dessalinização

As amostras de proteínas foram diluídas 1:1 (v/v) em tampão bicarbonato de amônio (100 mmol.L^{-1} , pH 7,8) contendo 10% acetonitrila, e adicionou-se uma solução de tripsina (Sequencing Grade Modified Trypsin, Promega) em ácido acético 50 mmol.L^{-1} em proporção 50:1 m/m (substrato/enzima). As soluções assim resultantes foram agitadas levemente por 16 h a 37°C . A seguir, as soluções foram dessalinizadas em cartuchos de extração em fase sólida (SPE) Oasis HLB (Waters), com capacidade de 1 cc/30 mg, mediante o seguinte procedimento: condicionamento do cartucho com 1 mL acetonitrila (0,1% ácido fórmico, v/v), seguido de 1 mL água (0,1% ácido fórmico, v/v), carregamento da amostra, dessalinização com 1 mL água (0,1% ácido fórmico v/v) e eluição dos peptídeos com 1 mL acetonitrila (0,1% ácido fórmico, v/v). A cada solução resultante foram adicionados $100 \mu\text{L}$ de H_2O e as mesmas foram então concentradas até cerca de $50 \mu\text{L}$ em um concentrador SpeedVac Savant SPD131DDA (Thermo Scientific). As soluções resultantes foram centrifugadas (10 min a $17000 \times g$) e o sobrenadante transferido para *vials* apropriados para análise por LC-MS.

3.15.2 Identificação de peptídeos por cromatografia líquida acoplada a espectrometria de massas (LC-MS)

As análises dos peptídeos presentes nos digestos protéicos por LC-MS foram realizadas em um cromatógrafo Waters nanoAcquity UPLC acoplado a um espectrômetro Waters Synapt HDMS, equipado com fonte de nanoESI. De 2 a $5 \mu\text{L}$ de amostra aquosa foram injetados pelo auto-injetor do sistema UPLC e direcionados a uma pré-coluna (Waters Symmetry C_{18} , $20 \text{ mm} \times 180 \mu\text{m}$ d.i., partículas de $5 \mu\text{m}$), onde foram dessalinizadas por 3 min com fluxo de $5,0 \mu\text{L min}^{-1}$ de 97:3 $\text{H}_2\text{O/MeCN}$ com 0,1% ácido fórmico v/v, sendo então transferidas para a coluna analítica (Waters BEH130 C_{18} , $100 \text{ mm} \times 100 \mu\text{m}$ i.d., partículas de $1,7 \mu\text{m}$)

e eluídas com com fluxo de $1,0 \mu\text{L min}^{-1}$ de acordo com o seguinte gradiente linear (Tabela III-13):

Tabela III-13: Gradiente linear de eluição dos peptídeos

Tempo (min)	% A (H ₂ O com 0,1% ácido fórmico v/v)	% B (MeCN com 0,1% ácido fórmico v/v)
0	97	3
40	70	30
50	20	80
55	20	80
56	97	3
60	97	3

A detecção dos peptídeos foi feita de forma *online* pelo espectrômetro de massas, configurado para operar em modo de aquisição dependente de dados (DDA), contendo uma função de MS (*fullscan* de m/z 200 a 2000), três funções de MS/MS e uma função de padrão externo de calibração (*lockmass*). Como demais parâmetros do espectrômetro de massas estão voltagem do capilar 3,0 kV, voltagem do cone 30 V, temperatura da fonte 100°C, fluxo do gás de nanoESI 0,5 L.h⁻¹, energias de colisão das celas *Trap* e *Transfer* respectivamente em 6 e 4 eV, detector em 1700 V. Os espectros de MS (*fullscan*) e MS/MS (espectro de íons fragmentos por dissociação induzida por colisão) foram adquiridos a uma taxa de 1 espectro s⁻¹. Previamente às análises, o instrumento foi calibrado com oligômeros de ácido fosfórico (solução de H₃PO₄ 0,05% v/v em H₂O/MeCN 50:50 (v/v)) de m/z 90 a 1960. Argônio a uma pressão de $9,7 \cdot 10^{-3}$ mbar foi utilizado como gás de colisão.

3.15.3 Processamento dos dados de LC-MS e busca em banco de dados

As corridas de LC-MS foram processadas pelo software ProteinLynx Global Server v.2.2. (Waters) e analisadas por busca em banco de dados empregando-se o sistema MASCOT v.2.2. (Matrix Science Ltd).¹¹¹ Como parâmetros da busca foram selecionados digestão com tripsina com até 1 sítio de clivagem ignorado, oxidação em resíduos de metionina como modificação variável e tolerância de massa de peptídeos e fragmentos ambos em $\pm 0,1$ Da. As buscas foram feitas em versão modificada do banco de dados NCBI nr contendo as sequências das proteínas *Xac*.¹¹²

3.16. Espectrometria de Massas (Footprinting) e Cross-Linking

Os experimentos de Footprinting foram realizados em quatro etapas.

- 1) Experimento das proteína com o DEPC
- 2) Digestão com tripsina
- 3) Análise e interpretação do dados
- 4) Checagem manual do espectros

3.16.1. Reação com o DEPC

O dietil pirocarbonato (DEPC) foi solubilizado em etanol na concentração final de $7,0 \times 10^{-4}$ mol.dm⁻³. Uma alíquota desta solução foi adicionada a solução contendo proteínas da *Xac* (XACb0033 e FlgN) em uma razão molar 1:10, proteína: DEPC. A reação foi mantida sob temperatura ambiente por 1 hora.

¹¹¹ <http://www.matrixscience.com>

¹¹² http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=ProgSelectionGuide#db

3.16.2 Digestão proteolítica das proteínas

As amostras foram submetidas à digestão enzimática utilizando tripsina (Promega), na razão de 1:50 (m/m) tripsina:proteína, durante 18 h a 37 °C. Após isso, estas amostras foram centrifugadas a 10 krpm durante 10 minutos, diluídas a concentração final de $10 \times 10^{-6} \text{ mol.dm}^{-3}$ e submetidas à análise por MS e MS/MS.

3.16.3 Análise por espectrometria de massa

As análises foram feitas por UPLC-ESI-MS e MS/MS no equipamento Xevo G1 QToF (Waters Corp.). As proteínas e peptídeos foram dessalinizados na coluna *trap*, e então direcionados para coluna de separação (C_{18}) onde foram eluídos aumentando-se a concentração de acetonitrila (ACN). Na análise dos peptídeos, a aquisição dos espectros de MS/MS foi realizada por meio de um experimento do tipo *DDA* (análise dependente dos dados), onde a cada segundo o equipamento adquire um espectro de massa. No caso da presença de espécies multicarregadas, as 3 mais intensas foram selecionadas e então fragmentadas na câmara de colisão, contendo gás argônio, (energia de colisão definida pela m/z e carga do precursor). Os espectros de massa (.raw) foram processados utilizando o programa Proteinlynx (Waters Co.), onde foi feita a deconvolução, deisotopização e correção dos sinais utilizando o ácido fosfórico como referência. Os espectros corrigidos foram salvos no formato .*pk1*.

3.16.4 Checagem manual dos espectros

Esses arquivos foram carregados no programa Mascot (Matrix Science) para a confirmação da identidade das proteínas e determinação do peptídeos modificados pelo DEPC, para tanto foram utilizados os seguintes parâmetros de busca:

Banco de dados: NCBIInr contendo as sequências das proteínas *Xac*¹¹²

Modificação Fixa: nenhuma

Modificações Variáveis: Oxidação da metionina, +72,02 u (DEPC) em His, Tyr, Arg e Lys.

Erro MS: 0,1 Da

Erro MS/MS: 0,1 Da

Ao final, os espectros de fragmentação dos peptídeos encontrados modificados, foram manualmente confirmados.

Para os experimentos de Cross-linking foram testados três agentes de ligação cruzada (ALC): DSS (Disuccinimidyl suberate), EDC (1-ethyl-3-(3-dimethylaminoisopropyl) carbodiimide) e SDAD (Succinimidyl 2-([4,4'-azipentanamido]ethyl)-1,3'-dithiopropionate).

3.17. Fluorescência

As medidas de fluorescência foram realizadas em espectrofluorímetro Cary Eclipse (Varian), utilizando uma cubeta de quartzo com 1 cm de caminho óptico e as amostras de proteínas em concentrações de 2-8 $\mu\text{mol.L}^{-1}$ em tampão fosfato de sódio (pH 8,0, 25 mmol.L^{-1}). As amostras foram excitadas na região de 280 nm, fazendo uma varredura da emissão de 300 a 500 nm, utilizando a abertura da fenda de 5 nm em ambos os casos.

Para as análises de interação proteína-proteína o tampão fosfato de sódio 25 mmol.L^{-1} , pH 8,0 foi usado como branco. Para estes ensaios, as proteínas foram preparadas em concentrações apresentadas na Tabela III-14.

Tabela III-14: Concentrações utilizadas para os ensaios de interações proteína-proteína

Proteína	Ensaio (1:1)	Ensaio (2:1)
FlgN 4 μ M	FlgN 4 μ M + FlgK 4 μ M	FlgN 8 μ M + FlgK 4 μ M
FlgK 4 μ M	-----	-----

Os dados obtidos no espectrofluorímetro foram tratados no Origin 8 ® (OriginLAB Corporation®) e os gráficos obtidos no Origin foram então padronizados.

3.17.1 Interações proteína-ligante

Como a Hsp90 da laranja possui 6 triptofanos em sua sequência primária, foi analisado se a fluorescência de emissão destes resíduos é modificada. Todas as análises foram realizadas na presença dos ligantes em espectrofluorímetro VARIAN CARY ECLIPSE com cubeta de quartzo de 10 mm de caminho ótico em tampão fosfato de sódio 100 mmol.L⁻¹. As concentrações e o comprimento de onda de excitação e emissão máxima dos ligantes e da Hsp90 estão mostrados na Tabela III-15. A proteína foi titulada com adições de 1 μ L (Tabela III-15) e as amostras foram excitadas na região de 280 nm, fazendo uma varredura da emissão de 300 a 500 nm, utilizando a abertura da fenda de 5 nm em todos os casos.

Tabela III-15. Concentração e comprimento de onda de excitação e emissão máximos em ensaios de interação ATP, ADP, ATP γ S e geldanamicina

Proteína e ligantes	Concentração	λ_{ex} (nm)	λ_{em} (nm)	Concentração usada para titulação	Volume adicionado (incrementos de 1 μ L)	Emissão Máxima (nm) Proteína + Ligante
Hsp90 da laranja	2,67 μ mol.L ⁻¹	277	341	2,67 μ mol.L ⁻¹ (2,21 μ mol.L ⁻¹ para geldanamicina)		
geldanamicina	1,78 mmol.L ⁻¹	364	515	1,78 mmol.L ⁻¹	30	340
ADP	13,00 mmol.L ⁻¹	290	390	80,00 mmol.L ⁻¹	25	340
ATP	13,00 mmol.L ⁻¹	290	388	80,00 mmol.L ⁻¹	15	340
ATP γ S	7,30 mmol.L ⁻¹	287	375	80,00 mmol.L ⁻¹	30	341

Todos os ensaios procederam do seguinte modo: a proteína Hsp90 da laranja 2,67 μ mol L⁻¹ (2,21 μ mol.L⁻¹ na titulação com geldanamicina) em tampão tris-HCl (10 mmol.L⁻¹ pH 7,4) foi titulada com os nucleotídeos (água destilada) e com o ligante geldanamicina (DMSO). Neste ensaio foi descontado como branco tampão tris-HCl (10 mmol.L⁻¹ pH 7,4) para a titulação com os nucleotídeos e DMSO para a titulação com geldanamicina. Para o ATP γ S e geldanamicina foram adicionadas 30 alíquotas de 1 μ L de uma solução estoque de ATP γ S 80,00 mmol.L⁻¹ e geldanamicina 1,78 mmol.L⁻¹. Para o ATP e ADP foram adicionadas 15 e 25 alíquotas de 1 μ L de uma solução estoque de ATP 80,00 mmol.L⁻¹ e ADP 80,00 mmol.L⁻¹, respectivamente. Também verificada a emissão de fluorescência máxima dos nucleotídeos, geldanamicina e da proteína Hsp90 da laranja.

O cálculo das constantes de dissociação (Kd) para os compostos foi realizada através de Regressão Não Linear de Mínimos Quadrados, utilizando o software *Mathematica*, versão 6.0.0, Wolfram Research, Inc,¹¹³ através do pacote

¹¹³ Mathematica, versão 6.0.0, Wolfram Research, Inc.

NonlinearRegression, de modo que foi utilizada uma planilha contendo a concentração final do ligante após adição à solução de Hsp90 da laranja, e a respectiva supressão de fluorescência.

3.18. Dicroísmo Circular (CD)

Os espectros de CD foram medidos utilizando-se um espectropolarímetro JASCO J-720 (Tóquio, Japão), cubeta de quartzo de 1 mm de caminho óptico e proteínas diluídas até a condição na qual fornecessem sinal a 222 nm necessariamente entre -10 e -40 miligraus (intervalo no qual a relação sinal/ruído é baixa). Vale lembrar que as medidas foram feitas até atingir a voltagem máxima de 700 V na fotomultiplicadora. Todas as medidas foram feitas sob fluxo de nitrogênio de 10 L.min.⁻¹ Os resultados apresentam a média de 16 espectros no intervalo de 200-260 nm lidos a uma velocidade de 50 nm.min.⁻¹, 4 s de resposta, tendo como branco tampão fosfato de sódio (pH 8, 25 mmol.L⁻¹) acumulado 8 espectros.

O programa utilizado para o registro dos dados foi o Spectra Manager® (JASCO), sendo que para o tratamento e confecção do gráfico dos dados obtidos, utilizou-se os programas Origin 8® (OriginLAB Corporation®). A temperatura da cela foi mantida constante a 25°C através de um sistema interno de controle (Peltier Type Control System PFD 425S™, JASCO®) e controlador externo de temperatura NESLAB RTE Series (NESLAB®). Os valores obtidos na leitura de CD (miligraus) foram convertidos para elipticidade molar residual ($[\theta]_{EMR}$) que é definida pela seguinte equação:

$$[\theta]_{EMR} = (\theta \times 100 \times MM) / (C \times l \times n) \quad (\text{Equação 3.3})$$

onde θ é a elipticidade em graus, MM é a massa molecular da proteína (Da), C é a concentração da proteína (mg.mL⁻¹), l é o comprimento do caminho óptico (cm) e n é o número de resíduos de aminoácidos da respectiva proteína.

Os parâmetros usados para medida de CD das proteínas antes e após interação com as proteínas parceiras foram:

Sensitivity: 100 mdeg

Start: 260 nm
End: 190 nm
Data Pitch: 0,5 nm
Scanning mode: continuous
Scanning Speed: 50 nm.min.⁻¹
Response: 4 s
Band Width: 1 nm
Accumulation: 16 para a amostra e 8 para o branco

Para o cálculo das porcentagens de estruturas secundárias, utilizou-se o programa CDNN¹¹⁴ Deconvolution® Versão 2.1, com a base NNET_23 que conta com 23 espectros para a análise do espectro da proteína de interesse, por combinação linear dos espectros da base.

Os ensaios de desenovelamento térmico das proteínas foram realizados variando a temperatura de 20 – 90°C com uma rampa de aquecimento de 30°C.h⁻¹. Em seguida, o renovelamento térmico foi realizado de 90 – 20°C com uma rampa de resfriamento de 30°C.h⁻¹. Os parâmetros usados para o desenovelamento térmico ida (unfolding) e volta (folding) foram:

Wavelength: 222 nm
Start: 20°C (unfolding) e 90°C (folding)
End: 90°C (unfolding) e 20°C (folding)
Data Pitch: 0,5°C
Delay Time: 60 s
Temperature Slope: 30°C.h⁻¹
Sensitivity: standard (100 mdeg)
Response: 8s
Band width: 1nm

Adicionalmente um desenovelamento térmico foi realizado variando a temperatura de 20 - 90°C e obtendo-se, a cada 5°C, um espectro de CD (mesmo

¹¹⁴ Böhm, G.; Muhr, R.; Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng* **1992**, *5*, 191-195.

parâmetros para o espectro de CD à 25°C. Os gráficos foram confeccionados utilizando-se o programa Origin 8® (OriginLAB Corporation®). Neste programa, o valor do branco foi considerado no cálculo.

Os cálculos dos complexos foram efetuados utilizando um valor aproximado de massa molar como sendo a somatória de ambas massas molares das proteínas [63325(FlgK) + 12120(FlgN)]. Considerou-se como número de aminoácidos no caso do complexo 1:1 [624(FlgK) + 110(FlgN)] a somatória do número de aminoácidos das proteínas e no caso do complexo 2:1 [624+ (2x110)]. A concentração da proteína usada no cálculo foi da proteína FlgK para as interações FlgN/FlgK. As amostras usadas para a análise de CD foram às mesmas utilizadas para a análise de fluorescência (Tabela III-14).

3.19. Ressonância Magnética Nuclear (NMR): ^1H e HSQC $\{^{15}\text{N}/^1\text{H}\}$

As amostras de proteína em tampão fosfato (25 mmol.L⁻¹) foram preparadas com adição de D₂O (5%, v/v, Cambridge Isotope Laboratories) em tubo Shigueti (350 µL). Os experimentos de RMN de ^1H e $\{^{15}\text{N}, ^1\text{H}\}$ HSQC foram adquiridos no equipamento Varian INOVA 600 MHz utilizando a criosonda HCN a 298 K no LNBio (Laboratório Nacional de Biociências no CNPEM). Espectro de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC foi adquirido com 8 scans e o espectro de ^1H NMR com 32 scans. A água residual HDO foi suprimida utilizando a sequência de pulso Saturation e referenciada em 4,80 ppm. Os espectros foram processados utilizando o software VNMRJ da Varian. Para as proteínas marcadas com ^{15}N foram adquiridos os espectros de ^1H NMR e o mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC (Tabela III-16 e Tabela III-17) no espectrômetro Varian 600 MHz no LNBio (Laboratório Nacional de Biociências no CNPEM).

Tabela III-16. Condições testadas em ensaios de RMN para a FigN

Anexo	Tampão	Concentração ($\mu\text{mol L}^{-1}$)	Tratamento	Resultado
—	Fosfato de sódio 100 mmol.L^{-1} , pH 7,8	258	Quantidade equimolar de arginina e ácido glutâmico	Não houve mudança
4 e 5	Fosfato de sódio 5 mmol.L^{-1} , pH 8,0	21	25°C	Não houve mudança
6 e 7	Fosfato de sódio 5 mmol.L^{-1} , pH 8,0	21	35°C	Não houve mudança

Tabela III-17. Condições testadas em ensaios de RMN para a XACb0033

Anexos	Tampão	Concentração (μM)	Tratamento	Resultado
13 e 14	Fosfato de sódio 10 mmol.L^{-1} , pH 8,0	50	iodoacetamida e DTT	Não houve alteração
15	Fosfato de sódio 10 mmol.L^{-1} , pH 8,0	50	Mesma proteína do ensaio anterior, com adição de glicerol	Não houve alteração
16 e 17	Fosfato de sódio 30 mmol.L^{-1} , pH 5,0	50	Mudança de pH	Não houve alteração
—	Fosfato de sódio 100 mmol.L^{-1} , pH 7,8	132	Quantidade equimolar de arginina e ácido glutâmico	Não houve alteração
18 e 19	Fosfato de sódio 25 mmol.L^{-1} , pH 8,0	11	25°C	Não houve alteração
20 e 21	Fosfato de sódio 25 mmol.L^{-1} , pH 8,0	11	40°C	Não houve alteração

Para a preparação dos complexos proteicos (FlgN/FlgK e XACb0033/XACb0032) as duas proteínas foram purificadas separadamente. Concentrou-se elas até um volume de 2 mL, mediu-se a concentração pelo método de Bradford e em seguida, preparou-se os complexos na razão molar 2:1 (FlgN/FlgK) e (XACb0033/XACb0032). Por último estas proteínas foram concentradas em filtro AMICON de 3 kDa até uma concentração de $255 \mu\text{mol.L}^{-1}$ e $398 \mu\text{mol.L}^{-1}$ para os complexos FlgN/FlgK e XACb0033/XACb0032, respectivamente.

3.19.1 Ressonância Magnética Nuclear (NMR): STD e HETCOR

As análises de STD foram realizadas com ambas as proteínas de choque térmico Hsp90 da laranja e Hsp100 da cana de açúcar que foram clonadas e purificadas pelo grupo de Prof. Carlos H. I. Ramos.

Os ensaios de STD foram realizados com a proteína Hsp90 da laranja, ADP, ATP e ATP γ S nas concentrações $30 \mu\text{mol.L}^{-1}$ (em tampão Tris-HCl 10 mmol.L^{-1} , pH 8,0), 80 mmol.L^{-1} , 80 mmol.L^{-1} e 80 mmol.L^{-1} (todos dissolvidos em água destilada), respectivamente.

Nos experimentos de STD, as amostras foram preparadas na razão molar proteína/ligante correspondente de 1:100, em $600 \mu\text{L}$ de D_2O . Os experimentos foram realizados no espectrômetro Varian INOVA-500, equipado com sonda penta, operando na frequência do hidrogênio em 499,89 MHz, a 298 K. As medidas de STD empregaram a supressão da água residual HDO utilizando a seqüência de pulsos Saturation e como referencia interna 4,70 ppm (HDO). A frequência de saturação do pulso seletivo foi aplicada em 0 ppm para on-resonance e em 30 ppm para off-resonance. Um “train” de pulsos com formato gaussianos de 50 ms foram aplicados. Os espectros foram subtraídos internamente via ciclos de fase de cada scan. Os espectros de STD foram obtidos com 1024 scans. Todos os espectros foram processados pelo software Varian (VnmrJ). As soluções estoques para estes estudos foram: proteína Hsp100 da

cana de açúcar $120 \mu\text{mol.L}^{-1}$ em tampão tris-HCl 10 mmol.L^{-1} pH 8,0, Hsp90 da laranja $30 \mu\text{mol.L}^{-1}$ em tampão tris-HCl 10 mmol.L^{-1} pH 7,4, ATP, ADP, e ATP γ -S 80 mmol.L^{-1} em água destilada. Também foram preparadas amostras de nucleotídeos ATP, ADP, e ATP γ -S na concentração de 1 mmol.L^{-1} em 99,9 % de D₂O como referência. Para os ensaios de HETCOR as amostras foram preparadas em nanossonda (40 μL).

Para as análises de $\{^1\text{H}, ^{13}\text{C}\}$ HETCOR as amostras foram preparadas dissolvendo 5,0 mg de ATP γ S em 40 μL de D₂O, usado como controle, e dissolvendo 2,5 mg de ATP γ S em 20 μL de Hsp100 da cana de açúcar (solução estoque: 120 μL em tampão Tris-HCl 10 mmol.L^{-1} , pH 8,0) em 20 μL de D₂O. Os espectros de HETCOR foram adquiridos com 1024 scans. Nesta análise, espectrometro Varian Inova-500 equipado com nanossonda foi usado, operando na frequência do carbono 125,71 MHz a 298 K.

3.20. Bioinformática

Como ambas proteínas XACb0033 e FlgN não apresentam homologia com nenhum dos modelos de alta resolução determinados por técnicas experimentais, a estrutura atômica das proteínas foi predita por métodos computacionais usando o I-TASSER¹¹⁵ e QUARK.¹¹⁶ As estruturas protéicas obtidas pelo I-TASSER e QUARK auxiliaram propor modelos de estruturas 3D para as proteínas baseadas em restrições obtidas via análises de footprinting, SASA¹¹⁷ (do inglês Solvent

¹¹⁵ (a) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **2010**, *5*, 725-738. (b) Zhang, Y. I-TASSER server for protein 3 D structure prediction. *BMC Bioinform* **2008**, *9*, 40-47. (c) Zhang, Y. I-TASSER: Fully automated *protein structure* prediction in CASP8. *Proteins Struct Funct Bioinf* **2009**, *77*, 100-113.

¹¹⁶ (a) Xu, D.; Zhang, Y. QUARK Ab Intio Protein Structure Prediction I: Methodology developments (in preparation). (b) Xu, D.; Zhang, Y. QUARK Ab Intio Protein Structure Prediction II: Results of benchmark and blind tests (in preparation).

¹¹⁷ (a) Bernadó, P., Blackledge, M.; Sancho, J. Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophys J* **2006**, *91*, 4536-4543. (b) Estrada, J., Bernadó, P., Blackledge, M.; Sancho, J. ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. *BMC Bioinformatics* **2009**, *10*, 104.

Accessible Surface Area) e RMSD (desvio da media quadrática, do inglês Root Mean Square Deviation). Os cálculos de acessibilidade de água SASA foram calculados através do programa ProtSA,¹¹⁸ os dados de RMSD foram gerados pelo programa VMD 1.9.1beta1 e as estruturas para visualização e análises gráficas foi utilizado o programa PyMOL (The PyMOL Molecular Graphics System, verso 1.3, Schrödinger, LCC).

¹¹⁸ <http://webapps.bifi.es/ProtSA/protSA.html>

Capítulo IV - Resultados e Discussões

4. Resultados e Discussões

As proteínas estudadas, XACb0033, XACb0032, FlgN e FlgK foram clonadas e suas condições de expressão foram padronizadas anteriormente.^{100,31}

O projeto teve início com a preparação de células competentes e posterior transformação destas com o DNA recombinante das proteínas alvo pelo método de choque térmico.¹¹⁹ Após confirmação da inserção do DNA de interesse nas células pela transformação, foi conduzida a expressão em meio LB ou meio mínimo no caso das XACb0033 e FlgN. O pellet bacteriano foi lisado com tampão de lise e ultrassom. A mistura foi novamente centrifugada, e a parte insolúvel, denominada corpo de inclusão, foi dissolvido em uréia 8 mol.L⁻¹ e dialisada durante um período de 24 horas, com consecutivas trocas do meio da diálise. Em seguida, o sobrenadante foi separado dos demais materiais precipitados por centrifugação. Após o sobrenadante ser filtrado em filtro Milipore 0,22 µm, este foi purificado em coluna cromatográfica de exclusão. Nesta purificação foi utilizado tampão fosfato de sódio (25 mmol.L⁻¹ pH 8,0) como fase móvel. Amostras coletadas durante todo o processo de expressão e purificação foram analisadas por eletroforese SDS-PAGE 15%.

Um esquema geral da expressão das proteínas em larga escala em meio LB e meio mínimo está mostrado no fluxograma da Figura IV-1.

¹¹⁹ (a) Mandel, M.; Higa, A. Calcium-dependent bacteriophage DNA infection. *J Mol Biol* **1970**, *53*, 159-162. (b) Cohen, S.; Chang, A.; Hsu, L. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA. *Proc Nat Acad Sci USA* **1972**, *69*, 2110-2114.

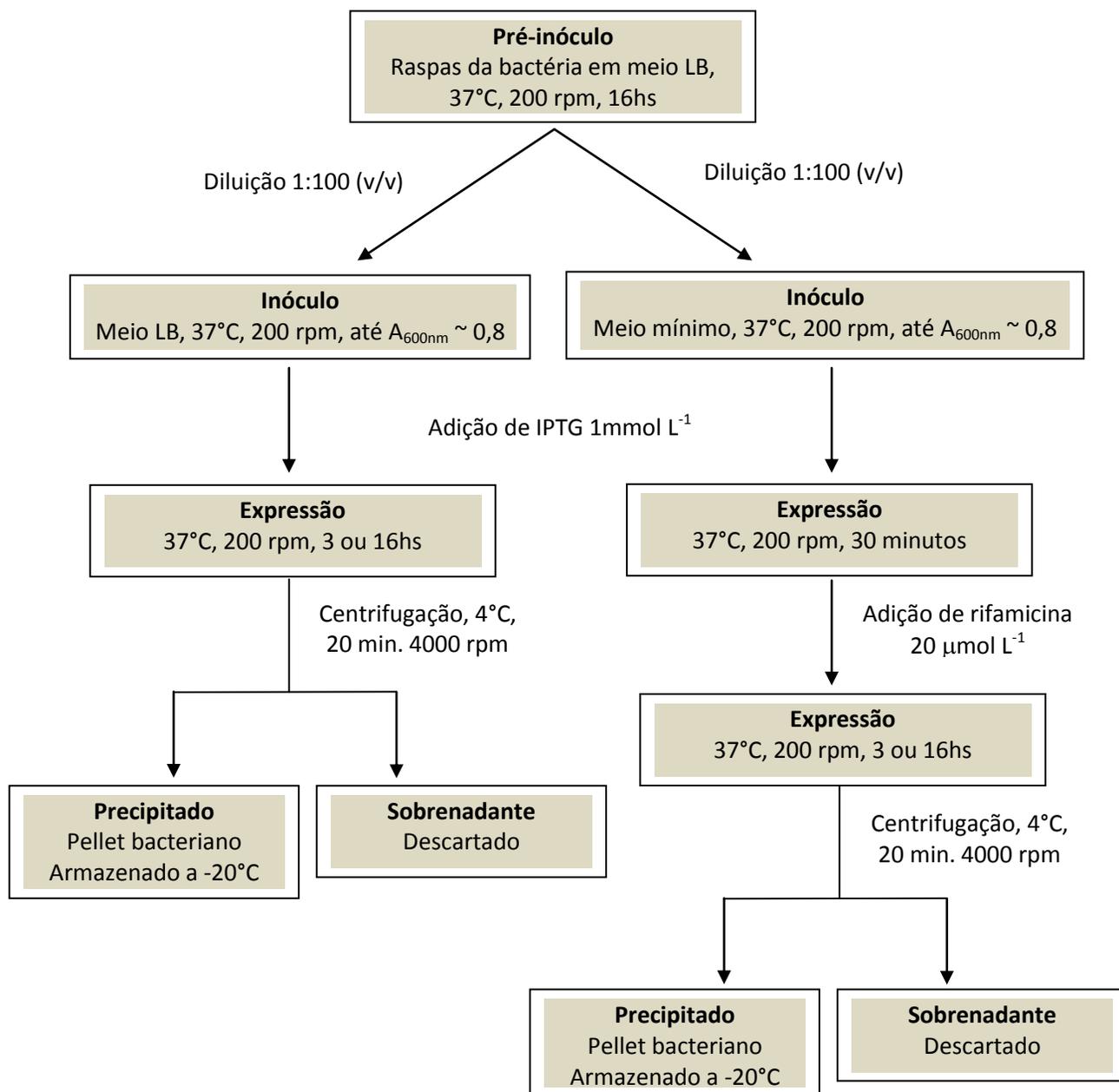


Figura IV-1: Fluxograma geral da expressão das proteínas em larga escala em meio LB e meio mínimo.

Os resultados para as XACb0033 e FlgN serão mostrados a seguir separadamente e os resultados discutidos individualmente.

4.1 A FlgN

A proteína FlgN, é constituída de 110 resíduos de aminoácidos em sua estrutura primária (Figura IV-2). Sua massa molecular é de aproximadamente 12,12 kDa, tem um pI teórico de 6,75 e é uma chaperona flagelar, como estudos relatados anteriormente.³¹



Figura IV-2: Sequencia primária da proteína FlgN. Resíduo de triptofano em destaque (vermelho).

A proteína FlgN expressa muito bem até 3 h mas, após este período ela começa a degradar. O gel da indução e lise esta ilustrado a seguir (Figura IV-3) e, como pode ser observado, a proteína FlgN se encontra majoritariamente no pellet bacteriano.

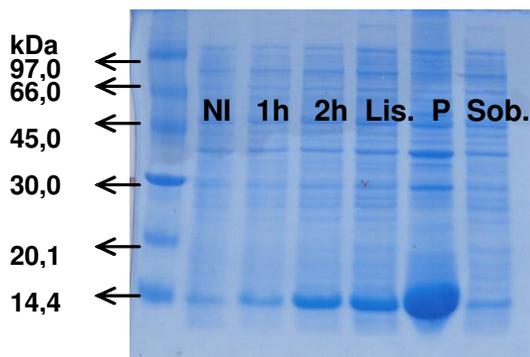


Figura IV-3: Gel de eletroforese SDS-PAGE (15%) da expressão da FlgN, tendo na sequência, da esquerda para a direita: marcador de peso molecular de proteína; amostras não induzidas (NI) e 1h, e 2h após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.)

A FlgN foi purificada em coluna de filtração em gel e amostras coletadas durante todo o processo de purificação foram analisadas por eletroforese SDS-PAGE 15% (Figura IV-4). A análise visual do gel nos fornece informação sobre o estado oligomérico desta proteína, ou seja, ela se apresenta como um monômero.

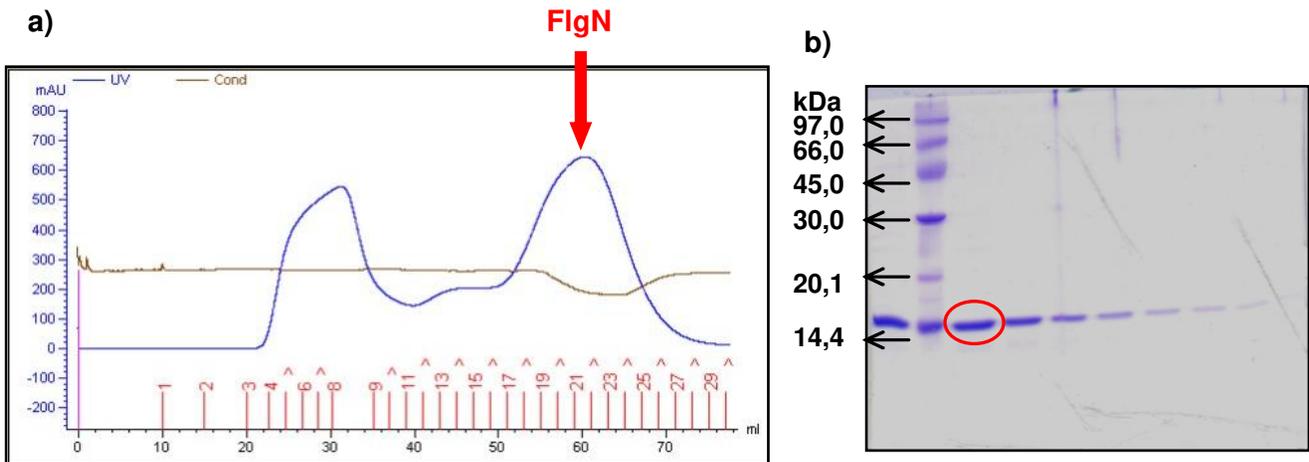


Figura IV-4: a) Cromatograma da purificação do sobrenadante do reenovelamento da FlgN, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína FlgN pura mostrada em destaque (círculo vermelho).

As amostras puras coletadas da coluna foram concentradas em filtros AMICON (Millipore) 3 kDa sob centrifugação a 4°C.

Como a proteína FlgN apresenta um resíduo de Trp em sua sequência primária, sua concentração foi determinada pelo método de Edelhoch¹⁰⁶⁻¹⁰⁷. Seu coeficiente calculado é igual a $6970 \text{ M}^{-1}\text{cm}^{-1}$. A concentração encontrada para a proteína FlgN foi de $78 \mu\text{mol L}^{-1}$.

A partir desta amostra foram realizados estudos estruturais de Dicroísmo Circular (Desenovelamento térmico), Fluorescência de Emissão, Espectrometria de Massas e Ressonância Magnética Nuclear.

A identificação da proteína foi feita empregando-se digestão tríplica, seguida de cromatografia líquida acoplada a espectrometria de massas (LC-MS) para separação e identificação dos peptídeos. A amostra de proteína digerida com

tripsina e dessalinizada em cartuchos de extração em fase sólida (SPE) de fase reversa foi analisada por espectrometria de massas. Dados então processados e submetidos a busca em banco de dados utilizando-se o MASCOT v.2.2. (Matrix Science Ltd.). Foram identificados peptídeos da FlgN, e com um score de **540** com uma cobertura de 68% (Figura IV-7) esta proteína foi identificada com sucesso. Dentre o conjunto de peptídeos identificados como sendo da FlgN, espectros representativos para os cinco com maior *score* são ilustrados nas Figuras IV-5 e IV-6.

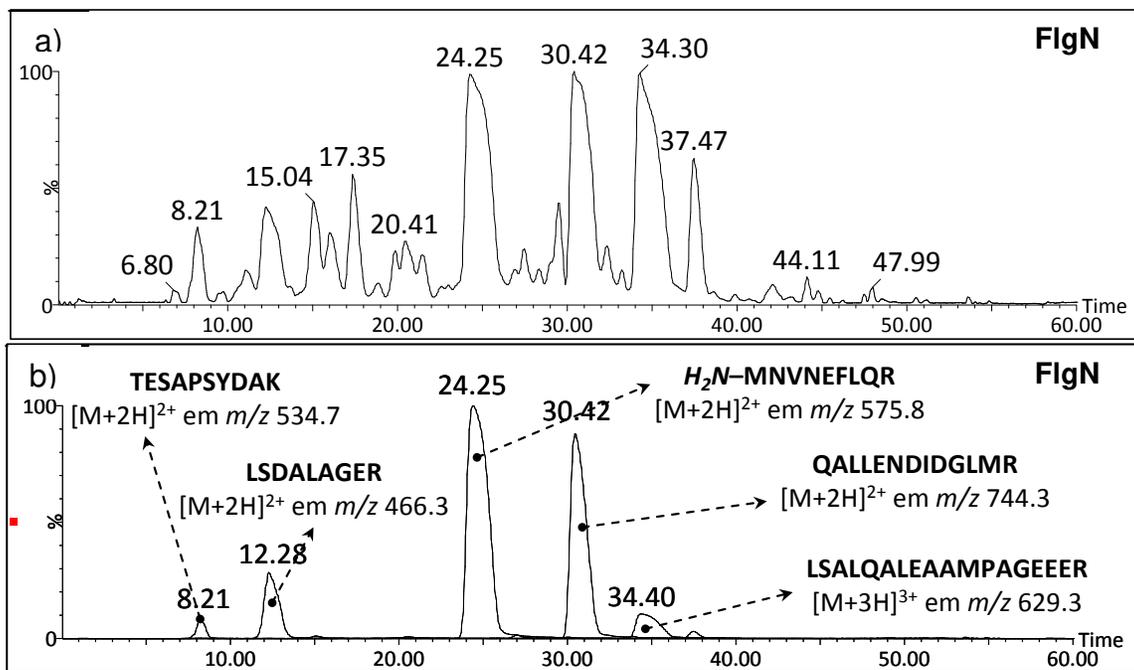


Figura IV-5: a) Cromatograma total (LC-MS) dos peptídeos da digestão trípica da FlgN e b) Cromatogramas extraídos para íons dos peptídeos com maior score no MASCOT oriundos da FlgN.

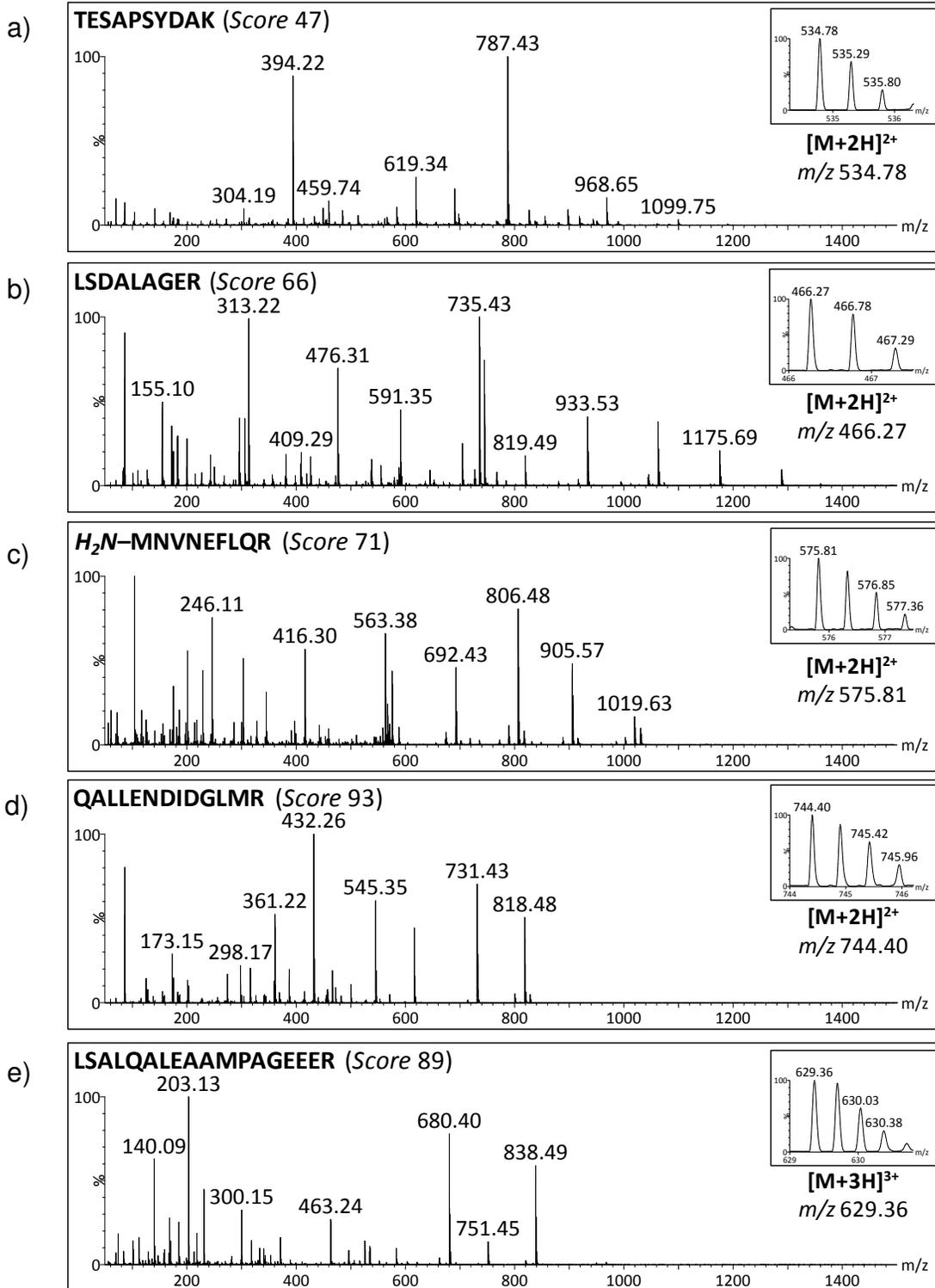


Figura IV-6: Espectros de MS/MS gerados a partir dos cinco peptídeos com “maior score” da proteína FlgN. a) MS/MS do peptídeo TESAPSYDAK. b) MS/MS do peptídeo LSDALAGER. c) MS/MS do peptídeo H₂N-MNVNEFLQR. d) MS/MS do peptídeo QALLENDIDGLMR. e) MS/MS do peptídeo LSALQALEAAMPAGEEER.

Nominal mass (M_r): **12114**; Calculated pI value: **6.74**
Matched peptides shown in **Bold Red**

MNVNEFLQRLSDALAGERQALLENIDGLMRHTQDKLS
ALQALEAAMPAGEEERLRELAEANRANGALLARRRREV
NWALRHLGRTESAPSYDAKGQSSVLRGGRSLAVA

Figura IV-7: Sequência primária da proteína FlgN. Os peptídeos encontrados para a FlgN (MS/MS) estão mostrados em destaque (vermelho), sendo obtidos 68% de cobertura da sequência.

Também foi obtido um espectro da proteína intacta para confirmar a massa exata desta proteína. Através da deconvolução foi obtida uma massa molar de 12120,7 Da (Figura IV-8) e comprovando que a proteína FlgN se comporta como um monômero.

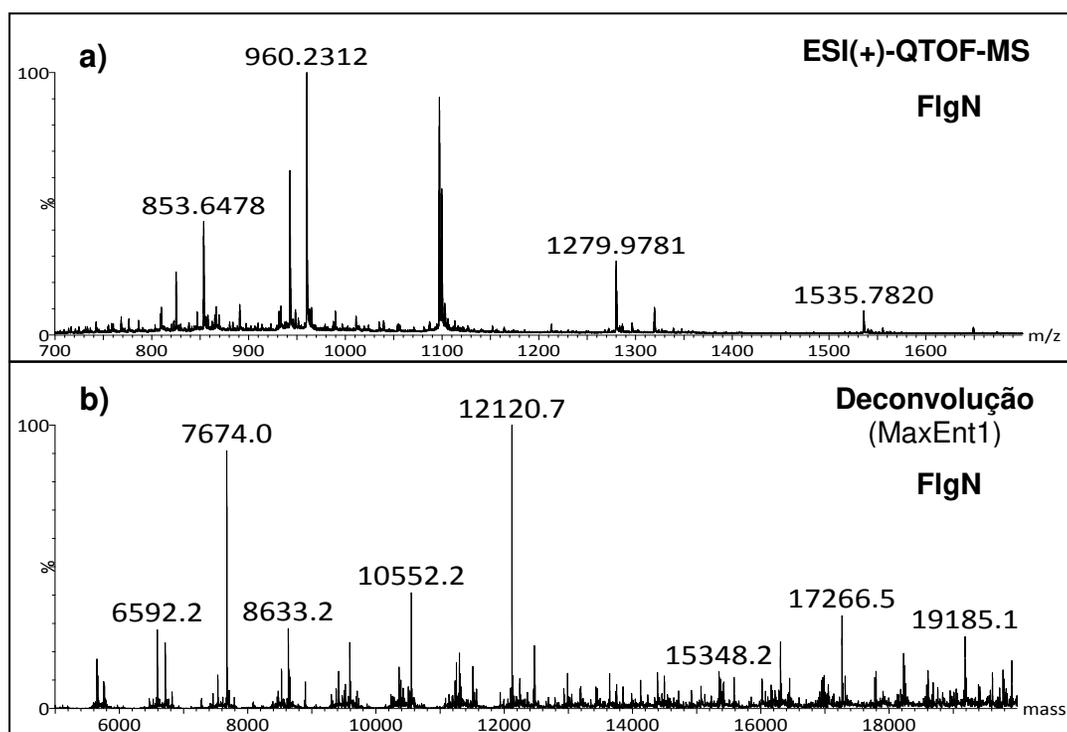


Figura IV-8: Espectros de massas da proteína FlgN. a) espectro de massa da proteína intacta e b) após processamento dos dados.

Esta proteína já teve o seu estado oligomérico determinado em trabalho anterior utilizando uma coluna de filtração em gel analítica calibrada e os resultados indicaram se tratar de uma proteína monomérica.⁶¹

Identificada a proteína FlgN, foram realizados experimentos de fluorescência de emissão, CD e NMR. Primeiramente serão apresentados experimentos de fluorescência de emissão, onde se observou um máximo em 349 nm (Figura IV-9), indicando resíduo de triptofano exposto ao solvente.⁷¹

A FlgN foi então submetida à análise por Dicroísmo Circular para avaliação da estrutura secundária.¹²⁰ Os dados obtidos por CD foram tratados no programa CDNN¹²¹ para cálculo da estrutura secundária da proteína FlgN e os resultados obtidos estão sumarizados na Figura IV-10. A proteína foi diluída para estas análises ($39 \mu\text{mol.L}^{-1}$).

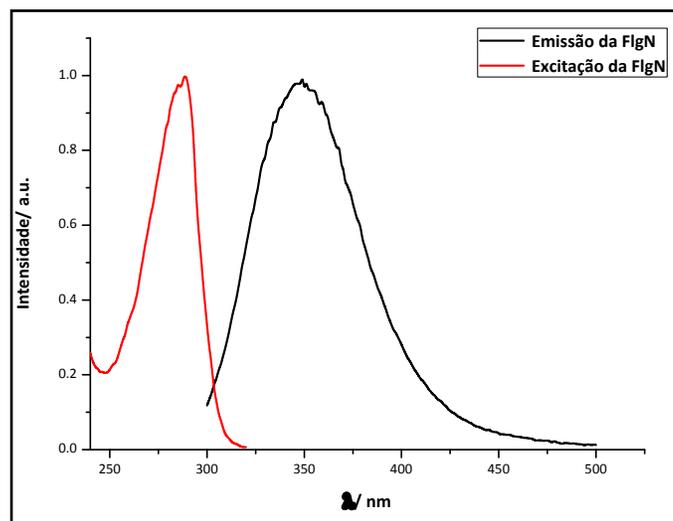


Figura IV-9: Espectros de fluorescência de excitação e de emissão da proteína FlgN com uma concentração de $78 \mu\text{mol.L}^{-1}$ em tampão fosfato de sódio (25mmol.L^{-1} , pH 8,0).

¹²⁰ Deléage, G.; Geourjon, C. An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum. *Comput Appl Biosci* **1993**, *9*, 197.

¹²¹ CDNN 2.1: Circular Dichroism Deconvolution, Neutral Networks 1999.

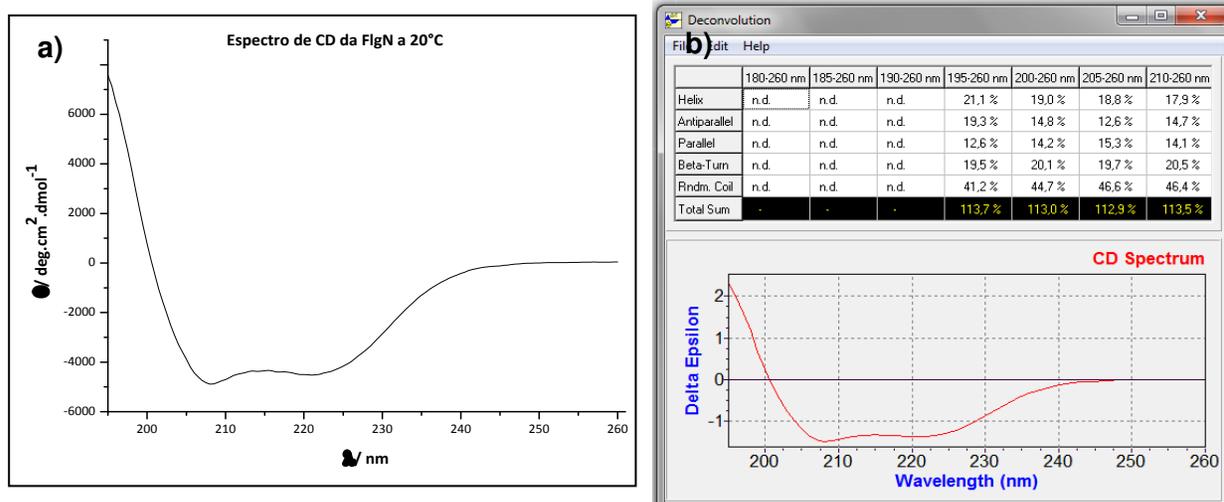


Figura IV-10: Ilustração dos resultados obtidos em análise de CD para a proteína FgN: a) Espectro de CD, b) Cálculo da estrutura secundária da proteína FgN obtida pelo programa CDNN a partir dos dados de CD. Abaixo, gráfico gerado pelo CDNN.

A partir dos dados de CD foi verificado que a proteína FgN se encontra parcialmente enovelada, ou ainda pode apresentar um equilíbrio entre as conformações, ou seja, uma troca conformacional. Além disso, esta proteína apresenta 21% conteúdo de α -hélice e este baixo conteúdo é coerente com uma baixa concentração, pois, o sinal de CD está relacionado com a concentração. A partir desta mesma proteína foram realizados ensaios de desenovelamento térmico e os resultados estão sumarizados na Figura IV-11.

Através destes ensaios foi confirmado que a proteína perde parte da sua estrutura secundária com o aumento da temperatura. Os gráficos sigmóidais (a, b e c) foram obtidos por “Ajuste linear sigmoidal de Boltzmann” no software OriginPro 8.0. A partir do gráfico (a) foi possível calcular um valor de T_m de 45°C, ou seja, é a temperatura em que a proteína perde 50% da sua estrutura secundária. Comparando-se os dois gráficos de desenovelamento e renovelamento (c) é observado que a proteína volta ao seu estado inicial após resfriamento por isso, esta proteína apresenta comportamento reversível.

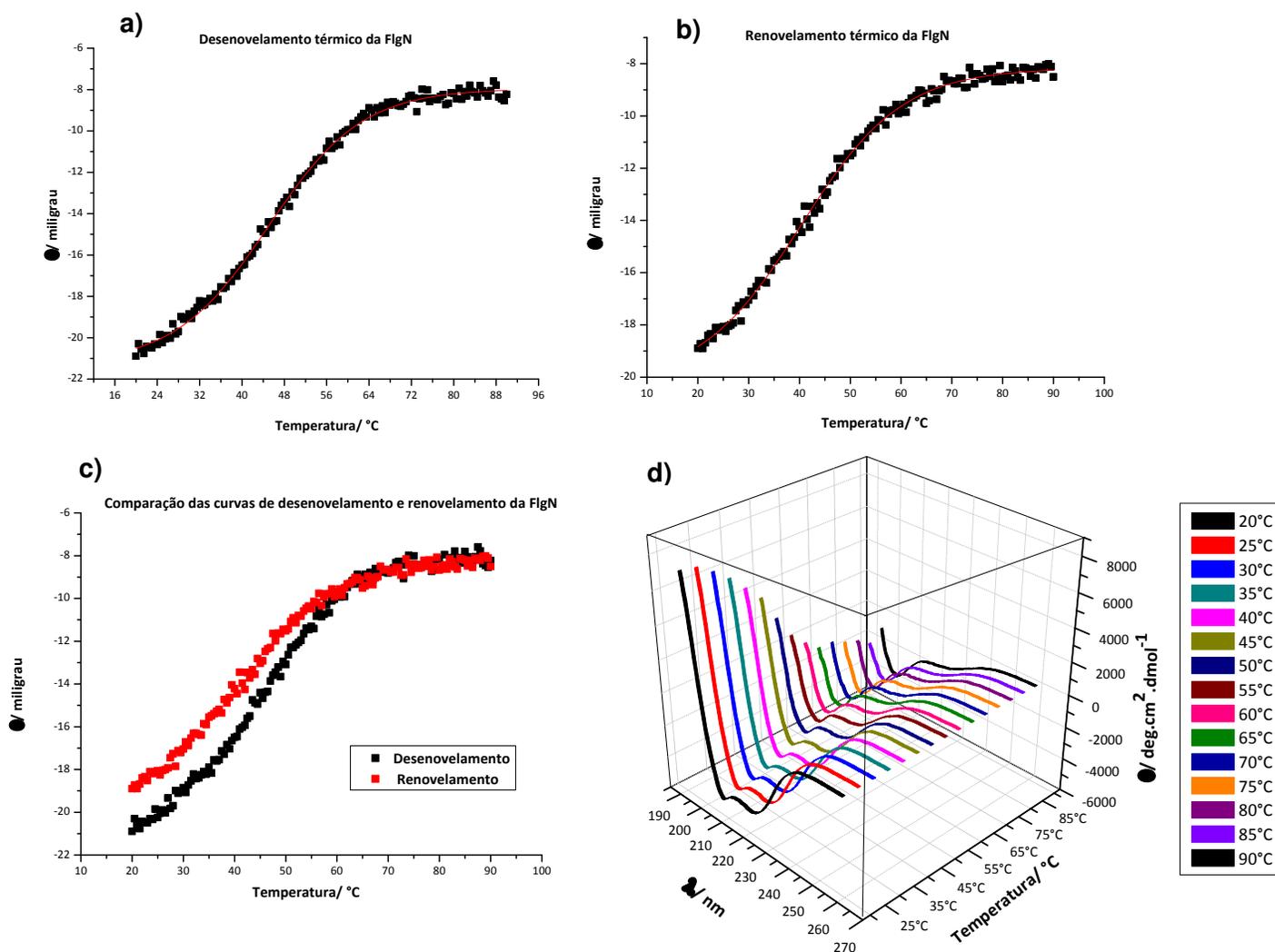


Figura IV-11: a) Desenovelamento térmico 20 – 90°C acompanhado à 222 nm, b) Renovamento térmico 90 – 20°C acompanhado à 222 nm, c) Espectros de desenovelamento e renovamento térmico da FlgN e d) Desenovelamento térmico (5°C de incremento).

A partir de outro lote da proteína FlgN purificada em uma concentração de $49 \mu\text{mol.L}^{-1}$ em tampão fosfato de sódio (100mmol.L^{-1} , pH 8,0) foi obtido o primeiro espectro de ^1H NMR (Figura IV-12).

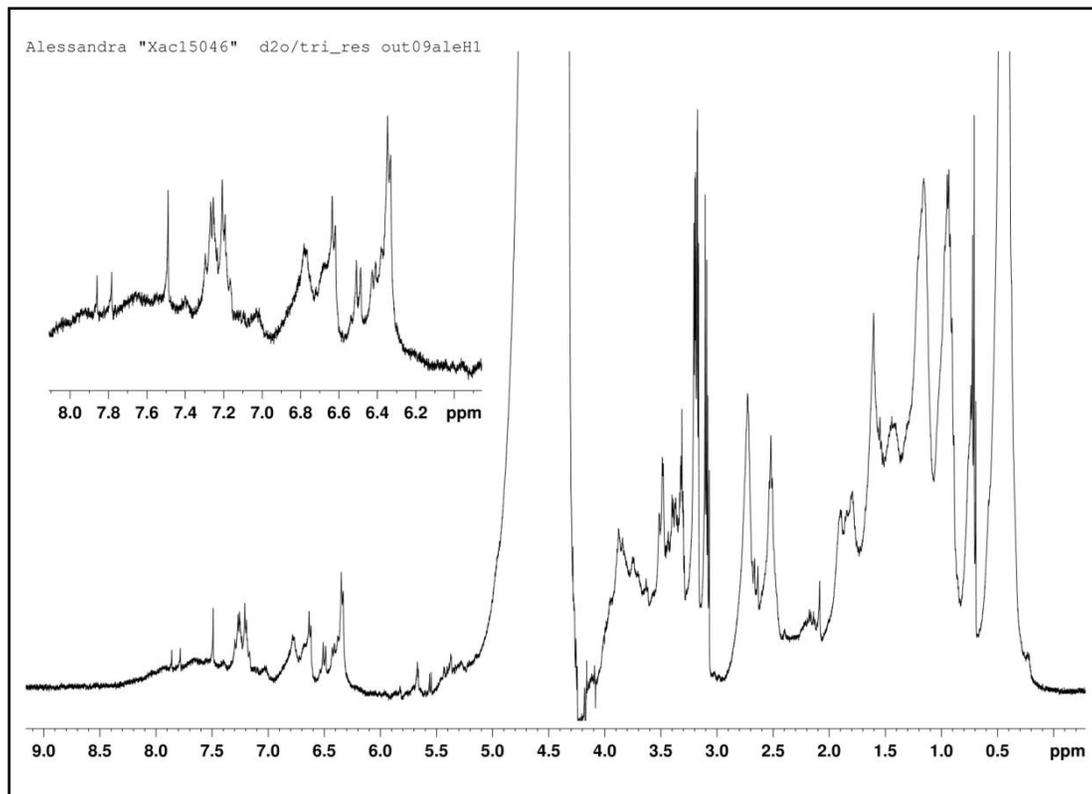


Figura IV-12: Espectro de ^1H NMR da FlgN ($49 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1}) com adição de D_2O (5% v/v) obtido no equipamento INOVA 500 MHz (499,98 MHz). Expansão da região de 6,0 à 8,0 ppm mostrado em evidência. A água residual foi referenciada em 4,70 ppm.

Como pode ser observado no espectro, a proteína FlgN apresentou-se agregada uma vez que os sinais vistos no espectro são alargados. Estes dados corroboram com os dados de CD pois foi verificado que esta proteína encontra-se com baixo conteúdo helicoidal, apresentando poucos sinais na região de 6,5-7,5 ppm característicos de α -hélice. Além disso, no espectro também é observado a presença de $\text{C}\alpha$ de estrutura secundária do tipo folhas β (5,0 – 5,5 ppm) sugerindo que uma troca conformacional possa estar ocorrendo.

Então se iniciou os testes de marcação isotópica da proteína FlgN.

O gel da indução e lise esta ilustrado a seguir (Figura IV-13). Tendo comportamento igual como no caso da expressão em LB, proteína foi purificada aplicando mesmo procedimento.

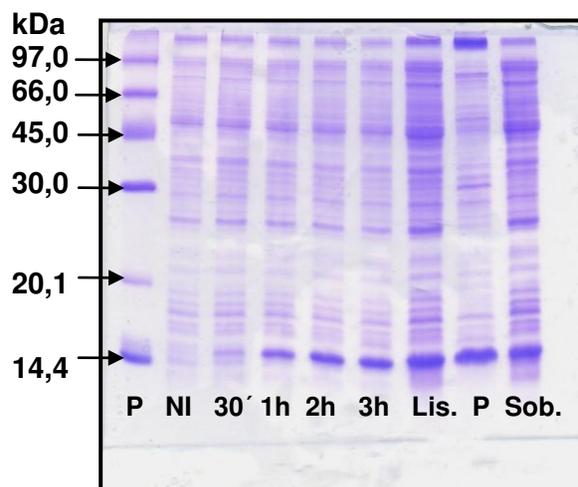


Figura IV-13: Gel de eletroforese SDS-PAGE (15%) da expressão da FlgN, tendo na sequência: marcador de peso molecular de proteína; amostras não induzidas (NI), 30 minutos após a adição de IPTG e 1h, 2h e 3h após a adição de rifamicina, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.)

A proteína foi purificada em coluna de filtração em gel e as amostras coletadas durante todo o processo de purificação foram analisadas por eletroforese SDS-PAGE 15% (Figura IV-14).

As amostras puras da proteína FlgN coletadas da coluna foram concentradas e a sua concentração determinada pelo método de Edelhoch. A concentração encontrada para a proteína FlgN foi de $198 \mu\text{mol.L}^{-1}$. Para esta proteína marcada com ^{15}N foram adquiridos os espectros de ^1H NMR (Figura IV-15 e Anexo 1) e de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC (Figura IV-16, Anexos 2 e 3) no espectrômetro Varian 600 MHz no LNBio (Laboratório Nacional de Biociências do CNPEM).

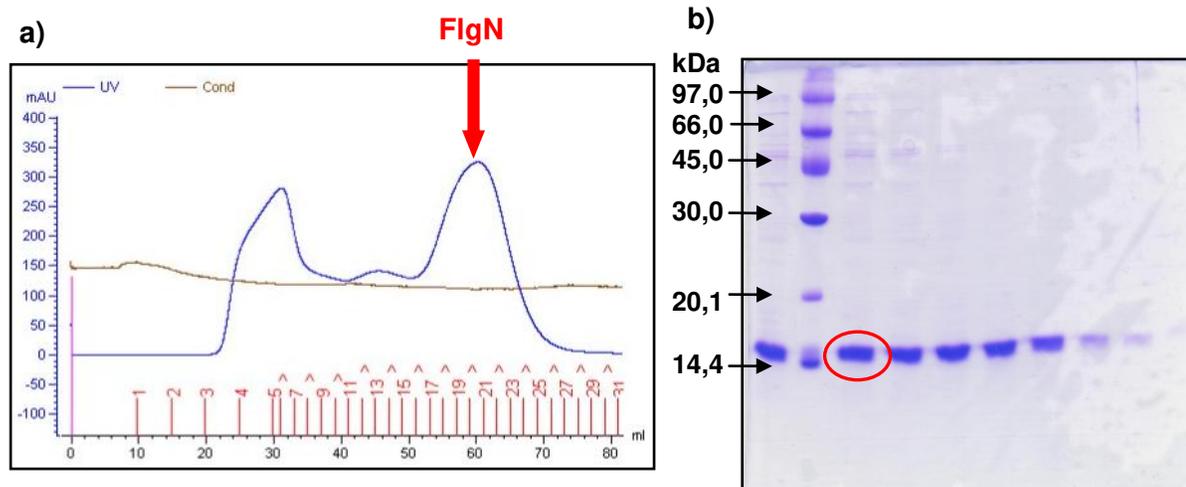


Figura IV-14: a) Cromatograma da purificação do sobrenadante do reenovelamento da FlgN, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína FlgN pura mostrada em destaque (círculo vermelho).

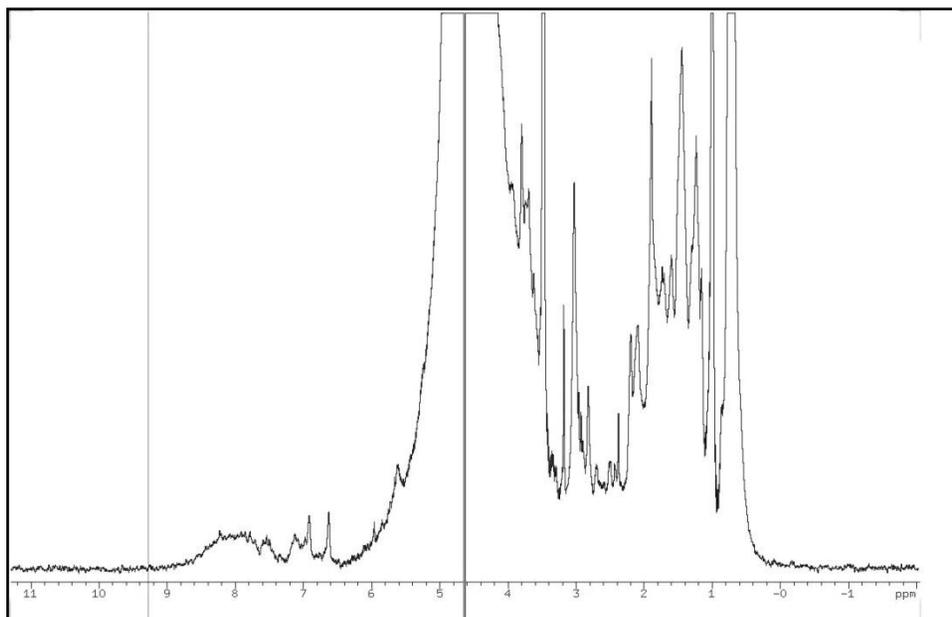


Figura IV-15: Espectro de ^1H NMR da FlgN.

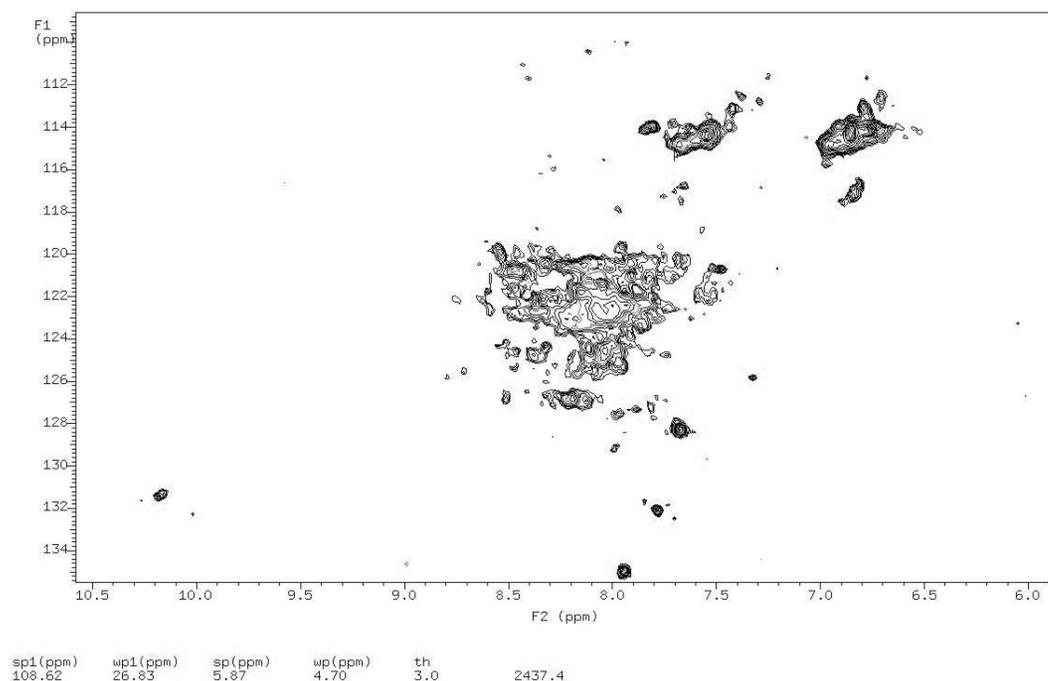


Figura IV-16: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN. A amostra proteica está em tampão fosfato de sódio 100 mmol L^{-1} pH 8,0 com adição de (D_2O 5%, v/v) e os espectros foram obtidos em equipamento INOVA 600 MHz.

Analisando-se os espectros observa-se que a proteína FlgN apresenta sinais muito aglomerados (Figura IV-16). Novamente estes dados corroboram com os dados de CD e ainda é possível concluir que esta proteína não apresenta estrutura terciária. Uma provável explicação pode ser atribuída a uma possível agregação desta proteína devido ao caráter anfipático do domínio C-terminal que cria sítios para interações proteína-proteína, ou, a proteína pode estar parcialmente enovelada.¹²² Neste mesmo ensaio, glicerol foi adicionado à amostra com o intuito de dispersar os sinais, porém o novo espectro indicou que amostra continua agregada. Para outro lote da proteína purificada foi testado o uso de quantidades equimolares de arginina e ácido glutâmico¹²³ pois, estudos indicaram

¹²² Fukuchi, S.; Hosoda, K.; Homma, K.; Gojobori, T.; Nishikawa, K. Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct Biol* **2011**, *11*, 1-10.

¹²³ Golovanov, A.; Hautbergue, G.; Wilson, S.; Lian, L. A Simple Method for Improving Protein Solubility and Long-Term Stability. *J Am Chem Soc* **2004**, *126*, 8933-8939.

que o uso de quantidades equimolares destes aminoácidos carregados auxilia na solubilização das proteínas (Tabela III-6, parte experimental). No entanto, em nenhum dos casos ocorreu significativa melhora nos mapas de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC.

Estudos anteriores indicaram que a proteína FlgK interage com a proteína FlgN.³¹ Por ser uma proteína parceira da FlgN, esperava-se que esta interação levasse a uma menor agregação por parte da proteína FlgN. Então, as interações entre estas proteínas foram estudadas por CD, fluorescência de emissão e NMR. Para isso, ambas proteínas foram purificadas separadamente.

A FlgK, é constituída de 624 resíduos de aminoácidos em sua estrutura primária (Figura IV-17). Sua massa molecular é de aproximadamente 63,32 kDa e tem um pI teórico de 4,78.

FlgK

```

MSIMSTGTSALIAFQRALSTVSHNVANINTEGYSRQRVEFATRPTDMGYAFVG
NGAKITDVGRVADQLAISRLLDGGELSRLLQQLSSLNRVDALYSNTATNVAGL
WSNFFDSTS AVSSNASSTAERQSM LDSGN SLATRFKQLNGQMDSL SNEVNSG
LTSSVDEVNRLTQQIAKLNGTIGSSAQAAAPDMLDQRDALVSKLVGFTGGTAVI
QDGGFMNVFTAGGQPLVVGTTSSKLTTVADPYQPTKLQVAMQTQGQNVLSLA
SSLGGQIGGLLEFRSSVLEPTQAELGRLAVGMAS TFNAGHSQGMDLYGAMGG
NFFNIGSPTVAANPSNAGSASLSASFNSVSAVDGQNVTL SFDGTNWKAINAST
GSAVPMTGTGTAADPLVLNGVSMVVG GTPASGDKFLLQPTAGLAGSL SVAITD
PSRIAAATPVKATATVANLGTGKISDVKVTNAQNPALLTPSSIEFIDANQYTIDGA
GPFAYTAGQTISANGWSFALDGAPKAGDTFGVGP MGAGSSDNGNAKLLANID
DAKALSGGTVTLNGALSGLTTSVGSAAARAASYSADAQKVINDQAQASRDSISG
VNLDEEAANMLKLQQA YQAAAQMISTADTIFQAILGAVR

```

Figura IV-17: Sequencia primária da proteína FlgK. Resíduos de triptofano em destaque (vermelho).

O DNA da FlgK foi cedido pelo grupo do Prof. Dr. Carlos Ramos. A expressão da proteína FlgK pode ser ilustrada na Figura IV-18. Essa proteína apresenta baixa expressão e ela é melhor visualizada após purificação.

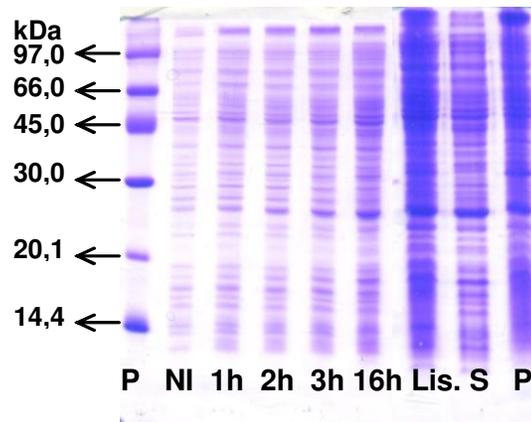


Figura IV-18: Gel de eletroforese SDS-PAGE (15%) da expressão da FlgK, tendo na sequência: marcador de peso molecular de proteína; amostras não induzidas (NI) e 1h, 2h, 3h e 16h após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (S).

A sua purificação foi realizada utilizando a cromatografia por afinidade obtendo-se a proteína FlgK (Figura IV-19).

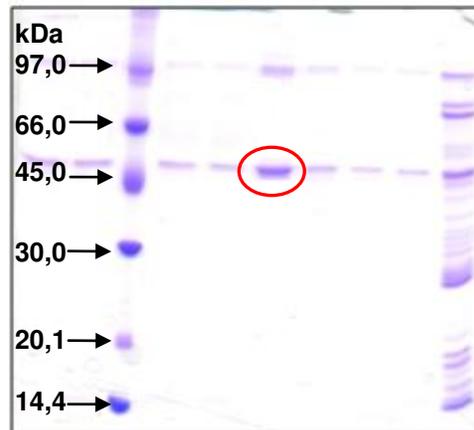


Figura IV-19: Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína FlgK pura mostrada em destaque (circulo vermelho). Na ultima lacuna esta a proteína aplicada na coluna.

Após a concentração desta proteína, sua concentração foi determinada pelo método de Bradford como sendo $127 \mu\text{mol.L}^{-1}$. A partir destes dados foram realizados ensaios de fluorescência para verificar a interação das proteínas. Para as proteínas livres e em complexo em razão molar de 1:1 ou 2:1 (FlgN:FlgK) foram

obtidos os espectros de fluorescência de excitação e emissão das proteínas FlgK e FlgN (Figura IV-20).

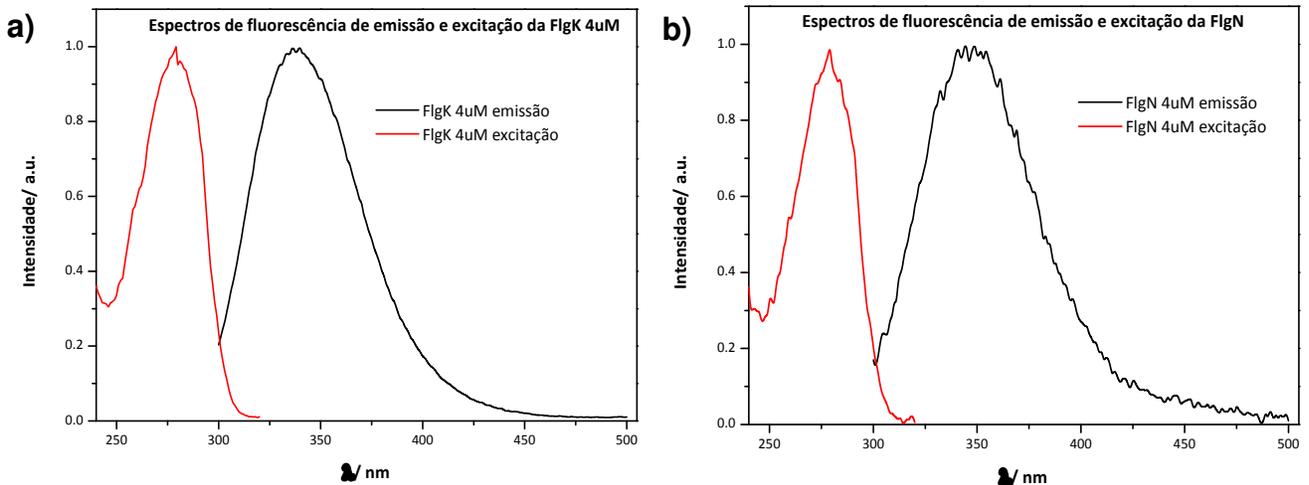


Figura IV-20: Espectros de fluorescência de excitação e emissão das proteínas a) FlgK e b) FlgN.

Os ensaios de interação foram realizados com as proteínas na proporção FlgN/ FlgK 1:1 e FlgN/ FlgK 2:1 sem padronização para melhor visualização (Figura IV-21). A partir dos espectros pode-se sugerir que a proteína FlgN interage melhor com a proteína FlgK na proporção de 2:1 pois ocorreu uma supressão da fluorescência maior neste caso.

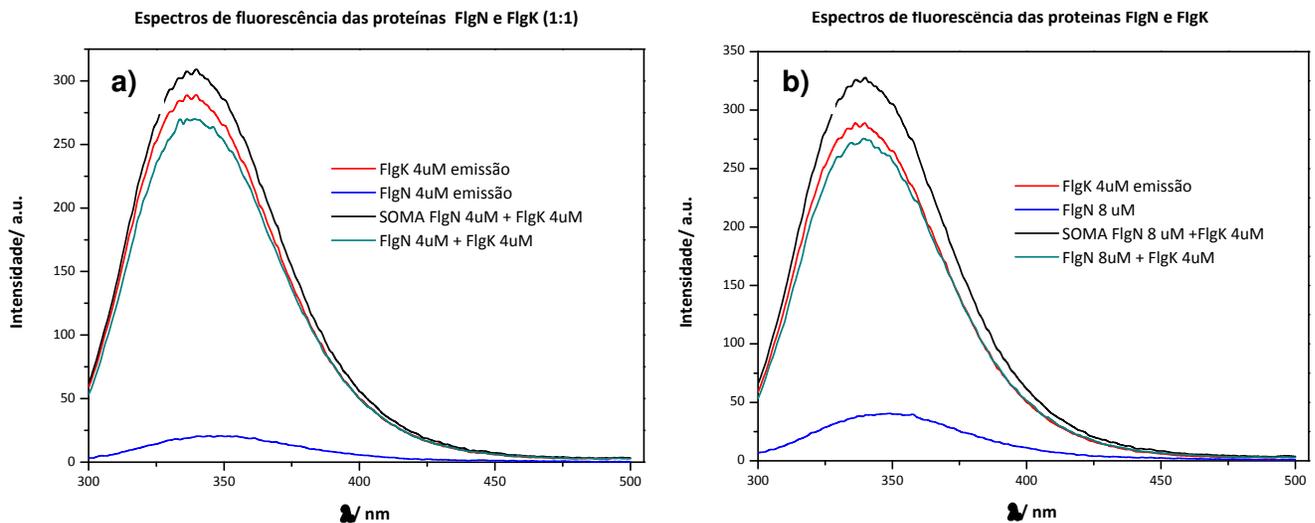


Figura IV-21: Espectros de fluorescência dos complexos de FlgN e FlgK na proporção molar a) 1:1 e b) 2:1 (FlgN/ FlgK).

A supressão da fluorescência pode ser atribuída à mudanças conformacionais na estrutura da proteína levando a uma menor exposição do resíduo de Trp por causa da interação com a proteína parceira e este resíduo pode estar participando também da interação com a outra proteína. Através dos dados de fluorescência pode-se sugerir que a interação na razão molar 2:1/ FlgN:FlgK foi a mais viável.

A partir das amostras usadas para os ensaios de fluorescência, foram realizados os ensaios de CD. Os espectros de CD obtidos para as proteínas estão mostrados na Figura IV-22. Estas amostras também foram analisadas por espectrometria de massas para a obtenção da massa do complexo, porém devido a degradação da amostra não foi possível concluir estes dados. Desse modo, os cálculos dos complexos foram efetuados usando um valor aproximado massa molar como a somatória de ambas massas molares das proteínas [63325(FlgK) + 12120(FlgN)]. Considerou-se como número de aminoácidos no caso do complexo 1:1 [624(FlgK) + 110(FlgN)] a somatória do número de aminoácidos das proteínas e no caso do complexo 2:1 [624+ (2x110)]. A concentração da proteína usada no cálculo foi a da proteína FlgK.

Pode-se observar que, assim como na fluorescência a interação é mais efetiva quando as proteínas estão interagindo na proporção 2:1. No caso na interação 1:1 não houve mudança tão significativa. Quando duas proteínas interagem mudanças substanciais não ocorrem nos componentes da estrutura secundária com a formação do complexo e estas mudanças no espectro de CD na região do UV-distante são muito pequenas, porém, podem ser um indicativo da interação. Para melhor visualizar esta interação seria melhor adquirir espectro de CD na região do UV-próximo (260-320 nm) sendo assim possível verificar mudanças na estrutura terciária.

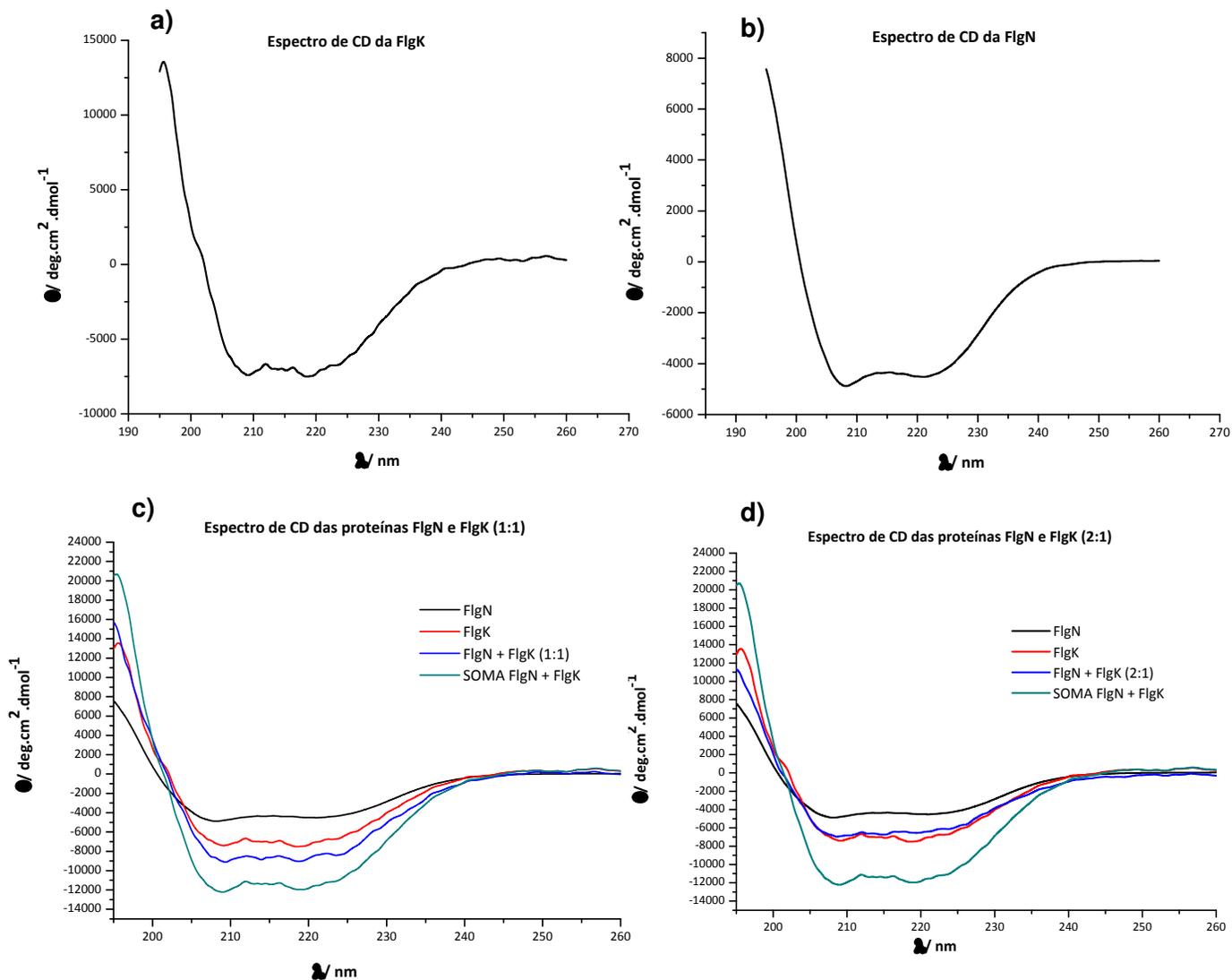


Figura IV-22: Espectros de CD. a) proteína FlgK livre. b) proteína FlgN livre. c) Espectros de CD das proteínas e do complexo FlgN + FlgK (1:1). d) Espectros de CD das proteínas e do complexo FlgN + FlgK (2:1). Tampão fosfato de sódio 25 mmol.L^{-1} , pH 8,0 foi usado como branco.

Após a purificação da proteína FlgN marcada com ^{15}N foram adquiridos os espectros de ^1H NMR e $\{^{15}\text{N}, ^1\text{H}\}$ HSQC no complexo com a FlgK (FlgN/FlgK, 2:1). É importante ressaltar que estas proteínas foram concentradas já complexadas. No entanto, novamente não houve mudanças significativas e a proteína FlgN continuou agregada (anexos 8 e 9). Os espectros adquiridos para a proteína FlgN livre e no complexo 2:1 estão ilustrados nas figuras IV-23 e IV-24.

Através dos espectros de NMR é verificado que a interação entre as duas proteínas FlgN/FlgK na proporção 2:1 levou a uma pequena melhora no sinal, porém a proteína FlgN ainda assim continua agregada. Através dos dados de massas foi observado que a proteína FlgN é uma proteína monomérica mas uma provável explicação para uma possível agregação desta proteína é devido ao caráter anfipático do domínio C-terminal que cria sítios para interações proteína-proteína.

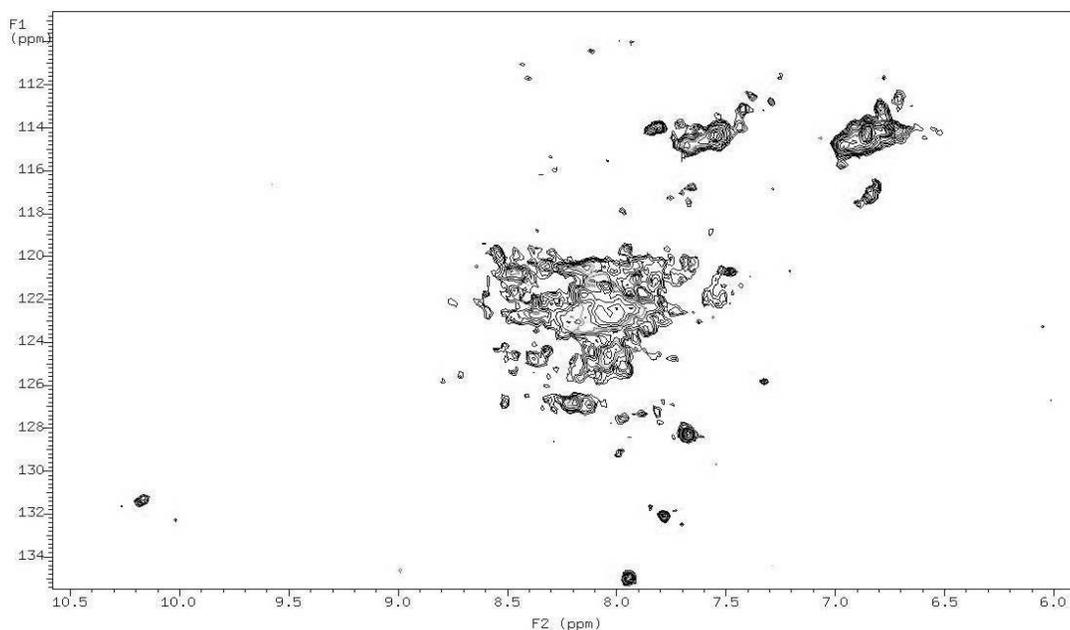


Figura IV-23: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN.

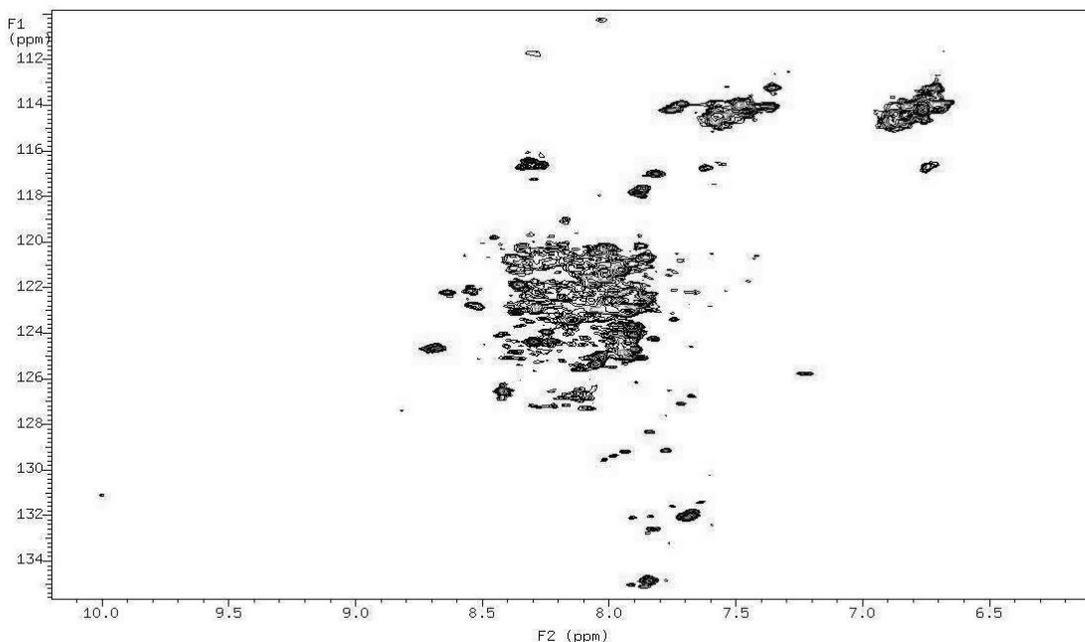


Figura IV-24: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo FlgN/FlgK (2:1).

Na tentativa de explicar melhor os dados, foram gerados modelos tridimensionais para as duas proteínas FlgN e FlgK utilizando o programa I-TASSER.¹¹⁵ O melhor modelo dentre os cinco gerados para a FlgK (*score* de -2,20) e dentre os cinco gerados para a FlgN (*score* de -2,51) estão ilustrados nas Figura IV-25 e 26, respectivamente.

A estrutura calculada pelo I-TASSER com o *score* de -2,51 para a proteína FlgN foi o modelo que melhor sobrepôs o envelope experimental obtido pelos dados de SAXS e que foi publicada recentemente.⁶²

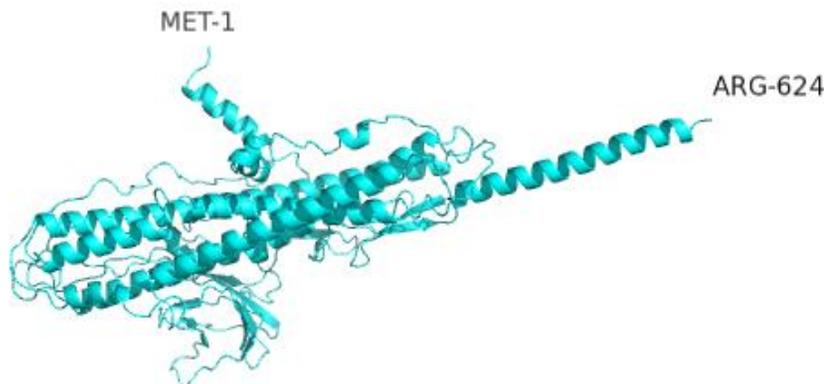


Figura IV-25: Representação em cartoon da estrutura da FlgK com N- e C-terminais marcados (I-TASSER C-score = -2,20).

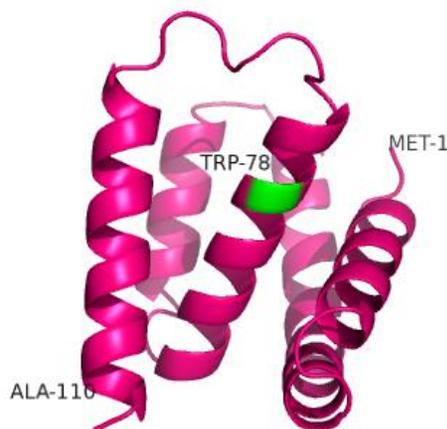


Figura IV-26: Representação em cartoon da estrutura da FlgN com N- e C-terminais marcados (I-TASSER C-score = -2,51) e o resíduo de Trp (W70) em verde (exposto ao solvente).

Através das estruturas tridimensionais geradas pelo I-TASSER é observado que a proteína FlgN apresenta uma estrutura praticamente do tipo α -hélice muito parecida com a estrutura da FliT de *Salmonella sp* e a proteína FlgK apresenta estrutura praticamente do tipo α -hélice.

Os dados de fluorescência corroboram com a estrutura proposta pelo I-TASSER, pois o único resíduo de triptofano da proteína FlgN, o Trp78, apresenta-se mais exposto ao solvente. No entanto, há discrepâncias entre os dados de CD

e NMR com a estrutura proposta pelo I-TASSER. Através dos dados de CD é observado baixo conteúdo α -hélice, que pode ser resultado da baixa concentração usada (amostra de fluorescência). Analisando-se os espectros de NMR 1D e 2D foi observado também esta baixa ocorrência de estrutura secundária. Uma explicação provável é uma possível agregação proteica durante a concentração da amostra ou ainda, uma troca conformacional. No mapa de contorno $\{^{15}\text{N}, ^1\text{H}\}$ HSQC desta proteína (Figura IV-23) também foi observado ausência de estrutura 3D. Na análise massas indicou que esta proteína se apresenta como um monômero. Para o estudo da interação entre as proteínas FlgN e FlgK acredita-se que possivelmente esteja ocorrendo uma interação entre as duas proteínas, que já foi comprovado em um estudo anterior³¹, mas, não é possível afirmar com veemência através dos dados de CD, fluorescência e NMR.

4.2 XACb0033

A proteína XACb0033 é constituída de 72 resíduos de aminoácidos em sua estrutura primária (Figura IV-27). Sua massa molecular é de aproximadamente 8,4 kDa, tem um pI teórico de 5,6.

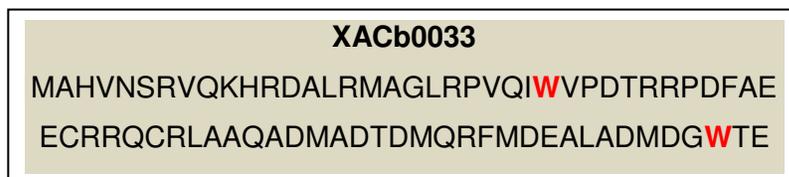


Figura IV-27: Sequencia primária da proteína XACb0033. Resíduos de triptofano em destaque (vermelho).

Assim como para a proteína FlgN, a proteína XACb0033 foi encontrada majoritariamente no pellet bacteriano (Figura IV-28) e expressa muito bem até 16hs. Como pode-se observar esta proteína se apresenta no gel como um monômero (8,4 kDa) e um dímero (16,8 kDa) também.

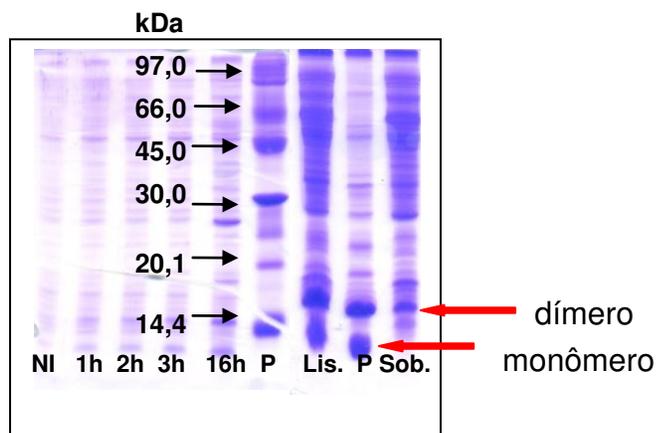


Figura IV-28: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0033, tendo na sequência: amostras não induzidas (NI) e 1h, 2h, 3h e 16hs após a adição de IPTG, marcador de peso molecular de proteína, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.). Os monômeros e dímeros estão indicados no gel.

A proteína XACb0033 foi purificada em coluna de filtração em gel e as amostras coletadas durante todo o processo de expressão e purificação foram analisadas por eletroforese SDS-PAGE 15% (Figura IV-29). A proteína XACb0033 foi purificada como um dímero.

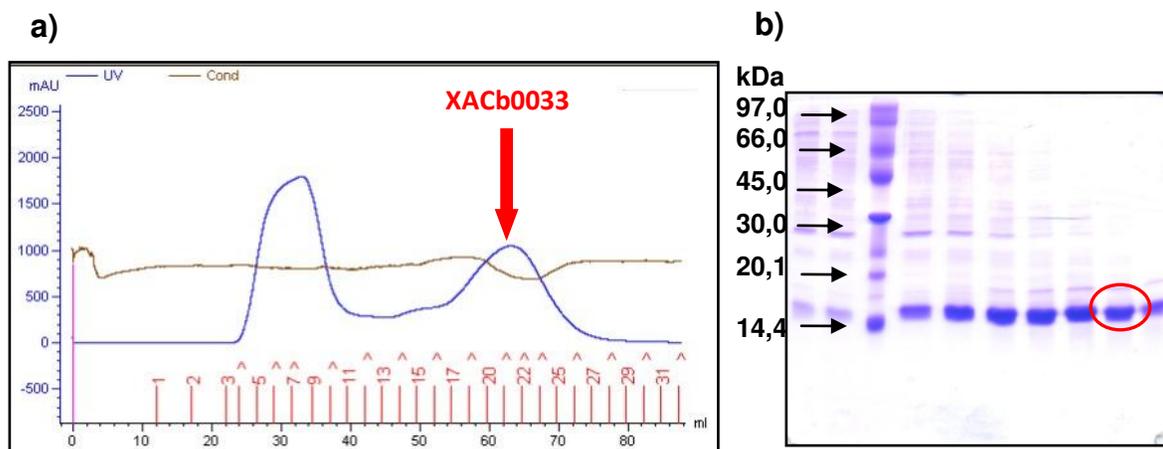


Figura IV-29: a) Cromatograma da purificação do sobrenadante do reenovelamento da XACb0033, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína XACb0033 pura mostrada em destaque (círculo vermelho).

As amostras puras da proteína XACb0033 coletadas foram concentradas e a sua concentração determinada pelo método de Edelhoch.^{106,107} A concentração encontrada para a proteína XACb0033 foi de $24,8 \mu\text{mol.L}^{-1}$.

A partir desta amostra foram realizados estudos estruturais de Dicroísmo Circular (Desenovelamento térmico), Fluorescência de Emissão e Espectrometria de Massas.

A identificação da proteína foi feita empregando-se digestão triptica, seguida de cromatografia líquida acoplada a espectrometria de massas (LC-MS) para separação e identificação dos peptídeos. A busca no MASCOT identificou vários peptídeos como sendo da XACb0033, o que permitiu a identificação desta proteína com um *score* de **257** com uma cobertura de 59% (Figura IV-32). Dentre

o conjunto de peptídeos identificados como sendo da XACb0033, espectros representativos para os três com maior *score* foram dispostos nas Figuras IV-30 e IV-31.

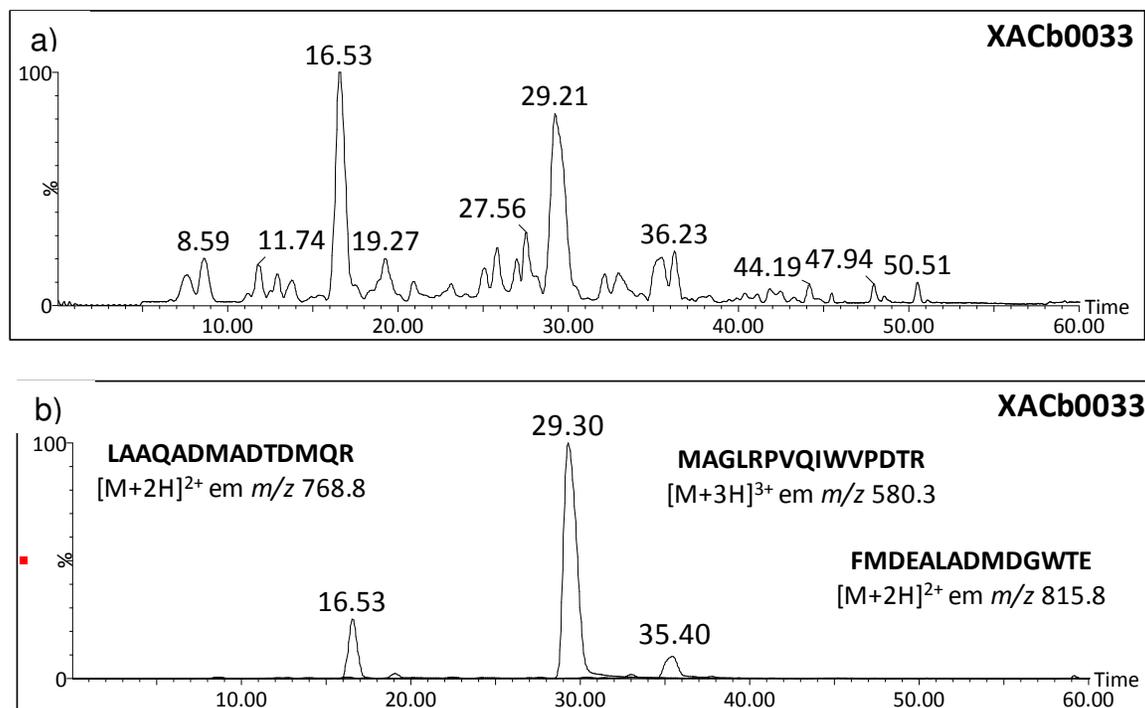


Figura IV-30: a) Cromatograma total (LC-MS) dos peptídeos da digestão trípica da XACb0033 e b) Cromatogramas extraídos para íons dos peptídeos com maior *score* no MASCOT oriundos da XACb0033.

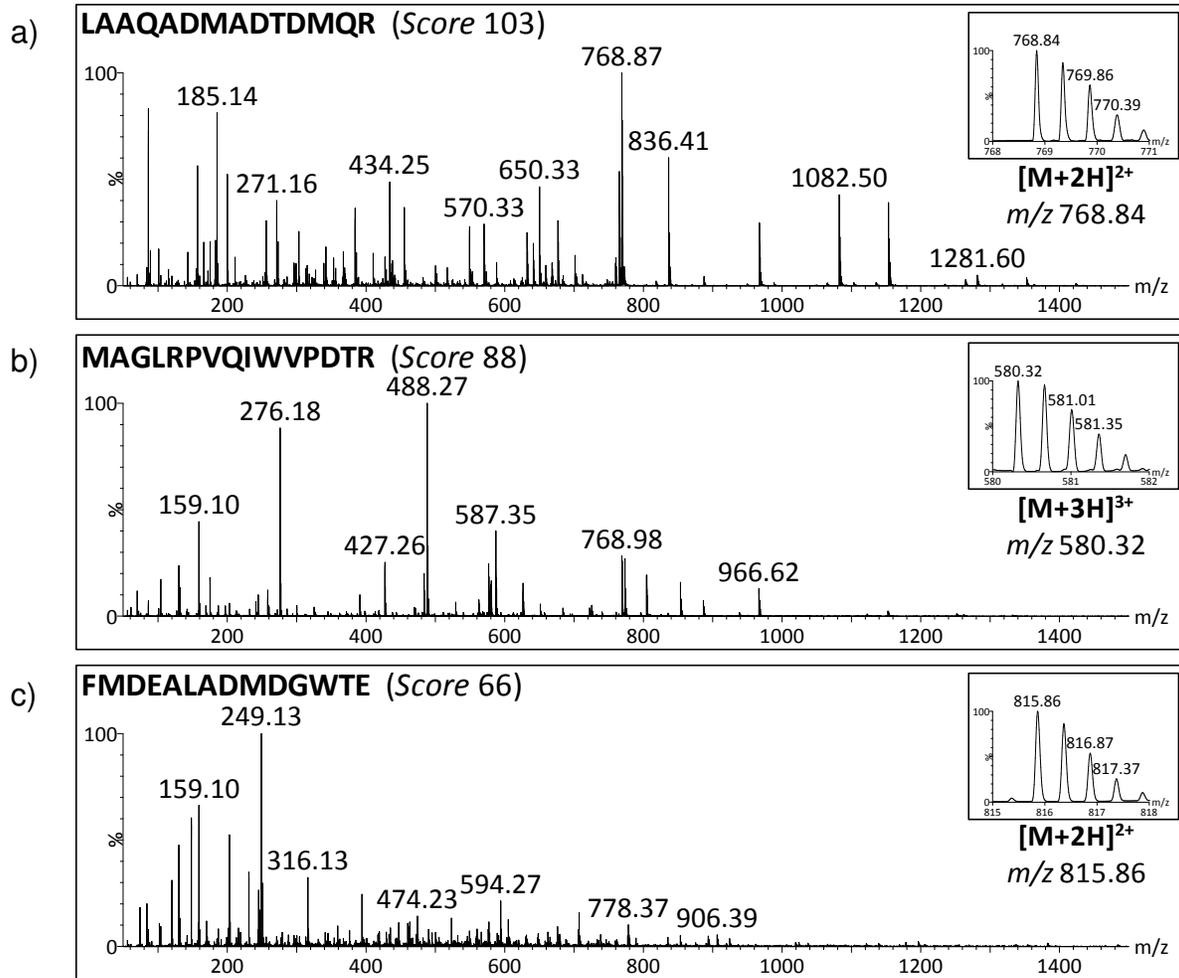


Figura IV-31: Espectros de MS/MS gerados a partir dos três peptídeos com maior score da proteína XACb0033. a) MS/MS do peptídeo LAAQADMADTDMQR. b) MS/MS do peptídeo MAGLRPVQIWVPDTR. c) MS/MS do peptídeo FMDEALADMDGWTE.

Nominal mass (M_r): **8413**; Calculated pI value: **5.61**
Matched peptides shown in **Red**

MAHVNSRVQKHRDALR**MAGLRPVQIWVPDTR**RRPDFAEI
CRRQCR**LAAQADMADTDMQR**FMDEALADMDGWTE

Figura IV-32: Sequência primária da proteína XACb0033. Os peptídeos encontrados para a XACb0033 (MS/MS) estão mostrados em destaque (vermelho), sendo obtidos 59% de cobertura da sequência.

Também foi obtido um espectro da proteína intacta para confirmar a massa exata desta proteína. Através da deconvolução foi obtida uma massa molar de 16570,0 e 8285,0 Da (Figura IV-33) comprovando que a proteína XACb0033 se comporta como um monômero e também como um dímero.

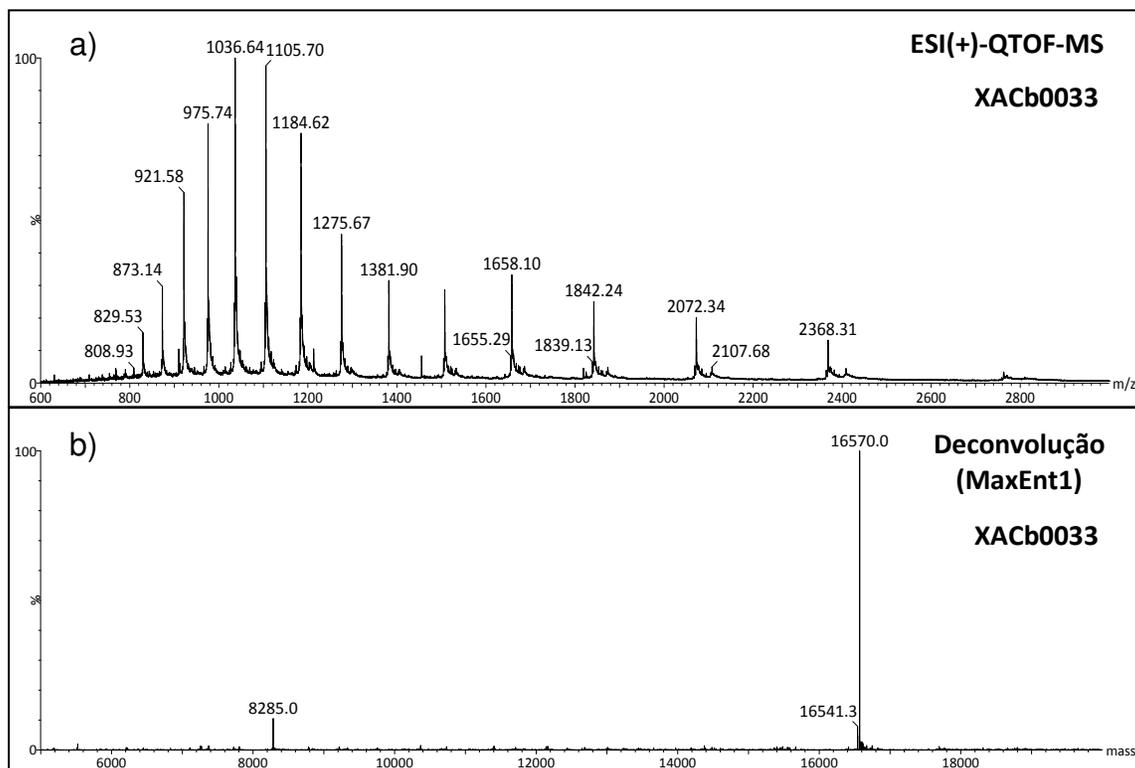


Figura IV-33: Espectros de massas da proteína XACb0033. a) Espectro de massa da proteína intacta e b) após processamento dos dados.

Através da análise do espectro deconvoluido observa-se que a proteína XACb0033 encontra-se praticamente como um dímero de 16570 Da mas, o monomero de 8285 Da também é observado. Neste espectro também é verificado que ocorreu a perda da primeira Met da proteína.

Para a proteína XACb0033 foram realizados experimentos de fluorescência de emissão, CD e NMR. Como supracitado, a proteína apresenta dois resíduos de triptofano em sua estrutura primária, e então, foram realizados experimentos de fluorescência de emissão, onde se observou um máximo em 346 nm (Figura IV-34), indicando resíduo de triptofano exposto ao solvente. Normalmente o máximo de emissão do triptofano livre em solução é na faixa de 350 nm e cai para 330 nm quando apresentam interações intramoleculares hidrofóbicas.⁷⁴

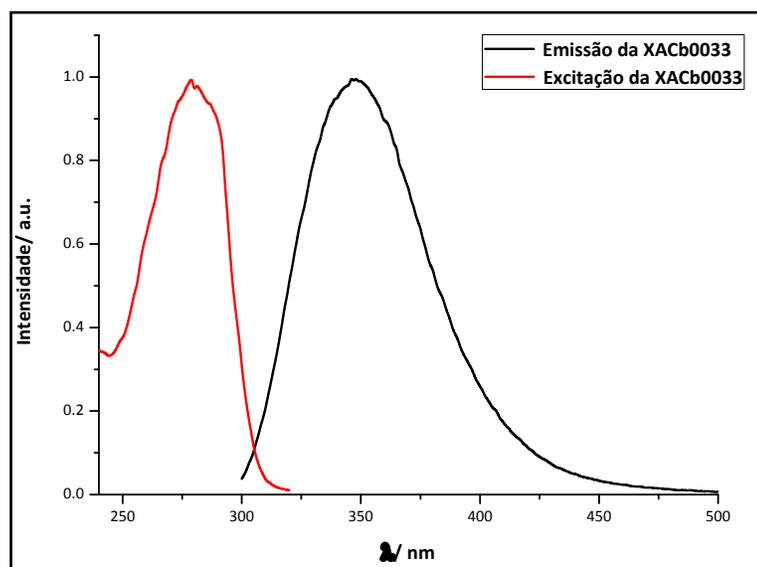


Figura IV-34: Espectros de fluorescência de excitação e de emissão da proteína XACb0033 com uma concentração de $24,8 \mu\text{mol.L}^{-1}$ em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0).

A XACb0033 foi então submetida à análise por Dicroísmo Circular para avaliação da estrutura secundária.¹²⁰ Os dados obtidos por CD foram tratados no programa CDNN¹²¹ para cálculo da estrutura secundária da proteína XACb0033 e os resultados obtidos estão sumarizados na Figura IV-35.

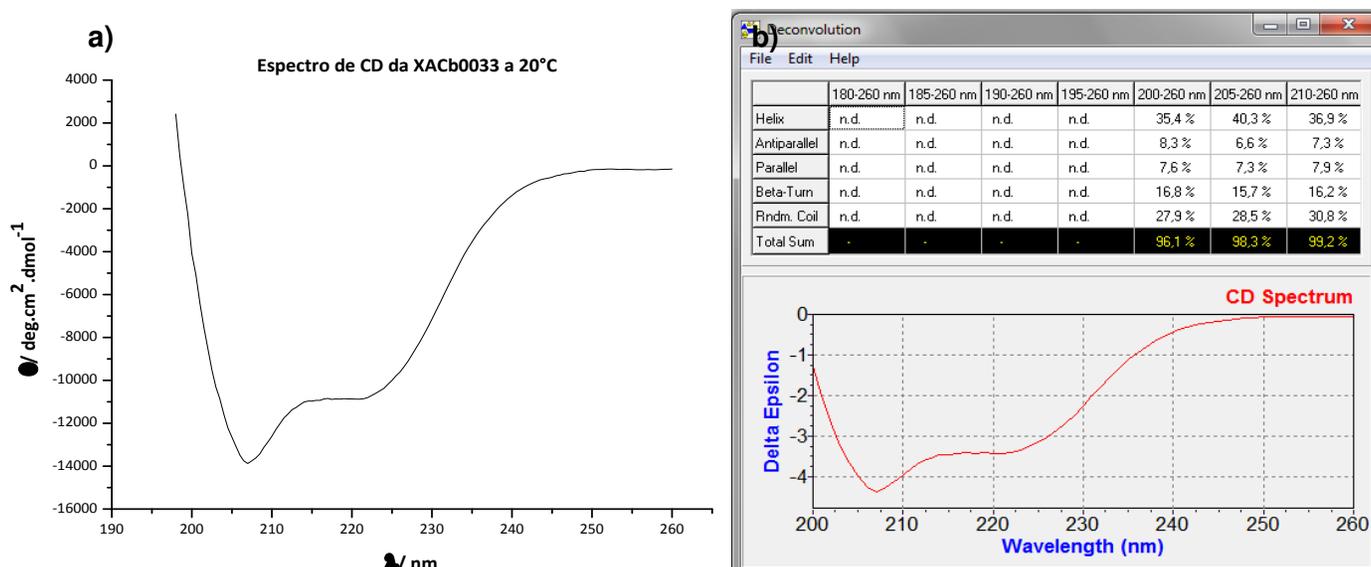


Figura IV-35: Ilustração dos resultados obtidos em análise de CD para a proteína XACb0033: a) Espectro de CD, (b) Cálculo da estrutura secundária da proteína XACb0033 obtida pelo programa CDNN a partir dos dados de CD. Abaixo gráfico gerado pelo CDNN. Para os cálculos de CD o valor do branco (tampão fosfato de sódio (25 mM, pH 8,0) foi descontado.

A partir dos dados de CD foi verificado que a proteína XACb0033 apresenta estruturas do tipo α -hélice. A partir desta mesma proteína foi realizado ensaio de desenovelamento térmico em que é observada a perda da estrutura secundária com o aumento da temperatura (Figura IV-36).

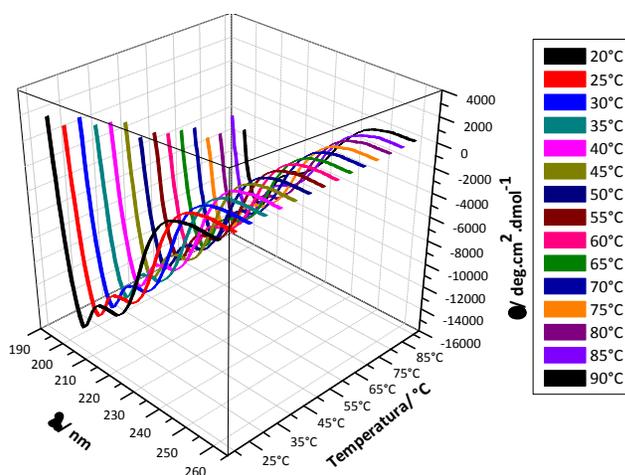


Figura IV-36: Desenovelamento térmico 20 – 90°C (5°C de incremento).

Então a partir destes dados se iniciou os testes de marcação isotópica da proteína XACb0033. O gel da indução e lise esta ilustrado a seguir (Figura IV-37) e, como pode ser observado, a proteína XACb0033 se encontra no pellet bacteriano.

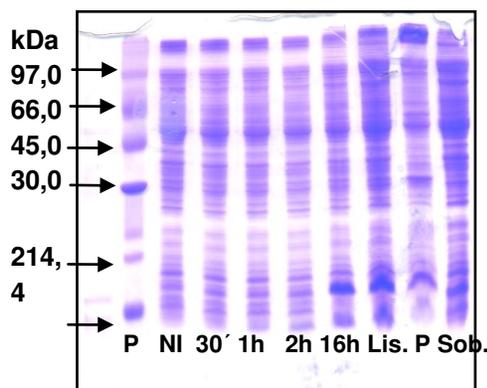


Figura IV-37: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0033, tendo na sequência: marcador de peso molecular de proteína; amostras não induzidas (NI), 30 minutos após a adição de IPTG e 1h, 2h e 16h após a adição de rifamicina, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.)

A proteína XACb0033 foi purificada em coluna de filtração em gel e as amostras coletadas durante todo o processo de expressão e purificação foram analisadas por eletroforese SDS-PAGE 15% (Figura IV-38).

As amostras puras da proteína XACb0033 foram concentradas e a sua concentração determinada pelo método de Edelhoch. A concentração encontrada para a proteína XACb0033 foi de $45 \mu\text{mol.L}^{-1}$. Para esta proteína marcada com ^{15}N foram adquiridos os espectros de ^1H NMR (Figura IV-39 e Anexo 10) e $\{^{15}\text{N}, ^1\text{H}\}$ HSQC (Figura IV-40, Anexos 11 e 12) no espectrômetro Varian 600 MHz no LNBio (Laboratório Nacional de Biociências do CNPEM).

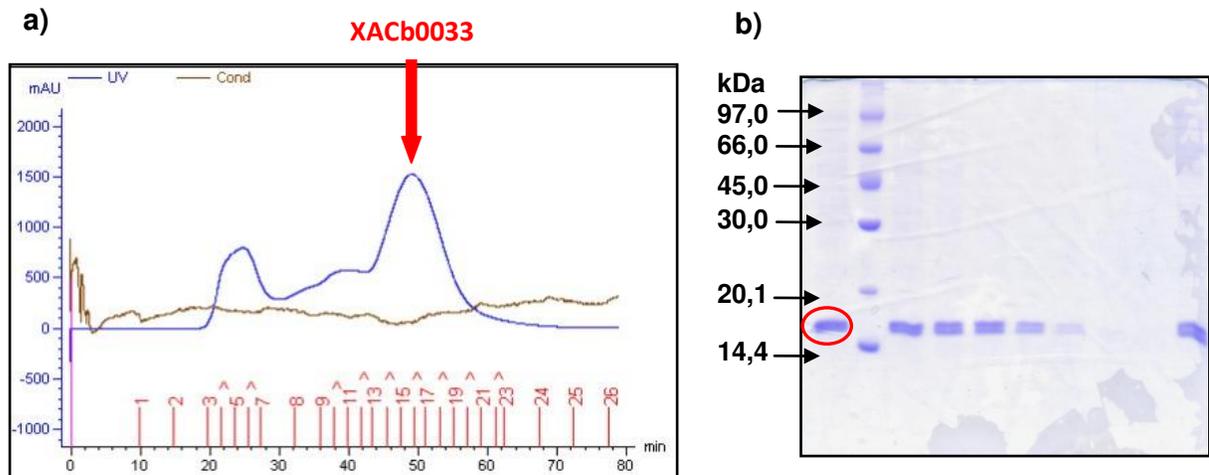


Figura IV-38: a) Cromatograma da purificação do sobrenadante do reenovelamento da XACb0033, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína XACB0033 pura mostrada em destaque (círculo vermelho).

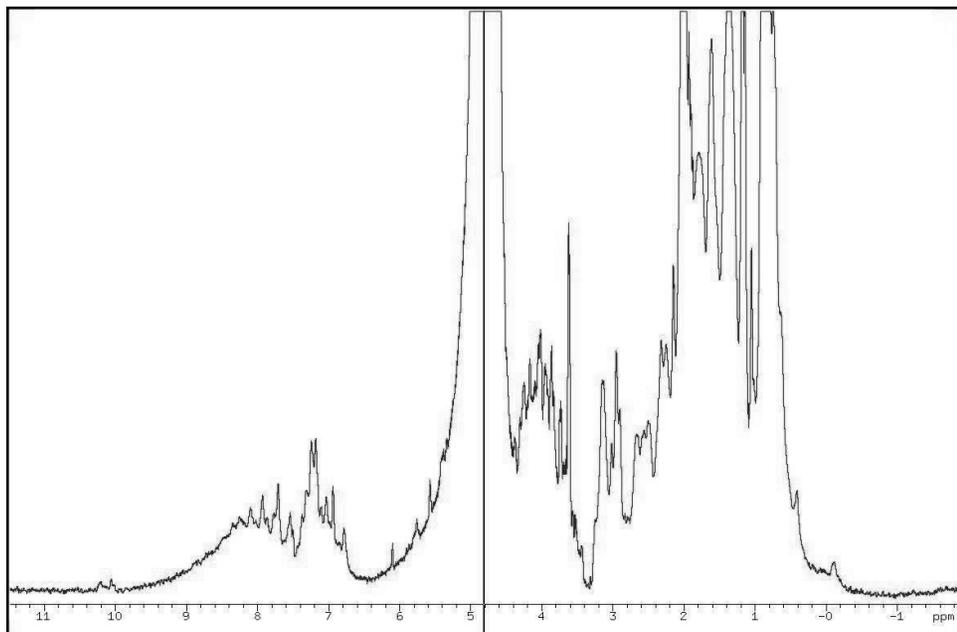


Figura IV-39: Espectro de ^1H NMR da XACb0033.

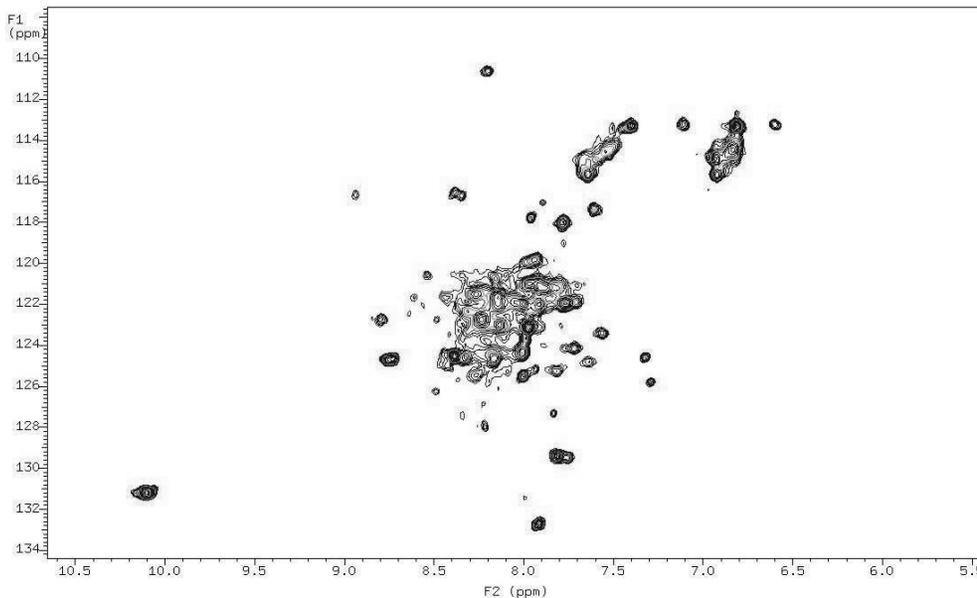


Figura IV-40: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033. A amostra proteica está em tampão fosfato de sódio 100 mmol L^{-1} com 50 mmol L^{-1} de NaCl com adição de (D_2O 5%, v/v) e os espectros foram obtidos em equipamento INOVA 600.

Analisando-se os espectros observa-se que a proteína XACb0033 apresenta sinais muito aglomerados (Figura IV-40), assim como para a proteína FlgN. Uma provável explicação pode ser atribuída a uma possível agregação desta proteína. No caso da proteína XACb0033 também foi observado ausência de estrutura 3D no mapa de contorno $\{^{15}\text{N}, ^1\text{H}\}$ HSQC. Neste mesmo ensaio, DTT (ditiotreitól) foi adicionado à amostra com o intuito de dispersar os sinais através da clivagem das ligações de dissulfeto, porém o novo espectro indicou que amostra continuou agregada. Outras purificações e tentativas foram realizadas com esta proteína marcada com ^{15}N (Tabela III-7, parte experimental), porém em nenhum dos casos ocorreu significativa melhora nos mapas de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC.

Estudos anteriores indicaram que a proteína XACb0032 é uma proteína parceira da proteína XACb0033. A XACb0033 e XACb0032 são chaperonas pequenas de até 12 kDa e formam um complexo XACb0033/XACb0032.^{3b}

Acredita-se que a chaperona XACb0032 esta envolvida no sistema de secreção do tipo IV da *Xac*. No caso da bactéria *Agrobacterium tumefaciens*, a chaperone (VirE1) esta envolvida na transferência de proteínas independentemente de DNA, mediando a transferência da proteína VirE2. A VirE1 é uma proteína de 7,5 kDa que interage diretamente com o domínio interno da VirE2 estabilizando-a e prevenindo que a VirE2 sofra agregação.^{3b} Reunindo as características citadas acima e as propriedades físicas de VirE1, acredita-se que as proteínas XACb0032 e XACb0033 possam interagir e formar um complexo não agregado. A XACb0033 tem características que poderiam classifica-la como uma VirE2 da *Xac*.

Por ser uma proteína parceira da XACb0033, esperava-se que a interação XACb0033/XACb0032 diminuísse a agregação por parte da proteína XACb0033. Como as chaperonas possuem superfície de interação hidrofóbica, talvez esta interação possa diminuir interações hidrofóbicas da XACb0033. Então, a interação destas proteínas foi estudada por CD, fluorescência de emissão e NMR. Para isso, ambas as proteínas foram purificadas separadamente.

A proteína XACb0032, é constituída de 107 resíduos de aminoácidos em sua estrutura primária (Figura IV-41). Sua massa molecular é de aproximadamente 11,54 kDa e tem um pI teórico de 9,0.



Figura IV-41: Sequência primária da proteína XACb0032.

O gel da indução e lise esta ilustrado a seguir (Figura IV-42) e, como pode ser observado, a proteína XACb0032 se encontra no sobrenadante.

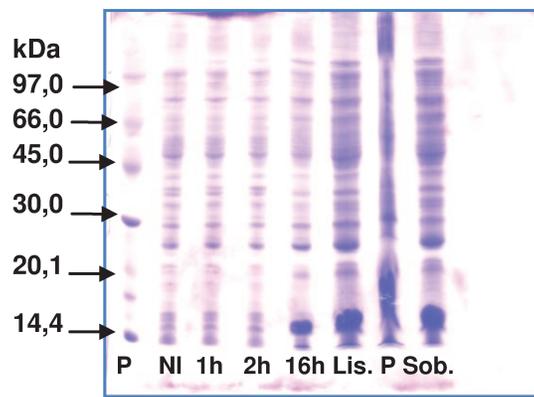


Figura IV-42: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0032, tendo na sequência: marcador de peso molecular de proteína, amostras não induzidas (NI) e 1h, 2h e 16hs após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.)

A XACb0032 foi encontrada no sobrenadante e como o próximo passo seria purificá-la em coluna de troca iônica CM- Sepharose, a proteína foi dialisada com tampão acetato (50 mmol.L^{-1} pH 5,0).

Porém, após a diálise o meio protéico tornou-se esbranquiçado e posterior centrifugação (4°C , 15000 rpm e 20 minutos) levou a proteína para o pellet (Figura IV-43).

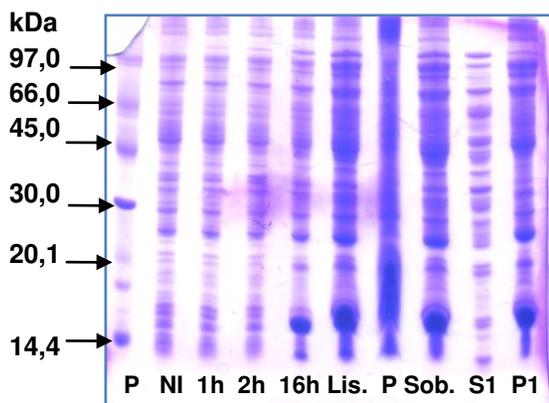


Figura IV-43: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0032, tendo na sequência: marcador de peso molecular de proteína, amostras não induzidas (NI), 1h, 2h e 16hs após a adição de IPTG, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.), sobrenadante após a mudança de tampão (S1) e pellet após a mudança de tampão (P1).

Como a proteína migrou para o pellet, pET23a-XACb0033 DNA-1 foi adicionado a proteína na tentativa de torná-la solúvel. Também, houve a tentativa em solubilizar a XACb0032 adicionando a *Xac* 306 genômico, DNA-2 (Figura IV-44). Estes complexos XACb0032/DNA foram dialisadas com tampão fosfato de sódio (100 mmol.L^{-1} , pH 8,0) usando uma membrana de porosidade de 3000 Da durante um período de 24 horas, com consecutivas trocas do meio da diálise.

Em seguida o meio foi centrifugado (4°C , 15000 rpm e 20 minutos) e complexo da XACb0032 + DNA-1 foi encontrado no sobrenadante com poucas impurezas (Figura IV-44).

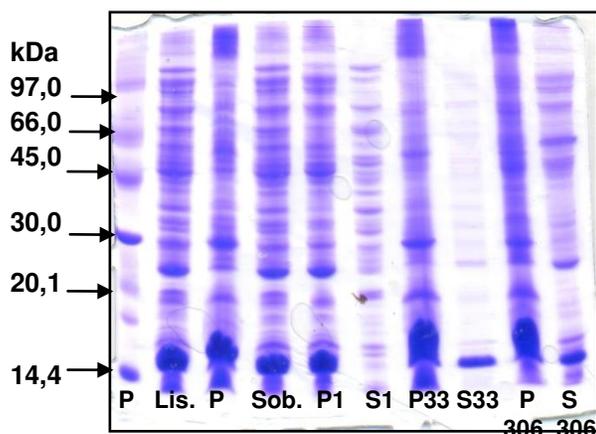


Figura IV-44: Gel de eletroforese SDS-PAGE (15%) da expressão da XACb0032, tendo na sequência: marcador de peso molecular de proteína, lisado (Lis.), precipitado da lise (P), sobrenadante da lise (Sob.), pellet após a mudança de tampão (P1), sobrenadante após a mudança de tampão (S1), pellet com DNA-1 (P33), sobrenadante com DNA-1 (S33), pellet com DNA-2 (P306) e sobrenadante com DNA-2 (S306).

A amostra contendo complexo XACb0032 + DNA-1 foi filtrada (filtro Milipore 0,22 μm) e medida a absorbância à 260 nm (DNA) e à 280 nm (proteína) e os resultados estão mostrados na Tabela IV-1.

Tabela IV-1: Absorbância da XACb0032 + DNA da XACb0033 à 280 nm e 260 nm

Absorbância à 280 nm	0,853
Absorbância à 260 nm	0,863
Razão 280 nm/ 260 nm	0,99

A partir dos dados da Tabela IV-1 pode-se concluir que a XACb0032 interage com o DNA-1 na proporção de 1:1. Então esta proteína foi purificada utilizando cromatografia de filtração em gel e as amostras coletadas durante todo o processo de purificação foram analisadas por eletroforese SDS-PAGE 15% (Figura IV-45).

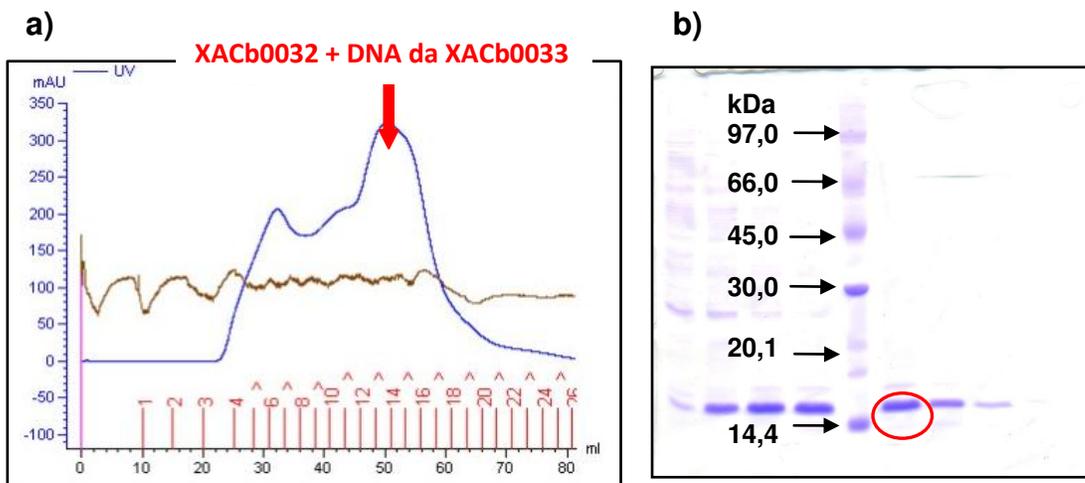


Figura IV-45: a) Cromatograma da purificação do sobrenadante do complexo da XACb0032 + DNA-1, com o pico correspondente a proteína indicado em vermelho. b) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: o marcador de peso molecular e as frações recolhidas da coluna, com a proteína XACb0032 pura mostrada em destaque (círculo vermelho).

Após a purificação das proteínas XACb0033 marcada com ^{15}N e XACb0032+DNA-1 também foram adquiridos os espectros de ^1H NMR e $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo XACb0033/XACb0032+DNA-1. Após a purificação, ambas as proteínas foram concentradas juntamente na razão molar 2:1 (XACb0033:XACb0032) e analisadas no equipamento INOVA 600. No entanto, novamente não houve mudanças significativas e a proteína XACb0033 continuou agregada. Os espectros adquiridos encontram-se nos anexos 22 e 23. Paralelamente a XACb0032 foi purificada também sem o DNA da XACb0033, porém a purificação é muito difícil e não obtivemos ela pura. Esta proteína também foi concentrada juntamente com a XACb0033 marcada com ^{15}N na razão molar 2:1 (XACb0033:XACb0032) e analisadas no equipamento INOVA 600. No entanto, novamente não houve alteração. Os espectros adquiridos encontram-se nos anexos 24 e 25. Os mapas de contorno $\{^{15}\text{N}, ^1\text{H}\}$ HSQC adquiridos para a proteína XACb0033 livre e complexada com a proteína XACb0032 estão ilustradas nas Figuras IV-46 e Figura IV-47. Do mesmo modo que para as proteínas FlgN e FlgK

complexadas, o complexo entre as proteínas XACb0033 e XACb0032 resultou em uma pequena melhora nos sinais mas os sinais estão menos intensos.

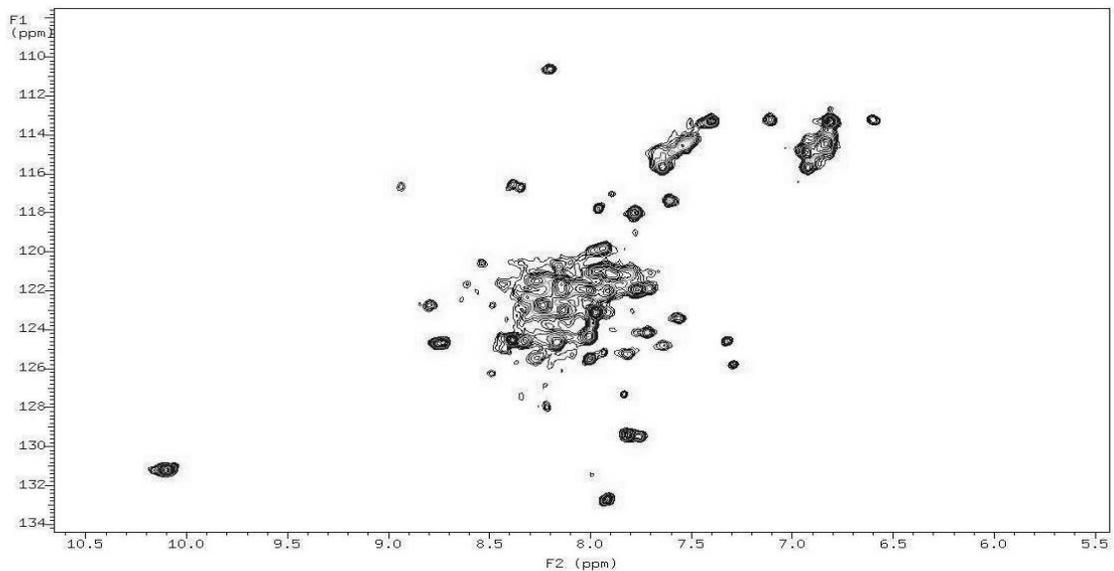


Figura IV-46: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da proteína XACb0033.

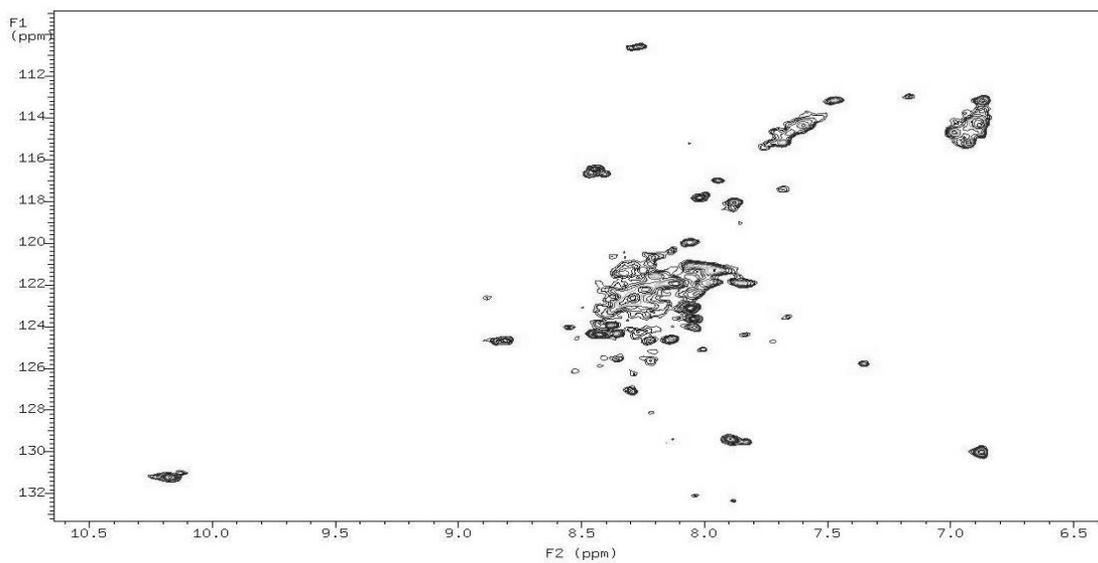


Figura IV-47: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo XACb0033/XACb0032 na razão molar 2:1.

Na tentativa de melhor explicar os resultados foram gerados modelos tridimensionais para as duas proteínas XACb0032 e XACb0033 utilizando o programa I-TASSER.¹¹⁵ O melhor dentre os cinco gerados para a XACb0033 foi obtida com um *score* de -2,41 (Figura IV-48b). Para a XACb0032 o programa I-TASSER gerou apenas um modelo com um *score* de 1,13 (Figura IV-48a).

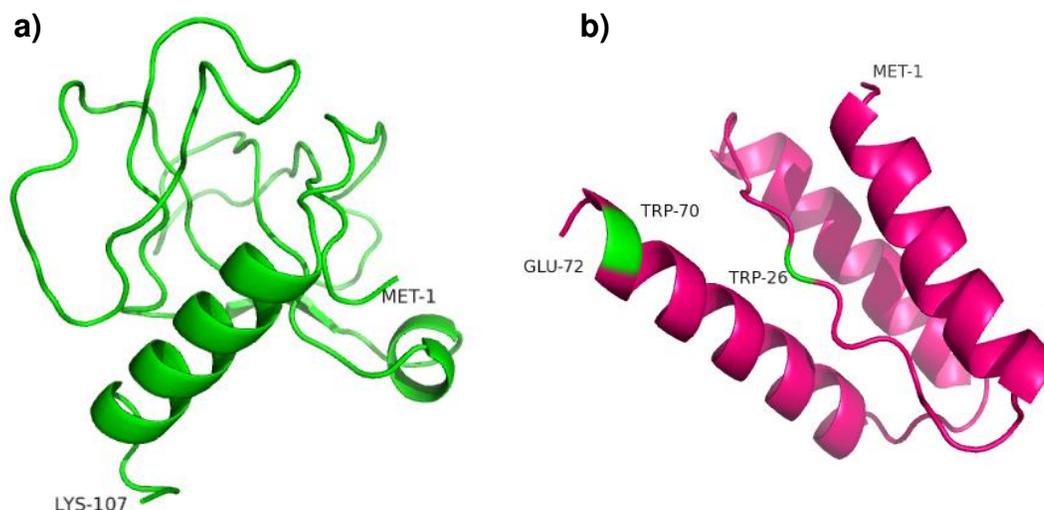


Figura IV-48: a) Representação em cartoon da estrutura da XACb0032 com N- e C- terminais marcados (I-TASSER C-*score* = 1,13, o único gerado pelo programa) e b) Representação em cartoon da estrutura da XACb0033 com N- e C- terminais marcados (I-TASSER C-*score* = -2,41, o valor de maior pontuação dentre os cinco gerados) e os resíduos de Trp (Trp70 e Trp26) em verde (exposto ao solvente).

A partir das estruturas geradas pelo I-TASSER é observado que a XACb0032 apresenta uma única hélice e o restante da estrutura apresenta-se desestruturada (estrutura randômica).

De acordo com o modelo da Figura IV-48b, tem-se que a proteína XACb0033 apresenta três hélices- α , intercaladas por segmentos desordenados. A estrutura conhecida de chaperonas de secreção mostra que o alto conteúdo helicoidal é uma similaridade estrutural entre elas.

Os dados de fluorescência corroboram com a estrutura proposta pelo I-TASSER, pois os dois resíduos de triptofanos da proteína XACb0033, Trp70 e Trp26, apresentam-se mais exposto ao solvente. Além disso, também foi verificado que o Trp26 encontra-se uma região randômica da proteína. No entanto, há discrepâncias entre os dados de CD e NMR com a estrutura proposta pelo I-TASSER assim como para a proteína FlgN. Através dos dados de CD é observado baixo conteúdo α -hélice, que pode ser resultado da baixa concentração usada (amostra de fluorescência). Analisando-se os espectros de NMR 1D e 2D foi observado também esta baixa ocorrência de estrutura secundária. Uma explicação provável é uma possível agregação proteica durante a concentração da amostra ou ainda, uma troca conformacional. No mapa de contorno $\{^{15}\text{N}, ^1\text{H}\}$ HSQC (Figura IV-46) também foi observado ausência de estrutura 3D.

4.3 Proposta de Estrutura para as proteínas FlgN e XACb0033

Nesta parte do trabalho as duas proteínas da *Xac* foram purificadas e analisadas pelas técnicas de ligação cruzada (*cross-linking*) e *footprinting* em colaboração com o grupo do Prof. Fábio C. Gozzo (IQ-UNICAMP).

Na análise de ligação cruzada foram testados três reagentes: DSS (Disuccinimidyl suberate), EDC (1-ethyl-3-(3-dimethylaminoisopropyl) carbodiimide) e SDAD (Succinimidyl 2-([4,4'-azipentanamido]ethyl)-1,3'-dithiopropionate). O DSS tem especificidade com a cadeia lateral de resíduos de lisina ou N-terminal. O EDC reage com a cadeia lateral de resíduos ácidos ativando-os e em seguida, leva a formação de uma ligação amida através de uma reação de condensação do resíduo de ácido com a cadeia lateral de resíduos de lisinas que estejam a uma distância de uma ligação de hidrogênio (3-7Å). Já o SDAD reage através do NHS com a cadeia lateral de um resíduo de lisina [- $(\text{CH}_2)_4\text{NH}_2$] e a reação inespecífica do carbeno formado pela saída de N_2 da porção diazirina.

No entanto, acredita-se que os experimentos de ligação cruzada não foram bem sucedidos em função da baixa ocorrência de resíduos de lisina (K) ao longo da sequência primária das proteínas (pI baixo), e da desativação dos poucos resíduos de lisina presentes, em função da alta ocorrência de resíduos de ácido glutâmico (E). Isso diminuiria o rendimento da reação com os reagentes de ligação cruzada, os quais são, em menor ou maior proporção, lisina-dependentes.

Em seguida, as proteínas FlgN e XACb033 foram submetidas ao experimento de *footprinting*. O reagente usado para a modificação química nesta etapa foi o dietil pirocarbonato (DEPC). Este reagente apresenta alta reatividade à resíduos de histidina, levando a uma modificação covalente no resíduo de histidina (His, H), mas também pode reagir com outros nucleófilos. Os espectros das Figuras IV-49-54 referem-se aos peptídeos modificados pelo DEPC, o que indica que estes encontram-se na superfície da proteínas, que pode ser utilizado em gerar as restrições para modelagem molecular.

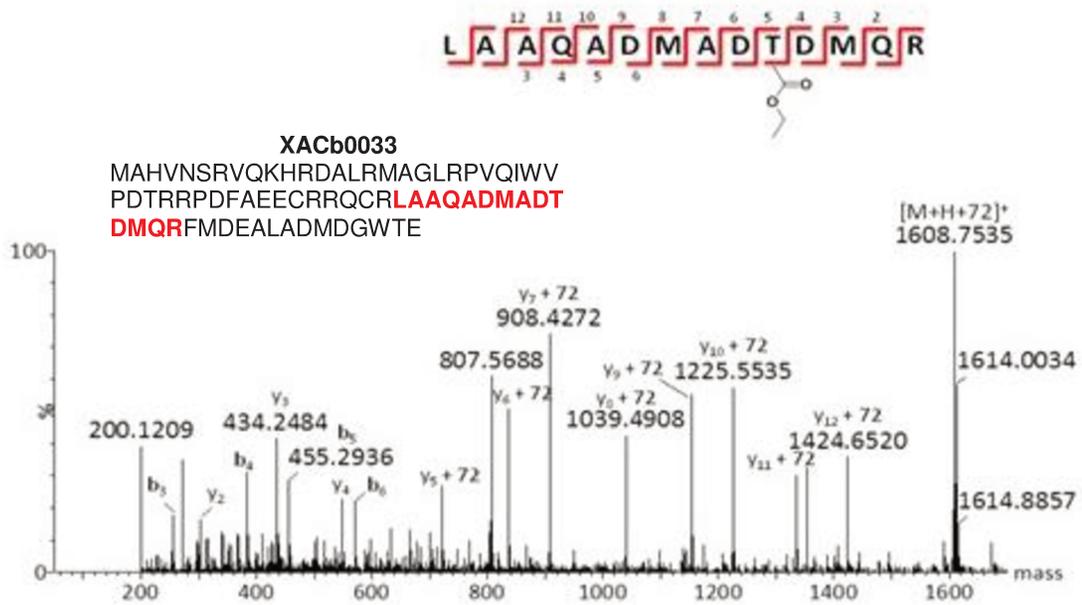


Figura IV-49: Espectro de MS do peptídeo **LAAQADMADT*DMQR** da XACb0033 que reagiu com o DEPC $[M+H+72]^+$.

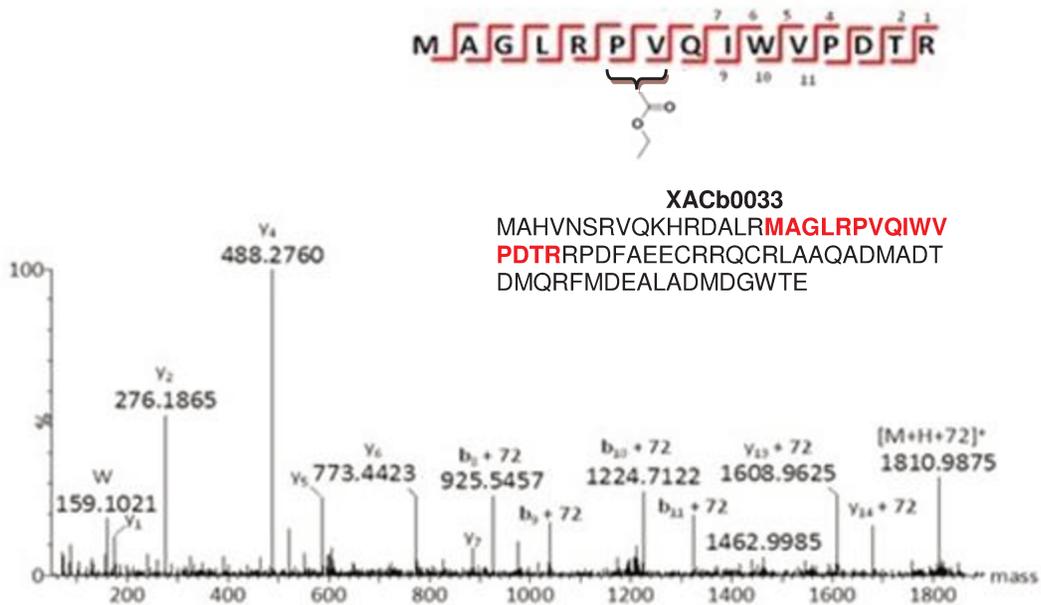


Figura IV-50: Espectro de MS do peptídeo **MAGLRP*V*QIWP DTR** da XACb0033 que reagiu com o DEPC $[M+H+72]^+$. O DEPC reagiu com o resíduo P ou V.

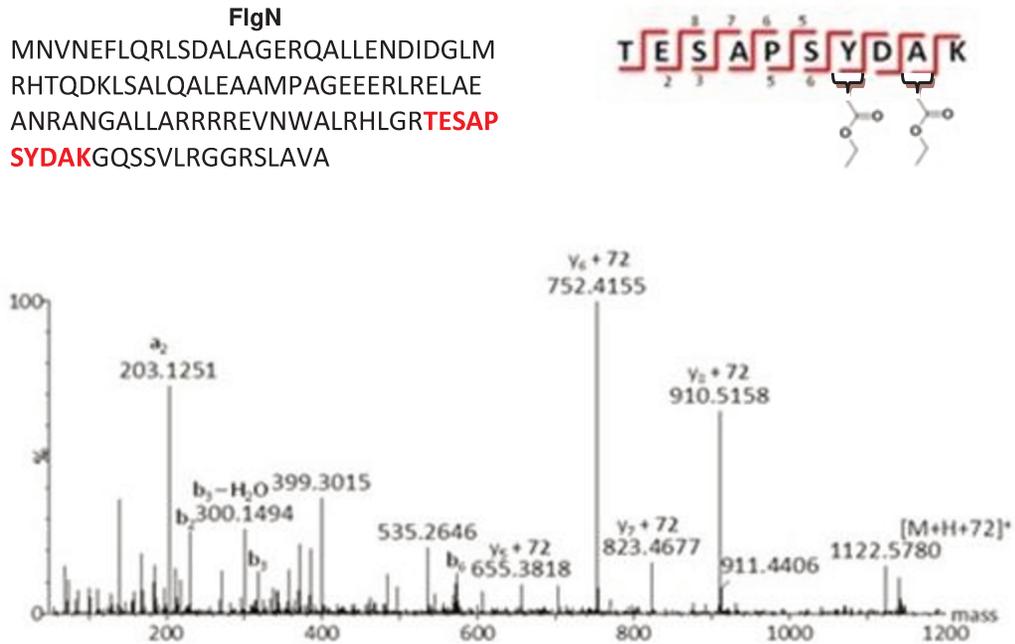


Figura IV-51: Espectro de MS do peptídeo **TESAPSY*DA*K** da FlgN que reagiu com o DEPC $[M+H+72]^+$. O DEPC reage com o resíduo Y ou A.

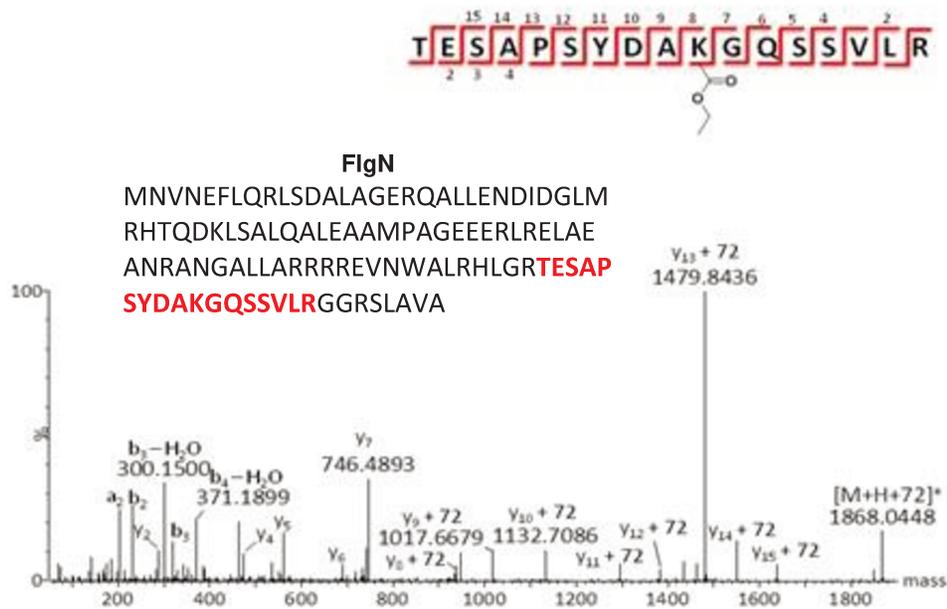


Figura IV-52: Espectro de MS do peptídeo **TESAPSYDAK*GQSSVLR** da FlgN que reagiu com o DEPC $[M+H+72]^+$.

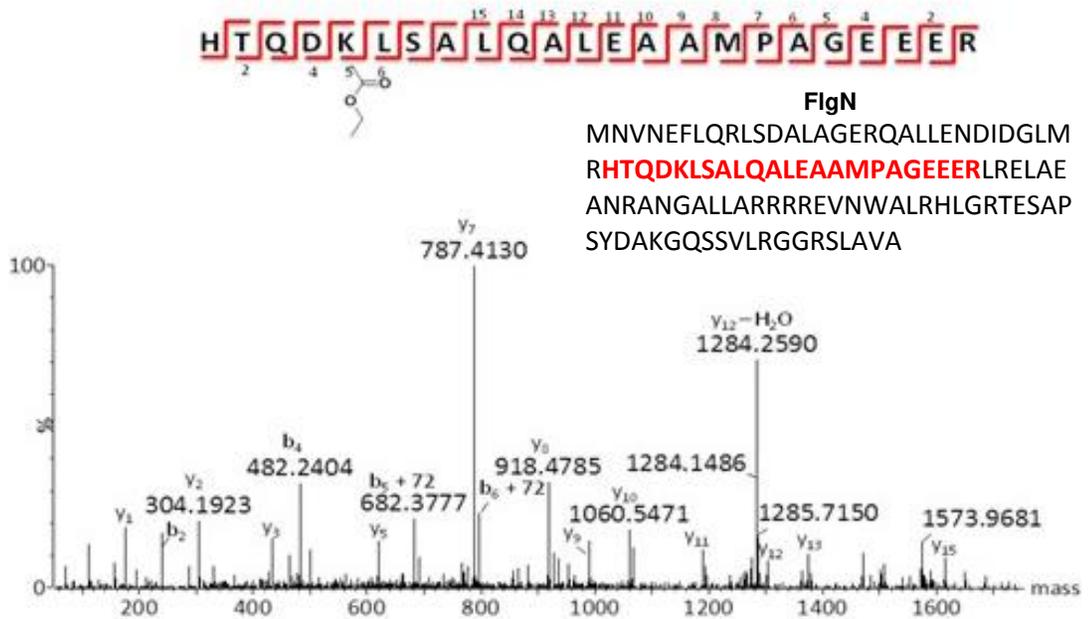


Figura IV-53: Espectro de MS do peptídeo **HTQDK*LSALQALEAAMPAGEEER** da FlgN que reagiu com o DEPC $[M+H+72]^+$.

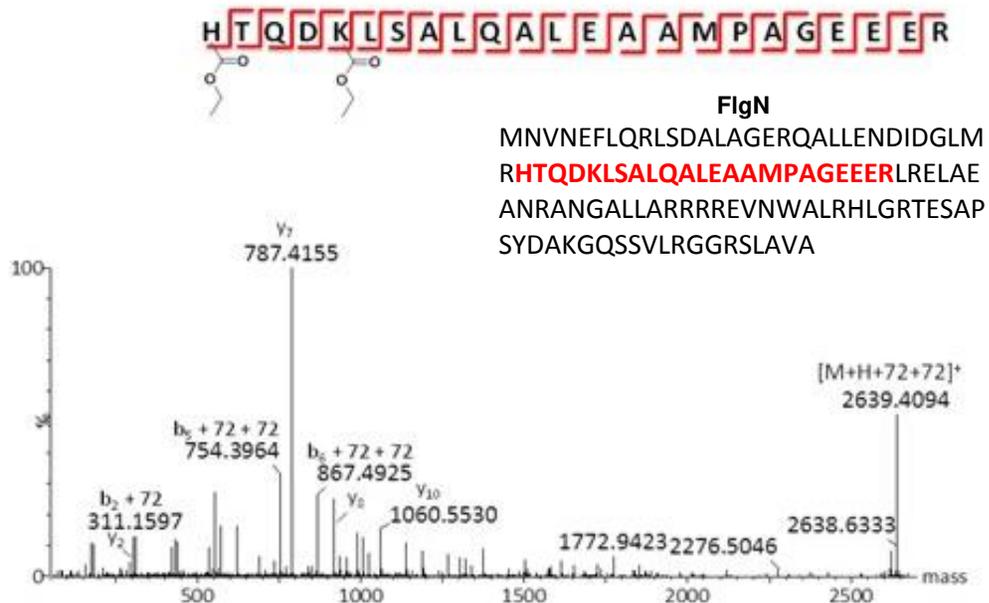


Figura IV-54: Espectro de MS do peptídeo **H*TQDK*LSALQALEAAMPAGEEER** da FlgN que reagiu com o DEPC $[M+H+72]^+$.

A partir dos dados obtidos foi difícil modelar uma estrutura 3D para as XACb0033 e FlgN, uma vez que poucos peptídeos modificados foram identificados após a reação com DEPC. Outro motivo é a ausência de modelos de alta resolução determinados por técnicas experimentais, pois as proteínas em estudo não apresentam homólogas ou apresentam baixa homologia com outras proteínas da família.

Com o objetivo de encontrar a melhor estrutura que condiz com os dados de *footprinting* foram gerados modelos de estrutura 3D para as proteínas XACb0033 e FlgN utilizando os programas computacionais I-TASSER¹¹⁵ e QUARK¹¹⁶ que se baseiam na sequência primária da proteína.

O I-TASSER utiliza o alinhamento com diferentes estruturas conhecidas para calcular uma estrutura tridimensional com base na sequência de aminoácidos. No caso da proteína FlgN, o I-TASSER gerou cinco modelos de estruturas, com um $TM = 0,42 \pm 0,14$ ($TM < 0,17$ indica modelos de topologia randômica) e, no caso da XACb0033 foram obtidas cinco estruturas *in silico*, com um $TM = 0,43 \pm 0,14$. TM é uma medida de similaridade estrutural entre duas estruturas. O QUARK é um programa que constrói um modelo tridimensional da proteína baseado somente na sequência de aminoácidos. Como ele não se baseia em estruturas conhecidas é apropriado para predição de proteínas que não apresentam homologia com outras proteínas. Para cada sequência de aminoácidos o QUARK gera 10 modelos de estruturas. A partir de 15 modelos de estrutura calculados para cada proteína (I-TASSER e QUARK), duas restrições foram utilizadas para tentar propor uma estrutura plausível. Uma restrição aplicada foi a SASA¹¹³ e a outra, obtida *via* RMSD (desvio da média quadrática) em relação a estrutura proposta pelo programa I-TASSER com maior score.

Os valores de SASA para peptídeos gerados no experimento de *footprinting* estão sumarizados na Tabela IV-2-7. Os modelos de 1-5 se referem aos modelos gerados pelo I-TASSER e de 6-15 os modelos gerados pelo QUARK. Os números em vermelho são valores de SASA em destaque dos possíveis modelos.

Tabela IV-2: Valores de SASA (\AA^2) para o peptídeo LAAQADMAD**T**DMQR gerado a partir da análise de footprinting da proteína XACb0033. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de *footprinting*

Peptídeo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Leu (L)	123	41	130	83	19	127	57	116	97	91	56	91	88	107	92
Ala (A)	5,4	52	28	1,2	36	18	0	34	1,0	2,1	1,8	1,0	0	0,7	0
Ala (A)	25	89	21	35	59	12	62	33	51	64	59	48	61	65	51
Gln (Q)	169	103	181	120	98	136	113	141	138	102	146	144	135	137	130
Ala (A)	62	70	59	48	0,4	70	39	80	41	79	49	20	32	23	25
Asp (D)	134	95	86	95	90	83	144	91	115	119	107	124	124	154	149
Met (M)	42	178	68	31	174	147	86	142	158	144	167	155	165	123	127
Ala (A)	44	25	75	47	43	50	62	53	33	14	61	30	63	7,7	33
Asp (D)	82	51	65	53	44	26	49	6,3	26	21	23	23	19	59	97
Thr (T)	95	74	109	105	90	56	86	77	81	82	72	76	75	94	78
Asp (D)	77	78	51	50	36	65	83	87	91	79	78	78	86	5,8	32
Met (M)	2,3	11	9	4,0	16	21	0,1	20	25	3,8	26	22	24	2,7	3,0
Gln (Q)	86	66	89	101	74	35	18	27	46	82	57	45	38	57	69
Arg (R)	157	165	102	142	104	148	145	163	158	151	154	150	154	119	138

Tabela IV-3: Valores de SASA (\AA^2) para o peptídeo MAGLR**PV**QIWVPDTR gerado a partir da análise de footprinting da proteína XACb0033. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de *footprinting*

Peptídeo 2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pro (P)	53	18	67	38	100	91	106	90	128	124	112	121	83	92	107
Val (V)	0,5	16	6,4	6,9	61	38	21	3,5	37	20	41	2,7	1,0	3,9	17
Gln (Q)	69	75	55	48	63	74	64	56	35	57	35	14	40	28	69
Ile(I)	1,6	21	56	9,4	95	0,7	2,7	0,3	2,0	0	67	8,9	3,7	7,8	0,2
Trp (W)	75	157	42	88	132	15	91	103	3,2	25	3,4	93	83	25	97

Tabela IV-4: Valores de SASA (\AA^2) para o peptídeo TESAPS**YDAK** gerado a partir da análise de footprinting da proteína FlgN. Os aminoácidos em vermelho são os resíduos que reagem com o DEPC na análise de *footprinting*

Peptídeo 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ser (S)	42	63	77	90	81	38	21	29	9,3	4,1	3,2	0	0,1	5,6	1,5
Tyr (Y)	196	22	164	163	48	93	92	66	92	116	102	65	27	111	81
Asp (D)	128	44	91	32	17	76	20	90	31	15	29	44	68	38	73
Ala (A)	24	56	69	70	64	70	97	104	101	45	85	66	110	96	102
Lys (K)	150	99	158	140	154	203	158	149	160	159	195	182	183	168	203

Tabela IV-5: Valores de SASA (\AA^2) para o peptídeo TESAPSYDA**K**GQSSVLR gerado a partir da análise de footprinting da proteína FlgN. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de *footprinting*

Peptídeo 2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Asp (D)	128	44	91	32	17	76	20	90	31	15	29	44	68	38	73
Ala (A)	24	56	69	70	64	70	97	104	101	45	85	66	110	96	102
Lys (K)	150	99	158	140	154	203	158	149	186	159	195	182	183	168	203
Gly (G)	45	0,6	61	33	16	58	43	4,4	53	53	60	47	33	31	17
Gln (Q)	27	109	36	19	57	94	101	39	150	116	69	138	95	100	155

Tabela IV-6: Valores de SASA (\AA^2) para o peptídeo HTQD**K**LSALQALEAAM PAGEEER gerado a partir da análise de footprinting da proteína FlgN. O aminoácido em vermelho é o resíduo que reage com o DEPC na análise de *footprinting*

Peptídeo 2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gln (Q)	63	114	94	99	118	103	109	103	93	111	104	106	58	107	91
Asp (D)	99	44	90	92	85	67	97	93	53	50	98	88	47	91	47
Lys (K)	16	6,8	48	75	9,6	74	40	69	6,9	37	32	71	1,4	27	0
Leu (L)	15	76	74	97	101	72	78	85	20	85	78	66	43	89	37
Ser (S)	64	60	71	63	56	64	61	54	55	64	68	68	39	61	61

Tabela IV-7: Valores de SASA (\AA^2) para o peptídeo **HTQDKLSALQALEA** AMPAGEEER gerado a partir da análise de footprinting da proteína FlgN. Os aminoácidos em vermelho são os resíduos que reagem com o DEPC na análise de *footprinting*

Peptídeo 2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
His (H)	67	39	135	127	45	66	71	83	52	34	81	85	28	49	50
Thr (T)	0	12	44	63	26	18	45	48	3,6	16	15	29	0	59	8,0
Gln (Q)	63	114	94	99	118	103	109	103	93	111	104	106	58	107	91
Asp (D)	99	44	90	92	85	67	97	93	53	50	98	88	47	91	47
Lys (K)	16	6,8	48	75	9,6	74	40	69	6,9	37	32	71	1,4	27	0

O experimento de SASA¹²⁴ mede a acessibilidade na superfície de Van der Waals da estrutura 3D. O raio de Van der Waals usado neste experimento foi o raio da água (aproximadamente 1,4 \AA) e, portanto, a superfície descrita é geralmente referida como a superfície acessível ao solvente. A partir de uma estrutura 3D, Alphasurf¹²⁵ calcula a área superficial exata usando um método analítico baseado na alpha shapes theory (descrito para um conjunto determinado de pontos). O ProtSA remove todos os átomos de hidrogênio antes de efetuar os

¹²⁴ a) Kim, R.; Choi, C. A linear function for the approximation of accessible surface area of proteins. *Protein Peptide Lett* **2006**, *13*, 549-553. b) Ojha, L.; Chaturvedi, A.; Bhardwaj, A. Role of solvent accessible surface area (ASA+) and Baladen 3D index ($j3D$) in binding affinity of Tibo derivatives. *Int J Pharm Science Res* 2011, *12*, 3177-3182. c) Lu, C.; Chen, S.; Bretaña, N.; Cheng, T.; Lee, T. Carboxylator: incorporating solvent-accessible surface area for identifying protein carboxylation sites. *J Comput Aided Mol Des* **2011**, *25*, 987-995. d) Wang, C.; Xi, L.; Li, S.; Liu, H.; Yao, X. A sequence-based computational model for the prediction of the solvent accessible surface area for α -helix and β -barrel transmembrana residues. *J Comput Chem* **2011**, 11-17. e) Faraggi, E.; Zhang, T.; Yang, Y.; Kurgan, L.; Zhou, Y. SPINE X: Improving protein secondary structure prediction by multistep learning coupled with prediction of solvent accessible surface area and backbone torsion angles. *J Comput Chem* **2011**, 259-267.

¹²⁵ Alphasurfer: parte do software ProGeom (<http://koehllab.genomecenter.ucdavis.edu/people/koehl/research/proshape>) para cálculos de SASA.

cálculos de SASA. Conjunto de raios atômicos de Van der Waals usados é apresentado em Naccess (<http://www.bioinf.manchester.ac.uk/naccess/>), que por sua vez foram tomados do Chothia.¹²⁶ Para átomos em resíduos diferentes dos 20 tipos padrões, um raio de 1,80 Å é usado.

A partir dos dados de SASA é esperado que os resíduos que estão na superfície da proteína enovelada estejam em maior contato com água. Sendo assim, os resíduos que estão mais expostos (com um valor de SASA maior) seriam os mais prováveis de reagir com o DEPC. Através dos dados de SASA (Tabela IV-2-7) foram observados os maiores valores (números indicados em vermelho) para o resíduos que estão diretamente ligados ao DEPC, resultado da análise de *Footprinting*.

Para a proteína XACb0033 as estruturas que condizem com os melhores valores de SASA são os modelos 1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14 e 15 e para a proteína FlgN condiz com as estruturas 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 e 15. Para o peptídeo TESAPSYDAK da proteína FlgN acredita-se que o DEPC reagiu com o resíduo de Y e para o peptídeo MAGLRPVQIWVPDTR da proteína XACb0033, o DEPC reagiu com o resíduo de P. Estes aminoácidos foram os selecionados pois são os que apresentam maior valor de SASA para cada peptídeo.

Outra restrição usada para propor modelos estruturais que melhor se encaixam a estes dados foi o cálculo dos valores de RMSD para as proteínas XACb0033 e FlgN. Como estrutura modelo para estes cálculos foram usados as estruturas do I-TASSER de melhor *C-score*. O valor de *C-score* determina a qualidade dos modelos preditos pelo I-TASSER e um valor de *C-score* maior significa um modelo de maior confiança.

¹²⁶ Chothia, C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* **1976**, *105*, 1–12.

Para a XACb0033 foram obtidos 5 modelos de estruturas gerados pelo I-TASSER, porém, a estrutura do modelo 1 teve um *score* de -2,41 (Figura IV-55) e foi escolhido como melhor modelo.

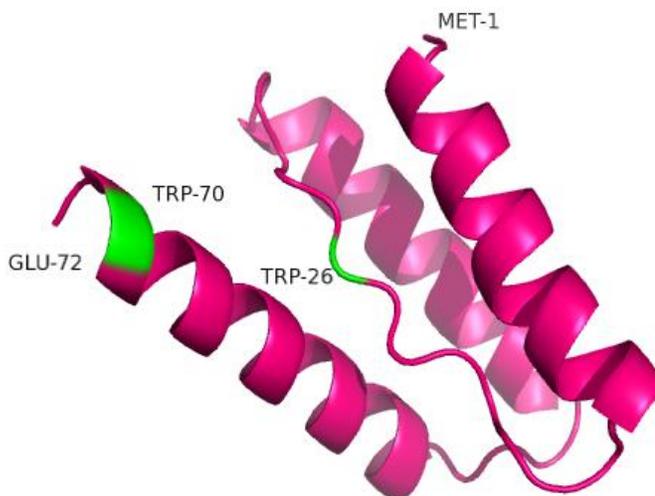


Figura IV-55: Representação em *cartoon* da estrutura da XACb0033 com N- e C-terminais marcados (I-TASSER C-*score* = -2,41, o valor de maior pontuação dentre os cinco gerados) e os resíduos de Trp (W70 e W26) em verde (exposto ao solvente).

Para a proteína FigN de 5 modelos de estruturas gerados pelo I-TASSER foi escolhido como melhor a estrutura de número 1 apresentando um *score* de -2,51 (Figura IV-56). Esta estrutura calculada pelo I-Tasser foi o modelo que melhor sobrepôs o envelope experimental obtido pelos dados de SAXS publicado recentemente.⁶²

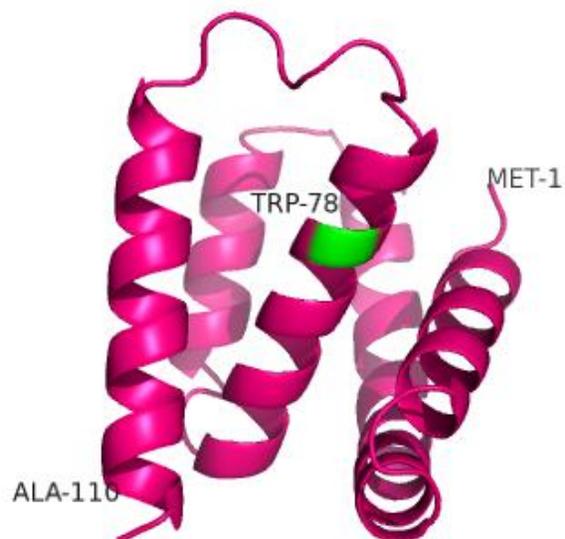
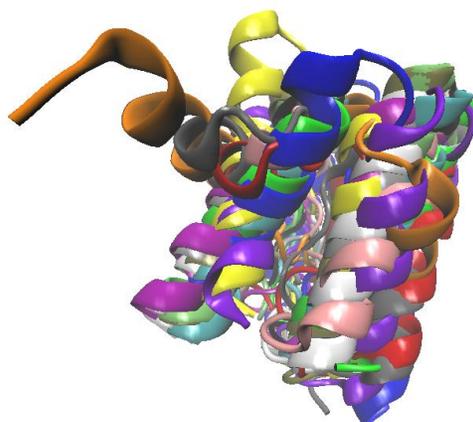


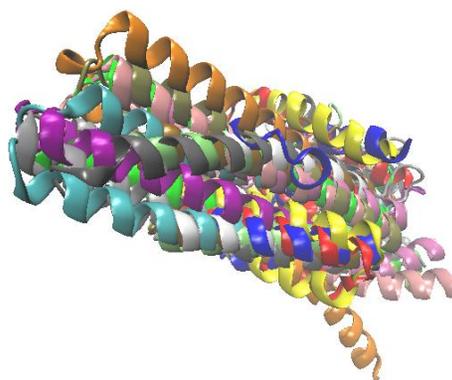
Figura IV-56: Representação em cartoon da estrutura da FlgN com N- e C-terminais marcados (I-TASSER C-score = -2,51) e o resíduo de Trp (W78) em verde (exposto ao solvente).

Estes modelos foram usados como referencia para o cálculo de RMSD. Neste cálculo os modelos foram alinhados com o modelo selecionado, no caso das XACb0033 e FlgN com modelos 1 e, em seguida foi calculado o desvio da média quadrática. O alinhamento e os cálculos foram realizados no programa VMD 1.9.1 e estão mostrados nas Figuras IV-57 e IV-58.



Modelo	RMSD (Å)	Modelo	RMSD (Å)
1	0,000	8	6,998
2	13,132	9	11,298
3	10,884	10	11,377
4	11,702	11	8,084
5	9,970	12	9,824
6	8,564	13	8,462
7	9,967	14	11,499
		15	9,451

Figura IV-57: Os 15 modelos alinhados gerados pelos programas I-TASSER e QUARK para a proteína XAcB0033 e ao lado valores de RMSD gerados no programa VMD usando-se como modelo a estrutura 1.



Modelo	RMSD (Å)	Modelo	RMSD (Å)
1	0,000	8	15,740
2	10,532	9	12,309
3	13,481	10	15,849
4	16,382	11	16,607
5	10,257	12	15,883
6	15,732	13	12,323
7	15,830	14	15,766
		15	11,973

Figura IV-58: Os 15 modelos alinhados gerados pelos programas I-TASSER e QUARK para a proteína FlgN e ao lado valores de RMSD gerados no programa VMD usando-se como modelo a estrutura 1.

Com base nos dados de RMSD, para a proteína XACb0033 os menores valores correspondem as estruturas 6, 8, 11 e 13. Dentre as opções selecionadas para a XACb0033 através dos dados de SASA o modelo que mais se aproxima ao escolhido (modelo 1 do I-TASSER) foi o modelo 6 obtido pelo QUARK. Já para a proteína FlgN, os menores valores condiz com as estruturas 2, 5, 9, 13 e 15. Dentre as opções selecionadas para a FlgN através dos dados de SASA, modelos 5 e 13 foram os mais similares com o modelo 1 (I-TASSER). As estruturas para os modelos selecionados estão apresentados nas Figuras IV-59 e 60.

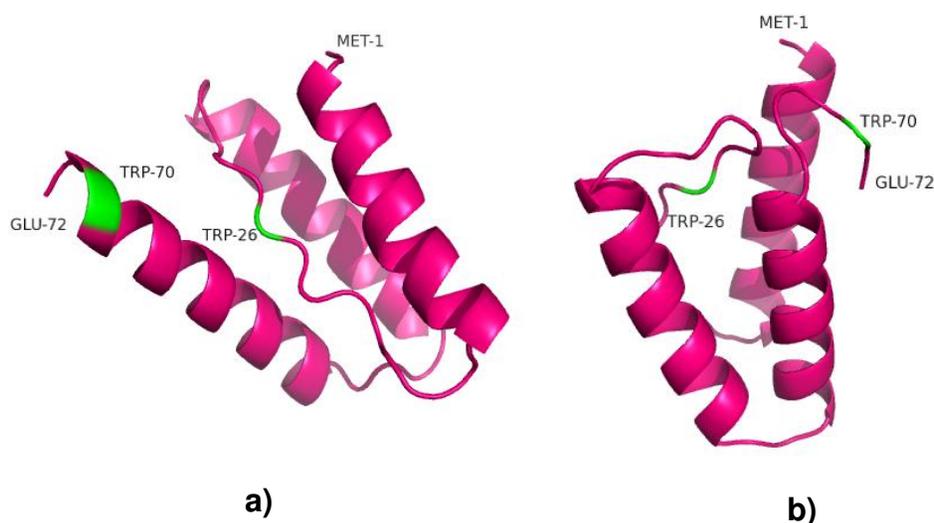


Figura IV-59: Representação em *cartoon* da estrutura da XACb0033 com N- e C-terminais marcados e os resíduos de Trp (W70 e W26) em verde. (a) Modelo 1 gerado pelo I-TASSER. (b) Modelo 6 gerado pelo Quark.

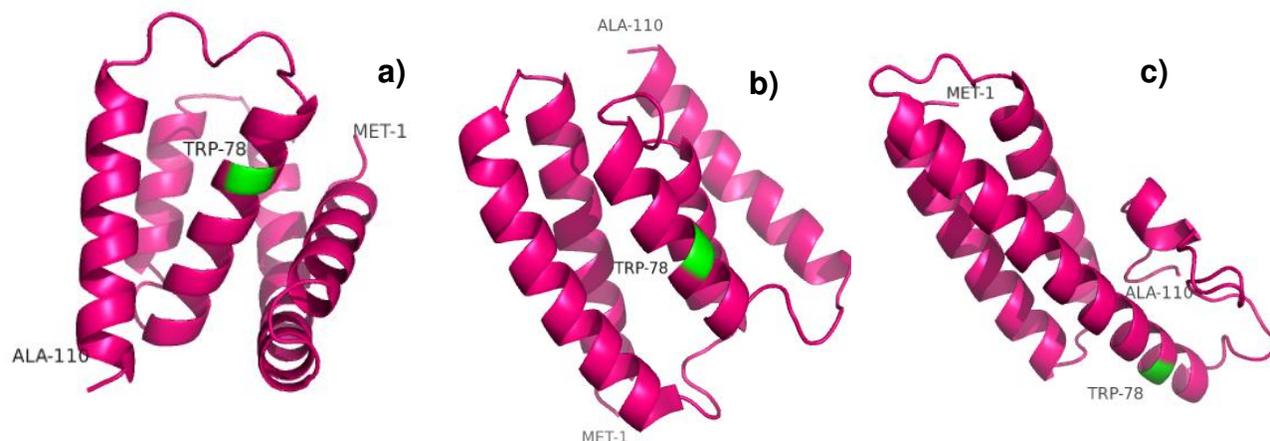


Figura IV-60: Representação em *cartoon* da estrutura da FlgN com N- e C-terminais marcados e o resíduo de Trp (W78) em verde. (a) Modelo 1 gerado pelo I-TASSER. (b) Modelo 5 gerado pelo I-TASSER. (c) Modelo 13 gerado pelo Quark.

Através das estruturas tridimensionais geradas pelos programas I-TASSER e QUARK, de acordo com os dados de SASA e restrições impostas (RMSD), foi observado que as proteínas FlgN e XACb0033 apresentam uma estrutura praticamente do tipo α -hélice.

A proteína XACb0033 apresenta dois resíduos de Trp em sua estrutura primária, Trp70 e Trp26, que se encontram em uma região desordenada da proteína no caso do modelo 6 gerado pelo programa QUARK e, Trp70 em região ordenada e Trp26 em região randômica, no caso do modelo 1 gerado pelo programa I-TASSER. Ambos modelos corroboram com os dados de fluorescência para esta proteína que apresentou um máximo em 346 nm, característicos de Trp mais expostos e mais acessíveis ao solvente.

Além disso, também foi verificado que esta proteína se comporta praticamente como um dímero em solução através de espectrometria de massas. No entanto, há discrepâncias entre os dados de CD e NMR com as estruturas propostas.

Através dos dados de CD foi observado baixo conteúdo α -hélice, e, além disso, o mesmo foi observado nos espectros de NMR 1D e 2D que indicaram agregação proteica ou troca conformacional além de ausência de estrutura 3D.

Já a proteína FlgN apresenta um único resíduo de Trp em sua estrutura primária, Trp78, que se encontra em uma região ordenada da proteína para todos modelos selecionados e além disso, este encontra-se mais exposto ao solvente e mais acessíveis ao solvente corroborando com os dados de fluorescência desta proteína. Através da análise de massas foi observado que esta proteína se comporta como um monômero em solução. No entanto, assim como para a XACb0033, há discrepâncias entre os dados de CD e NMR com as estruturas propostas.

Através dos dados de CD foi observado baixo conteúdo α -hélice, e, além disso, o mesmo foi observado nos espectros de NMR 1D e 2D que indicaram agregação proteica ou troca conformacional além de ausência de estrutura 3D.

4.4 Interações Proteínas e Ligantes

Na segunda parte do trabalho, foram estudadas as interações entre a proteína Hsp90 da laranja com os três ligantes ATP, ADP e ATP γ S utilizando método de STD-NMR. Estas interações e a interação da proteína Hsp90 com a geldanamicina foram estudadas aplicando a técnica de fluorescência de emissão, também. A função da Hsp90 da laranja, conforme a nossa hipótese, é de ser uma peça chave em re-estabelecimento de enovelamento correto das proteínas secretadas pela *Xac*. Nesta etapa, as ATPases como a Hsp100 e a Hsp90 utilizam a hidrólise do ATP como fonte de energia no processo de enovelamento.

Os dados obtidos para a proteína Hsp100 da cana de açúcar foram utilizados como padrão para interação positiva e de comparação com a Hsp90 da laranja. Os espectros para os ligantes ATP, ATP γ S e ADP estão apresentados nas Figuras IV-61-62.

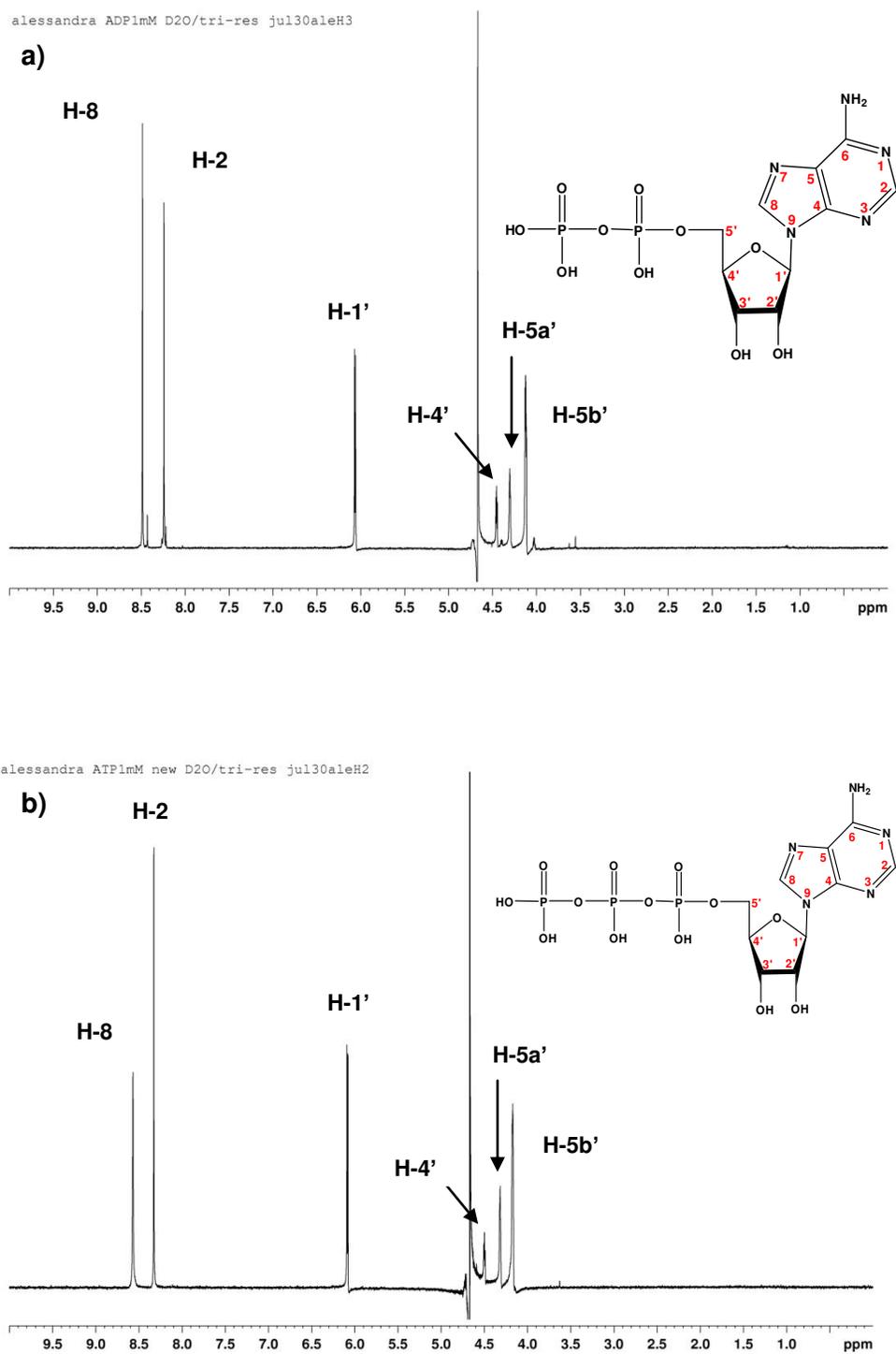


Figura IV-61: Espectros de ^1H NMR (499,99 MHz) em D_2O . (a) ADP $1 \text{ mmol}\cdot\text{L}^{-1}$ e (b) ATP $1 \text{ mmol}\cdot\text{L}^{-1}$.

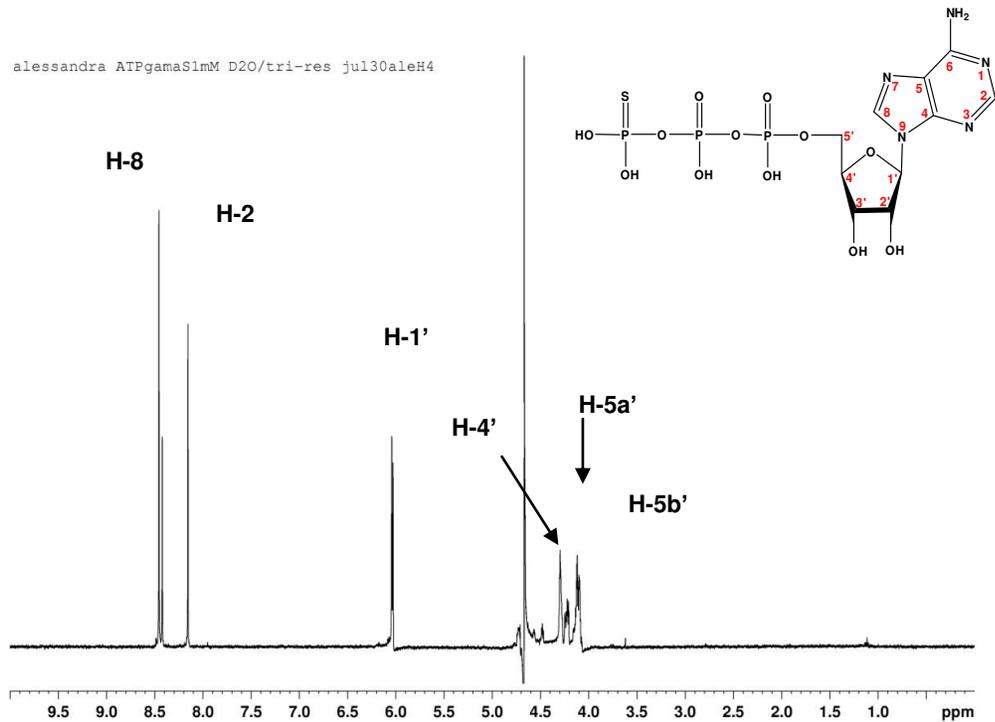


Figura IV-62: Espectros de ^1H NMR (499,99 MHz) em D_2O do $\text{ATP}_{\gamma}\text{S}$ 1 mmol.L^{-1} .

Os espectros de STD obtidos para a interação da proteína Hsp100 da cana de açúcar e $\text{ATP}_{\gamma}\text{S}$, ATP, e ADP estão ilustrados nas Figuras IV-63-65.

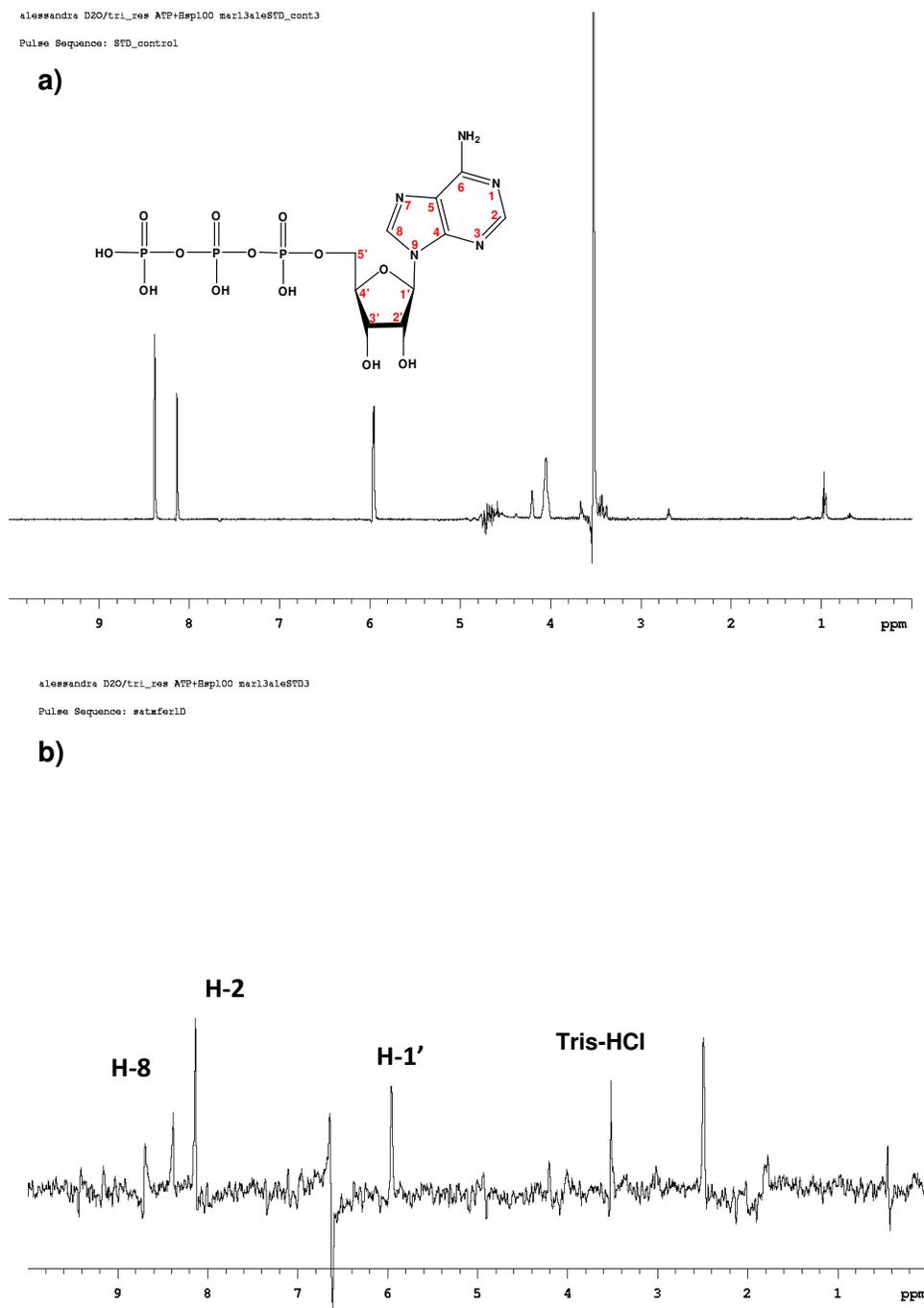


Figura IV-63: Espectros de NMR da proteína Hsp100 com o nucleotídeo ATP na razão molar 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.

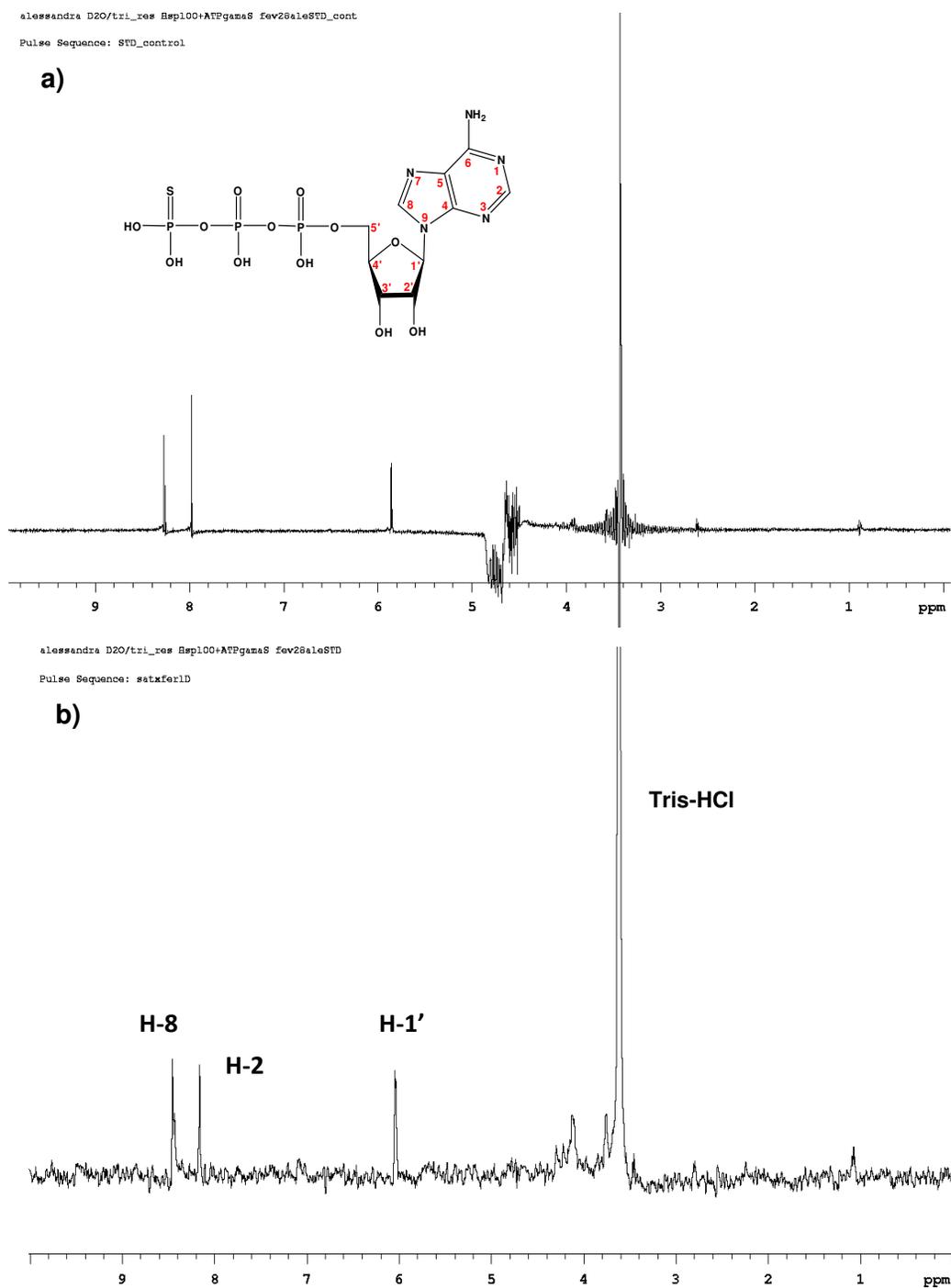


Figura IV-64: Espectros de NMR da proteína Hsp100 com o nucleotídeo ATP γ S na razão molar 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.

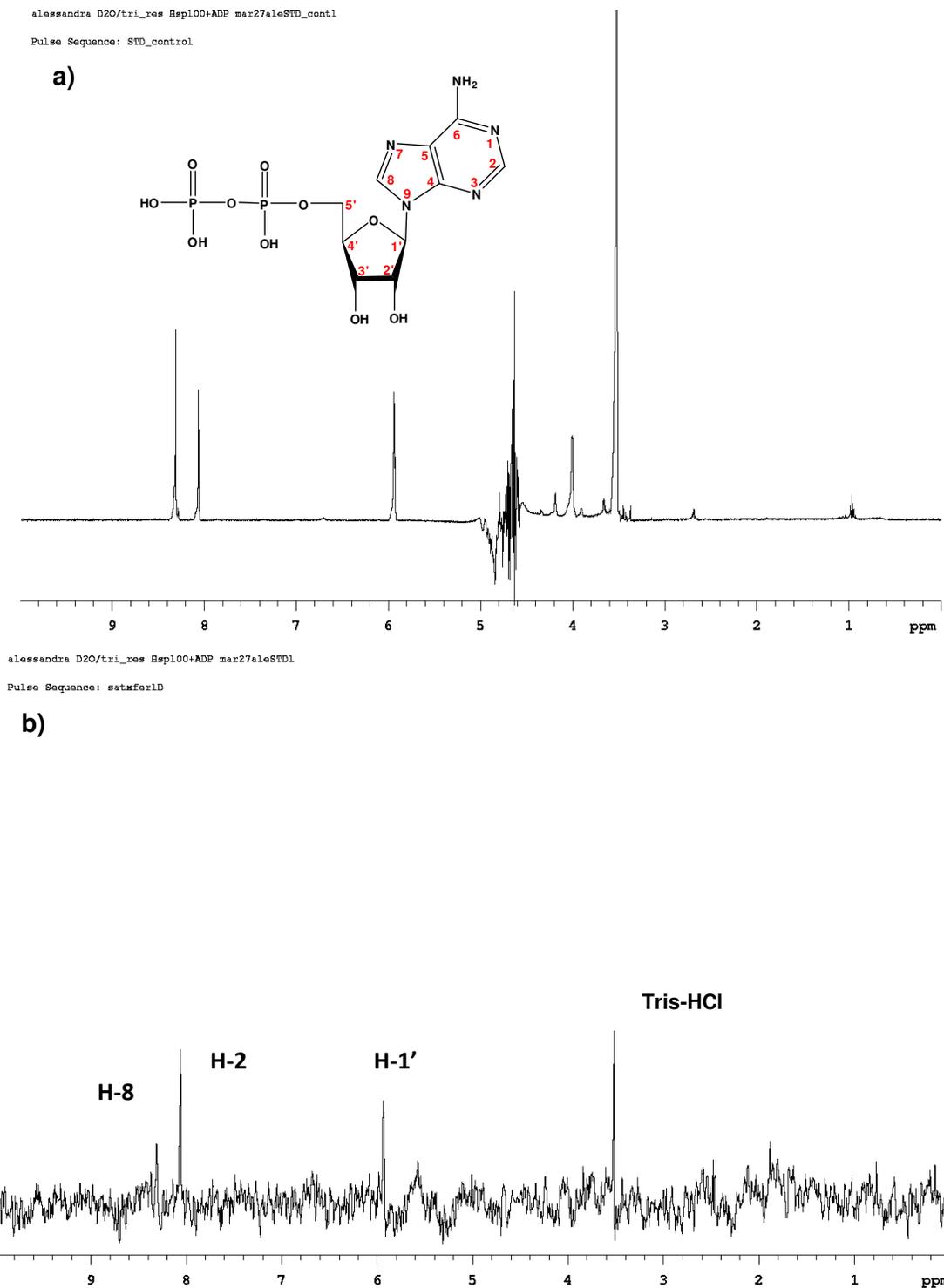


Figura IV-65: Espectros de NMR da proteína Hsp100 com o nucleotídeo ADP na razão molar 1:100. (a) Espectro de STD-controle e (b) Espectro de STD.

Comparando os espectros de STD (Figuras IV-63-65), podemos notar que apenas os hidrogênios H-8 e H-2 da adenosina e H-1' da ribose dos nucleotídeos interagem diretamente com a proteína analisada (Hsp100). Por isso, acredita-se que estes ligantes interagem com o sítio ligante da Hsp100 da cana de açúcar de modo que, os fosfatos estão orientados para fora do sítio ligante e a adenina e ribose encontram-se no bolsão do sítio ligante.¹⁰¹ Através dos resultados de STD foi possível calcular o fator de STD, que é um parâmetro que nos fornece informação acerca da porcentagem de interação dos hidrogênios do ligante que estão interagindo com a proteína. Sendo assim, o mapa de epítipo do ligante pode ser determinado. Os valores do fator STD para os experimentos realizados com as misturas: Hsp100/ATP, Hsp100/ADP e Hsp100/ATP γ S estão apresentados na Tabela IV-8. Os valores obtidos indicam que a Hsp100 interagiu efetivamente com todos os ligantes. Estes valores foram calculados do seguinte modo:

$$\text{Fator de STD (\%)} = (I_{0(\text{STDcontrol})} - I_{(\text{STD})}) / I_{0(\text{STDcontrol})} \quad (\text{Equação IV-1})$$

Em que $I_{0(\text{STDcontrol})}$ representa a intensidade do sinal no espectro de STD controle e $I_{(\text{STD})}$ representa a intensidade do sinal no espectro de STD.

Tabela IV-8: Mapa de epítipos obtido em interações da Hsp100 da cana de açúcar com ATP γ -S, ATP e ADP.

Hidrogênio	Hsp100+ATP γ S	Hsp100+ATP	Hsp100+ADP
H-8	99%	100%	100%
H-2	100%	100%	99%
H-1'	98%	100%	99%

As interações da proteína Hsp100 da cana de açúcar com o nucleotídeo ATP γ S também foram observadas no experimento de HETCOR através de mudança dos deslocamentos químicos de alguns átomos de carbono do ligante (IV-66).

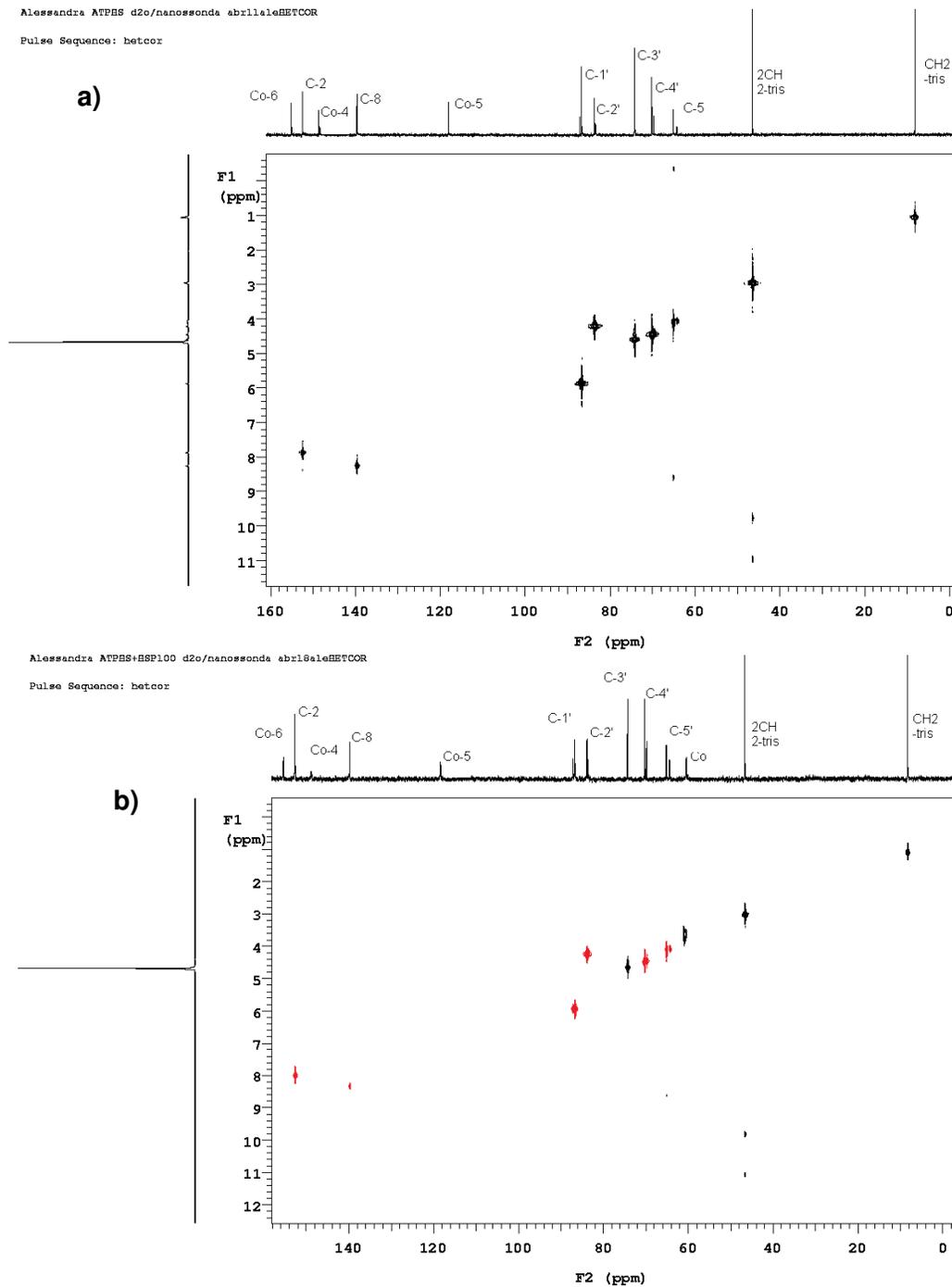


Figura IV-66: Mapas de contorno de experimentos HETCOR $\{^{13}\text{C}-^1\text{H}\}$. (a) $\text{ATP}\gamma\text{S}$ (40 μL de D_2O + 5,0 mg de $\text{ATP}\gamma\text{S}$) e (b) $\text{ATP}\gamma\text{S}$ + Hsp100 (20 μL de D_2O + 2,5 mg de $\text{ATP}\gamma\text{S}$ + 20 μL de Hsp100, 120 $\mu\text{mol}\cdot\text{L}^{-1}$). Em vermelho estão destacadas as correlações $^{13}\text{C}-^1\text{H}$ em $\text{ATP}\gamma\text{S}$ que sofreram deslocamentos após complexação com a Hsp100.

Como visto pelas representações na Figura IV-66, Hsp100 interage com a ATP γ -S modificando os deslocamentos químicos dos sinais (apresentados em vermelho na tabela IV-9) nos mapas de contorno de HETCOR. As mudanças observadas nos deslocamentos químicos estão resumidas na Tabela IV-9.

Tabela IV-9: Atribuição dos sinais de HETCOR obtidos em interações da Hsp100 com ATP γ -S em comparação com ATP γ -S. A diferença de deslocamentos químicos (CICS, Complexation-Induced Chemical Shift) foram calculados.

Atribuição	HETCOR - ATP γ S		HETCOR- ATP γ S + Hsp100		CICS	
	Carbono (ppm)	Hidrogênio (ppm)	Carbono (ppm)	Hidrogênio (ppm)	Carbono (ppm)	Hidrogênio (ppm)
C-5'	65,0	4,1	65,1	4,1	0,1	0
C-4'	70,1	4,2	70,2	4,2	0,1	0
C-3'	74,2	4,4	74,2	4,4	0	0
C-2'	83,6	4,6	83,8	4,5	0,2	0,1
C-1'	86,6	5,9	86,7	5,9	0,1	0
C₀-5	118,0	-----	118,3	-----	0,3	-----
C-8	139,6	8,2	139,8	8,0	0,2	0,2
C₀-4	148,6	-----	148,8	-----	0,2	-----
C₀-6	155,0	-----	155,4	-----	0,4	-----
C-2	152,4	7,8	152,6	8,3	0,2	0,5

Quanto à precisão dos dados de deslocamento químico adquiridos para o CICS, pode-se determinar o valor mínimo exigido de mudança para deslocamento químico tendo em conta a preparação dos experimentos. Como ambos os experimentos, na presença e ausência da proteína ligante, foram realizados no mesmo equipamento, em sequência, e com a manutenção da mesma temperatura da amostra, o único erro de medida esperado vem da resolução espectral utilizada no experimento. Para os valores do Carbono, utilizou-se uma janela espectral de 180 ppm, adquirida com 2048 pontos (dimensão direta), obtendo-se assim uma resolução de 0,087ppm por ponto, tornando válida qualquer diferença de deslocamento químico medida acima desse valor. Já para o Hidrogênio, utilizou-se uma janela espectral de 13ppm, adquirida com 128 pontos (dimensão indireta), obtendo desta forma uma resolução de 0,101ppm por ponto, garantindo assim que qualquer diferença de deslocamento químico medido acima deste valor representa interação da proteína com o ligante.

A partir dos resultados acredita-se que os ligantes interagem com a proteína Hsp100 através dos hidrogênios H-2 e H-8 da adenosina e H-1' da ribose. Os ligantes entram no sítio ligante da proteína com os fosfatos direcionados para fora do sítio.

A partir dos resultados positivos para interação entre a Hsp100 da cana de açúcar e os nucleotídeos, foram iniciados os experimentos com a Hsp90 da laranja. Os experimentos de STD e STD controle foram preparados do mesmo modo que para a Hsp100 e os resultados obtidos estão mostrados nas Figuras IV-67 e 68.

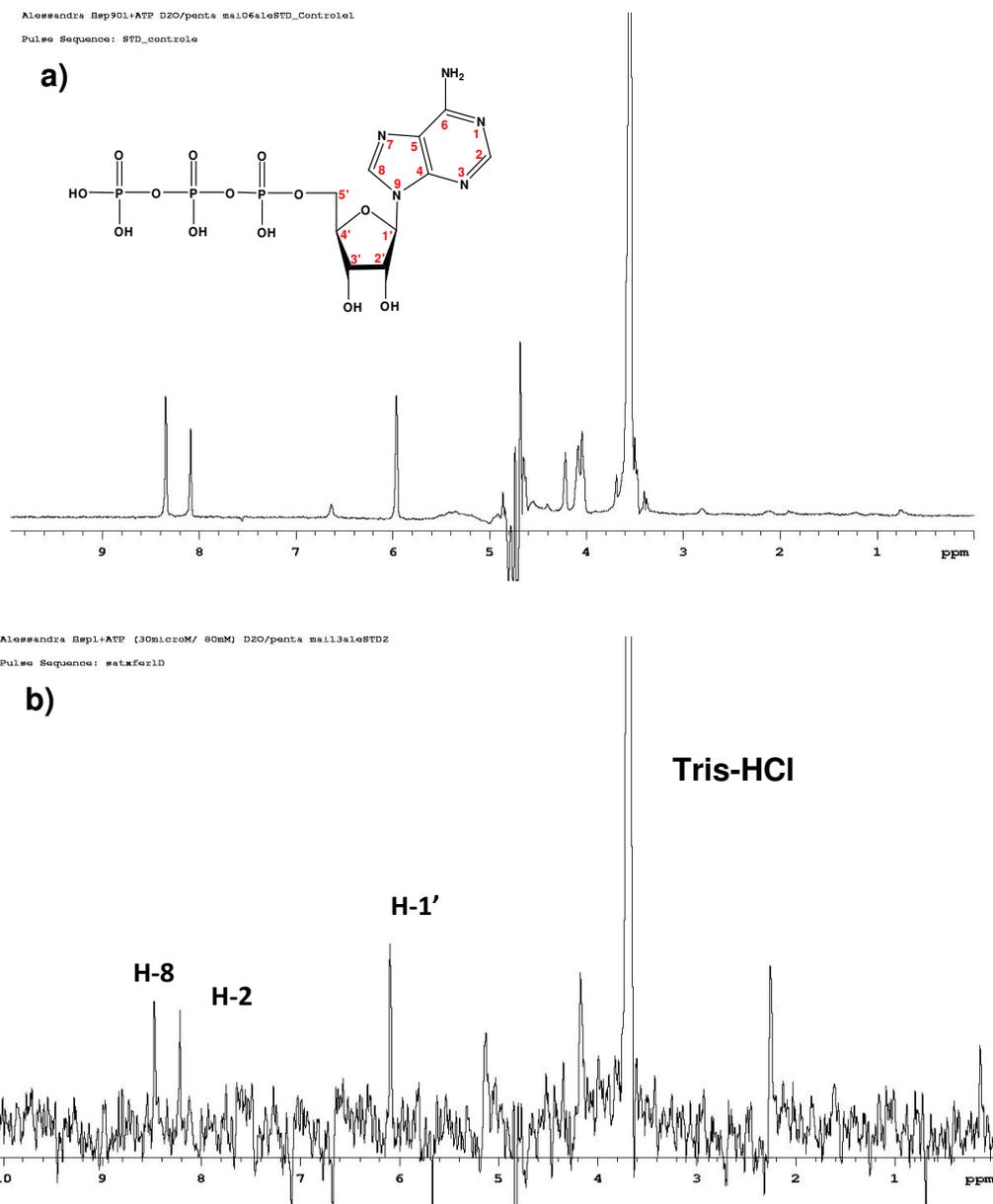


Figura IV-67: Espectros de NMR da proteína Hsp90 da laranja com o nucleotídeo ATP na razão molar 1:100. (a) Espectro de STD-control e (b) Espectro de STD.

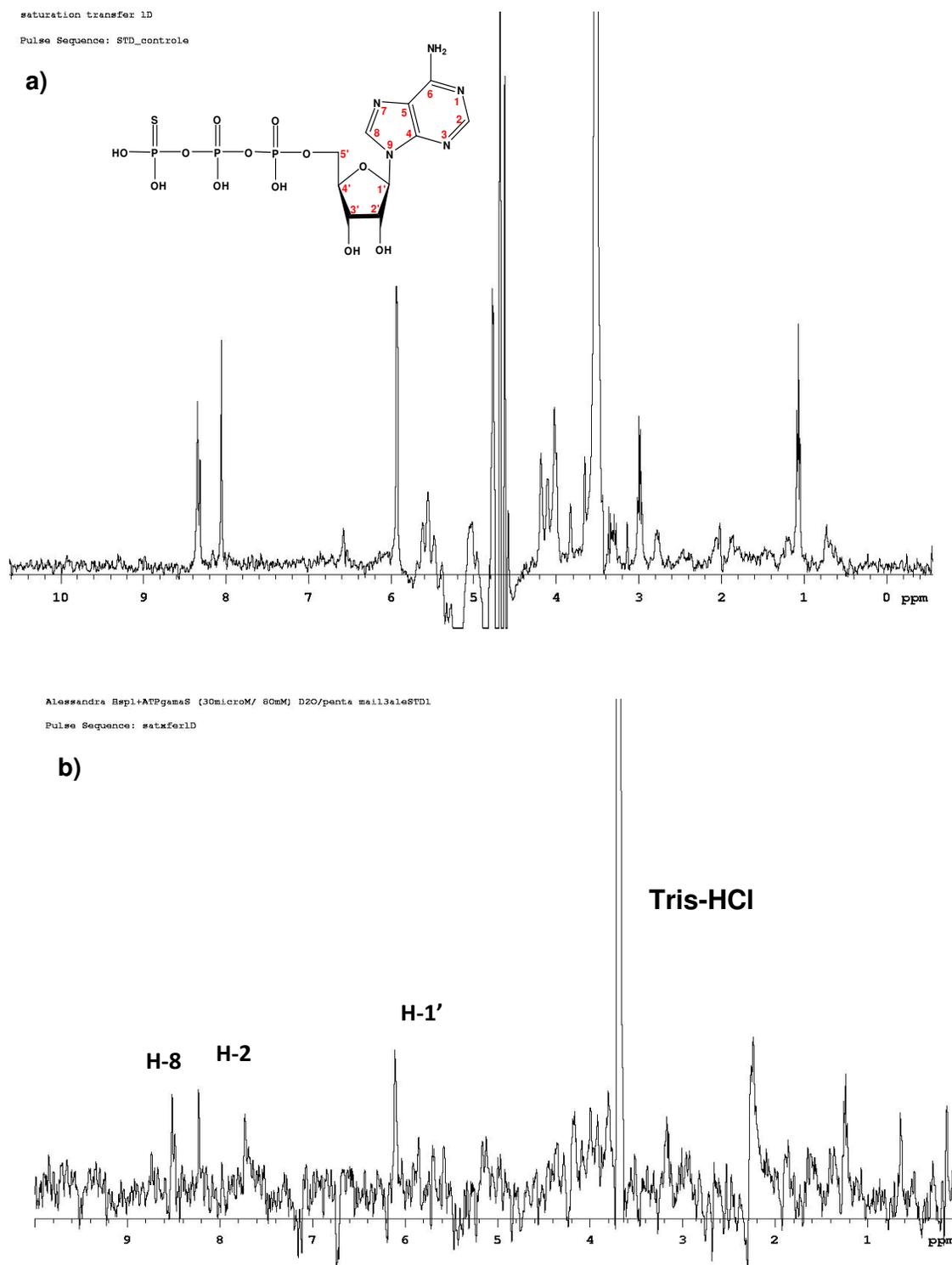


Figura IV-68: Espectros de NMR da proteína Hsp90 da laranja com o nucleotídeo ATP γ S na razão molar 1:100. (a) Espectro de STD-control e (b) Espectro de STD.

Os resultados dos experimentos de STD entre Hsp90 com os nucleotídeos supracitados, revelaram que os átomos de hidrogênio H-8 e H-2 da adenosina e H-1' da ribose também interagem com a proteína Hsp90. Em todos os experimentos calculou-se o mapa do epítipo e, em todos os casos, estes valores estavam próximos a 100% (Tabela IV-10). Para o ligante ADP não foi observada interação com a proteína Hsp90 da laranja e as interações entre a Hsp90 da laranja e a geldanamicina não foram realizadas, pois a geldanamicina apresenta baixa solubilidade em dimetilsulfoxido- d_6 (DMSO) e quando em contato com a proteína provoca precipitação. A interação com o ligante ADP não foi visualizada possivelmente por esta interação ser fraca ou muito forte, não havendo ligante livre em tempo de RMN (escala).

Tabela IV-10: Mapa de epítipo obtido em STD-NMR para os hidrogênios H-8, H-2 e H-1' obtidos em interações da Hsp90 da laranja com ATP γ S e ATP.

Hidrogênios	Hsp90+ATP γ S	Hsp90+ATP	Hsp90+ADP
H-8	99%	90,5%	----
H-2	98%	92%	----
H-1'	98%	90%	----

Em seguida, foi realizado um estudo da interação desta proteína com os mesmos ligantes utilizando a técnica de fluorescência de emissão. Como a Hsp90 da laranja possui 6 triptofanos em sua sequência primária (Figura IV-69), foi analisado se a fluorescência de emissão destes resíduos é modificada na presença dos ligantes. As concentrações e o comprimento de onda de excitação e emissão máxima dos ligantes e da Hsp90 da laranja estão mostrados na Tabela IV-11.

Hsp90laranja

MASETETFAFQAEINQLLSLIINTFYSNKEIFLRELISNSSDALDKIRFESLTDKSKLDAQPE
 LFIHIIPDKTNNLSIIDSIGGMTKADLVNNGTIARSGTKEFMEALAAAGADVSMIGQFGVG
 FYSAYLVAEKVIVTAKHNDDEQYI**W**ESQAGGSFTVTRDTSCELLGRGKITLHLKEDQLE
 YLEERRLKDLIKKHSEFISYPISL**W**IEKTTEKEISDDEDEEEKKDEEGKVEDVDEEKEKEE
 KKKKKIKEVSHE**W**SLVKNQKPI**W**MRKPEEITKEEYAAFYKSLTND**W**EEHLAVKHFSVEG
 QLEFKAILFVPKRAPFDLFDTRKKPNNIKLYVRRVFIMDNCEELIPEYLG FVKGIVDSEDL
 PLNISRETLQQNKILKVIRKNLVKKCIELFQEIAENKEDYNKFYESFSKNLKLGIHEDSTNK
 TKLAELLYHSTKSGDELTSKDYVTRMKEGQNDIYYITGESKKAVENSPFLEKLVKGGY
 EVLYMVAIDAIDEYAVGQLKEFEGKLVSATKEGLKLDESEDEKVVVETLKEKFEGLCKVIK
 DVLGDKVEKVVVSDRVVDSPPCLVTGEY**W**TANMERIMKAQALRDNSMAGYMSSKKT
 MEINPENPIMEELRKRADADKNDKSVKDLVLLLFETALLTSGFSLDDPNTFGNRIHRMLK
 LGLSIEEDAGDADADMPPLEDAADDAEGSKMEEVD

Figura IV-69: Sequencia primária da proteína Hsp90 da laranja. Resíduos de triptofano em destaque (vermelho).

Tabela IV-11: Concentração e comprimento de onda de excitação e emissão máximos em ensaios de interação ATP, ADP, ATP γ S e geldanamicina

Proteína e ligantes	Concentração	λ_{ex} (nm)	λ_{em} (nm)	Concentração usada para titulação	Volume adicionado (incrementos de 1 μ L)	Emissão Máxima (nm) Proteína + Ligante
Hsp90 da laranja	2,67 μ mol.L ⁻¹	277	341	2,67 μ mol.L ⁻¹ (2,21 μ mol.L ⁻¹ para geldanamicina)	-----	-----
geldanamicina	1,78 mmol.L ⁻¹	364	515	1,78 mmol.L ⁻¹	30	340
ADP	13,00 mmol.L ⁻¹	290	390	80,00 mmol.L ⁻¹	25	340
ATP	13,00 mmol.L ⁻¹	290	388	80,00 mmol.L ⁻¹	15	340
ATP γ S	7,30 mmol.L ⁻¹	287	375	80,00 mmol.L ⁻¹	30	341

Como pode ser observado a emissão máxima dos ligantes encontra-se em uma faixa de comprimento de onda não observado para a proteína Hsp90 sendo assim, interessante para o estudo (Figura 70 - 72).

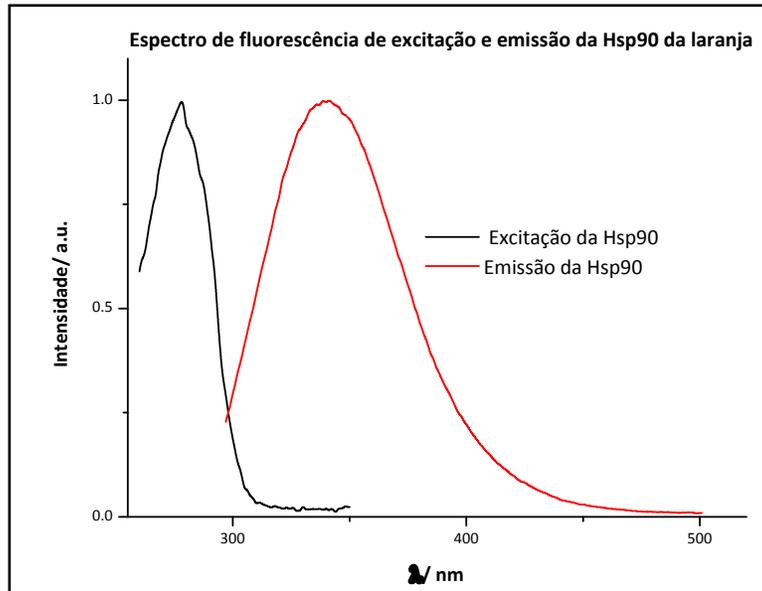


Figura IV-70: Espectro de fluorescência de excitação e emissão da Hsp90 da laranja $2,67 \mu\text{mol.L}^{-1}$.

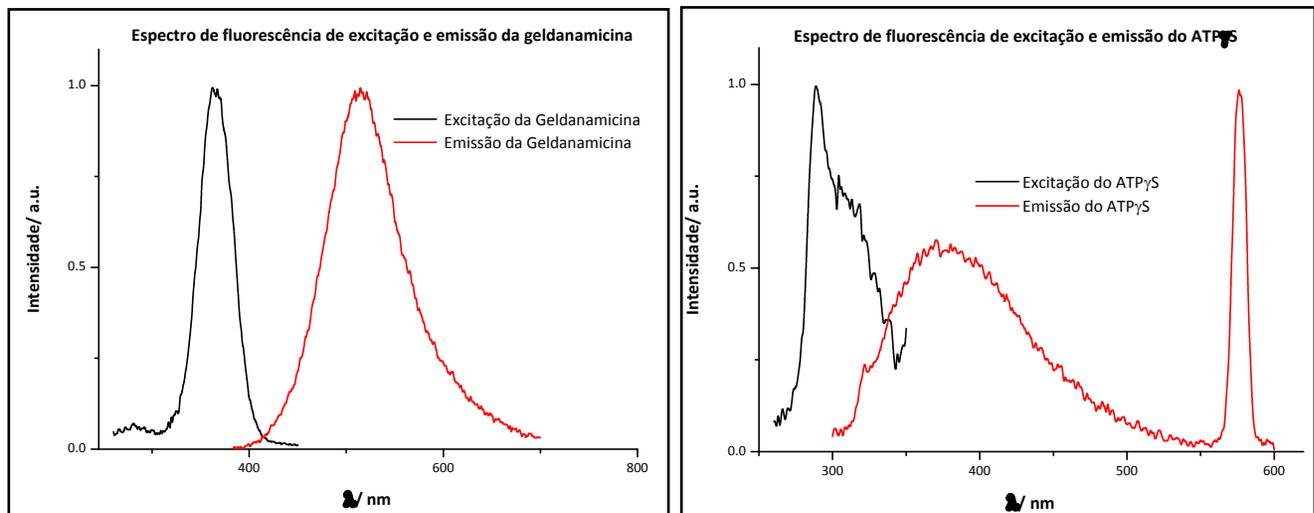


Figura IV-71: Espectros de fluorescência de excitação e emissão da geldanamicina $1,78 \text{ mmol.L}^{-1}$ e do ATP γ S $7,30 \text{ mmol.L}^{-1}$

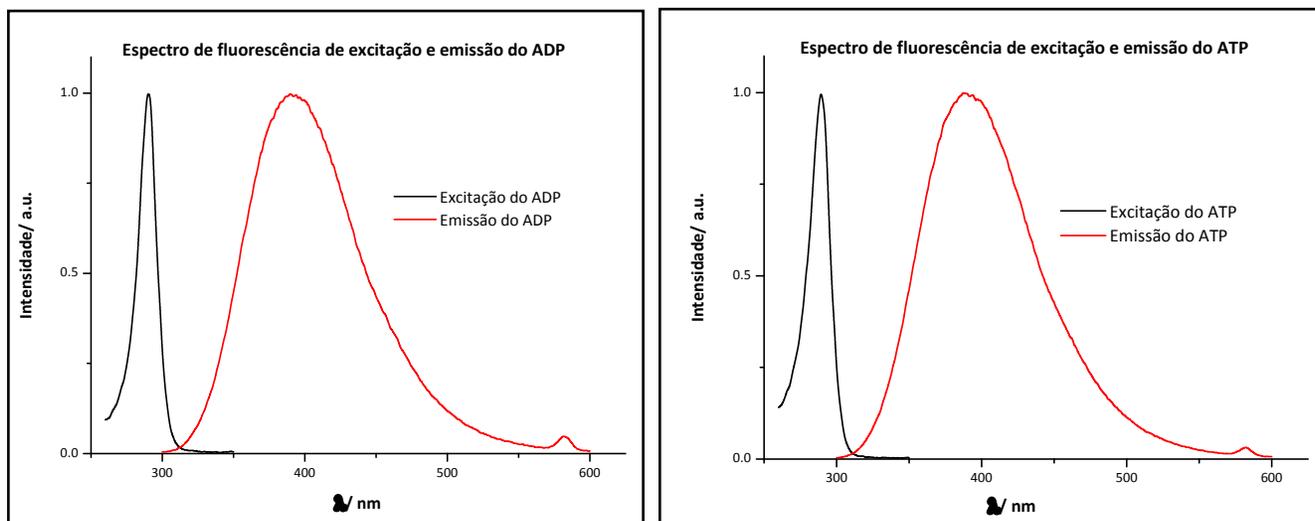


Figura IV-72: Espectros de fluorescência de excitação e emissão do ADP 13 mmol.L^{-1} e do ATP 13 mmol.L^{-1} .

Os resultados obtidos para a titulação da proteína Hsp90 da laranja com os ligantes encontram-se nas Figuras IV 73 - 76.

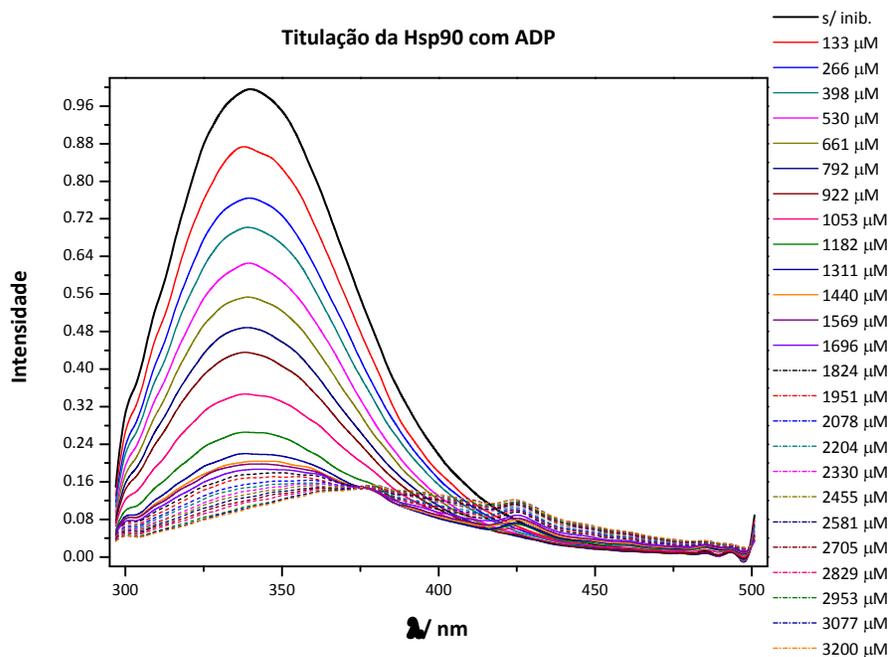


Figura IV-73: Espectros de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com ADP.

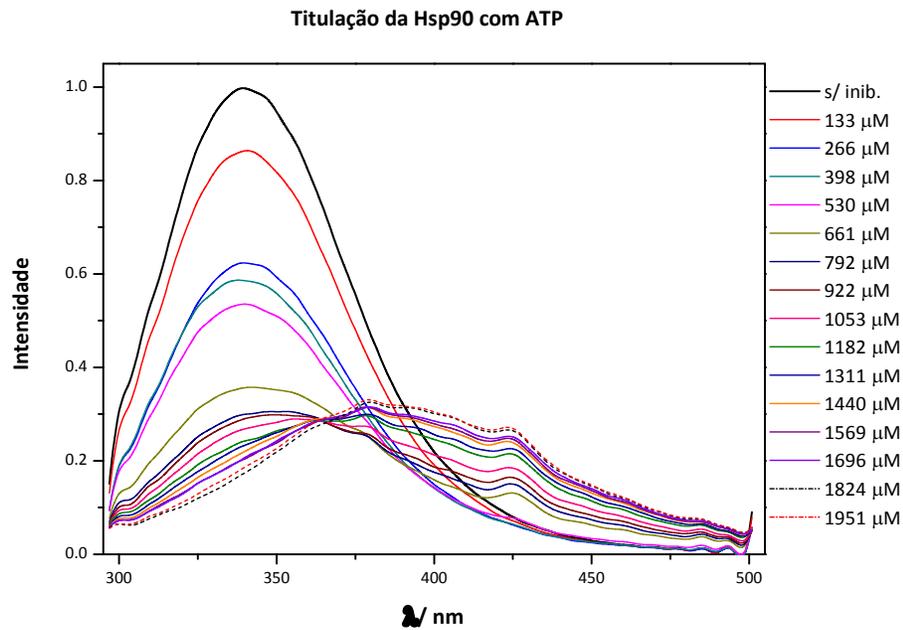


Figura IV-74: Espectros de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com ATP.

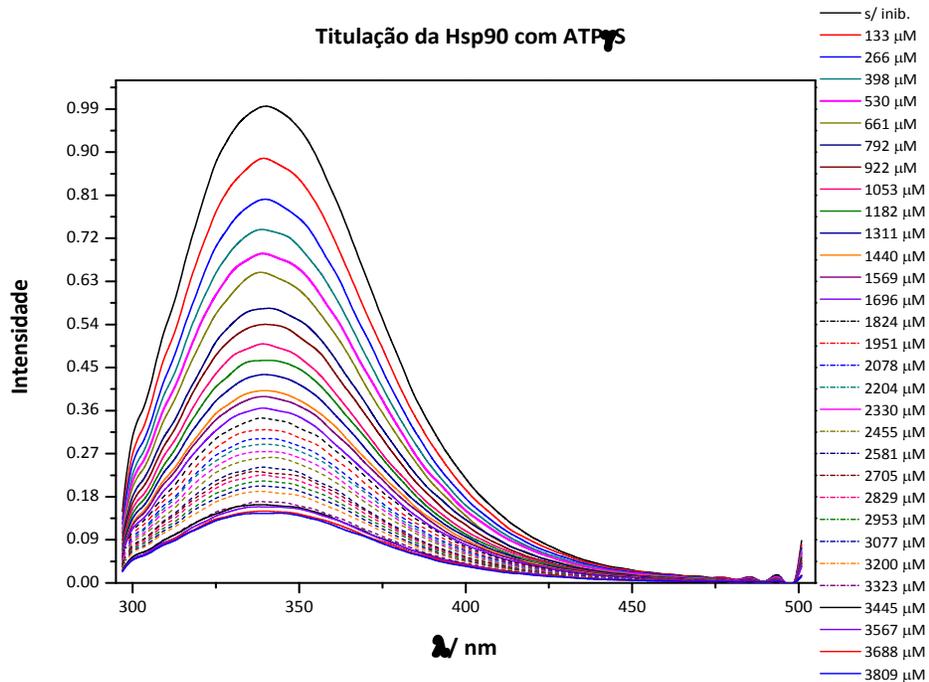


Figura IV-75: Espectros de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com ATP γ S.

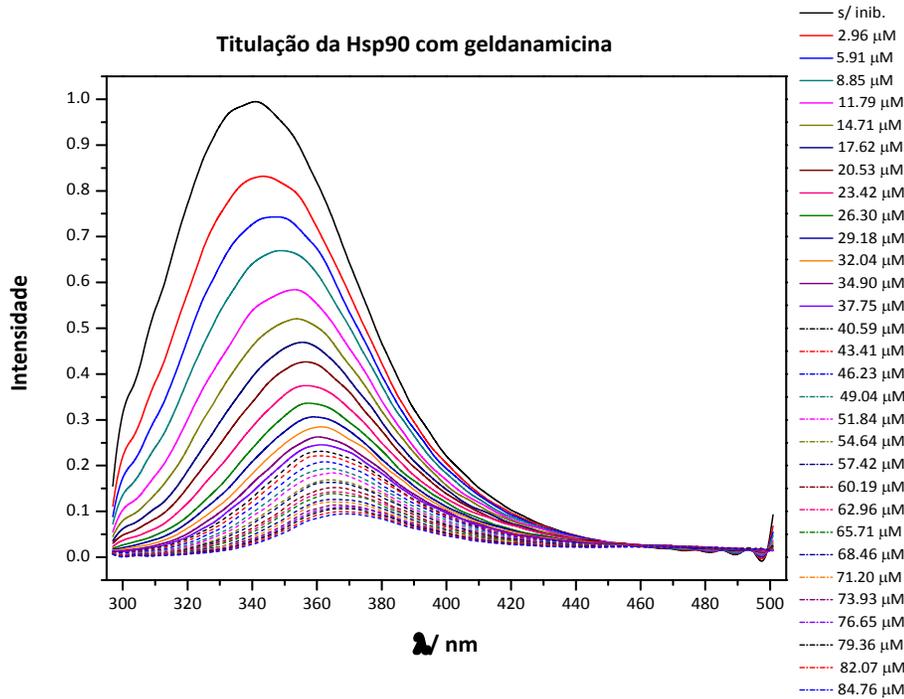


Figura IV-76: Espectros de fluorescência de emissão da proteína Hsp90 da laranja quando titulada com geldanamicina.

A partir dos resultados obtidos foi observado que houve a supressão de fluorescência de emissão (*quenching*) da Hsp90 em todos os casos como, também, ocorreu deslocamento de comprimento de onda (λ_{\max}) de aproximadamente 30 nm da Hsp90 quando titulada com a geldanamicina (Figura IV-76). No caso da titulação com ADP (Figura IV-73) e ATP (Figura IV-74), também foi verificado que há interação destes ligantes com a Hsp90. A interação da Hsp90 com o ligante ADP não foi observada através da técnica de STD, porém esta interação foi verificada como efetiva na fluorescência de emissão. Uma possível explicação é que o ADP liga e se desliga rapidamente quando interage com a Hsp90 e a proteína perde a conformação ativa, se torna aberta. Já o ATP, um ligante natural da Hsp90, interagiu com esta proteína resultando em uma alta supressão da fluorescência em pequenas concentrações do ligante sugerindo uma interação Hsp90/ ATP bastante efetiva. Também foi verificada a influência do

branco, ou seja, titulação da proteína Hsp90 da laranja com DMSO, porém, não houve considerável supressão da fluorescência.

A supressão de fluorescência observada para a interação Hsp90/ ligantes pode ter ocorrido através de fenômenos de Transferência de Energia de Ressonância (RET). Sendo assim, os resíduos de triptofano da proteína quando excitados transferem energia para os ligantes. O RET é um fenômeno específico porque o ligante tem que se ligar especificamente no sítio ligante para formar um complexo não fluorescente. Durante a titulação da geldanamicina, por exemplo, a geldanamicina forma um complexo não fluorescente com a proteína, mas, a proteína não ligada a geldanamicina continua fluorescendo. Por isso, conforme aumenta a quantidade de ligante, ele se liga especificamente no sítio ligante da proteína e ocorre o *quenching*.

A partir dos dados de fluorescência de emissão da proteína titulada com os ligantes foi construído um gráfico de *quenching* versus a concentração do ligante. A curva de supressão da fluorescência para todos os ligantes ATP γ S e geldanamicina estão mostradas na Figura IV-77.

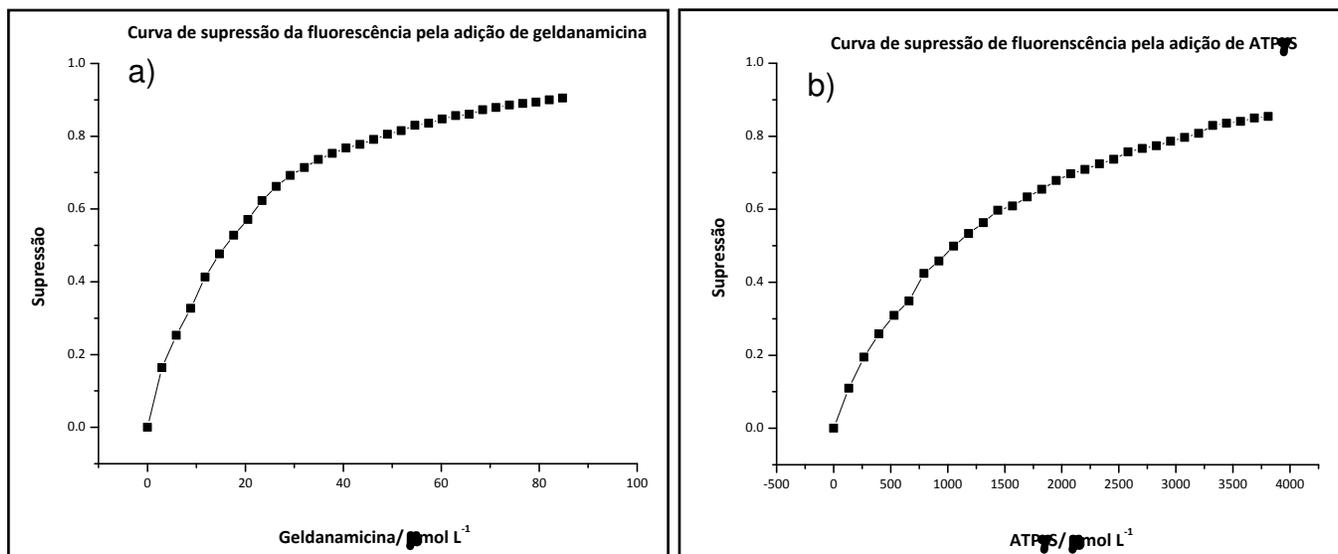


Figura IV-77: Gráficos de Stern-Volmer para os ligantes (a) geldanamicina; (b) ATP γ S frente a sua proteína alvo Hsp90 da laranja.

A partir dos dados fluorimétricos foram calculadas as constantes de dissociação, um importante fator na avaliação destas interações. A presença de seis triptofanos na estrutura primária da Hsp90 da laranja faz com que esta não apresente comportamento fluorescente homogêneo, de modo a observar-se desvio negativo de linearidade, apresentando então um comportamento hiperbólico em gráficos de Stern-Volmer, utilizados na determinação de K_d . Para este cálculo usou o ajuste dos dados já normalizados de supressão de fluorescência (*quenching*) em função da concentração do ligante. A curva de supressão da fluorescência usada para o cálculo de K_d para a geldanamicina e $ATP\gamma S$ estão mostradas na Figura IV-77.

Os valores de K_d calculados para complexos Hsp90/ligantes foram 9,4; 585; 250 e 566 $\mu\text{mol.L}^{-1}$, para a geldanamicina, $ATP\gamma S$, ATP e ADP, respectivamente.

As constantes de dissociação determinadas indicam média afinidade entre os ligantes, particularmente aqueles com os mais baixos valores de K_d , geldanamicina e ATP, e a Hsp90 da laranja. Além disso, foi observado que a geldanamicina apresenta um valor de K_d menor comparado ao ligante natural da Hsp90, o ATP, sugerindo que a geldanamicina é um potencial inibidor desta enzima.

Na tentativa de mostrar tridimensionalmente a interação da Hsp90 da laranja com os ligantes, usou-se como modelo a estrutura cristalográfica de uma Hsp90 em que o sítio ligante para geldanamicina (PDB 1yet)⁵⁷ e ATP (PDB 1am1)⁵⁸ já haviam sido determinadas.

Através do alinhamento do sítio ligante das duas Hsp90 usando o programa PyMOL foi observado alto grau de alinhamento (Figura IV-78).

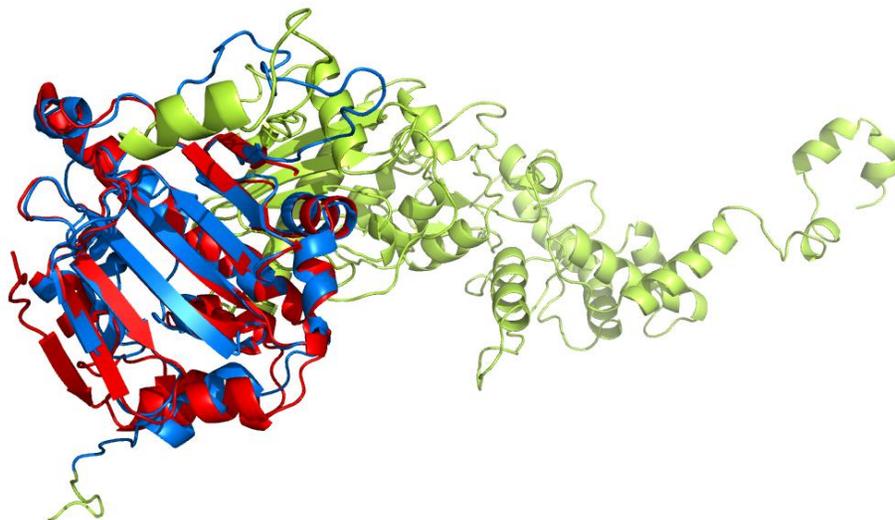


Figura IV-78: Representação em *cartoon* da estrutura da Hsp90 da laranja em verde (I-TASSER C-score = -2,82) com seu sítio ligante em azul e em vermelho, o sítio ligante da 1yet (PDB).

A partir do sítio ligante da 1yet⁵⁷ (PDB) e da 1am1⁵⁸ (PDB) com os seus ligantes orientados (geldanamicina e ATP) especialmente pode-se propor estruturas para Hsp90 em complexos com estes ligantes. Para isso, a estrutura tridimensional da Hsp90 com o ligante geldanamicina, por exemplo, foi carregado no PyMOL juntamente com a estrutura tridimensional da Hsp90 da laranja gerada pelo programa I-TASSER. Após alinhamento, a estrutura da Hsp90 que continha o ligante foi escondida e com isso, obteve-se a estrutura tridimensional da Hsp90 da laranja com o ligante geldanamicina. As figuras tridimensionais foram desenhadas no I-TASSER (Figura IV-79) e as figuras com corte transversal foram desenhadas no Chimera 1.5.2 (Figuras IV-80).

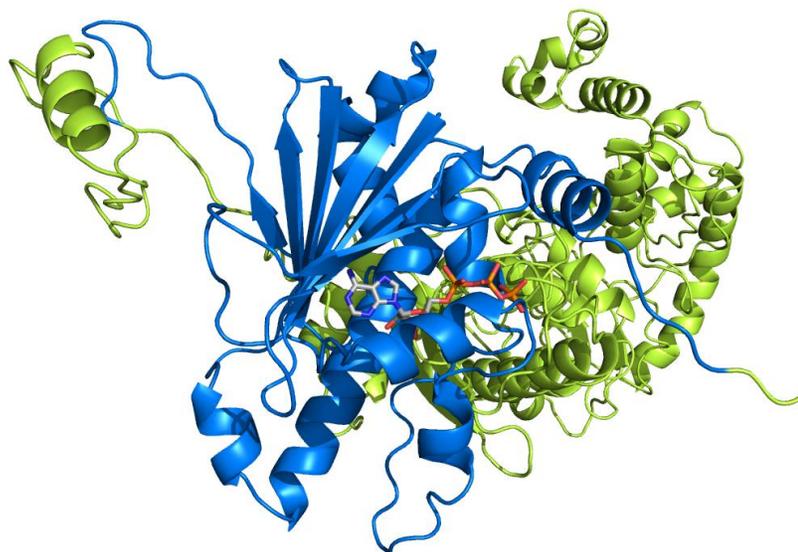


Figura IV-79: Representação em *cartoon* da estrutura da Hsp90 da laranja em verde com seu sítio ligante mostrado em azul com a estrutura do ATP em vermelho.

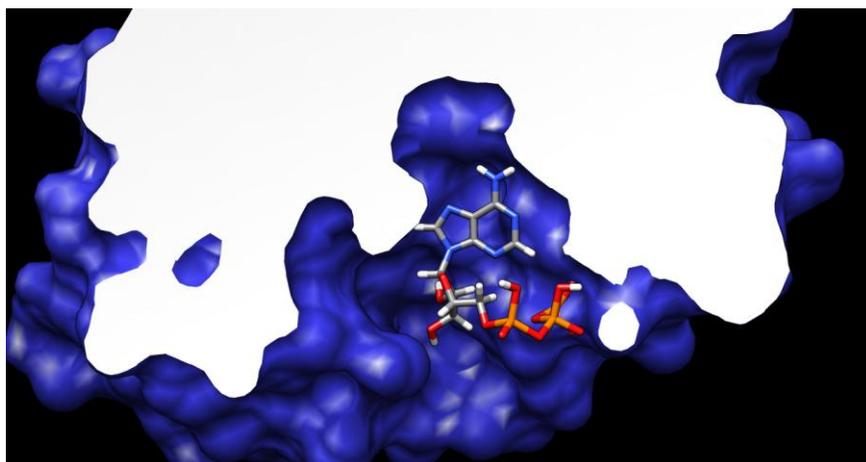


Figura IV-80: Corte transversal da superfície do sítio ligante da Hsp90 da laranja com a estrutura do ATP.

A parte branca na figura de corte transversal é o corte transversal da superfície da proteína (se a proteína fosse um objeto sólido, cheio de cavidades, a parte branca seria a porção sólida e o sítio ligante a cavidade). Através dos dados de STD, fluorescência e propostas de estrutura acredita-se que a proteína Hsp90 da laranja faz interações hidrofóbicas e ligações de hidrogênio com os hidrogênios da adenina e com o hidrogênio da pentose dos ligantes.

O domínio ligante da Hsp90 tem um bolsão de 15 Å de profundidade formado pelos resíduos 9-232 e os ligantes entram e interagem com estes resíduos da Hsp90 através dos átomos de hidrogênio H-8 e H-2 da adenosina e H-1' da ribose de modo que os seus fosfatos estão orientados para fora do sítio ligante da Hsp90.

Analisando-se a estrutura proposta é observado que realmente a parte da adenina e ribose se encontram dentro do sítio ligante da Hsp90 da laranja e esta informação foi confirmada através dos dados de STD. Na STD foi verificado quais hidrogênios da adenina e ribose estavam interagindo com a proteína e a fluorescência confirmou esta interação através da supressão da fluorescência da proteína. O valor de K_d determinado também se mostrou coerente com os dados de ressonância e dentro da faixa esperada para STD.

Capítulo V - Conclusões

5. Conclusões

Este trabalho visou as análises biofísicas das chaperonas de secreção FlgN e XACb0033 da *Xac* que, por dados de espectrometria de massas apresentaram-se como monômero (FlgN) e dímero (XACb0033). As análises de dicroísmo circular e cálculo de estrutura secundária confirmaram baixo conteúdo helicoidal (α -hélice) para as proteínas FlgN e XACb0033 enquanto que o desenovelamento térmico da proteína FlgN acusou a perda de 50% da estrutura secundária (T_m) em torno de 45°C. A análise de fluorescência de emissão revelou que os resíduos de Trp de ambas as proteínas estão expostos ao solvente.

Os dados de NMR de 1D (^1H NMR) e 2D $\{^1\text{H}, ^{15}\text{N}\}$ HSQC indicaram proteínas agregadas ou com troca conformacional e ausência de estrutura 3D. As análises de estrutura 3D destas proteínas foram realizadas por espectrometria de massas, aplicando a técnica de *Footprinting* e utilizando-se como restrições SASA e RMSD para os modelos estruturais gerados aplicando ferramentas de bioinformática (I-TASSER e QUARK). Os dados de fluorescência corroboraram com a estrutura proposta, em que é possível verificar resíduos de Trp mais expostos ao solvente. No entanto, os dados de CD e NMR conflitam com a estrutura proposta. NMR revelou ausência de estrutura 3D e o CD e NMR baixo conteúdo helicoidal. Muito provavelmente a estrutura proposta pode estar correta, pois como ambas proteínas são proteínas de corpos de inclusão, o seu envolvimento correto pode ser prejudicado e quando concentradas formado agregados proteicos dificultando a visualização de estrutura secundária.

As interações entre as proteínas FlgN com FlgK foram sugeridas também através das análises de CD e fluorescência, porém esta interação não foi suficiente para diminuir a agregação protéica da proteína FlgN nas análises de NMR.

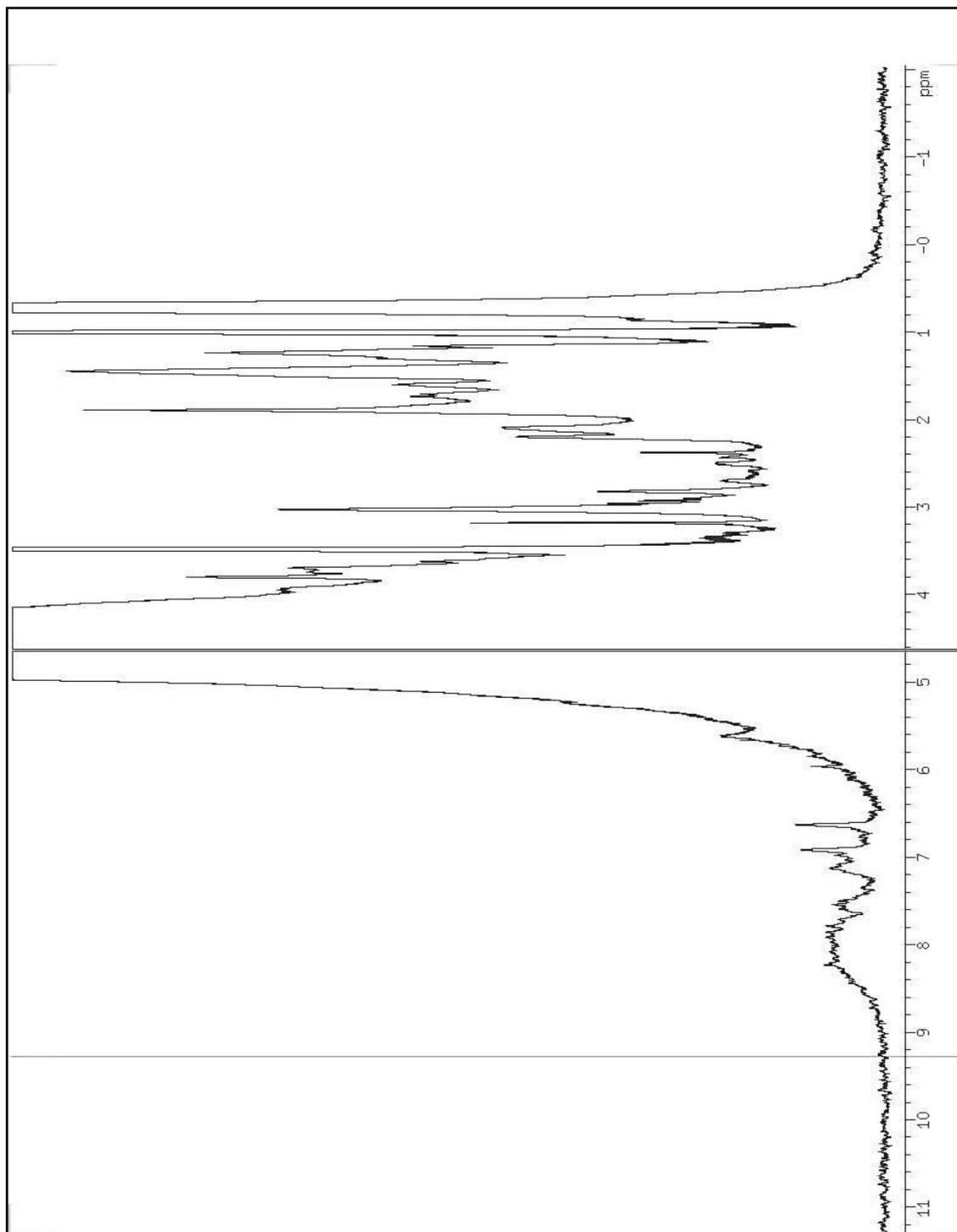
Na segunda parte do trabalho foram estudadas as interações entre a proteína Hsp90 da laranja e ligantes (ATP, ADP, ATP γ S e geldanamicina) através das técnicas de STD e fluorescência de emissão. A proteína Hsp100 da cana de

açúcar foi usada como modelo positivo de interação entre chaperonas e nucleotídeos na análise de STD revelando interação forte com os ligantes ATP, ADP e ATP γ S. A interação da Hsp100 da cana de açúcar e ATP γ S também foi observada no experimento de HETCOR através da mudança no deslocamento químico de alguns carbonos do ATP γ S. Para a proteína Hsp90 da laranja a análise de STD revelou interações com os ligantes ATP γ S (100%) e ATP (90%) através dos átomos de hidrogênio H-2 e H-8 da adenina e H-1' da ribose. No caso do ligante ADP não foi observada interação e os ensaios com a geldanamicina não foram realizados devido à baixa solubilidade da geldanamicina em H₂O, D₂O ou DMSO.

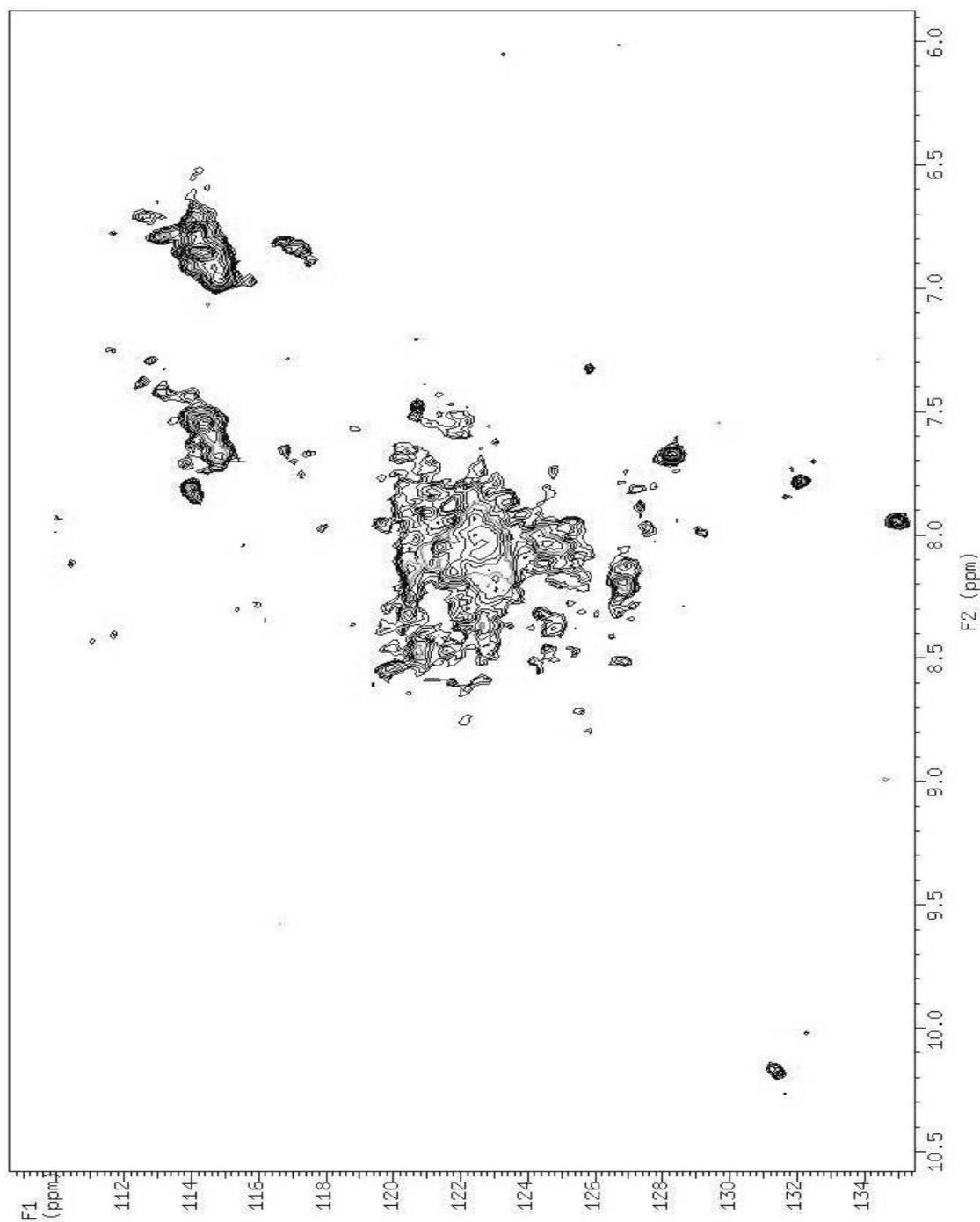
As análises de fluorescência de emissão da Hsp90 da laranja na presença dos ATP, ADP e geldanamicina foram indicativas para interações, uma vez que os resíduos de Trp da Hsp90 sofreram supressão de fluorescência em todos os casos e, também, deslocamentos de aproximadamente 30 nm para vermelho (*redshifts*) no caso de geldanamicina. A partir dos dados fluorimétricos foram calculados os valores de K_d dos ligantes, sendo eles: 9,4; 585; 250 e 566 $\mu\text{mol.L}^{-1}$, para a geldanamicina, ATP γ S, ATP e ADP, respectivamente. As constantes de dissociação determinadas indicaram grande afinidade entre os ligantes e a Hsp90, e, para a interação da Hsp90 da laranja e geldanamicina foi observado um valor de K_d menor sugerindo que a geldanamicina é um potente inibidor desta proteína.

Baseando-se em dados estruturais cristalográficos das proteínas 1yet (PDB) e 1am1 (PDB) com a geldanamicina e ATP, foram propostos modelos para estrutura 3D do sítio ligante da proteína Hsp90 da laranja na presença dos seus ligantes ATP, ADP, ATP γ S e a geldanamicina e foi verificado que os dados de STD e fluorescência corroboraram com o modelo proposto.

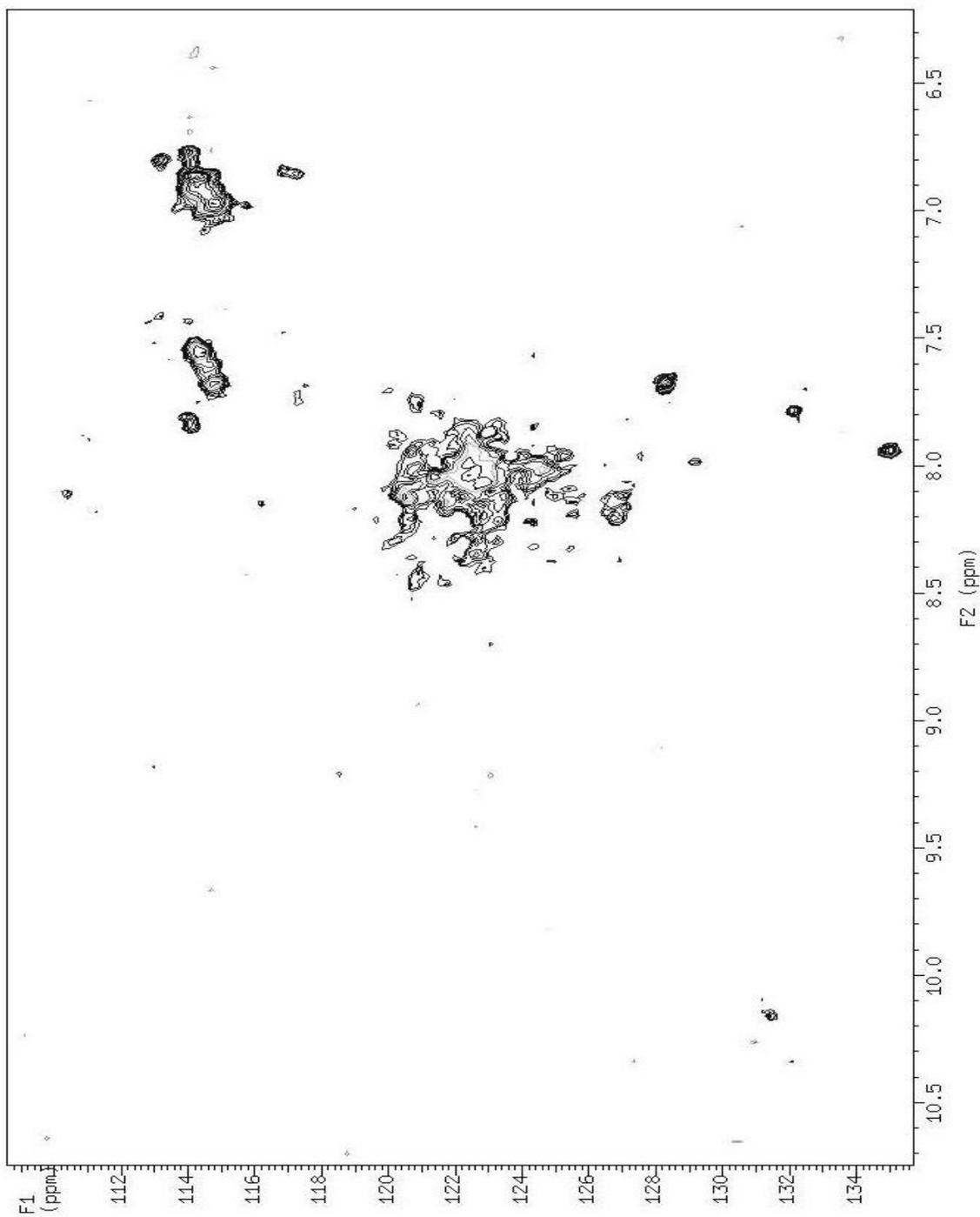
ANEXOS
Espectros de NMR das proteínas XACb0033 e FlgN



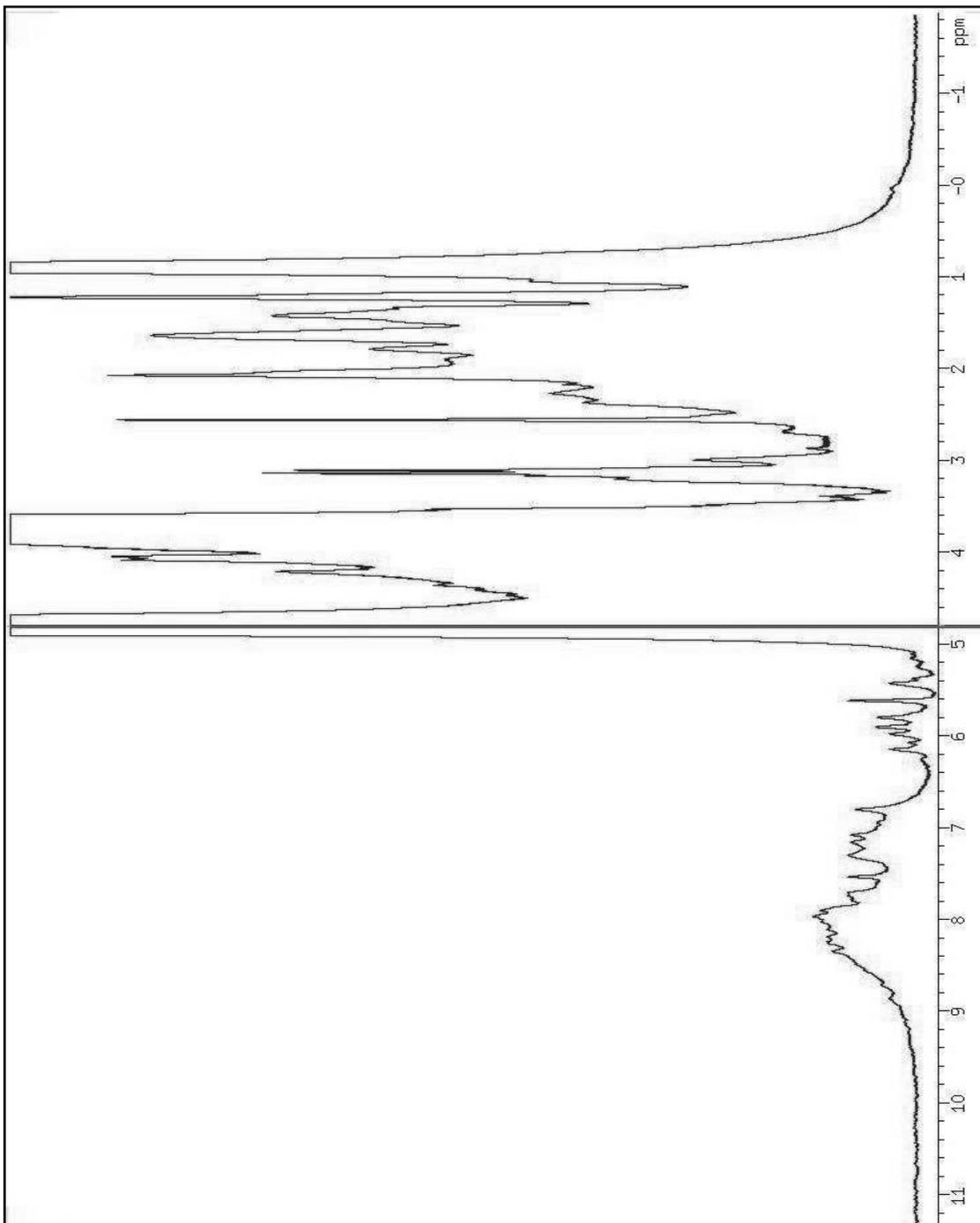
Anexo 1: Espectro de ^1H NMR da FlgN ($198 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



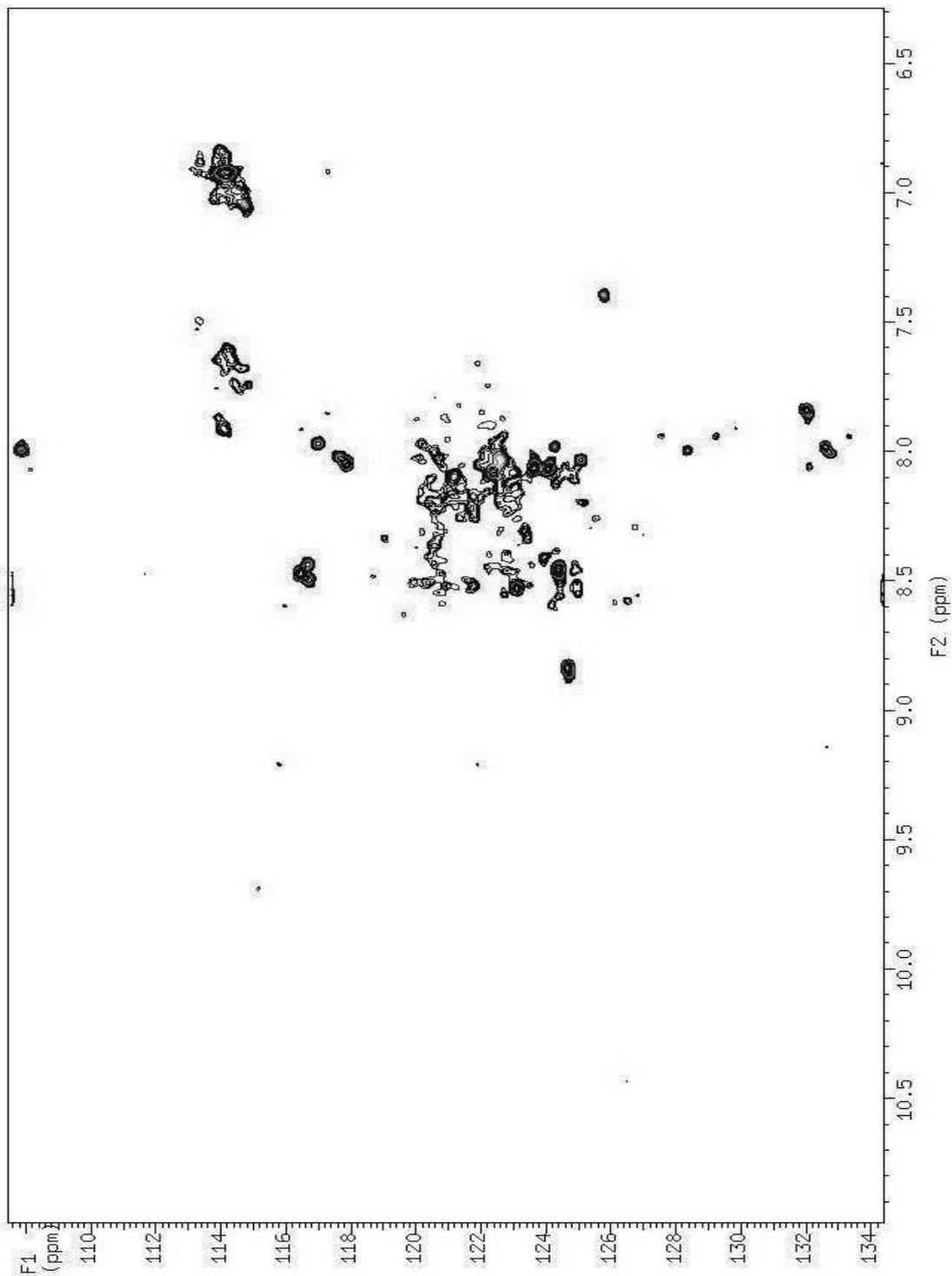
Anexo 2: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($198 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



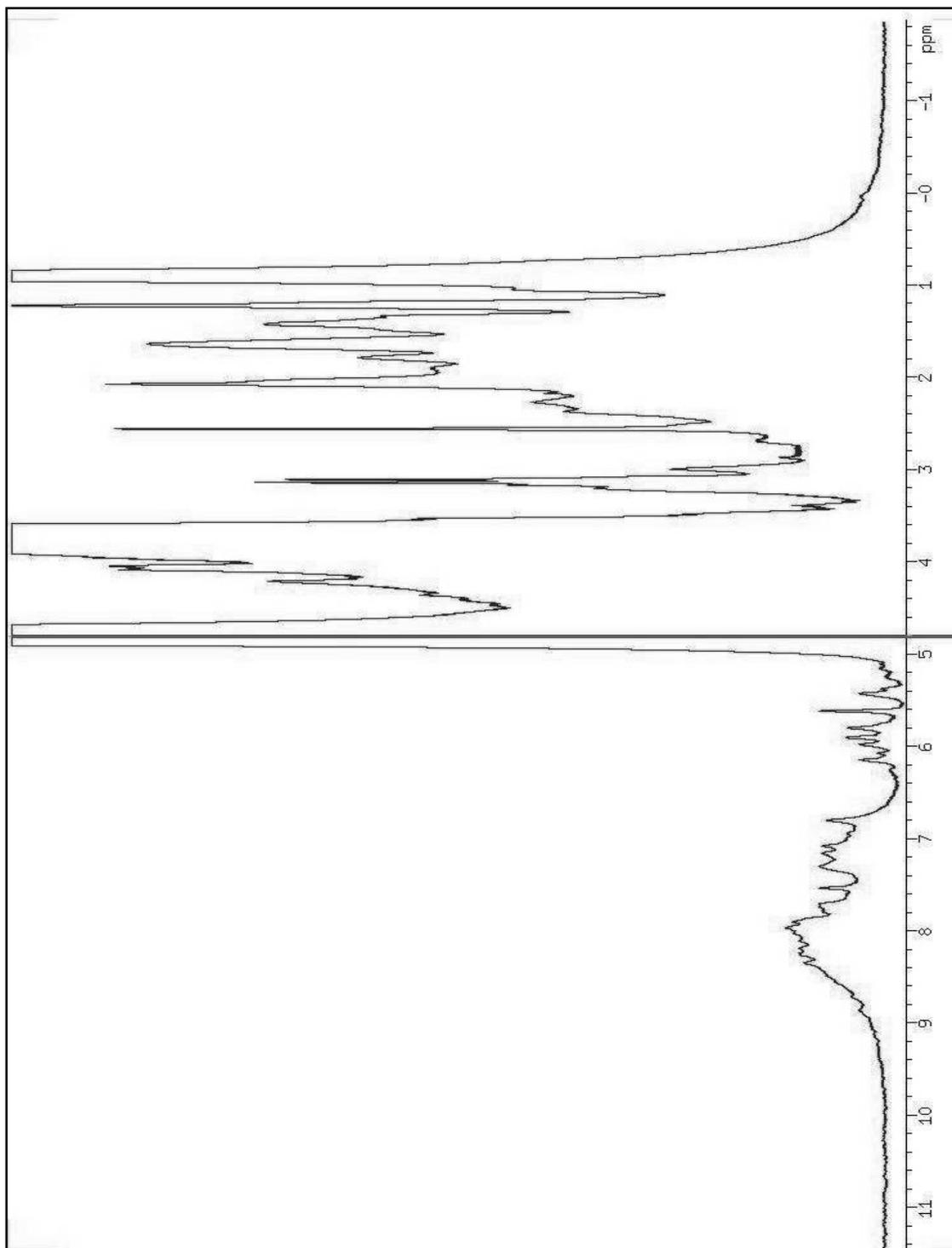
Anexo 3: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($198 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz. Após adição de glicerol.



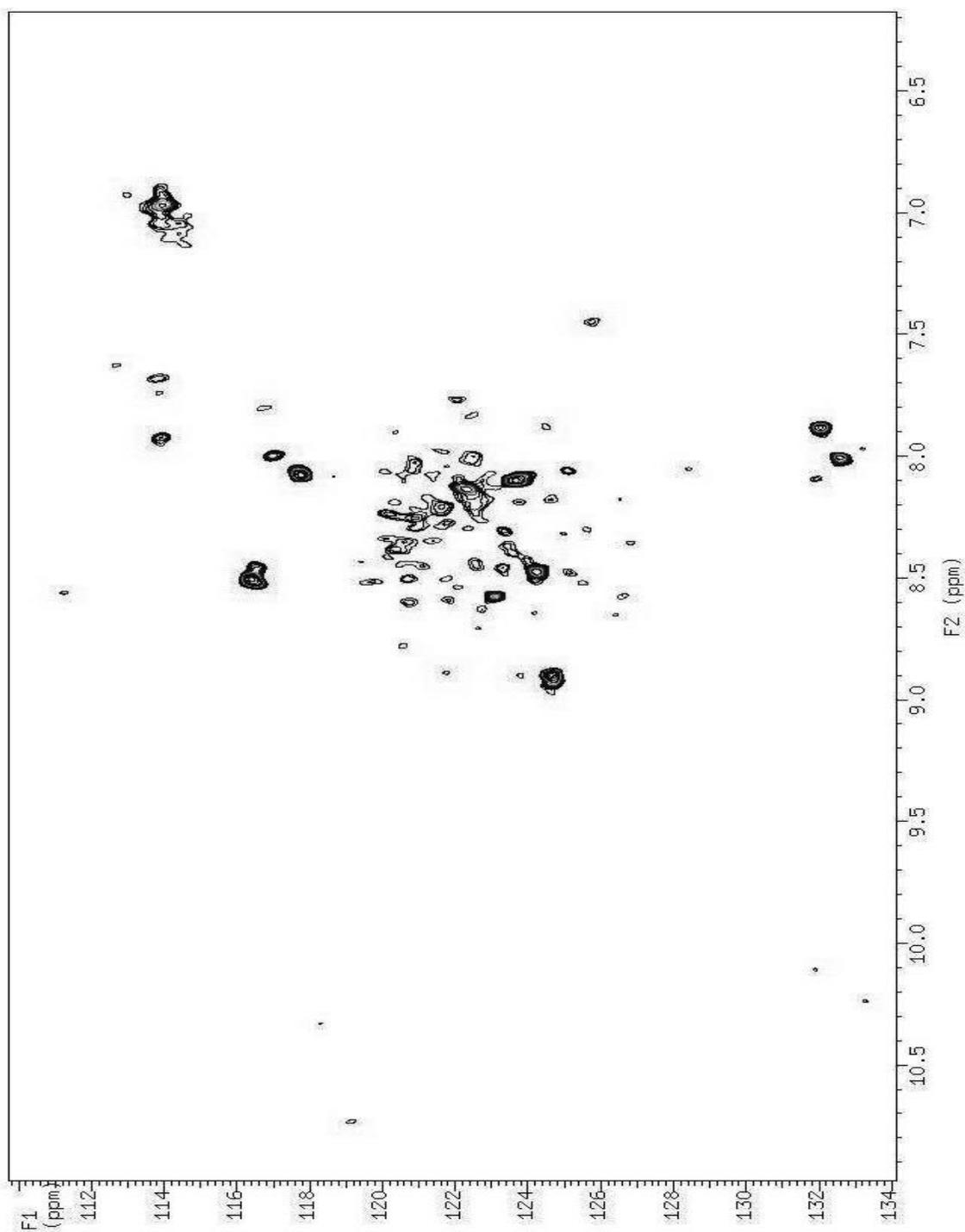
Anexo 4: Espectro de ^1H NMR da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



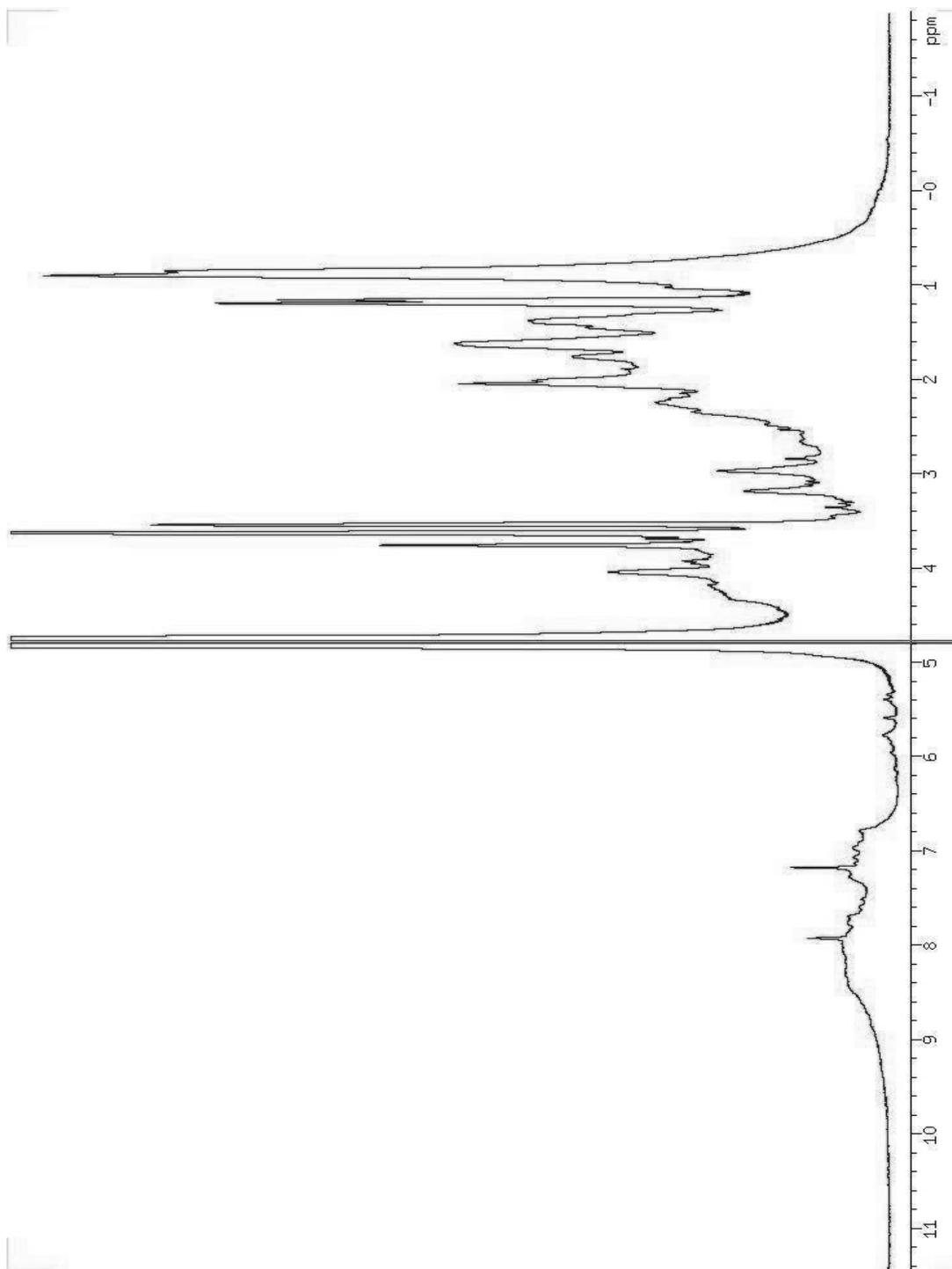
Anexo 5: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5 mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



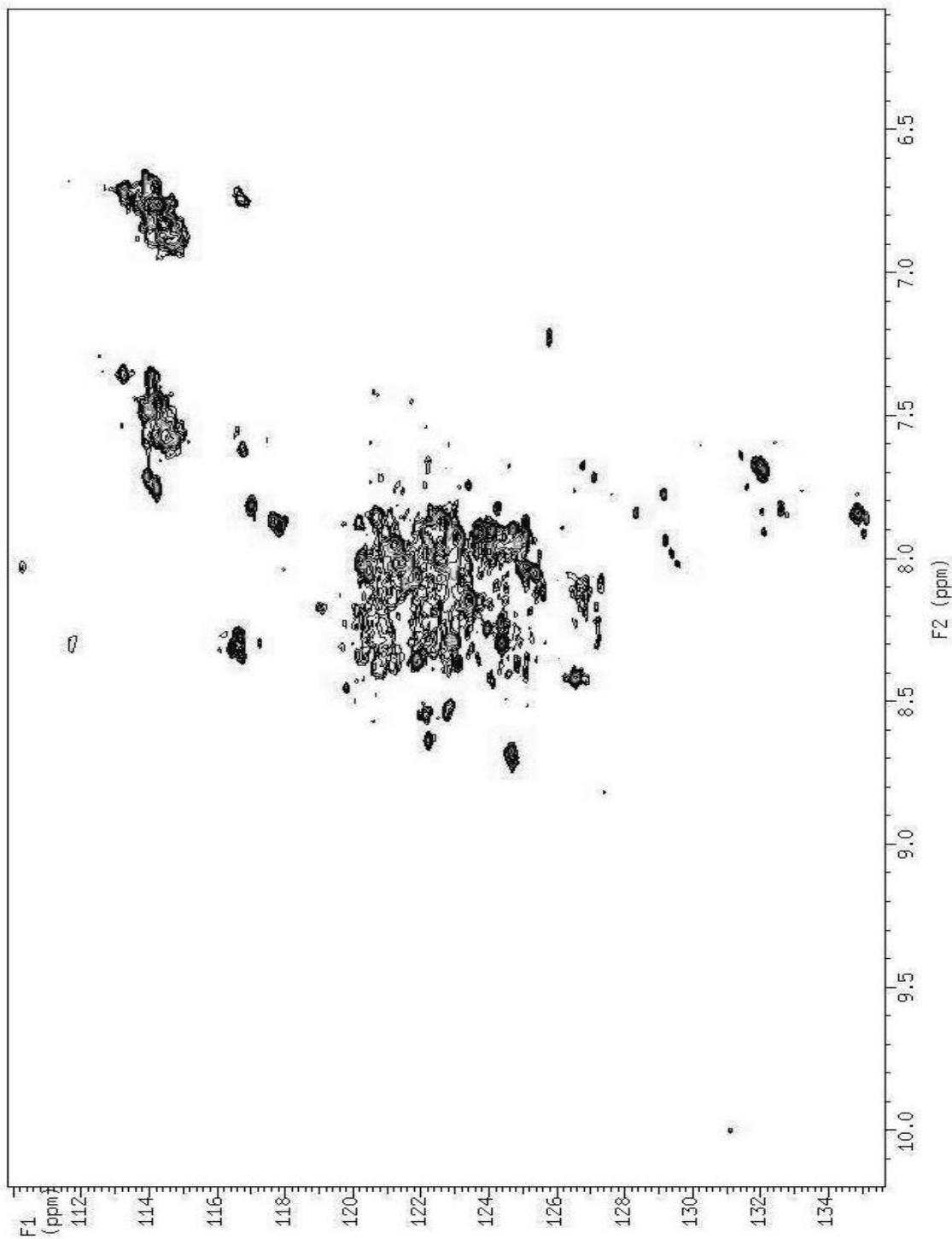
Anexo 6: Espectro de ^1H NMR da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5 mmol.L^{-1} , pH 8,0 a 35°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



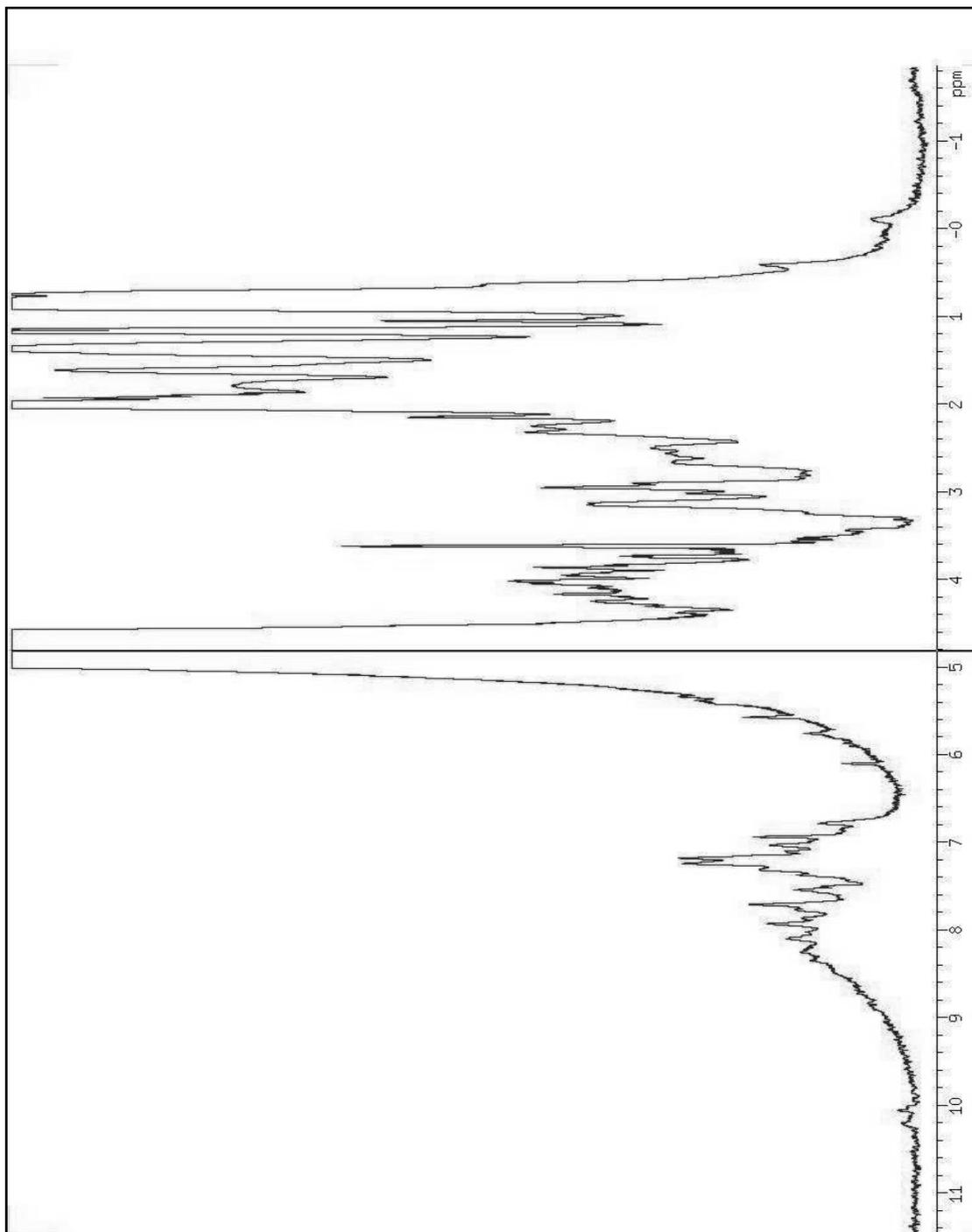
Anexo 7: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da FlgN ($21 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (5mmol.L^{-1} , pH 8,0 a 35°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



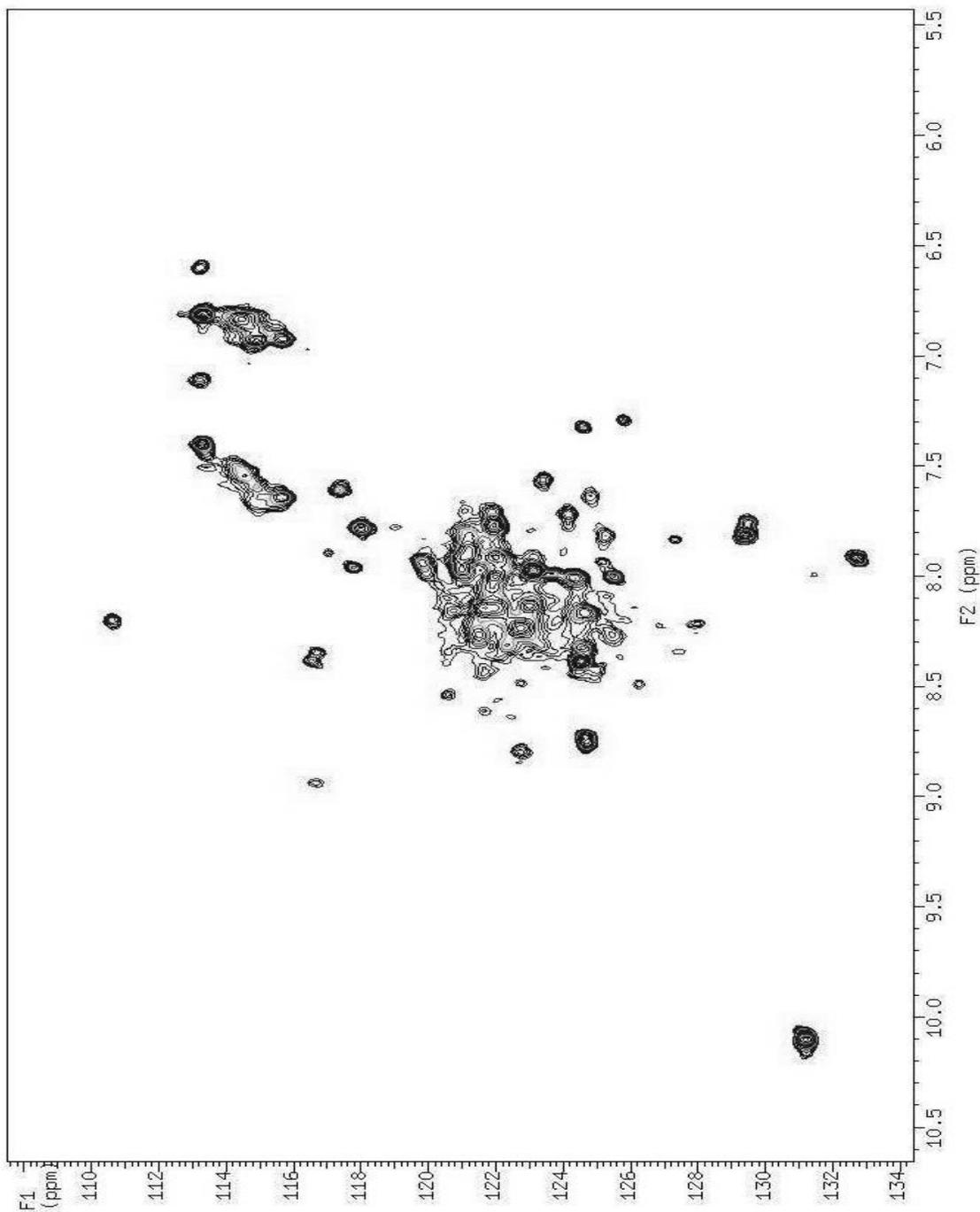
Anexo 8: Espectro de ^1H NMR do complexo FlgN + FlgK na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



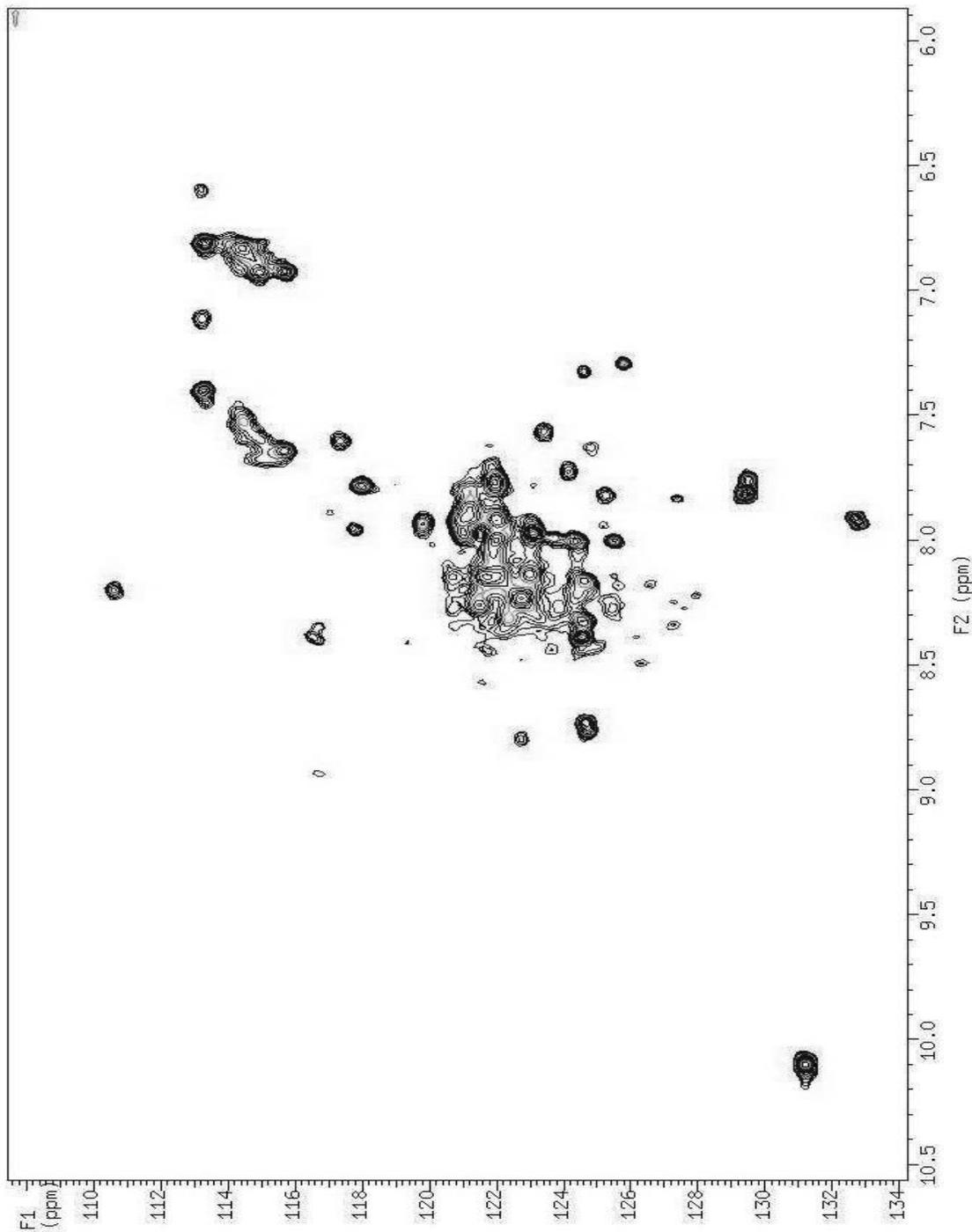
Anexo 9: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo FlgN + FlgK na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



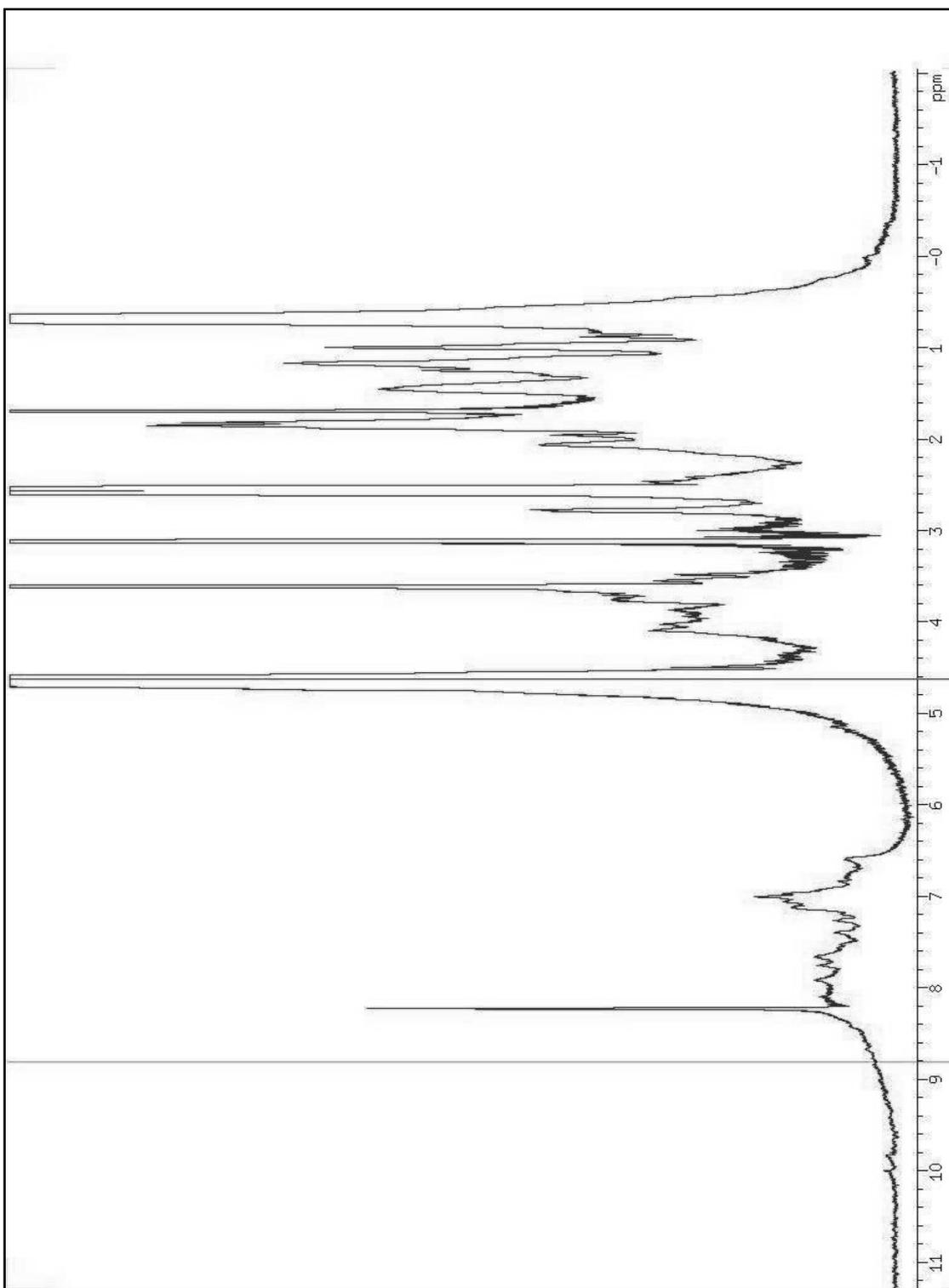
Anexo 10: Espectro de ^1H NMR da XACb0033 ($45 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1}) com 50mmol.L^{-1} de NaCl, (pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



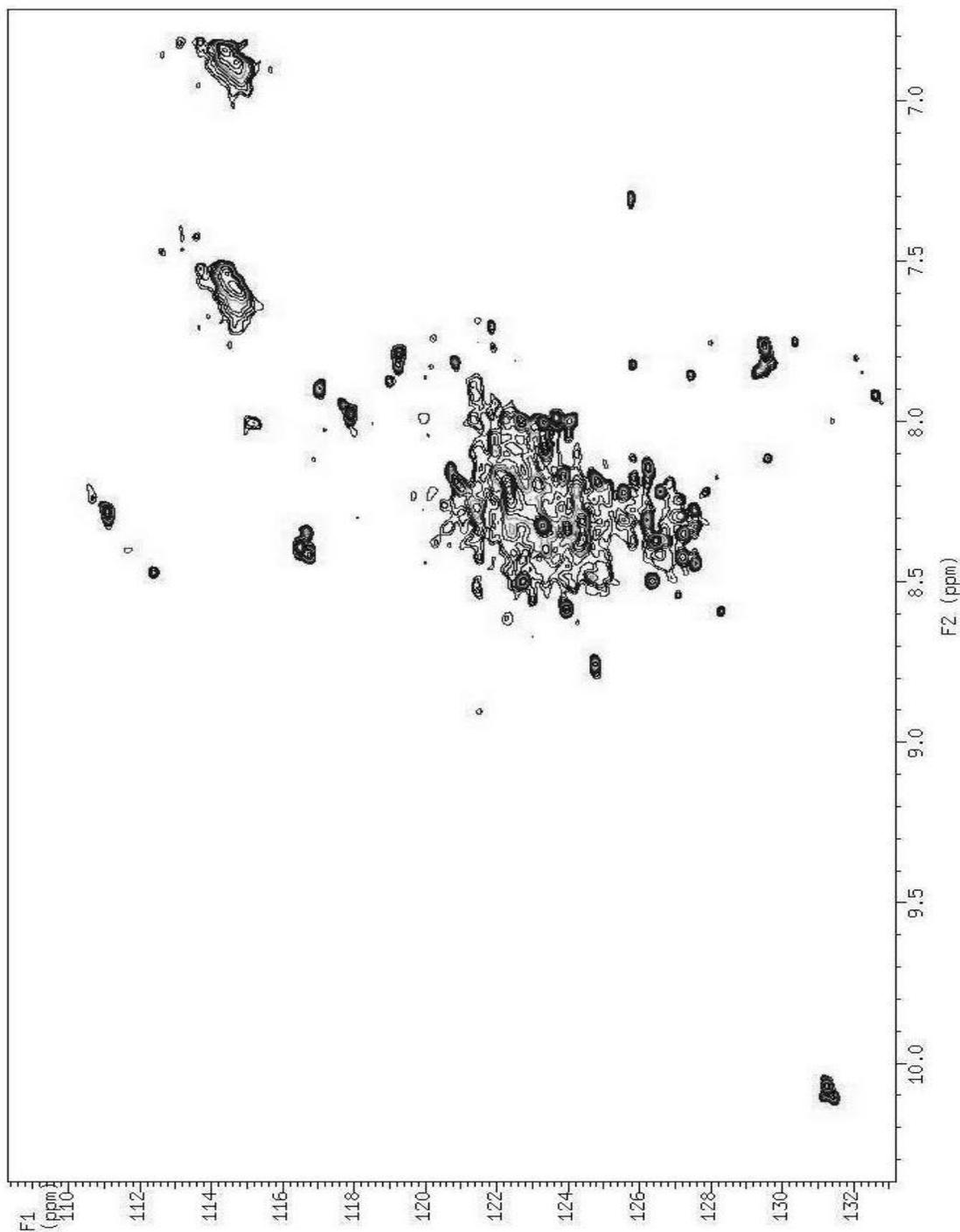
Anexo 11: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 ($45 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1}) com 50mmol.L^{-1} de NaCl, (pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



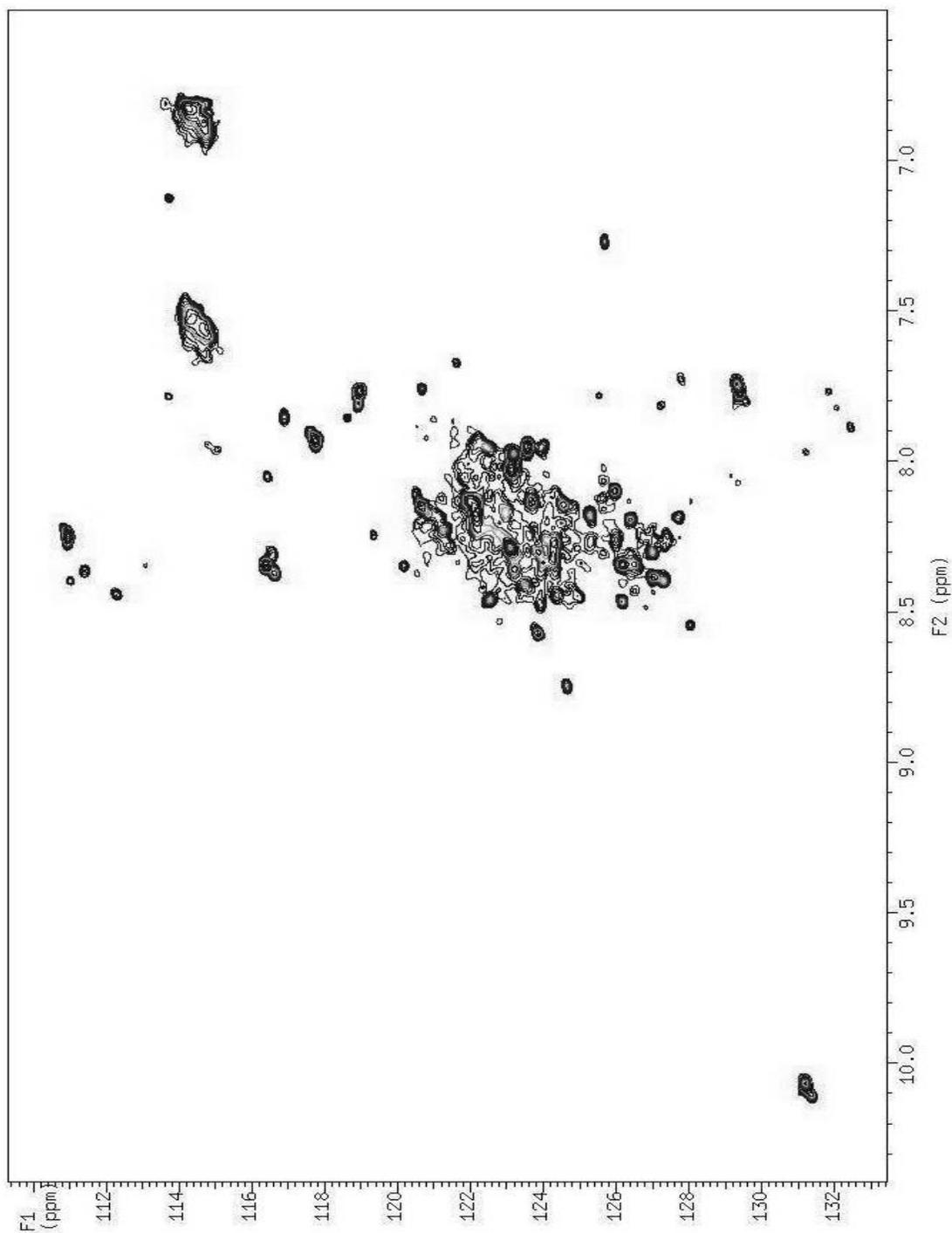
Anexo 12: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 ($45 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (100mmol.L^{-1}) com 50mmol.L^{-1} de NaCl, (pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz. Após adição de DTT.



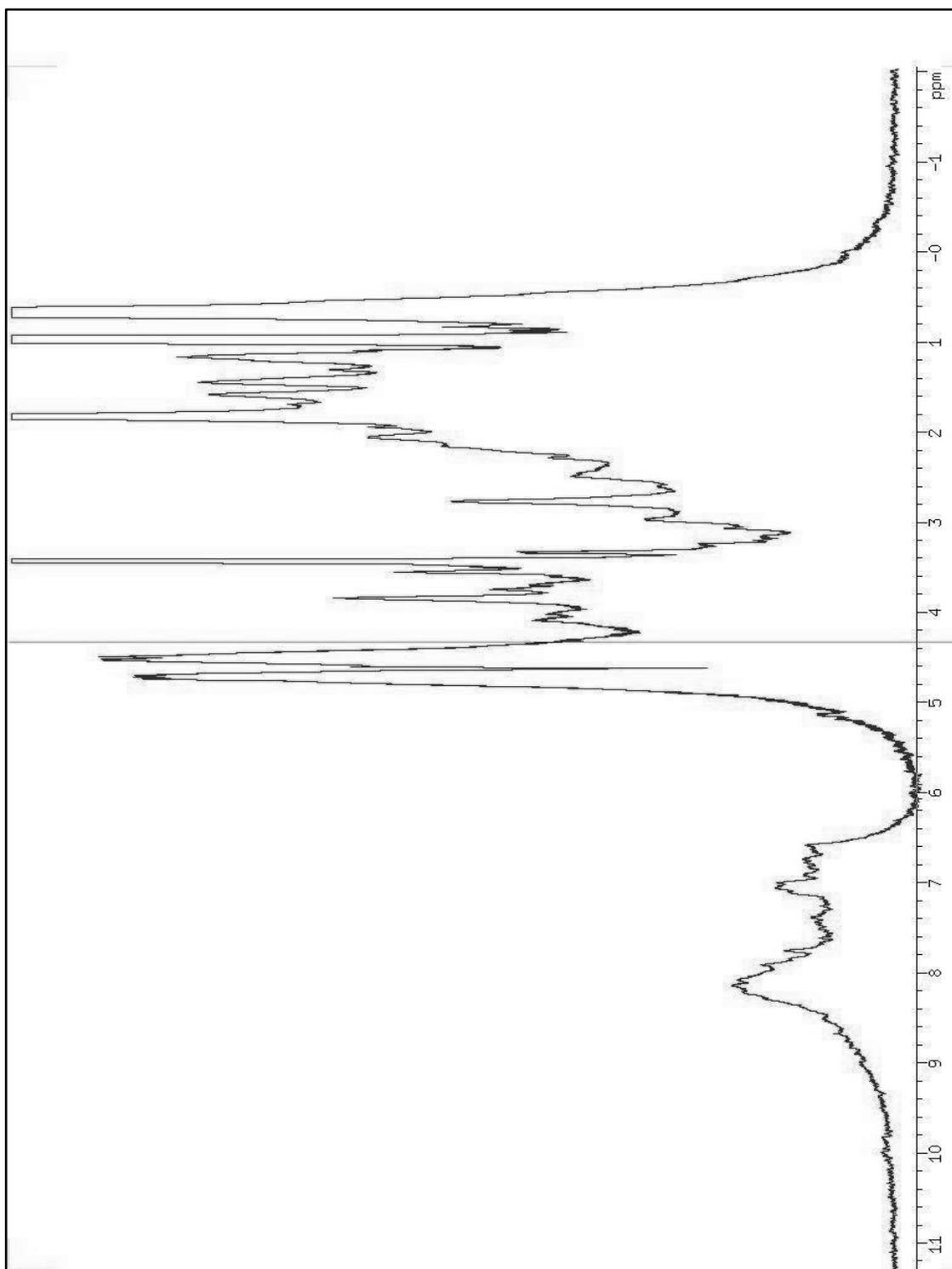
Anexo 13: Espectro de ^1H NMR da XACb0033 ($50 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (10mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



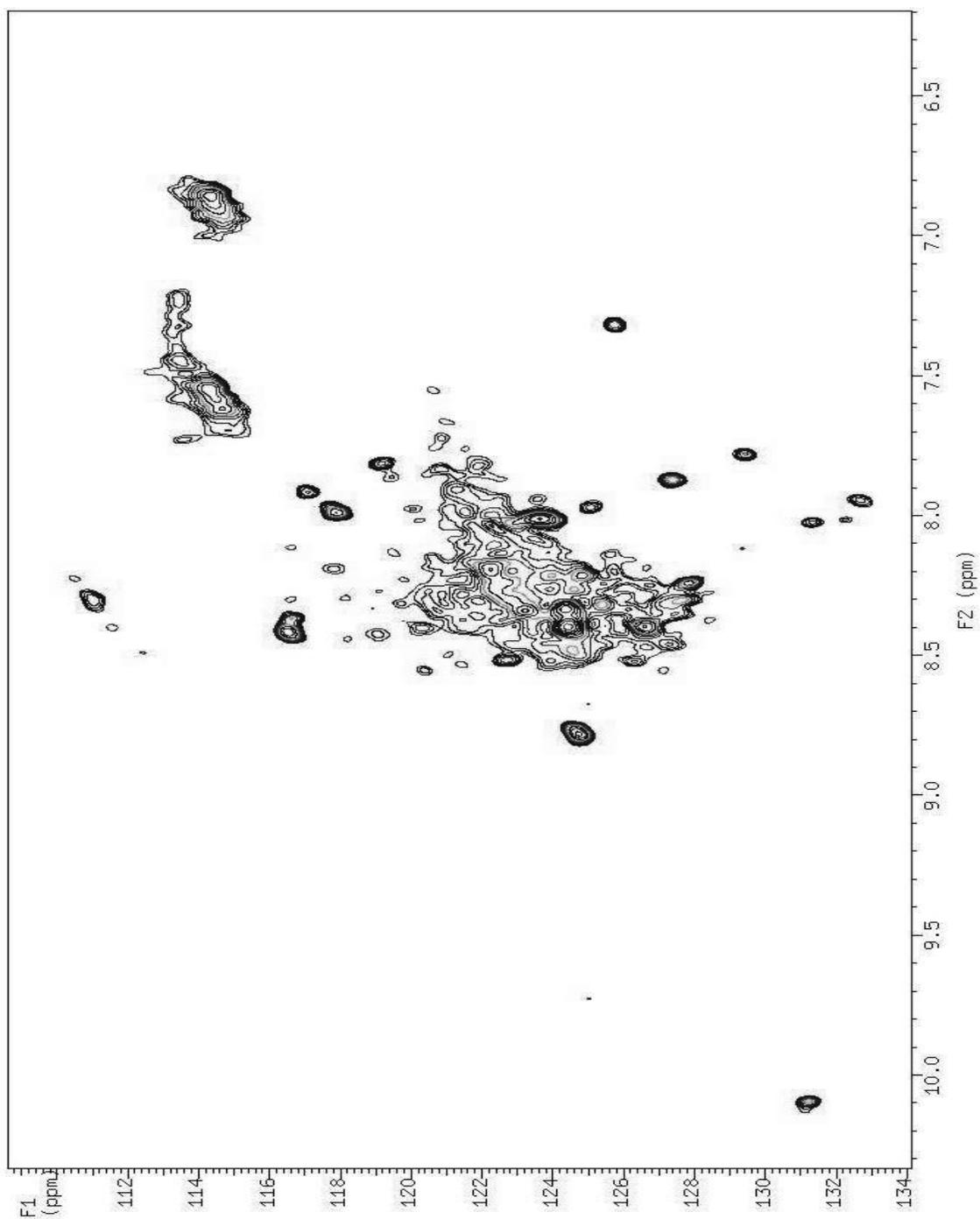
Anexo 14: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XAcB0033 ($50 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (10mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



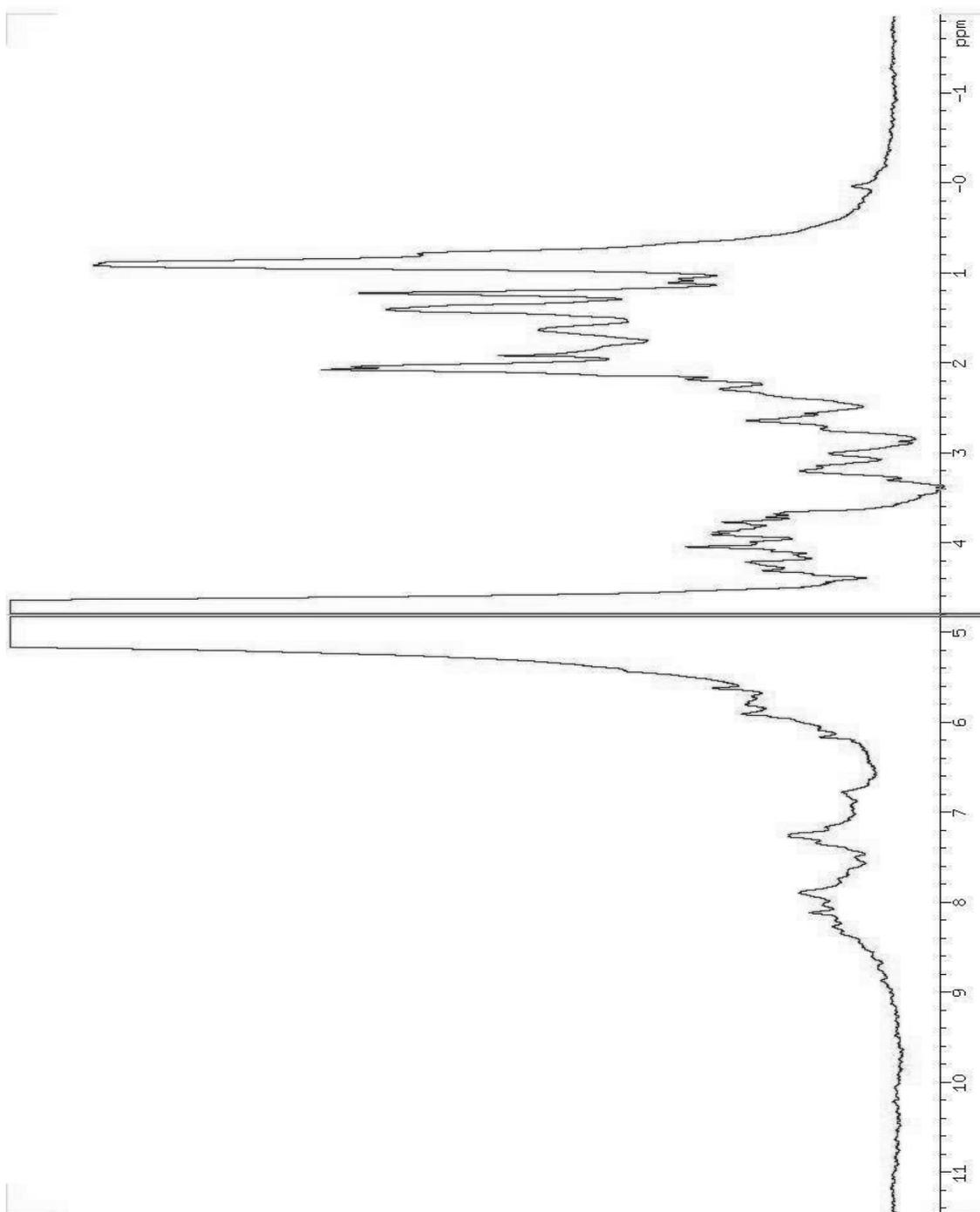
Anexo 15: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 ($50 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (10mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz. Após adição de glicerol.



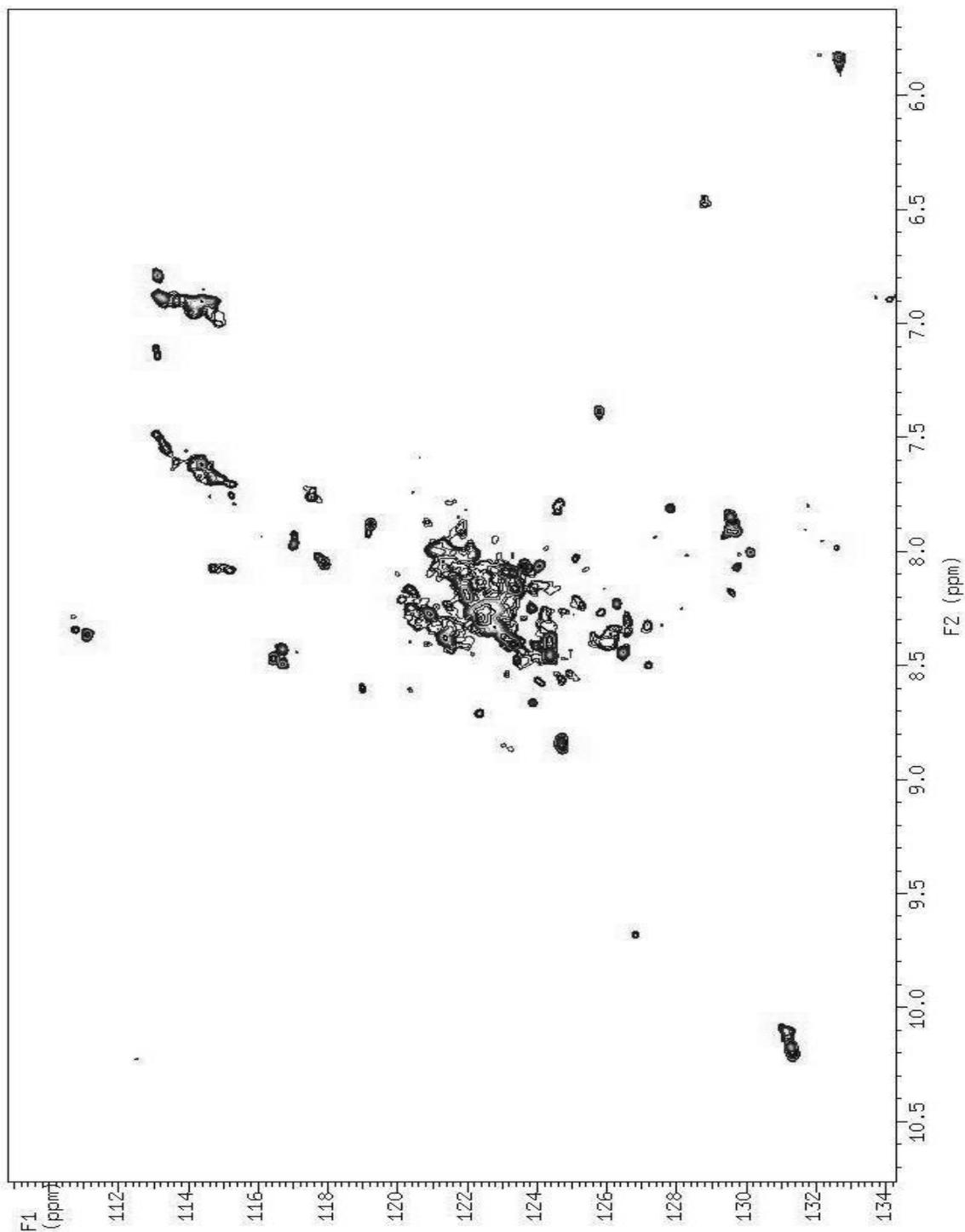
Anexo 16: Espectro de ^1H NMR da XACb0033 ($50 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (30mmol.L^{-1} , pH 5,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



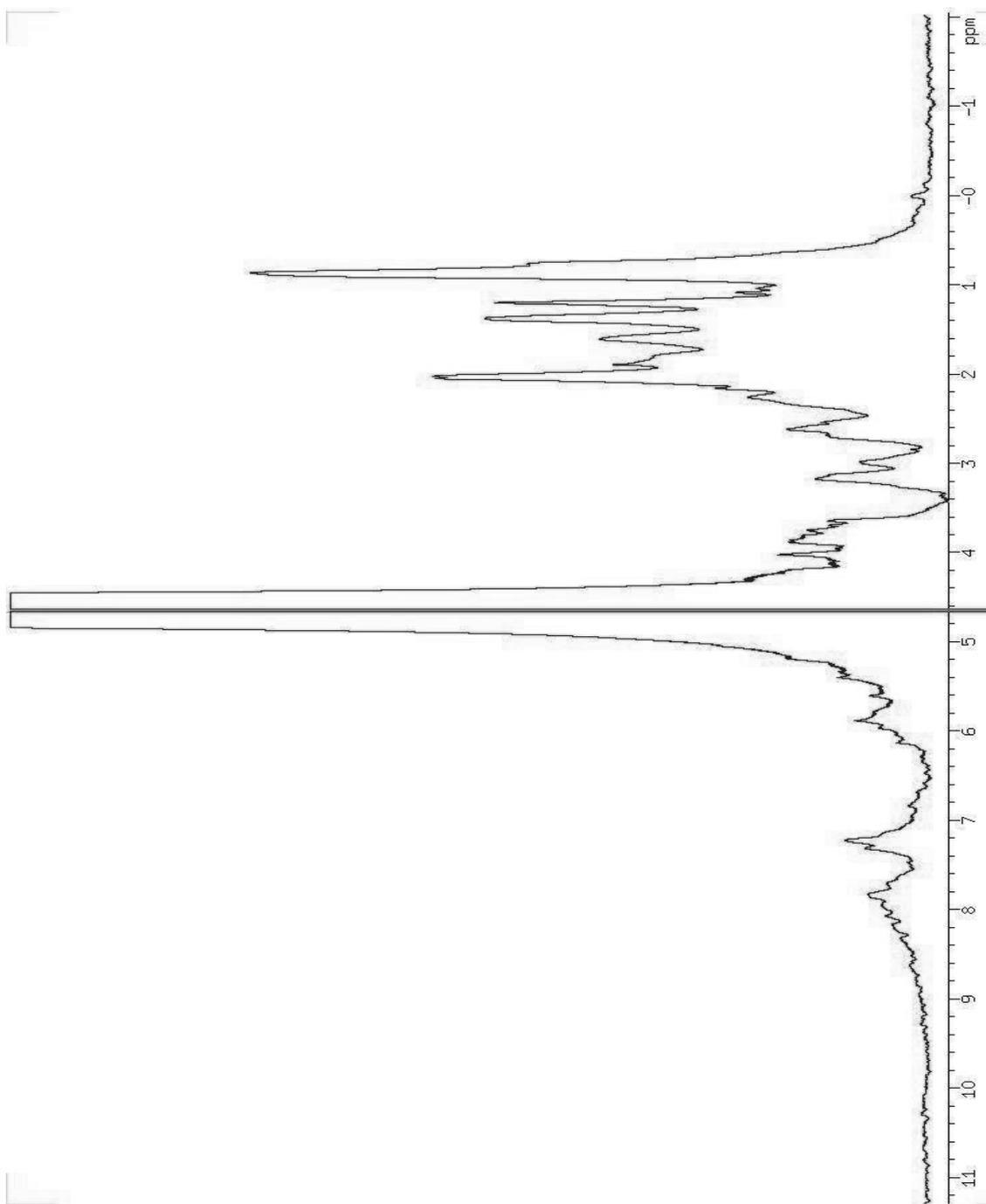
Anexo 17: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XACb0033 ($50 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (30mmol.L^{-1} , pH 5,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



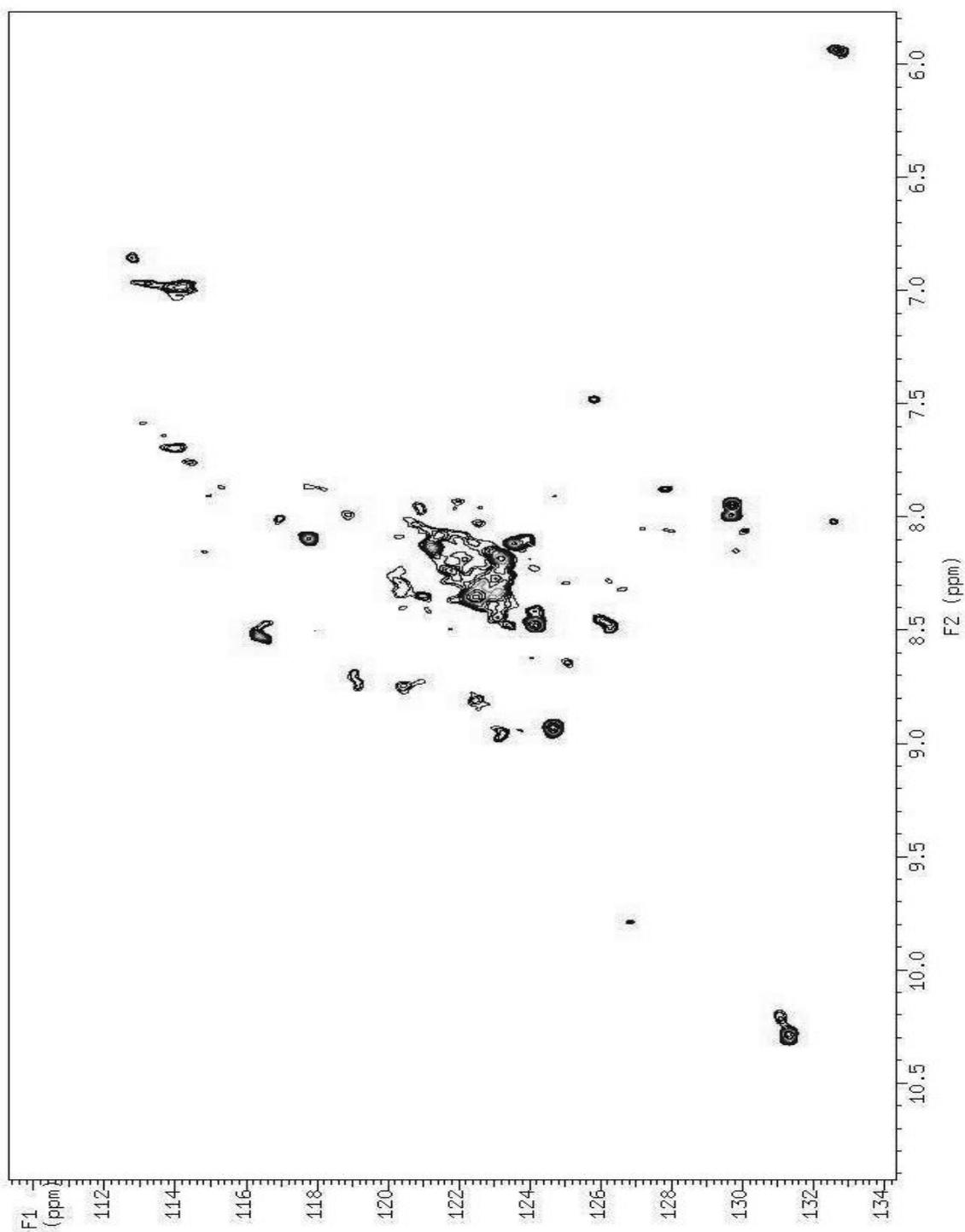
Anexo 18: Espectro de ^1H NMR da XACb0033 ($11 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



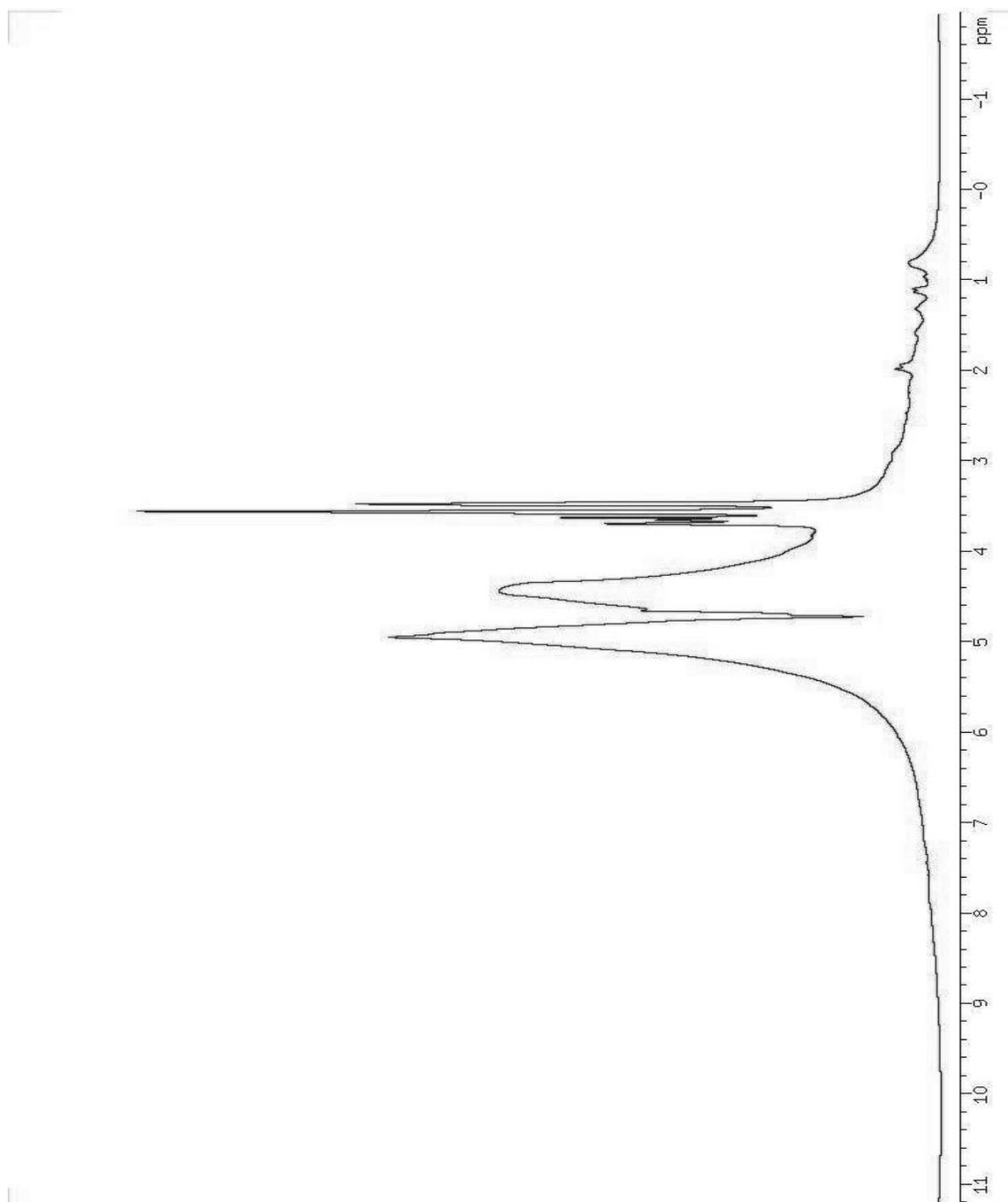
Anexo 19: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XAc0033 ($11 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25mmol.L^{-1} , pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



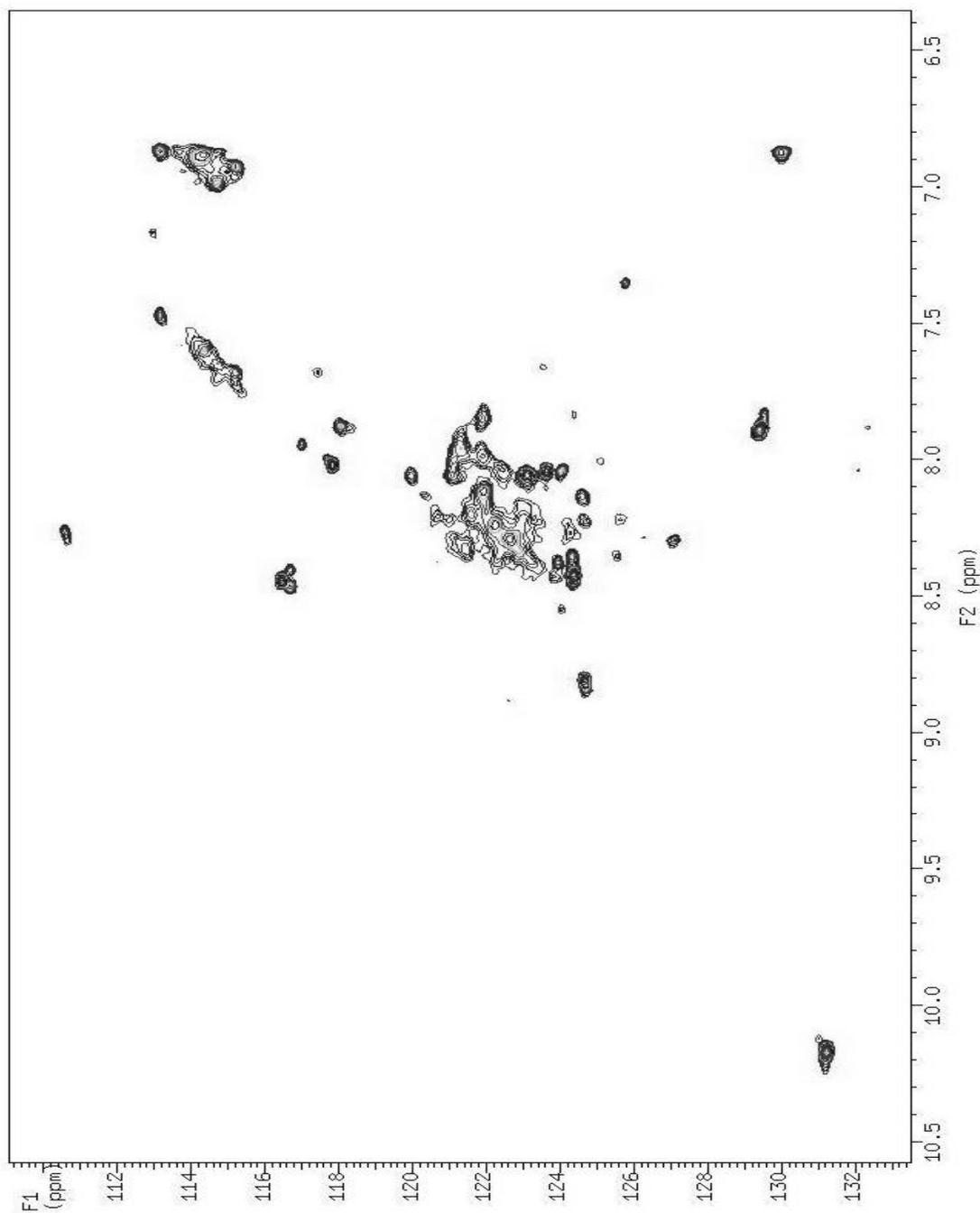
Anexo 20: Espectro de ^1H NMR da XACb0033 ($11 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25mmol.L^{-1} , pH 8,0 a 40°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



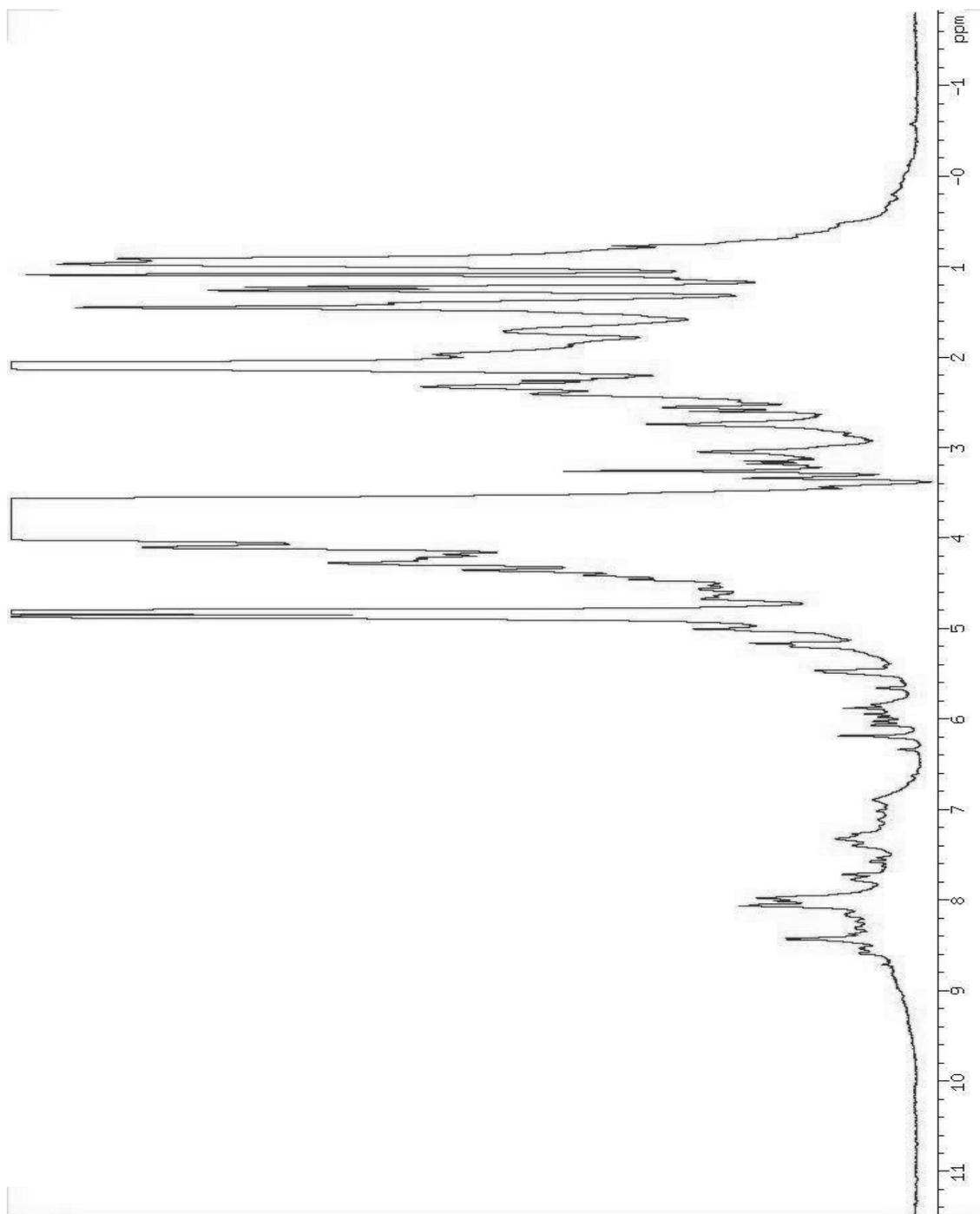
Anexo 21: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC da XAcB0033 ($11 \mu\text{mol.L}^{-1}$) em tampão fosfato de sódio (25mmol.L^{-1} , pH 8,0 a 40°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



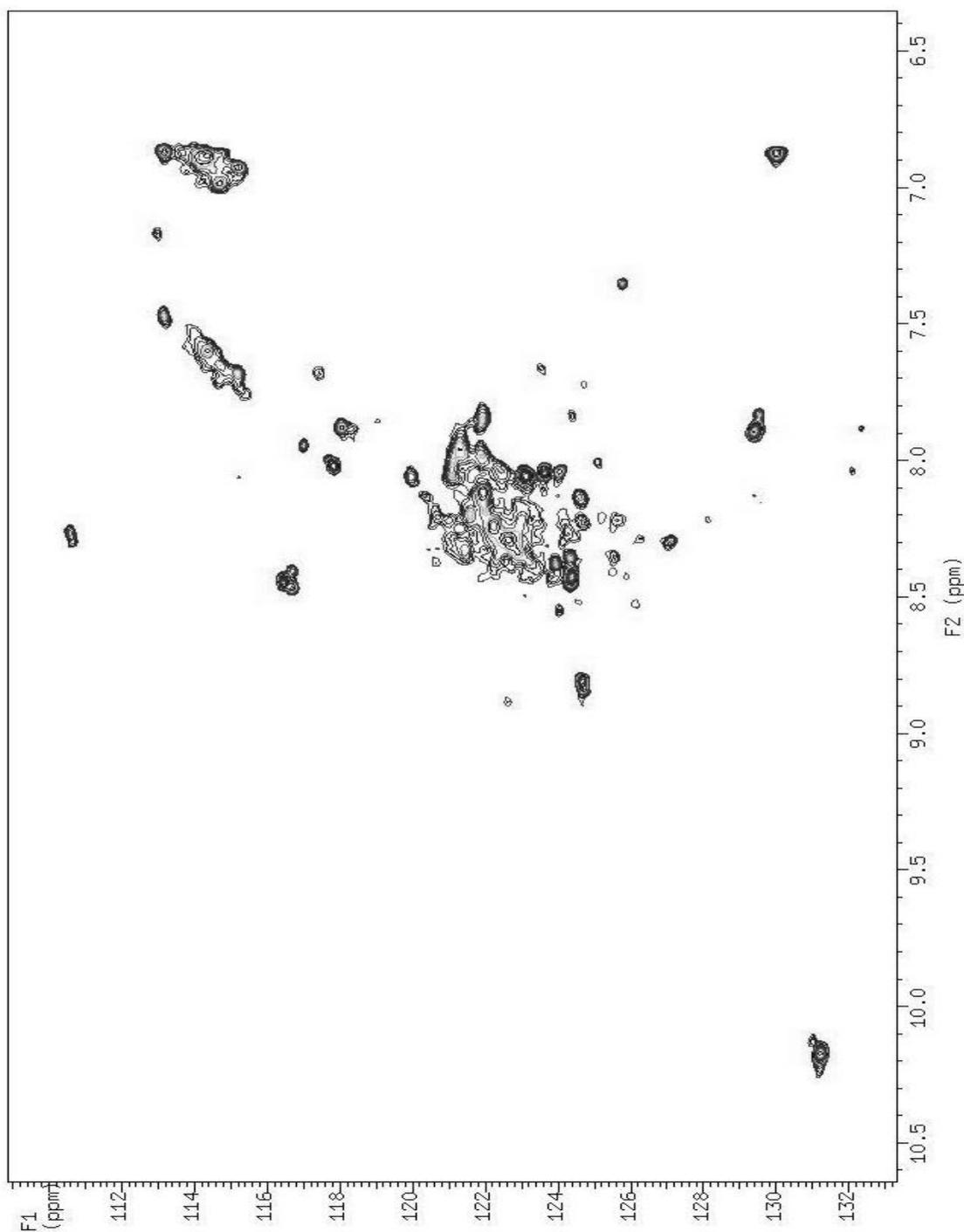
Anexo 22: Espectro de ^1H NMR do complexo XACb0033 + XACb0032+DNA da XACb0033 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



Anexo 23: Mapa de contorno de $\{^{15}\text{N},^1\text{H}\}$ HSQC do complexo XACb0033 + XACb0032+DNA da XACb0033 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



Anexo 24: Espectro de ^1H NMR do complexo XACb0033 + XACb0032 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.



Anexo 25: Mapa de contorno de $\{^{15}\text{N}, ^1\text{H}\}$ HSQC do complexo XACb0033 + XACb0032 na razão molar 2:1 em tampão fosfato de sódio (25 mmol.L^{-1} pH 8,0 a 25°C) com adição de D_2O (5% v/v) obtido no equipamento INOVA 600 MHz.

ANEXOS – ARTIGOS

Metabolic Profiling of Human Blood Serum from Treated Patients with Bipolar Disorder Employing ^1H NMR Spectroscopy and Chemometrics

Alessandra Sussulini,^{†,‡} Alessandra Prando,^{§,||} Danilo Althmann Maretto,^{‡,⊥} Ronei Jesus Poppi,^{‡,⊥} Ljubica Tasic,^{§,||} Cláudio Eduardo Muller Banzato,[#] and Marco Aurélio Zezzi Arruda^{*,†,‡}

Group of Spectrometry, Sample Preparation and Mechanization (GEPAM), National Institute of Science and Technology for Bioanalytics, Organic Chemistry Department, Chemometrics Laboratory in Analytical Chemistry, and National Institute of Science and Technology for Structural Biology and Bioimaging, Institute of Chemistry, University of Campinas (Unicamp), P.O. Box 6154, 13083-970 Campinas, and Department of Psychiatry, Faculty of Medical Sciences, Unicamp, P.O. Box 6111, 13081-970 Campinas, SP, Brazil

Metabolic profiling employing hydrogen nuclear magnetic resonance (^1H NMR) spectroscopy and chemometric analysis of human blood serum samples taken from the control group ($n = 25$) and patients with bipolar disorder ($n = 25$) was performed to identify molecular changes related to the disorder and to different drug treatments: lithium ($n = 15$) versus other medications ($n = 10$). This strategy showed significant potential for exploring pathophysiological and toxicological features involved in bipolar disorder. The investigated groups (control and patients with bipolar disorder under different treatments) could be distinguished according to their metabolic profiles, and the main differential metabolites found were lipids, lipid-metabolism-related molecules (acetate, choline, and *myo*-inositol), and some key amino acids (glutamate, glutamine). Our results suggest that some of the 24 identified metabolites may be linked to lithium- and other-medication-provoked metabolic changes or may even be directly related to the disorder. Thus, these findings may contribute to paving the way for future studies aiming at identifying potential biomarkers for bipolar disorder.

Bipolar disorder, formerly known as manic-depressive psychosis, is one of the most debilitating and common psychiatric disorders worldwide. It is characterized by recurrent mood disturbances that comprise periods of depression (abnormally depressed mood, loss of interest or pleasure in activities that usually are pleasurable, decreased energy or increased fatigability, among others), mania (marked elevated mood, increased energy and activity during at least 1 week), hypomania (elevated mood, increased energy and activity during at least 4 days), and mixed

states (symptoms of both mania and depression). The absence of depressive or manic episodes is called euthymia (normal mood).¹ Bipolar disorder is further categorized into subtypes that include bipolar I (one or more episodes of mania with or without major depressive episodes) and bipolar II (one or more episodes of hypomania as well as at least one major depressive episode).²

Diagnosing bipolar disorder can be challenging sometimes, due to the heterogeneity of the clinical presentation, the unclear boundaries with other mental disorders (hence, a skilled differential diagnosis is required), and, last but not least, a late occurrence of the first episode of mania/hypomania, after recurrent episodes of depression.³ It is still not known what causes bipolar disorder, although a variety of biochemical, genetic, and environmental factors seem to be involved in both causing and triggering bipolar episodes. As of now, there is no independent test to confirm the disorder; diagnosis still relies on clinical expertise and judgment. Very often the individuals with bipolar disorder are misdiagnosed as unipolar depressive (because the first hypomanic or manic episode may only come later on, after one or several depressive episodes), leading to inadequate treatments and outcome. Not to mention that treatment with antidepressant drugs for patients with bipolar depression may provoke a switch into hypomania or mania. Therefore, increased accuracy in diagnosing bipolar disorder is the key to improving the mental health and treatment of patients with the disorder, which could possibly be attained by identifying differential biomolecules that reflect pathophysiological processes in the presence of the illness.^{4–6}

In this work, a metabonomics study employing ^1H NMR and chemometrics was performed to detect molecular changes in

* To whom correspondence should be addressed. E-mail: zezzi@iqm.unicamp.br. Fax: +55 19 35213023.

[†] GEPAM, Institute of Chemistry.

[‡] National Institute of Science and Technology for Bioanalytics, Institute of Chemistry.

[§] Organic Chemistry Department, Institute of Chemistry.

^{||} National Institute of Science and Technology for Structural Biology and Bioimaging, Institute of Chemistry.

[⊥] Chemometrics Laboratory in Analytical Chemistry, Institute of Chemistry.

[#] Department of Psychiatry, Faculty of Medical Sciences.

(1) Malhi, G. S.; Adams, D.; Lampe, L.; Paton, M.; O'Connor, N.; Newton, L. A.; Walter, G.; Taylor, A.; Porter, R.; Mulder, R. T.; Berk, M. *Acta Psychiatr. Scand.* **2009**, *119* (Suppl. 439), 27–46.

(2) American Psychiatric Association. *Diagnostic and Statistical Manual for Mental Disorders*, text revision (DSM-IV-TR), 4th ed.; American Psychiatric Association: Washington, DC, 2000.

(3) Depression and Bipolar Support Alliance. Perceptions and Impact of Bipolar Disorder: How Far Have We Really Come? Presented at the Fourth International Conference on Bipolar Disorder, Pittsburgh, PA, June 14–16, 2001.

(4) Phillips, M. L.; Vieta, E. *Schizophr. Bull.* **2007**, *33*, 893–904.

(5) Marmol, F. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2008**, *32*, 1761–1771.

human blood serum samples by comparing the metabolic profiles of healthy subjects, patients being treated for bipolar disorder with lithium,^{6–8} and patients being treated for bipolar disorder with other medications not including lithium. The medications routinely used in bipolar disorder treatment include mood stabilizers (lithium, valproic acid, carbamazepine), second-generation antipsychotics (olanzapine, risperidone, quetiapine), antidepressants (especially—and carefully—selective serotonin reuptake inhibitor drugs—the SSRIs—such as fluoxetine, paroxetine, sertraline, and citalopram), and anxiolytics (clonazepam, diazepam).¹ Although lithium is the most widely used drug in many cases for bipolar disorder treatment (both to treat current episodes and to prevent further ones), the precise neurobiological mechanisms through which lithium exerts its clinical effects are not clear, and some results found in the literature are contradictory.^{5,8,9} Therefore, two groups of bipolar disorder patients (treated with lithium or not) were studied to evaluate lithium effects on blood serum metabolomics.

Metabolomics is defined as the measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification. With metabolomics, changes in endogenous metabolite levels that may result from disease processes, drug toxicity, or gene function have been evaluated in cells, tissues, or biological fluids.^{10–16} Latent biochemical information obtained from metabolomics may be used for diagnostic or prognostic purposes. Such information reflects actual biological events rather than the potential for disease which gene expression data provide.¹⁷ Different analytical platforms, based on mass spectrometry or NMR spectroscopy, are currently used for such studies. While mass spectrometry is more sensitive and specific in comparison to NMR, it relies on the separation of the analytes prior to detection using GC, HPLC, or CE. In this sense, NMR spectroscopy is more closely a universal detector in that the sample can be analyzed directly and many kinds of small metabolites can be measured at the same time.^{12,14,16}

Hydrogen nuclear magnetic resonance (¹H NMR) spectroscopy already enabled a large number of biofluid constituents to be identified and catalogued.^{10–12} ¹H NMR has an exceptional reproducibility and is quantitative to the extent that a given peak area is directly proportional to the concentration of the corresponding metabolite, which has allowed it to become a well-established technique used in metabolomics studies.¹³ Various alterations in the metabolite levels present

in the brain of patients with psychiatric disorders using NMR spectroscopy have been reported.¹⁸ For bipolar disorder investigations, in vivo hydrogen nuclear magnetic resonance spectroscopy (¹H MRS),¹⁹ hydrogen magnetic resonance imaging (¹H MRSI),²⁰ and ¹H NMR spectroscopy-based metabolomics have been applied.²¹

In this rather exploratory work, metabolic profiling has been performed for searching molecular changes in human blood serum related to bipolar disorder and the lithium treatment (the treatment that has the longest record of efficacy^{6–8}). Blood serum samples were chosen for being obtained by a minimally invasive method and for showing spectroscopic profiles robust to small variations in sample collection and handling relative to biological differences.²² Our results pointed out that the three investigated groups (the control and bipolar disorder patients under two different treatments) can be distinguished according to their metabolic profiles and the identified metabolite alterations could guide future studies on biomarker discovery for this disorder.

MATERIALS AND METHODS

Serum Collection and Storage. This study was approved by the local Ethics Committee (Hospital de Clínicas, University of Campinas, Brazil), and the subjects gave their written informed consent before sample collection. All blood samples were taken in the afternoon (between 14 and 16 h). Blood was drawn into Vacutainer tubes, immediately placed on ice, allowed to clot for at least 30 min, and centrifuged at 1500g for 15 min. The obtained serum was aliquoted, transferred into polypropylene tubes containing 0.01% (m/v) sodium azide, and stored at – 80 °C until assayed. The maximum period of storage was two weeks.

Fifty serum samples were collected and classified into three groups: the control group, constituted by 25 samples of subjects without bipolar disorder, bipolar disorder patients currently under treatment with lithium group, constituted by 15 samples, and bipolar disorder patients under treatment with other drugs than lithium group, constituted by 10 samples. Bipolar disorder patients were all in the euthymic state, previously identified as bipolar I, and under treatment in the psychiatric outpatient clinic (Hospital de Clínicas, University of Campinas, Brazil). No participants have other concomitant diseases such as cancer, AIDS, or hepatic, endocrinological, or metabolic diseases. The summary of the collected sample characteristics is presented in Table 1, and the description of the medications used for the treatment of bipolar disorder patients is described in Table 2. Bipolar disorder is associated with significantly higher prevalences of tobacco smoking behavior compared with the general population.²³ Such higher prevalence of tobacco smoking was found in our sample of bipolar patients, which means that this variable was not controlled in our study.

- (6) Oswald, P.; Souery, D.; Kasper, S.; Lecrubier, Y.; Montgomery, S.; Wyckaert, S.; Zohar, J.; Mendlewicz, J. *Eur. Neuropsychopharmacol.* **2007**, *17*, 687–695.
- (7) Gitlin, M. *Mol. Psychiatry* **2006**, *11*, 227–240.
- (8) Maj, M. *Bipolar Disord.* **2003**, *5*, 180–188.
- (9) Detera-Wadleigh, S. D.; MacMahon, F. J. *Biol. Psychiatry* **2006**, *60*, 106–114.
- (10) Lenz, E. M.; Wilson, I. D. J. *Proteome Res.* **2007**, *6*, 443–458.
- (11) Moco, S.; Bino, R. J.; de Vos, R. C. H.; Vervoort, J. *Trends Anal. Chem.* **2007**, *26*, 855–866.
- (12) Wishart, D. S. *Trends Anal. Chem.* **2008**, *27*, 228–237.
- (13) Maher, A. D.; Crockford, D.; Toft, H.; Malmödin, D.; Faber, J. H.; McCarthy, M. I.; Barrett, A.; Allen, M.; Walker, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2008**, *80*, 7354–7362.
- (14) Dunn, W. B.; Ellis, D. I. *Trends Anal. Chem.* **2005**, *24*, 285–294.
- (15) Kaddurah-Daouk, R.; Krishnan, K. R. *Neuropsychopharmacology* **2009**, *34*, 173–186.
- (16) Nicholson, J. K.; Lindon, J. C. *Nature* **2008**, *455*, 1054–1056.
- (17) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. *Nat. Rev. Drug Discovery* **2002**, *1*, 153–161.

- (18) Sanacora, G.; Rothman, D.; Krystal, J. H. *Neuroscientist* **1999**, *5*, 192–196.
- (19) Öngür, D.; Prescott, A. P.; Jensen, J. E.; Cohen, B. M.; Renshaw, P. F. *Psychiatry Res.: Neuroimaging* **2009**, *172*, 44–48.
- (20) Bertolino, A.; Frye, M.; Callicott, J. H.; Mattay, V. S.; Rakow, R.; Shelton-Repella, J.; Post, R.; Weinberger, D. R. *Biol. Psychiatry* **2003**, *53*, 906–913.
- (21) Lan, M. J.; McLoughlin, G. A.; Griffin, J. L.; Tsang, T. M.; Huang, J. T. J.; Yuan, P.; Manji, H.; Holmes, E.; Bahn, S. *Mol. Psychiatry* **2009**, *14*, 269–279.
- (22) Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C. *Anal. Chem.* **2006**, *78*, 4307–4318.
- (23) Diaz, F. J.; James, D.; Botts, S.; Maw, L.; Susce, M. T.; de Leon, J. *Bipolar Disord.* **2009**, *11*, 154–165.

Table 1. Collected Sample Characteristics^a

	control	bipolar patients using lithium	bipolar patients not using lithium
sample size	<i>n</i> = 25	<i>n</i> = 15	<i>n</i> = 10
age, years (mean ± sd)	28 ± 5	40 ± 13	42 ± 17
no. of each gender (female/male)	14/11	9/6	7/3
duration of illness, years	NA	1–28	1–20
duration of current treatment, months	NA	2–240	1–120
no. with treatment with antipsychotics	NA	5	3
no. with treatment with mood stabilizers	NA	15	10
no. with treatment with antidepressants	2	1	2
no. with treatment with anxiolytics	NA	5	5
lithium doses, mg (mean ± sd)	NA	911 ± 325	NA
fraction smoking, %	0	26.7	20

^a Abbreviations: sd, standard deviation; NA, not applicable.

Table 2. Description of the Medications Used for the Treatment of Bipolar Disorder Patients

sample no.	bipolar patients using lithium	bipolar patients not using lithium
1	lithium and carbamazepine	risperidone, clonazepam, fluoxetine, and diazepam
2	lithium and olanzapine	carbamazepine, valproic acid, and diazepam
3	lithium and valproic acid	valproic acid
4	lithium and valproic acid	valproic acid
5	lithium, olanzapine, and clonazepam	valproic acid
6	only lithium	quetiapine
7	lithium and risperidone	valproic acid and lamotrigine
8	only lithium	valproic acid and clonazepam
9	lithium, fluoxetine, and clonazepam	valproic acid, clonazepam, and olanzapine
10	only lithium	valproic acid and citalopram
11	lithium, diazepam, and chlorpromazine	
12	lithium and clonazepam	
13	lithium and lorazepam	
14	only lithium	
15	lithium, risperidone, and valproic acid	

Sample Preparation and Acquisition of ¹H 1D (*T*₂-Edited) and 2D NMR (TOCSY) Data.

For NMR spectroscopic analyses, serum samples were thawed and centrifuged at 12300*g* for 10 min at 4 °C to separate any precipitate. Aliquots of 250 μL of the supernatants were diluted with 350 μL of D₂O and placed in 5.0 mm diameter NMR tubes. A simple dilution procedure was employed to avoid additional sample preparation steps because one of the intentions of the proposed methodology is to be applied in bipolar disorder diagnosis and/or lithium treatment monitoring. All the NMR experiments were carried out without spinning at 499.89 MHz and 25 °C on an INOVA-500 (*B*₀ = 11.7 T) spectrometer (Varian, Palo Alto, CA) equipped with a 5 mm inverse triple-resonance probe. Standard one-dimensional (1D) PRESAT spectra were acquired using a 90° pulse sequence, with 128 scans, a pulse length of about 10 μs, and a recycle delay of 2 s. The 1D spin-echo spectra were recorded using the CPMG (Carr–Purcell–Meiboom–Gill) sequence of $D-[-90^\circ-(\tau-180^\circ-\tau)n-ACQ]$, where a fixed total spin-spin relaxation delay $2n\tau$ of 100 ms was used to attenuate the broad NMR signals from slowly tumbling molecules such as proteins and retain those from low molecular weight compounds and some lipid components. Typically, 64 transients and 32K data points were collected with a spectral width of 12 kHz, an acquisition time of 1.64 s, and a relaxation delay of 4 s. The free induction decay (FID) was zero-filled to 64K, and an exponential line-broadening function of 0.3 Hz was applied to the FID prior to Fourier transformation. All spectra were carefully phase and baseline corrected and referenced to the methyl peak of lactate at 1.33 ppm (3 H, d, ³*J* = 7 Hz), present

in all spectra, since TSP is not a suitable reference for the serum samples due to interaction-induced line broadening. All spectra were processed with VNMR software (Varian). To confirm the assignments made from 1D ¹H NMR spectra, some blood serum samples were also examined using 2D ¹H–¹H TOCSY spectra with solvent suppression. The spectra were acquired with a 1.5 s relaxation delay, 1.5 s water signal suppression, and 6030.5 Hz spectral width for the ¹H dimensions. For each 2D spectrum, 256 increments with 64 transients per increment were collected and extended to 4K data points using linear prediction and zero filling approaches. The TOCSY experiments used an MLEV-17 spin-lock scheme for ¹H–¹H transfers with a mixing time of 90 ms at a spin-lock strength of 8 kHz. The signal assignments were based on the literature and/or Madison Metabolomics Consortium Database^{24,25} and are indicated on the *T*₂-edited spectrum and confirmed by the 2D fully assigned ¹H–¹H TOCSY spectrum.

Chemometrics Analyses of ¹H NMR Spectral Data. ¹H NMR data were transported to a data matrix, and chemometrics analyses, based on interval principal component analysis (iPCA) and partial least-squares discriminant analysis (PLS-DA), were performed using MATLAB 6.5 software (The Mathworks, Natick, MA). iPCA was performed for the spectral region of chemical shifts ranging from –0.5 to +4.4 ppm. The principle

(24) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. *Nat. Biotechnol.* **2008**, *26*, 162–164.

(25) *Madison Metabolomics Consortium Database*; University of Wisconsin: Madison, WI. <http://mmcd.nmr.fam.wisc.edu/>. (Accessed July 20, 2009).

of this algorithm is to split the spectra into smaller equidistant regions and, afterward, calculate PCA models for each interval and present the results in multiple score plots.²⁶ PLS-DA is intended to give a data overview and can be helpful in exploratory studies and interpretation (e.g., when looking for clustering among the samples). It can also be used as a supervision method of classification. The analyzed spectral area ranged from -0.5 to $+4.4$ ppm, where information and reduced noise were obtained. The chosen preprocessing method was the orthogonal signal correction (OSC), and unnecessary information was eliminated. In the OSC procedure, the X matrix was corrected by a subtraction of variation orthogonal to the y vector calibration.²⁷ Considering this work, y was a vector containing the class corresponding to each sample. This vector was used for both preprocessing OSC and PLS-DA algorithms. All variables were mean-centered, the spectra were normalized, and a "leave one out" cross-validation was performed. After that, another validation of the model was done by splitting at random the samples into calibration and validation sets and building a new PLS-DA model using the calibration set and the same conditions of the previous one. Then a class prediction of the validation samples was done.

Chemicals. All chemical reagents were of analytical grade. Deuterium oxide (D_2O ; 99.9% D) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA), and sodium azide was purchased from Sigma (St. Louis, MO).

RESULTS

1H NMR Spectral Data Analysis Using Chemometrics.

The chemometrics tools were combined with NMR spectroscopy in the analyses of human blood serum samples to evaluate changes in the metabolic profiles in the presence of bipolar disorder and to compare drug treatments with and without lithium. The spectral region between 4.5 and 4.8 ppm was not considered because of the residual water signal (HDO, 4.7 ppm). The region above 4.8 ppm was also not considered for further analyses because it did not show important differences among the groups.

To explore all the potential differences in the metabolic profiles of the studied groups, the NMR spectra were first pretreated with the standard normal variate (SNV) for correcting spectral noise and background effects that caused baseline shifting and tilting in the spectra, then segmented into 0.6 ppm intervals, and finally subjected to iPCA. On the basis of the iPCA results, it was possible to observe a distinction of the samples into two groups: the control and bipolar disorder patients (independently of the drug treatment). Some samples presented a different behavior than the majority of the group, but it is a difficult task to determine exactly the reason for the metabolic profile differentiation, as we do not have the means for accounting for such outliers. One interesting point concerns the subject from sample 18, a male control classified together with the bipolar disorder patient group. This subject does not take any medications, nor has he had a psychiatric history, but his younger sister has bipolar disorder.

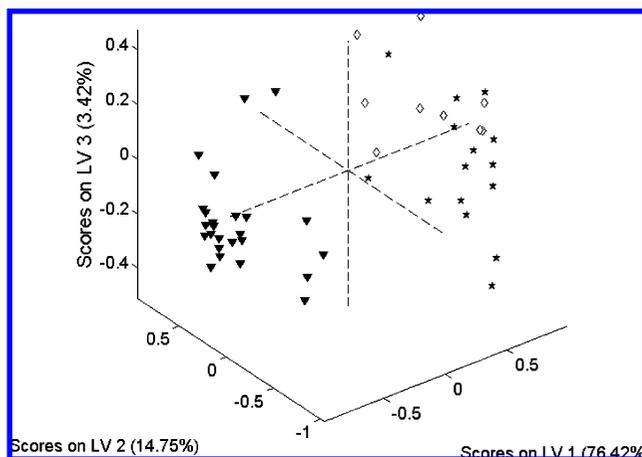


Figure 1. PLS-DA score plot for the chemical shift interval from -0.5 to $+4.4$ ppm. Samples from the control group, bipolar patients treated with lithium, and bipolar patients not treated with lithium are represented with the following symbols: ∇ , \star , and \diamond , respectively.

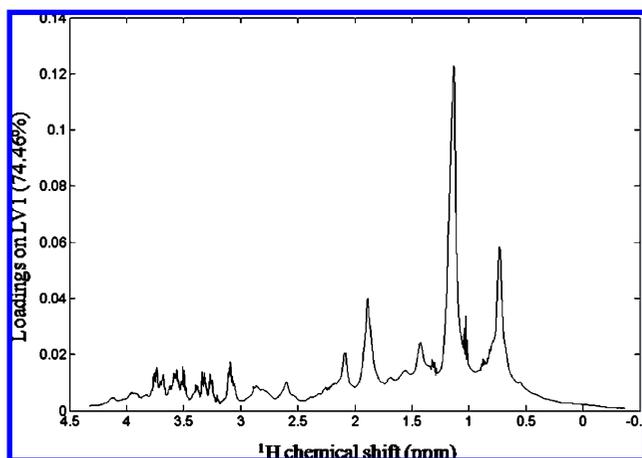


Figure 2. PLS-DA loading plot on the first latent variable for the chemical shift interval ranging from -0.5 to $+4.4$ ppm.

This observation can be in agreement with previous researches that evidence a genetic characteristic of the disease.²⁸

Considering the iPCA results and the previous observations about the spectral regions not relevant for analyses, PLS-DA was performed. For such analyses, the spectra were first normalized and then processed with OSC²⁹ preprocessing, used to remove information within the NMR data not correlated to the target variables, by applying restricted principal component analysis. This data-filtering method is particularly important to human metabolomic studies like this one, because of the great variability in human populations, when compared to studies involving laboratory-controlled animals.

Figure 1 shows the PLS-DA score plots for the best model built, which was the one using 10 latent variables (LVs) with a variance of 98.74% in the X -block (spectra) and 62.05% in the Y -block (classes). The score plot allowed the visualization of the relations among the samples in the plane model to estimate whether there are any clustering, trends, and/or outliers.³⁰ On the basis of the PLS-DA results, the three groups, i.e., the control

(26) Nørgaard, L. *The iToolbox for MATLAB*; KVL: Frederiksberg, Denmark. <http://www.models.life.ku.dk>. (Accessed July 27, 2009).

(27) Sena, M. M.; Chaudhry, Z. F.; Collins, C. H.; Poppi, R. J. *J. Pharm. Biomed. Anal.* **2004**, *36*, 743–749.

(28) Kato, T. *Neurosci. Res.* **2001**, *40*, 105–113.

(29) Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175–185.

(30) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–479.

Table 3. Chemical Shift Assignments for Metabolites Identified in ¹H NMR Spectra (500 MHz) of Serum Samples^a

peak ^b	chemical shift/ppm	molecule	assignment
1	0.99	valine/lipids/ lipoproteins/mobile lipids	CH ₃
2	1.04	valine/broad peak underneath proteins	CH ₃
3	1.33	lactate, lipids, lipoproteins, mobile lipids	CH ₂
4	1.48	alanine	CH ₃
5	1.57	lipids (mainly vldl's ^c)	CH ₂ CH ₂ CO
6	1.72	lipids	CH ₂ CH ₂ CH=CH
7	1.93	lipids	CH₂CH=CH
8	1.95	glycoprotein lipids	CH₂CH₂CH=CH
9	1.99	proline	β-H
10	2.04	glutamine	β-H
11	2.12	acetate	CH₃
12	2.24	valine	β-H
13	2.29	proline	β-H
14	2.30	glutamate	γ-H
15	2.44	asparagine	β-H
16	2.6 and 2.64	lipids	CH=CHCH₂CH=CH
17	2.81, 2.83, and 2.85	albumin lysyl	ε-CH ₂
18	3.03	lysine/creatine	δ-H (lysine)
19	3.21	choline	δ-CH₂
20	3.24	arginine/glucose	δ-H (arginine)
21	3.37	proline/glucose	δ-H (proline)
22	3.54	myo-inositol	H1, H3
23	3.77	arginine	α-H
24	3.79	lysine	α-H

^a In bold are highlighted the metabolites with higher loading values (see Figure 2). ^b According to Figure 4d. ^c Very low density lipids.

(group I), bipolar treated with lithium (group II), and bipolar treated with other medications not including lithium (group III), showed trends to form distinct groups, as illustrated in Figure 1. The samples from group I had scores below zero in LV1, staying apart from the other samples, which had scores above zero. Despite the fact that the samples from group II and group III had not formed individual groups, it is possible to note different trends of each set. Samples from group II had a trend to have negative values in LV2, while samples from group III had a trend to have positive values in LV2.

The loading plot (Figure 2) highlights the most significant variables by describing the influence and relation among the variables in the model plane.³⁰ Therefore, it is possible to conclude that the separation among the groups is due mainly to the peaks with chemical shifts of 0.99, 1.04, 1.33, and 1.93 ppm. Such peaks refer to valine, lactate, lipids, and lipoproteins. Other very important chemical shifts are listed in Table 3.

To validate the model, the samples were split at random into calibration and validation sets. Five, three, and two samples were selected and respectively correlated to the control, bipolar treated with lithium, and bipolar treated with other medications not including lithium for validation sets. A new PLS-DA model was built using the calibration set applying the same conditions of the previous model. To build it, the spectral calibration set was first corrected with OSC. Once the correction was done, validation or prediction data could be adjusted through the function $NEWX = X - X(NW) \text{inv}[(NP')(NW)](NP')$, where $NEWX$ is the OSC-corrected validation or prediction matrix, X is the validation or prediction original matrix, NW is the weights matrix, NP' is the loads matrix, and NT is the scores matrix that were used in making the correction. Inv is the inverse of the square matrix $[(NP')(NW)](NP')$. The OSC algorithm³¹ gives all these parameters.

The new model was built using the same number of latent variables as the previous model after the normalization of all

spectra. The variance was 98.90% in the X-block (spectra) and 93.94% in the Y-block (classes). A leave one out cross-validation was performed, with errors of 0%, 20.24%, and 15.62%, respectively, for the three groups. Figure 3 shows the predictions for the three classes of the studied groups. It can be seen that the predictions were precise to classify samples among the bipolar disorder patient groups and the control group. Two mistakes occurred in the prediction of group II, where sample 43, which actually belongs to group I, was predicted and sample 47 was not predicted.

Differential Metabolite Identification. Figure 4 shows the ¹H NMR spectra for a sample of each one of the studied groups. The same pattern was observed in all spectra, but specific peaks appeared with different intensities for each group. These results were also demonstrated by the PLS-DA analysis. The NMR spectra contain broad peaks from molecules with high molecular mass, such as lipids and proteins, which resulted in a rugged baseline that caused overlapping of some low molecular mass compounds and disabled their identification. The chemical shift is perhaps the most important parameter of an NMR spectrum. It is directly proportional to the electron density surrounding the nucleus and is used to obtain structural information about the molecules present in a sample.³² In Figure 4d, the marked peaks refer to the metabolites that differed between the groups studied. These metabolites were identified as lipids, lipid-metabolism-related molecules, and amino acids (Table 3) by comparing their chemical shifts to those previously reported in the literature.^{33,34} Combining the results from the loading plot (Figure 2) with the information obtained from the

(32) de Oliveira, P. R.; Tasic, L.; Rocco, S. A.; Rittner, R. *Magn. Reson. Chem.* **2006**, *44*, 790–796.

(33) Nicholson, J. K.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Lindon, J. C. *Anal. Chem.* **1995**, *67*, 793–811.

(34) Tang, H.; Wang, Y.; Nicholson, J. K.; Lindon, J. C. *Anal. Biochem.* **2004**, *325*, 260–272.

(31) *PLS Toolbox 3.5 for MATLAB*; Eigenvector Research: Wenatchee, WA, 2005.

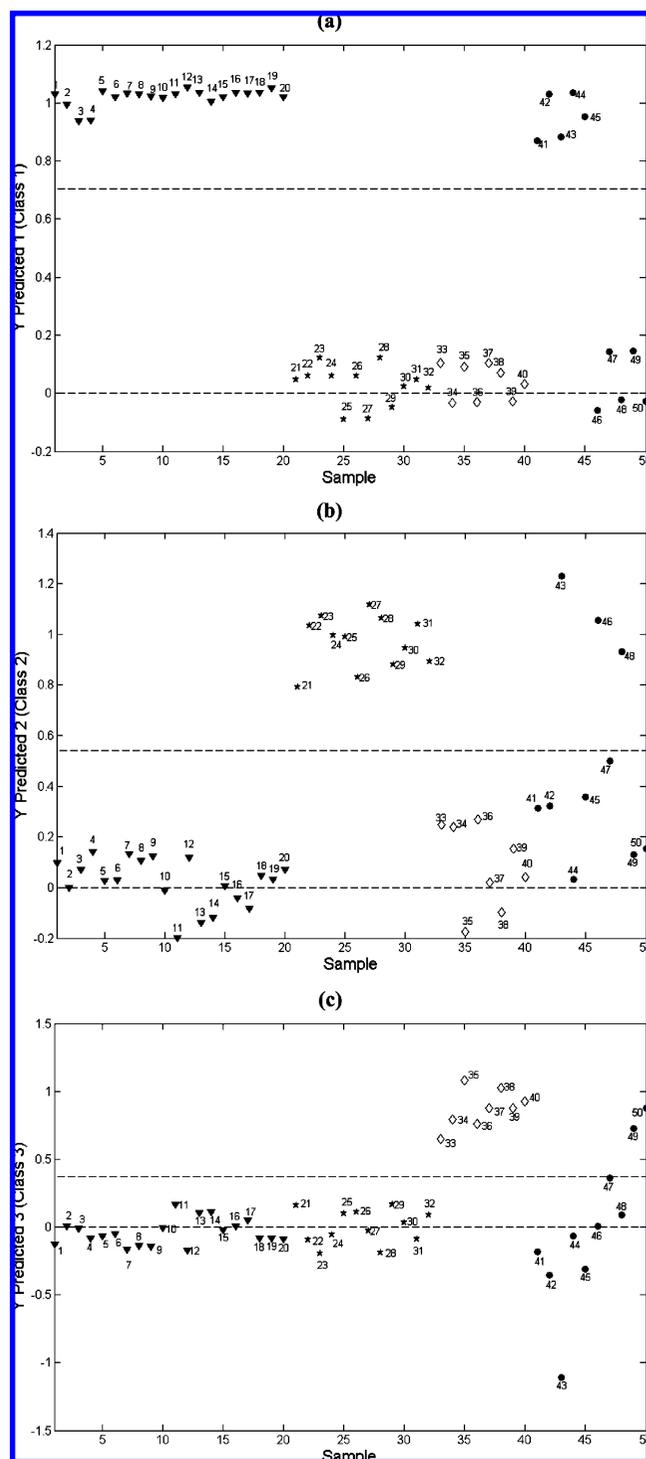


Figure 3. Predictions for (a) class 1 (control group), (b) class 2 (bipolar patients treated with lithium), and (c) class 3 (bipolar patients not treated with lithium). Samples 1–40 are from the calibration set: class 1 (▼), class 2 (★), and class 3 (◇). All the samples marked with a solid circle are from the validation set. Samples 41–45 belong to class 1, 46–48 to class 2, and 49 and 50 to class 3. The dashed line above zero (a–c) is the threshold. A color code can be seen in Figure S-2 in the Supporting Information.

NMR spectra (Figure 4 and Table 3), significant potential biomarkers for distinguishing bipolar disorder patients from healthy controls and also bipolar patients under treatment using lithium from those not undergoing lithium treatment are possible to

assign. Such molecules, shown in bold (Table 3), are glycoprotein lipids, mono- and polyunsaturated lipids, acetate, choline, and *myo*-inositol.

DISCUSSION

A metabonomic approach was employed to evaluate the metabolic profile of blood serum from patients with treated bipolar disorder. Those with the disorder were readily distinguished from control subjects. Moreover, bipolar patients under treatment using lithium were distinguished from those not treated with this drug by comparing ^1H NMR metabolic profiles (illustrated in Figure 1). The metabolic profiles of healthy subjects (control group) and bipolar disorder patients under treatment using lithium (15 subjects) or not using lithium (10 subjects) were compared. For that purpose, two different control groups were considered: one age-matched with the bipolar patients (data shown in Figure S-1, Supporting Information) and another with a higher sample number as described in Table 2. The characteristics of the first group were age average of 31 ± 5 years (age averages among groups not differing with statistical significance at 5% probability as indicated by the Tukey test³⁵), sample number of 15 (9 female and 6 male), and all nonsmokers. The results showed that age was a nonrelevant parameter considering bipolar disorder metabolic profile changes in the present case.

As mentioned before, the loading plot (Figure 2) highlighted the most significant variables with the highest loading values, which enabled the identification of potential biomarkers for bipolar disorder: glycoprotein lipids, mono- and polyunsaturated lipids, acetate, choline, glutamate, and *myo*-inositol, mainly.

Also using ^1H NMR spectroscopy-based metabonomics analysis, Lan et al.²¹ identified molecular changes in postmortem brain tissue of bipolar disorder patients and in rat brain tissue after chronic treatment with lithium or valproate. Glutamate levels were increased in postmortem brains of bipolar patients, while the glutamate/glutamine ratio was decreased following valproate treatment, and γ -aminobutyric acid levels were increased after lithium treatment. Creatine and *myo*-inositol levels were increased in the postmortem brain, but decreased in the presence of the medications.

Lipids. Lipid level changes associated with bipolar disorder were previously reported in the literature. As an example, Atmaca et al.³⁶ observed decreased serum cholesterol and leptin levels in bipolar disorder patients with manic episodes and in patients with bipolar disorder I in full remission. More recently, Ozbulut et al.³⁷ evaluated cholesterol, leptin, and ghrelin levels in euthymic patients with bipolar disorder that received lithium maintenance monotherapy and found that decreased serum ghrelin and increased total cholesterol levels in the patients under lithium treatment were detected when compared with those of the controls. Ozbulut et al.³⁷ suggest that ghrelin and total cholesterol might be associated with lithium treatment and lithium-induced improvement of symptoms such as food intake and sleep–wake regulation, but not with weight gain. Schwarz et al.³⁸ used a high-

(35) Miller, J. C.; Miller, J. N. *Statistics and Chemometrics for Analytical Chemistry*, 4th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, 2005.

(36) Atmaca, M.; Kuloglu, M.; Tezcan, E.; Ustundag, B.; Bayik, Y. *Neuropsychobiology* **2002**, *46*, 176–179.

(37) Ozbulut, O.; Guler, O.; Kurt, E.; Alatas, G.; Serteser, M.; Gecici, O. *Neuro. Psychiatry Brain Res.* **2007**, *14*, 127–130.

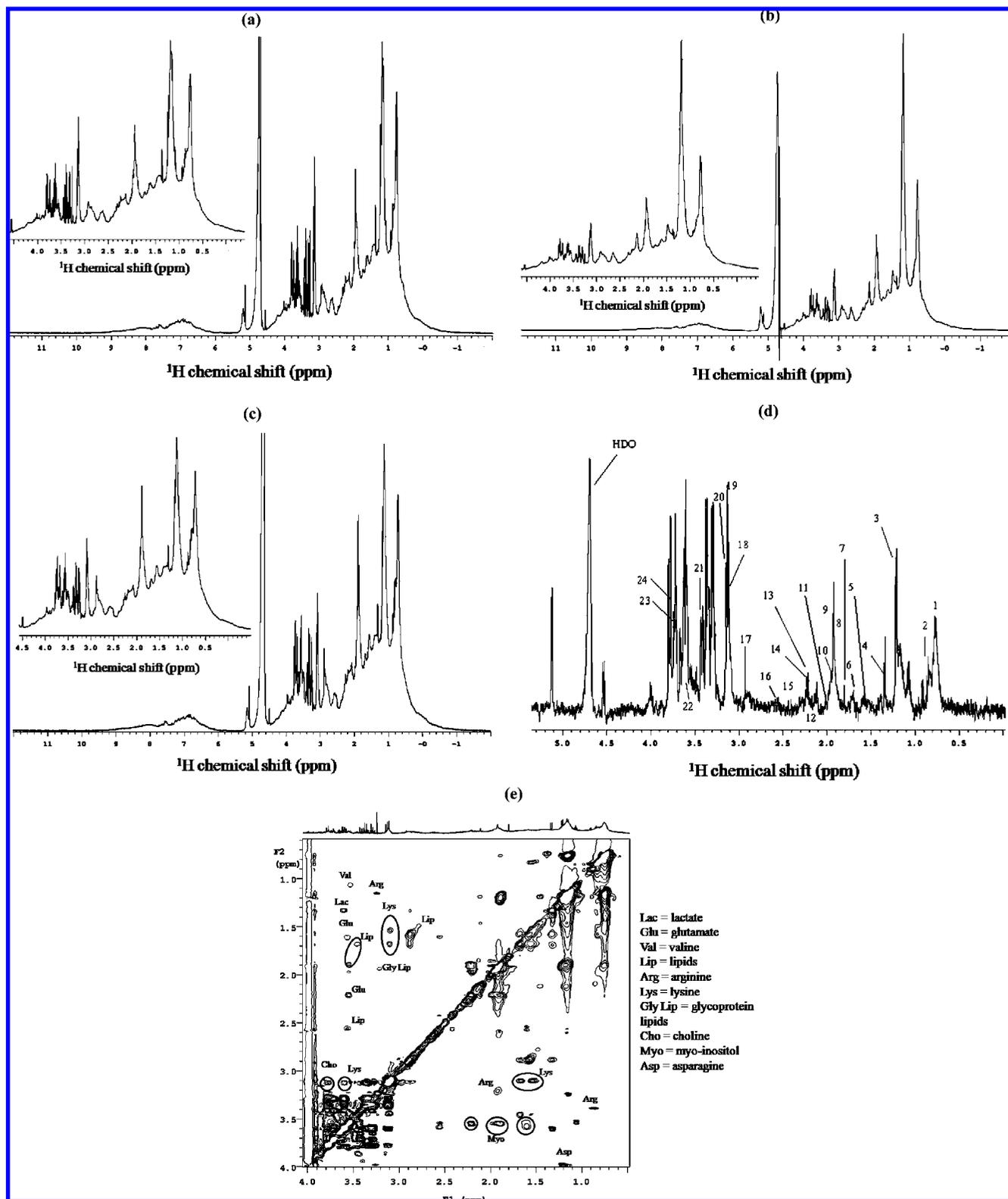


Figure 4. ^1H NMR spectra for a blood serum sample of (a) a control subject, (b) a bipolar disorder patient treated with lithium, and (c) a bipolar disorder patient not treated with lithium. The expanded -0.5 to $+4.4$ ppm region used in chemometrics is shown in the upper left corners (a–c). The peak assignments described in Table 3 are shown in (d) on the T_2 -edited spectrum of a control sample. (e) ^1H – ^1H TOCSY NMR spectrum of human blood serum recorded (control sample) using MLEV-17 as a spin-locking scheme with a mixing time of 90 ms. Some important metabolites are indicated.

throughput mass spectrometry approach (UPLC–MS) to analyze samples of gray and white matter and red blood cells and compared subjects with schizophrenia and bipolar disorder to

control subjects. Significant alterations in the levels of free fatty acids and phosphatidylcholine were detected. Such differences suggest that lipid abnormalities may be an intrinsic feature of both

schizophrenia and bipolar disorder that is reflected by significant changes in the central nervous system as well as in peripheral tissues. In addition, a recent study from some of the authors of this paper (in preparation), using independent proteomic profiling techniques on the same samples, showed a consistent and significant alteration in the levels of apolipoprotein A-I, which is a component of the high-density lipid fraction. Therefore, these current studies support the hypothesis of lipids as potential biomarkers related to bipolar disorder and/or treatment using lithium.

The lipid-metabolism-related molecules found were acetate, glutamate, choline, and *myo*-inositol. Acetate is formed in the body by the metabolism of certain substances, particularly in the liver in the oxidation of lipids. Glutamate is the most common precursor of the brain neurotransmitter (GABA) and is always excitatory usually due to simple receptors that increase the flow of positive ions by opening ion channels. As reported in the literature for postmortem brain tissue samples,²¹ glutamate levels increased in bipolar patient serum samples independent of the lithium treatment.

It is known that lipid concentrations in serum are altered in smokers compared with nonsmokers.³⁹ Nevertheless, a constant behavior of lipid levels in the bipolar disorder patient groups (containing some smokers) in relation to the control group (containing only nonsmokers) is possible to assign. As mentioned in the Materials and Methods, this variable was not controlled in the present study, but it was shown to be a nonrelevant factor, since the lipid levels changed in the same way inside each investigated group.

Choline. Choline is a natural amine found in the lipids that make up cell membranes, and it is the precursor of the neurotransmitter acetylcholine, which acts in the cholinergic neurotransmission.⁴⁰ Choline, lithium, and bipolar disorder are linked by interactions at several levels. Clinically, there is evidence that the choline precursor lecithin (phosphatidylcholine) is moderately effective in some patients with mania. Furthermore, lithium exerts a potent and specific inhibitory effect on human choline transport.⁴¹

***myo*-Inositol.** *myo*-Inositol is a sugar involved in the regulation of neuronal osmolarity, the metabolism of membrane-bound phospholipids, and the phosphoinositide secondary messenger pathway.⁴² Changes in *myo*-inositol levels may reflect increased inositol monophosphatase (IMPase) activity, which would lead to an increase in the levels of *myo*-inositol containing compounds in patients treated with lithium.⁴³ In the present work, *myo*-inositol was found with an increased level in the serum of bipolar patients treated with lithium compared to the control group, and with a lower level in the serum of bipolar patients treated with drugs not including lithium. In the recent literature,²¹ this molecule was also found with an increased level in postmortem brain tissue of bipolar disorder patients, but with a decreased level in the brain tissue of rats treated with lithium. Comparing these results, a discrepancy

between *myo*-inositol levels found in human samples (increased) and animal samples (decreased) when taking into account the lithium treatment effects is possible to observe.

Amino Acids. Differential amino acid (proline, glutamine, valine, asparagine, arginine, and lysine) levels were also found when comparing control subjects to bipolar patients under the different treatments. This suggests an alteration in the patients' amino acid metabolism. The metabolites *N*-acetylaspartate, choline, *myo*-inositol, glutamate/glutamine, and creatine separately were reported in the literature for euthymic, maniac, and depressed adult and child/adolescent bipolar patients by ¹H MRS analyses in specific cerebral regions.⁴² This further supports our results in that we found such metabolites also differing in blood serum samples.

CONCLUSIONS

Metabolic profiling using ¹H NMR spectroscopy and chemometrics analyses proved to be an innovative strategy to differentiate healthy subjects from bipolar ones and also to distinguish bipolar patients according to treatment (using lithium or not). A possible limitation to this study could be the small sample size. However, even with this limitation a clear distinction among the studied groups was observed, where those samples belonging to the same group closely showed the same results. To the best of our knowledge, this is the first time that such methodology was used with these purposes employing human blood serum samples. The results found in this work for blood serum samples (systemic level) corroborate those found in the recent literature²¹ for postmortem brain tissue samples (local level), especially in terms of glutamate and *myo*-inositol level changes. The proposed methodology allows the detection of multiple differential metabolites simultaneously, providing a general pattern for the disease and for a specific treatment, with the advantage of requiring serum samples only, which are obtained through a minimally invasive strategy. A distinctive pattern in the metabolic profile of serum from bipolar disorder patients could even come to play a diagnostic role in the future, and the differential patterns comparing the patients treated with lithium or not could possibly indicate relevant drug action pathways.

It is important to note that the mechanism of metabolic changes in human blood serum of bipolar patients, considering treatment with lithium or not, should be further studied. ¹H NMR spectroscopy analysis provides only a static picture of the metabolites measured, and it does not allow the determination of rates of metabolism or changes in the metabolic pathways that may be altered in the presence of the disease and with lithium treatment. The differential metabolites addressed in this work could also guide further studies on the pathophysiology of bipolar disorder and mechanisms of action of lithium treatment and also on biomarker discovery.

(38) Schwarz, E.; Prabakaran, S.; Whitfield, P.; Major, H.; Leweke, F. M.; Koethe, D.; McKenna, P.; Bahn, S. *J. Proteome Res.* **2008**, *7*, 4266–4277.

(39) Craig, W. Y.; Palomaki, G. E.; Haddow, J. E. *Br. Med. J.* **1989**, *298*, 784–788.

(40) Siegel, G. J.; Albers, R. W.; Brady, S.; Price, D. L. *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, 2nd ed.; Academic Press: London, 2006.

(41) Stoll, A. L.; Sachs, G. S.; Cohen, B. M.; Lafer, B.; Christensen, J. D.; Renshaw, P. F. *Biol. Psychiatry* **1996**, *40*, 382–388.

(42) Yildiz-Yesiloglu, A.; Ankerst, D. P. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2006**, *30*, 969–995.

(43) Forester, B. P.; Finn, C. T.; Berlow, Y. A.; Wardrop, M.; Renshaw, P. F.; Moore, C. M. *Bipolar Disord.* **2008**, *10*, 691–700.

ACKNOWLEDGMENT

Prof. Dr. Fred Y. Fujiwara is acknowledged for very valuable help involving the NMR spectrometer software and acquiring the T_2 -edited spectra. We also thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, São Paulo, Brazil), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brasília, Brazil), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brasília, Brazil) for financial support and fellowships.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review July 7, 2009. Accepted October 12, 2009.

AC901502J

Structural Insights on Two Hypothetical Secretion Chaperones from *Xanthomonas axonopodis* pv. *citri*

Juliana Fattori · Alessandra Prando ·
Leandro H. P. Assis · Ricardo Aparicio ·
Ljubica Tasic

Published online: 28 May 2011
© Springer Science+Business Media, LLC 2011

Abstract Several Gram-negative bacterial pathogens have developed type III secretion systems (T3SSs) to deliver virulence proteins directly into eukaryotic cells in a process essential for many diseases. The type III secretion processes require customized chaperones with high specificity for binding partners, thus providing the secretion to occur. Due to the very low sequence similarities among secretion chaperones, annotation and discrimination of a great majority of them is extremely difficult and a task with low scores even if genes are encountered that codify for small (<20 kDa) proteins with low pI and a tendency to dimerise. Concerning about this, herein, we present structural features on two hypothetical T3SSs chaperones belonging to plant pathogen *Xanthomonas axonopodis* pv. *citri* and suggest how low resolution models based on Small Angle X-ray Scattering patterns can provide new structural insights that could be very helpful in their analysis and posterior classification.

Keywords *Xanthomonas axonopodis* pv. *citri* (*Xac*) · Type III secretion chaperones · Flagellar chaperones · Small angle X-ray scattering (SAXS)

Abbreviations

<i>Xac</i>	<i>Xanthomonas axonopodis</i> pv. <i>citri</i>
T3SSs	Type III secretion systems
T4SSs	Type IV secretion systems
CBD	Chaperone binding domain
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Tris	Tris(hydroxymethyl) aminomethane
EDTA	Ethylene diamine tetra acetic acid
CD	Circular dichroism
SAXS	Small Angle X-ray Scattering
pI	Isoelectric point

1 Introduction

The Gram-negative bacterial pathogens of animals and plants, such as *Xanthomonas axonopodis* pv. *citri*, for example, have evolved a sophisticated arsenal of proteins, commonly directly injected into the host cells by one of the type III (T3SSs) and/or type IV (T4SSs) secretion systems [56]. The *Xac* genome [14] with 4,313 predicted proteins has around 30% of these with unknown function and classified as hypothetical. Among the hypothetical proteins, some were identified as potential secretion chaperones belonging to T3SS [1, 2] while T3SS was pointed out as crucial for bacterial pathogenicity and the initiation of disease [1, 2, 13, 24]. Besides cell-to-cell contact, molecular chaperones for secretion in T3SS are involved in the assembly of extracellular filaments or pili [9, 15, 44]. T3SS is comprised of more than 20 proteins [12] that form inner and outer membrane ring structures, an extracellular needle structure with pore-forming proteins at the distal tip that engage a host cell membrane, an ATPase at the base, with energetic and chaperone-effector recruitment roles, and a suite of chaperones to coordinate the assembly and

J. Fattori · A. Prando · L. H. P. Assis · L. Tasic (✉)
Chemical Biology Laboratory, Department of Organic
Chemistry Institute, University of Campinas (UNICAMP),
P.O. Box 6154, Campinas, SP 13083-970, Brazil
e-mail: ljubica@iqm.unicamp.br

R. Aparicio
Structural Biology Laboratory, Department of Physical
Chemistry, Chemistry Institute, University of Campinas
(UNICAMP), P.O. Box 6154, Campinas, SP 13083-970, Brazil

function of the apparatus during infection [11]. The narrow opening of the needle channel dictates that, once recognized by the secretion systems, substrates must be unfolded before secretion [56]. It is known that most secreted proteins possess a secretion signal located within the first 20–30 amino acid residues and an additional layer of specificity may be conferred by secretion chaperones that specifically bind to some of the secreted proteins. Although the details of the signaling pathway involved in these effectors delivery are still unknown, it is accepted that chaperone binding prevents non-productive pre-secretory associations of substrates [1, 2, 29, 57].

Almost all secretion chaperones from T3SS share common features such as low molar masses (<20 kDa), low isoelectric point, the presence of an amphipathic helix, and their corresponding genes are usually located adjacent, in the operon, to the gene of the virulence factor they bind to [22]. These chaperones have little or no sequence similarity but current literature groups them into three classes (I–III) based on their physical interactions with the binding partners [9, 11, 13, 41, 45, 52]. Class I chaperones bind to translocated effectors at a chaperone-binding domain located in the amino terminus of the effector. Chaperones belonging to this class have a structural fold of five β -strands and three α -helices (α - $\beta\beta\beta$ - α - $\beta\beta$ - α), form homodimers and bind to the CBD in a horseshoe-like structure. Class II chaperones bind to translocon proteins that make up the secretion pore in the host target membrane and class III chaperones bind to the extracellular filament proteins (or flagellin rod in the orthologous flagellar system) that polymerize into a helical structure following secretion from the bacterial cell. Secondary structure predictions suggest that class III chaperones adopt an extended alpha helical structure, which is confirmed by the crystal structure of the CsaA chaperone in enteropathogenic *E. coli* that binds the EspA filament protein [58]. Despite the lack of sequence identity, the structures of chaperones from class I, for example, reveal a dimeric organization with very similar α/β folds, and with the CBD of their binding partners showing mostly nonglobular substrate organization around the surface of the chaperone [53]. On the other hand, flagellar chaperones belonging to class III, differ from other chaperones structurally and in binding to the C-termini of their cognate substrates. However, the surface stabilization of secondary structure elements of the substrate may reflect a mode of recognition similar to class I chaperones [20].

In this work we present the structural characterization of two hypothetical T3SS secretion chaperones belonging to the *Xac*, named XAC0419 and FlgN. The FlgN was previously identified as a possible flagellar chaperone [33] and this hypothesis was supported by our findings, once it consists entirely of α -helices like others flagellar

chaperones and shows similar fold to most chaperones from the class III chaperones.

2 Materials and Methods

2.1 Protein Expression

For protein expression, cells with the recombinant plasmids were grown in Luria–Bertani medium at 37 °C, 200 rpm, during 16 h. Ten milliliter of these overnight cultures were added per liter of the same medium and grown at 37 °C, 200 rpm, until the $A_{600\text{nm}}$ reached ~ 0.8 . Then protein expressions were induced with IPTG 1 mmol L⁻¹ and grown at 37 °C, 200 rpm, during 3 h (FlgN) or 16 h (XAC0419). Cells were harvested by centrifugation at 4500 rpm, 4 °C, 15 min, and stored (–80 °C) prior to purification.

2.2 Protein Purification

Frozen cells were thawed, suspended in a lysis buffer (100 mmol L⁻¹ Tris-HCl, pH 8.0, 100 mmol L⁻¹ NaCl, and 1 mmol L⁻¹ EDTA; 15 mL L⁻¹) and disrupted by sonication. The lysates were centrifuged at 15,000 rpm 4 °C for 30 min. XAC0419 was encountered in supernatant after lysis and was purified in two steps by ion exchange and size exclusion chromatography. FlgN was encountered insoluble after lysis and the pellet was suspended in a solubilization buffer (50 mmol L⁻¹ sodium phosphate buffer, pH 8.0, 150 mmol L⁻¹ NaCl, 8 mol L⁻¹ urea), incubated for 30 min at room temperature with continuous stirring, diluted twice and dialysed at 4 °C in the equilibration buffer (50 mmol L⁻¹ sodium phosphate buffer, pH 8.0, 150 mmol L⁻¹ NaCl) [47]. After dialysis, the suspension was centrifuged (30 min, 15,000 rpm, 4 °C), and the soluble fraction containing FlgN was purified by size exclusion chromatography. The purity of the proteins was analyzed by SDS–PAGE (15%). Physical–chemical parameters were calculated from the protein sequence using the ExPASy proteomics server [25, 26]. The concentration of FlgN was determined by UV–Vis spectroscopy [19] applying the calculated molar absorption coefficient as follows: $\epsilon_{280\text{nm}} = 6,990 \text{ L mol}^{-1} \text{ cm}^{-1}$ [25] while the concentration of XAC0419 was determined using the Bradford method [8].

2.3 Analytical Gel Filtration

The molar mass of proteins and their oligomerization states were estimated by analytical gel filtration with a SuperdexTM 200 prep grade resin (GE Healthcare) [21], and a low molar mass Gel Filtration Calibration Kit (LMW, GE

Healthcare) was utilized to plot a calibration curve of partition coefficient (k_{av}) versus molar mass. After calibration, solutions containing XAC0419 and FlgN were loaded onto the column using the same conditions utilized during the calibration [32].

2.4 Spectroscopic Measurements

Circular dichroism measurements were carried out on a JASCO J-720 and/or J-810 spectropolarimeters (Tokyo, Japan) equipped with a Peltier-Type temperature controller (Control System PFD 4255). The CD spectra were taken in cuvettes of 10 mm path length using 2–4 μM protein samples in 20 mM sodium acetate or phosphate buffers (pH 5.0 or 8.0, respectively). Two types of experiment were performed and in temperature stability experiments, 4–80 °C and 80–4 °C ranges in steps of 5 °C min^{-1} , applying the scan speed of 20 nm min^{-1} from 200 to 260 nm were used. The second type of experiments was conducted using 4 $\mu\text{mol L}^{-1}$ protein samples at 20 °C. The protein α -helical contents were determined applying the computer method CDNN [7]. The emission fluorescence measurements were acquired in a spectrofluorimeter (Carey Eclipse, Varian) using the 2.6 $\mu\text{mol L}^{-1}$ protein sample in phosphate buffer (50 mmol L^{-1} , pH 8.0). The excitation wavelength was 280 nm with bandpass of 5 nm. Lifetimes were measured with the modulation frequency range from 10 to 200 MHz on multi frequency cross-correlation phase and modulator fluorometer ISS K2 by applying the excitation at 295 nm that was selected through a 310 nm filter (Edmund Industrial Optics). Ficoll 400 in aqueous solution with lifetime of 0 ns was used as a reference. The instrument program ISS K2 was used in data treatment. All CD and fluorescence spectra were baseline corrected with the buffer and were the average of at least three independent experiments.

2.5 SAXS Experiments

Protein samples were prepared in concentrations of 3.8 and 7.9 mg mL^{-1} . SAXS data were collected at the D02A-SAXS2 beamline of the National Synchrotron Light Laboratory (LNLS, Campinas-SP, Brazil), equipped with a MARCCD detector. The X-ray wavelength and sample-to-detector distance were 1.488 Å and 1079.7 mm, respectively, corresponding to the overall q range $0.0153 < q < 0.3056 \text{ \AA}^{-1}$ ($q = 4\pi\sin\Theta/\lambda$, where 2Θ is the scattering angle). The temperature of the sample holder was kept constant at 25 °C using a water-bath temperature controller. Scattering patterns for protein samples and buffers were collected alternatively with exposure times of 600 s, optimized to reduce radiation damage. For each protein sample, successive frames were recorded. Data reduction

comprised radial integration of the 2D-SAXS patterns to 1D scattering profiles using the program Fit2D [28], followed by averaging of individual curves with PRIMUS [36, 37].

The radius of gyration, R_g , and the scattering intensity at zero angle, $I(0)$, were estimated using the Guinier approximation $I(q) = I(0)\exp(-q^2 R_g^2/3)$, valid for small angles ($q < 1.3/R_g$) [27]. A convenient q range was chosen in order to diminish potential aggregation effects noted at very low angles. AUTOGNOM [36] was used to obtain the distance distribution function, $P(r)$, molecule anisometry and maximum intramolecular distance, along with more accurate estimates for R_g and $I(0)$. Kratky plots ($q^2 I(q) \times q$) calculated from the scattering data were used to assess the conformational state of the proteins in solution [18, 46].

Molar mass was estimated within an error of about 10% using a standard (bovine serum albumin and/or lysozyme) collected in the same experimental conditions for calibration [43]. An alternative method based on single curves measured on a relative scale [23] was also used. For each protein, a low resolution model was obtained from the experimental curve using the program DAMMIF to generate twenty independent dummy bead models which were subsequently averaged with DAMAVER/DAMFILF [54]. For a clearer visualization, the coordinate file resulting from this procedure was masked using NCSMASK [10, 16], resulting in the final low resolution three-dimensional envelope which represents the protein molecules.

2.6 Modeling

A BLAST search against the Protein Data Bank using default parameters resulted in E-values ~ 1 , essentially meaning that no significant alignments were obtained [3]. In fact, for both analyzed proteins, XAC0419 and FlgN, no homologues with known three-dimensional structure are available. Thus, in the absence of high resolution models determined by experimental techniques, atomic structures for FlgN and XAC0419 were predicted by computational methods using I-TASSER [48, 59, 60]. The results from I-TASSER assays aided in evaluation of the low resolution envelopes recovered from the SAXS experimental curves. Examination of the predicted models for similarity and structure alignment for classification within the SCOP [4, 42] database (1.73 archive) was done with PDBeFold [39]. Superpositions were carried out with SUPCOMB [35, 38]. All steps were executed with auxiliary semi-automated shell scripts under Linux. GNUPLOT (gnuplot, version gnuplot 4.4 patchlevel CVS) was employed for data analysis and plotting, RASMOL [49, 50] and PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LCC) were used for graphical analysis.

A PQR file containing per-atom charge and radius was generated using the AMBER force field with PDB2PQR [17] and electrostatic properties were calculated with the Adaptive Poisson-Boltzmann Solver (APBS) plug-in of PyMOL [6].

3 Results

The proteins XAC0419 (NP_640774.1) and FlgN (NP_642316.1) were successfully expressed in *E. coli* BL21(DE3)pLysS strain at 37 °C. XAC0419 was encountered soluble in the supernatant after cell lysis and was purified using cation exchange chromatography. A further purification step was executed using size-exclusion chromatography (Fig. 1). This protein was found in both monomer and dimer forms, and both were purified as indicated in Fig. 1a, b. In comparison to previously published data [33], FlgN was found mostly insoluble after lysis (Fig. 1c), and was purified from pellet only after urea refolding [47]. Refolding resulted in obtaining the soluble FlgN and purification was conducted by size-exclusion chromatography.

Gel filtration was used to estimate the molar masses of XAC0419 and FlgN, applying Gel Filtration Calibration Kit LMW (GE Healthcare) for calibration. The experimental parameters were: geometric column volume (V_C) of 58.8 mL and the void volume of 23.2 mL. After analyzing the two standard protein mixtures (Gel Filtration Calibration Kit LMW, Materials and methods), the protein elution volumes (V_e) were obtained. The partition coefficients (k_{av}) were calculated and then the calibration curve was plotted with a 0.99 adjusted R-square. The target proteins, XAC0419 and FlgN, were submitted to gel filtration using the same conditions as the standards and the estimated molar masses were 11 kDa for XAC0419 and 12 kDa for FlgN. It is important to mention that gel filtration doesn't provide exact molar masses for proteins but only an approximation, and these experiments were conducted mainly to estimate the proteins oligomerization (monomers in both cases).

The purified proteins exhibited CD spectra characteristic for folded proteins (Fig. 2) with α -helix secondary structure elements as judged by the molar ellipticities at 222 nm, especially in the case of FlgN (Fig. 2b). Deconvolution of the spectra indicated α -helix contents of around 30% for XAC0419 in monomeric form, 54% for XAC0419 dimer (Fig. 2a) and 99.0% for FlgN (Fig. 2b). Additional CD experiments for XAC0419 and FlgN were carried out in the temperature range 4–80 °C in steps of 5 °C (data not shown). CD spectra did not change significantly with temperature until 65 °C for XAC0419 and 55 °C for FlgN, indicating stable proteins up to these temperature limits.

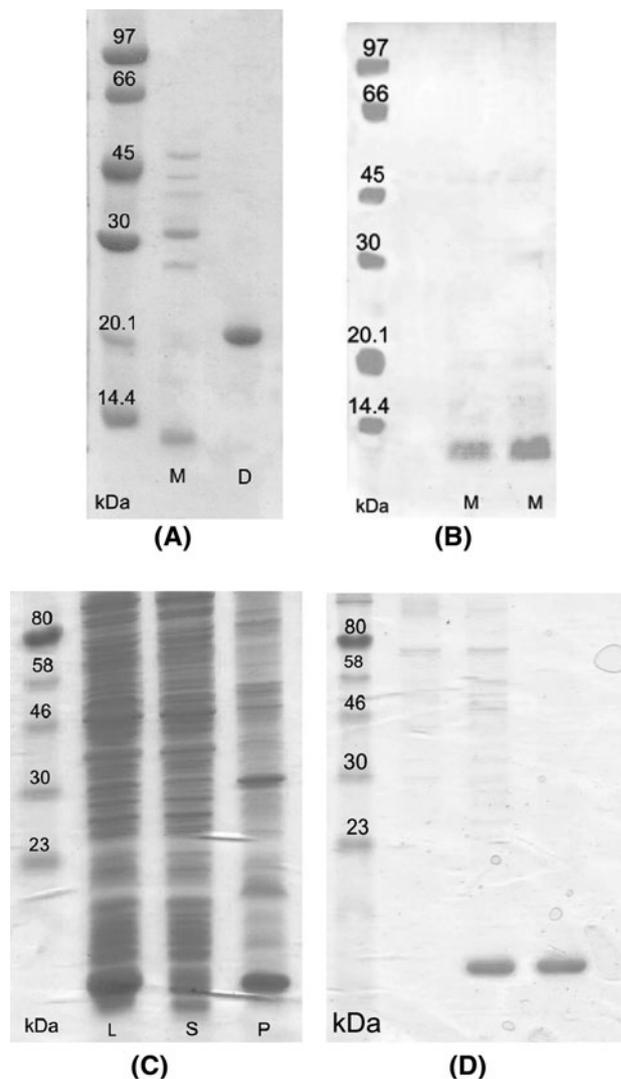


Fig. 1 SDS-PAGE electrophoresis gels. **a** XAC0419 after first purification procedure showing the pure dimer (D) and monomer (M) fractions; **b** XAC0419 after second purification procedure showing the pure monomer (M); **c** FlgN after lysis with samples showing L—total lysate, S—supernatant of lysis and P—pellet of lysis; **d** FlgN purification results showing pure protein. All of gels present on the left the molecular protein ladder with the respective molecular mass values in kDa

Both proteins exhibited reversible behavior in opposite temperature direction (80–4 °C). The emitted fluorescence spectrum of FlgN is shown in Fig. 3a and presents the normalized, background-corrected data measured at 25 °C. A characteristic fluorescence emission maximum at 352 nm suggests that the single tryptophan residue in FlgN (W90) is exposed to solvent, while dynamic fluorescence experiments indicated three life times for this Trp residue estimated at 5.7, 2.3 and 0.5 ns and with populations of 60.8, 30.2 and 9% ($\chi^2 = 1.023$), respectively, pointing for three different conformers for FlgN, one prevalent (Fig. 3b).

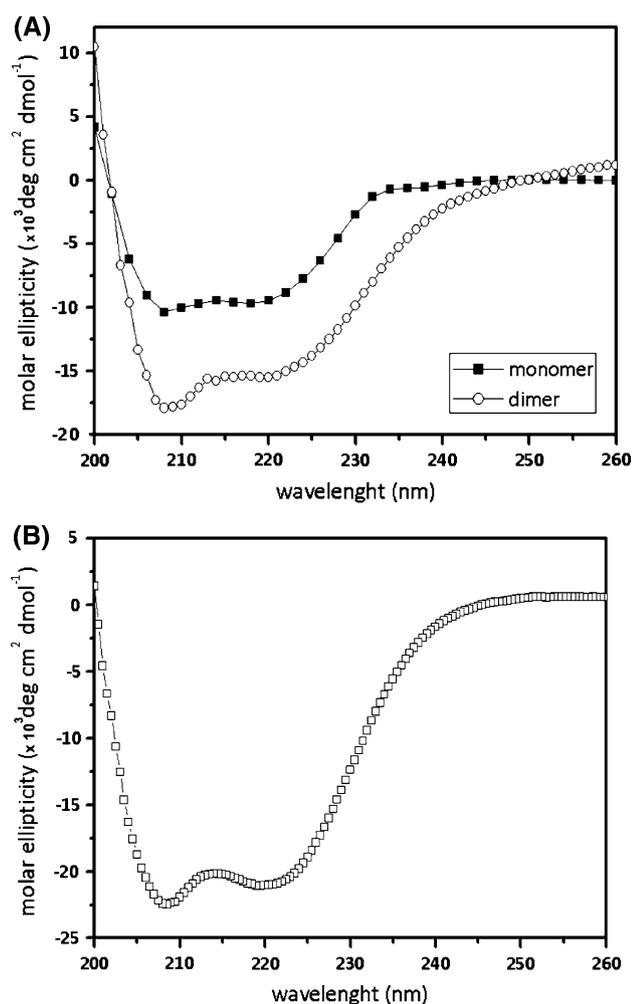


Fig. 2 Spectroscopic data. **a** CD spectra of the XAC0419 at $4 \mu\text{mol L}^{-1}$ in both forms: (*open circle*) dimeric and (*filled square*) monomeric; **b** CD spectra of the FlgN at $4 \mu\text{mol L}^{-1}$ (*open square*). All of them recorded as described in Materials and methods

SAXS data were collected for both proteins after confirming the correct folding by CD spectroscopy and sample monodispersity by size-exclusion chromatography, from which a first estimate of molar mass was obtained. Figure 4a presents the final SAXS curves after data reduction (Materials and methods) and the corresponding Guinier regions, where the expected linear behavior for monodisperse samples is observed. Radii of gyration estimated from the linear regression (Guinier plot) were 14.2 and 22.5 Å for XAC0419 and FlgN, respectively. Using a standard for calibration and the scattering intensity at zero angle resulting from Guinier analysis, molar masses of XAC0419 and FlgN were estimated to be 9.0 and 14.6 kDa, respectively, which is in reasonably good agreement with the previously obtained data. Distance distribution functions and Kratky plots obtained from the experimental curves are shown in Fig. 4b, c, respectively.

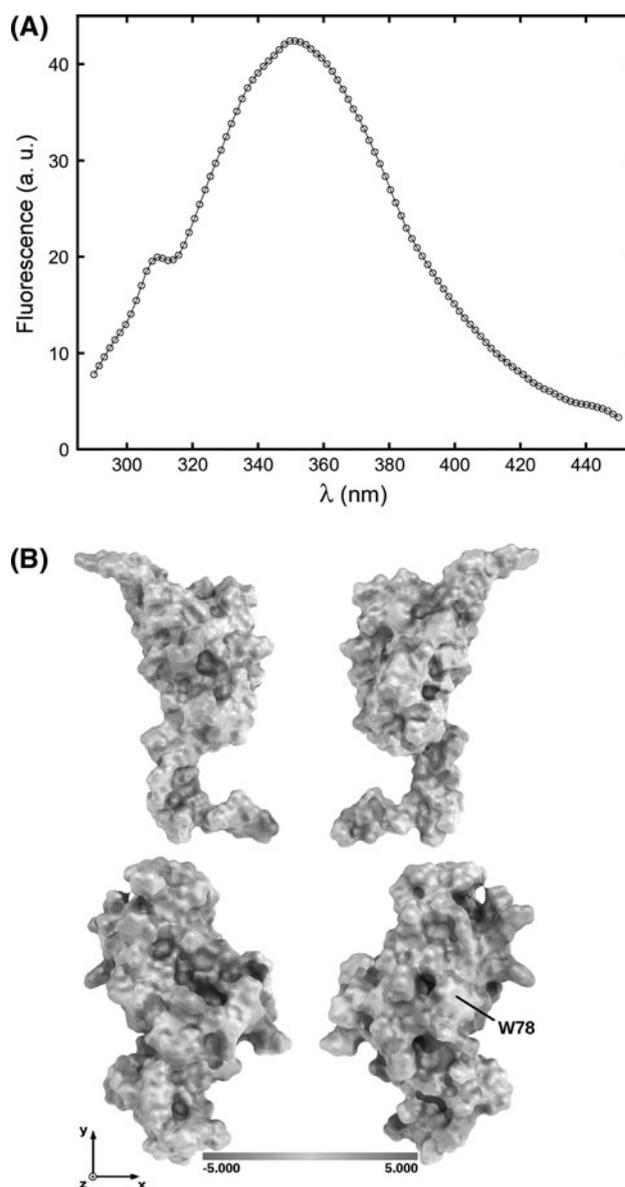


Fig. 3 Spectroscopic and bioinformatics data for target proteins. **a** Emitted fluorescence spectrum of FlgN measured at 25 °C. **b** The electrostatic surface potentials (310 K) for the predicted models of XAC0419 (*top row*) and FlgN (*bottom row*). The right column is rotated clockwise by 180° around the y-axis. Electrostatic potentials are shown as multiples of kT/e , where k is the Boltzmann's constant, T is the temperature and e is the electron charge. The surface region occupied by the single tryptophan residue (W78) in FlgN is marked in the figure (*lower right*)

Rg values derived from the distance distribution function were 13.5 Å for XAC0419 and 20.9 Å for FlgN, in agreement with those obtained by Guinier analysis. The molar masses obtained from the curves on a relative scale [23], 8.6 and 13.0 kDa for XAC0419 and FlgN, respectively, are comparable to those anticipated from the Guinier analysis of the SAXS curves using a standard for calibration, and from the size-exclusion chromatography

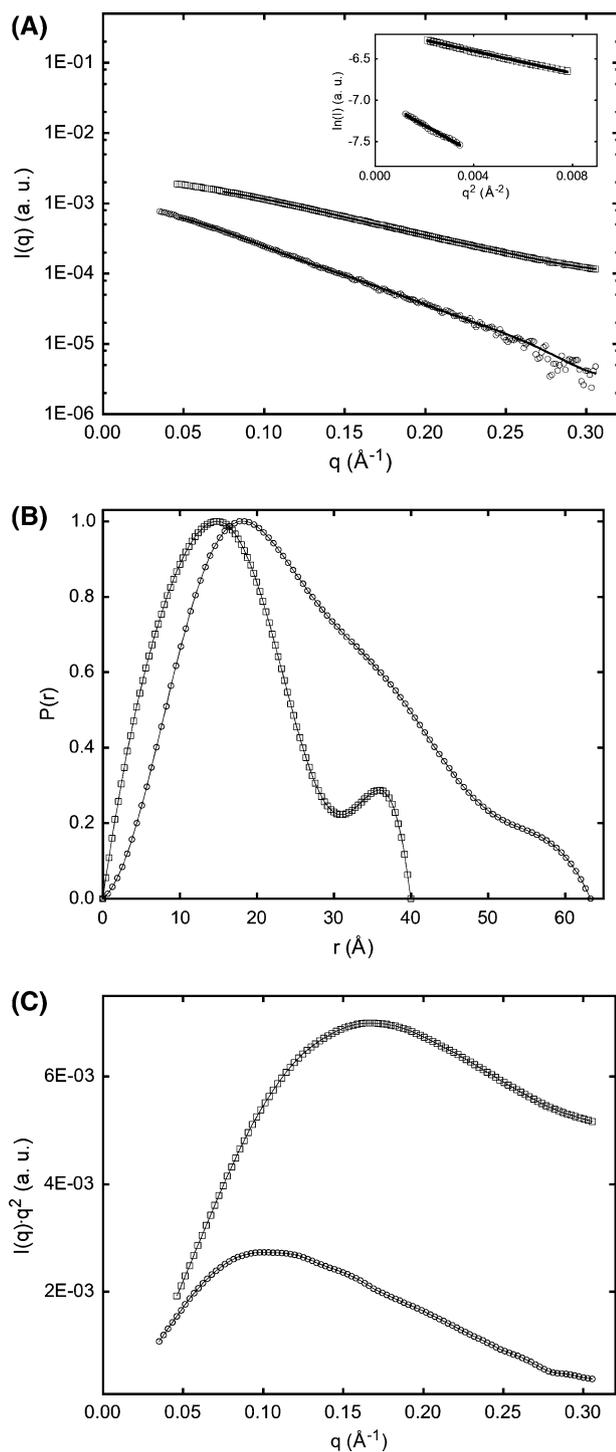


Fig. 4 SAXS results. **a** SAXS scattering curves corresponding to the average of 3 consecutive frames of XAC0419 (*open square*) and FlgN (*open circle*). In each case, the solid line represents the curve fitting corresponding to the distance distribution function obtained with AUTOGNOM [36]. The *inset* shows the linear fitting obtained in the respective Guinier regions. **b** Distance distribution functions obtained from the respective scattering curves using the program AUTOGNOM and normalized to the maximum of each individual curve. The resulting integral parameters are quoted in the text. **c** Kratky plots normalized to $I(0) = 1$

data. The three estimates agree, within experimental error, with the masses calculated from the protein sequences, 10.8 kDa for XAC0419 and 12.1 kDa for FlgN, indicating that, in both cases, monomers were present in solution.

Experimental low resolution models were obtained from the SAXS scattering curves along with theoretically predicted coordinate models both for XAC0419 (Fig. 5) and FlgN (Fig. 6), according to the procedure described in the previous section.

4 Discussion

The present work applies different experimental and theoretical techniques to tackle the challenging task of corroborating genome original annotation of the target proteins as secretion chaperones. The two target proteins, XAC0419 and FlgN, were classified as conserved hypothetical and possible secretion chaperones belonging to T3SS [1, 2, 33]. The chromosomal gene codifying for XAC0419 (NP_640774.1) is small and exhibits amphipathic domains (amino acid residues 67–91) [1, 2], common features shared with other known T3SS secretion chaperones as illustrated in Fig. 7. The flagellar chaperone FlgN also exhibits the common secretion chaperone features: low molar mass, low pI (6.7) and an amphipathic domain on the C-terminal [1, 2, 33].

Although sharing common properties, including similar 3D folds within the same class of T3SS chaperones [22], these proteins, in general, do not exhibit primary structure homology, making it difficult to assure their function based on sequence data. Furthermore, structural data are relatively scarce in comparison with the number of proteins thought to function as secretion chaperones. Circular dichroism confirmed that both analyzed proteins possess helical secondary structure elements, in agreement with a predicted amphipathic helix on the XAC0419 and FlgN C-terminals [1, 2], a common feature to known T3SS secretion chaperones. CD data analysis also revealed that FlgN is a stable protein comprised entirely of α -helices (99% utilizing CDNN software) as described for most class III T3SS chaperones [22]. Fluorescence experiments revealed that the unique FlgN W78 residue is exposed to solvent as its maximum emission was at 352 nm. Tryptophan residues exposed to solvent (water) are known to exhibit maximum fluorescence emission at longer wavelengths, while W residues involved in intermolecular interactions exhibit it at shorter wavelengths (around or below 330 nm) [56]. As already described, a common feature of secretion chaperones is the presence of an amphipathic domain composed of α -helices where polar and nonpolar amino acids are localized on opposing sides [1, 2]. Indeed, the electrostatic surface potentials calculated

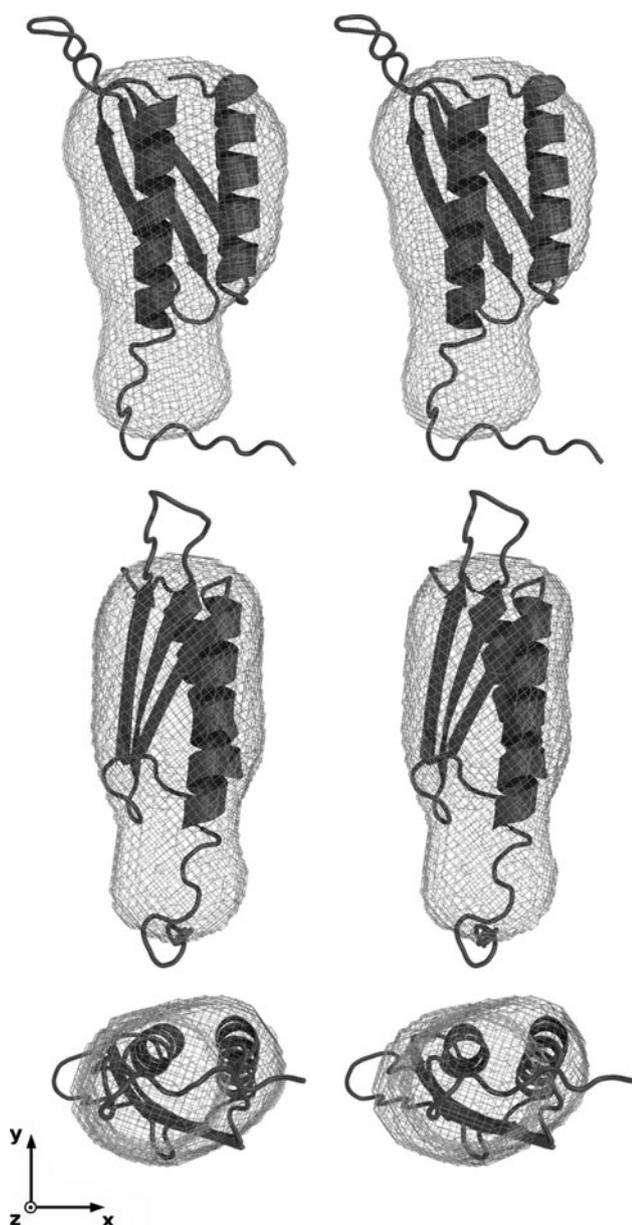


Fig. 5 Stereo views showing the superposition of the molecular envelope of XAC0419 derived from the experimental curves (*in light gray*) superposed onto a cartoon representation of the theoretical atomic model (*in dark gray*) obtained by computational modeling. The middle and bottom rows are rotated clockwise by 90° around the y-axes and counterclockwise by 90° around x-axes, respectively. Drawings were prepared with PyMOL and edited using GIMP (<http://www.gimp.org>) under Linux

for the predicted three-dimensional models were examined and showed predominantly hydrophobic areas interspersed with positively and negative charged regions, suggesting continuous patches for protein–protein interactions (Fig. 3b). Another relevant observation is that an inspection of the unique tryptophan residue of FlgN has shown that it lies in the exterior of one of the α -helices in the predicted model (Fig. 3b), which could be related to the



Fig. 6 Stereo view of the molecular envelope of FlgN recovered from the scattering curve (*in light gray*) superposed onto a cartoon representation of the atomic model predicted by computational modeling techniques (*in dark gray*). The middle and bottom rows are rotated clockwise by 90° around the y-axes and counterclockwise by 90° around x-axes, respectively. Figure was drawn using PyMOL and edited using GIMP (<http://www.gimp.org>) under Linux

observed fluorescence spectrum, typical for a tryptophan residue exposed to solvent.

Three independent estimates of the molar mass of the proteins in solution were obtained. With good agreement within experimental error, they indicate that, in both cases, the proteins preferentially assume a monomeric form in solution under the analytical conditions used, which include low protein concentrations and absence of reducing

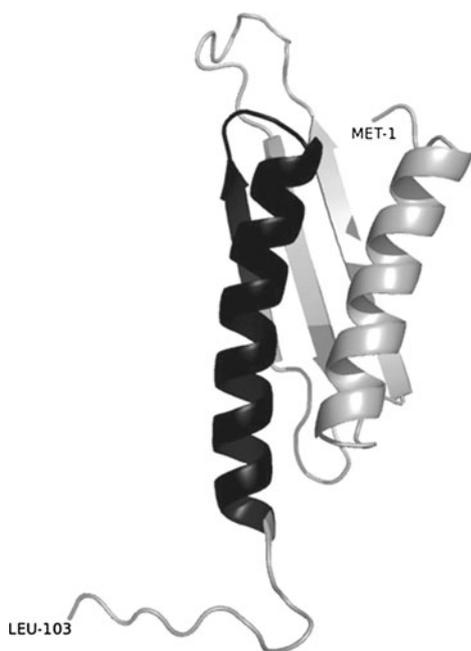


Fig. 7 Cartoon representation of the atomic model for XAC0419 as predicted by I-TASSER. This is the second scored model (C-score -1.36) drawn using PyMOL with N- and C-terminal residues marked (Met-1 and Leu-103), and the predict amphipathic helix (67–91) showed in dark gray

agents. In addition to the CD analysis, which confirmed the presence of secondary structural elements, the Kratky plots shown in Fig. 4c exhibit a well-defined maximum characteristic of compact proteins folded in a stable tertiary structure [18, 40, 46, 51].

For different protein concentrations and buffer conditions, a nonlinear dependence of the Guinier region of the scattering curves was observed for the first points, corresponding to very low q values, a probable indication of a slight sample inhomogeneity and small amounts of aggregation. Interestingly enough, this fact may bear a relationship to the amphipathic character of the C-terminal domain [1, 2, 33], which may provide patches for protein–protein interactions, as discussed below. It is worth noting that pruning a few points in the beginning of the scattering curves did not compromise further analysis. In the present case, proteins that are relatively small in size and molar mass were studied, for which the information content extends to higher angles. The convenient low resolution cut-off chosen in each case resulted in a reliable and consistent data processing.

The non-symmetrical distance distribution functions (Fig. 4b), with a bell shape bend to the left and a peak at lower distances, show that XAC0419 and FlgN share an elongated shape, the later exhibiting a greater maximum distance, corresponding to the estimated higher radius of gyration. The envelopes recovered from the experimental

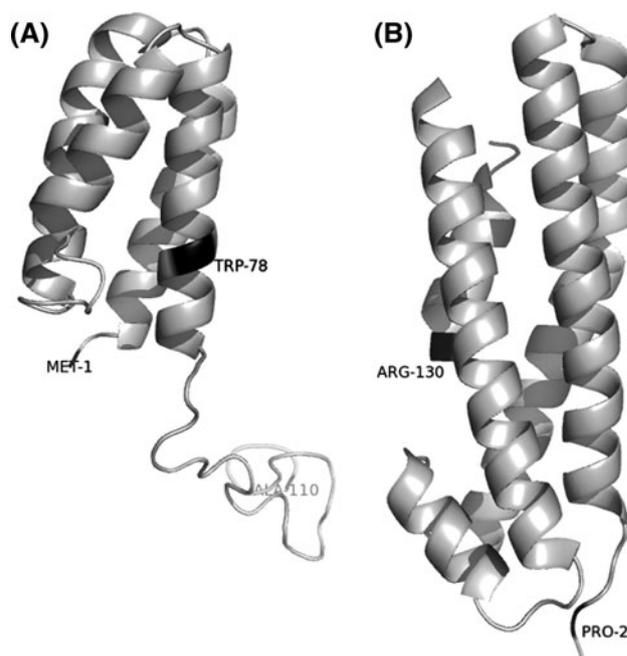


Fig. 8 Cartoon representation of the atomic model predicted for FlgN (I-TASSER) and PDB structure of the FlgN from *Pseudomonas aeruginosa*. **a** This is the third scored model with I-TASSER C-score of -2.66 and drawn using PyMOL with N- and C-terminal marked (Met-1 and Ala-110), and showing once again the W78 exposed to solvent in dark gray. **b** Crystallographic structure of FlgN from *Pseudomonas aeruginosa* (PDB entry: 2fup, deposited in 2007, data not published), which presents 49% of similarity and 27% of identity to *Xac* FlgN

curves, shown in Figs. 5, 6, are prolate as expected and give more details on the three-dimensional structure of these proteins.

Due to the intrinsic low resolution of the SAXS technique, atomic coordinate models are normally used to interpret the resulting envelopes, whether derived by high resolution experimental techniques or predicted by computational methods [35, 40, 46, 55]. In the present case, experimental high resolution models were not available, for either of the proteins under study or for homologues. Although efforts were made, attempts to obtain samples at higher concentrations were unsuccessful, which limited the use of high resolution techniques such as X-ray Diffraction Crystallography. For these reasons, the protein structure prediction was employed in order to obtain coordinate models through the I-TASSER server [48, 59, 60].

The predicted top models had estimated accuracies (TM-score) of 0.75 ± 0.10 and 0.52 ± 0.15 for XAC0419 and FlgN respectively, suggesting a high quality prediction in both cases [48, 59]. For each protein, four other models were predicted, which share the same underlying structural characteristics, differing mainly in loop or unordered regions. The closest matches resulting from XAC0419 structure alignments belong to the SCOP family d.204.1.1

of ribosome binding protein Y [4, 42], whose fold consists mainly of antiparallel β -sheets, with segregated α and β regions. The predicted FlgN models contain a compact domain composed of α -helices and a structural similarity search of SCOP has shown that they match the SCOP family k.9.1.1 of de novo designed single chain three-helix bundle. DALI [30] search with the XAC0419 and FlgN gave similar results about fold families (SCOP). It is important to mention that in this search FlgN (Fig. 8a) was also correlated to FliT chaperones and also with the FlgN from *Pseudomonas aeruginosa* (Fig. 8b, PDB entry: 2fup), once more indicative of similar 3D structures for this class of proteins.

It is very interesting that the best fitting to the experimental envelopes were not obtained for the I-TASSER models with the highest scores, but with the second and third best models for XAC0419 and FlgN, respectively. Remarkably, except for probable mobile or intrinsically unordered regions in the proteins, an impressive similarity between the envelopes and the predicted models was observed in both cases, as illustrated in Figs. 5, 6. An excellent agreement is observed for the core region comprising the α -helices and β -sheets predicted for XAC0419, although loop regions may assume a more globular, compact shape in solution than what was anticipated by the predicted atomic model. In the case of FlgN, the compact domain composed of α -helices superposes very well onto a core portion of the experimental envelope and, at the same time, an unordered, although globular region seems to be present, based on the predicted atomic model. It is important to mention that the experimental low resolution envelopes and the coordinate models were obtained by completely independent approaches. Furthermore, as in the present case, there are various examples where SAXS data have been used as a valuable tool to compare and select compatible models of proteins [5, 35, 40, 46].

Although there are some structural data on secretion chaperones, almost all of them present class I T3SS chaperones as dimers [22, 57]. Surprisingly, our results indicate that XAC0419 and FlgN are monomeric in solution, which is an opening for further structural investigations on this kind of proteins. Structural data of class III secretion chaperones or flagellar proteins are very limited and some of the published data are not in agreement with each other [22]. Our results indicate that FlgN is a helical protein, whose three-dimensional folding assumes a prolate shape in agreement to the data expected for the most flagellar chaperones, with crystallographic data on FliS from *Aquifex aeolicus* [20], and with size-exclusion chromatography data on FliT from *Salmonella* [31, 34]. On the other hand, the hypothetical XAC0419 protein also elongated in shape and with 2α - 3β fold (α - $\beta\beta\beta$ - α), looks very similar in structure to known class I T3SS chaperones [11, 22].

While it is well established that classes I and II of T3 secretion chaperones share significant structural features, the same has not been verified for class III T3SS chaperones and should be investigate.

Furthermore, the molecular envelopes obtained for XAC0419 (Figs. 5 and 7) and FlgN (Figs. 6 and 8) are similar to the other T3 secretion chaperones from the same class. Thus, we can state that even though secretion chaperones have no primary structure homology, they may share the cited common features and similar tertiary structure folds within the T3SS chaperones class. As far as we know, these are the first low-resolution structural data obtained for T3SS secretion chaperones in *Xac*.

Acknowledgments The authors thank the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP) and *Conselho Nacional de Pesquisa e Desenvolvimento* (CNPq) for financial support, LNLs for beamline time and the SAXS2 beamline staff for helpful discussions during data collection. The authors are also very grateful to Professor Carol Collins (IQ-UNICAMP) for helpful suggestions about English grammar and style.

References

- Alegria MC, Souza DP, Andrade MO, Docena C, Khater L, Ramos CHI, da Silva AC, Farah CS (2005) *J Bact* 187: 2315–2325
- Khater L, Santos T, Alegria MC, Docena C, da Silva AC, Ramos CHI (2005) *Genet Mol Biol* 28:321–327
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402
- Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG (2008) *Nucleic Acids Res* 36:D419–D425
- Aparicio R, Fischer H, Scott DJ, Verschuere KH, Kulminkskaya AA, Eneiskaya EV, Neustroev KN, Golubev AM, Polikarpov I (2002) *Biochemistry* 41:9370–9375
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) *Proc Natl Acad Sci USA* 98:10037–10041
- Böhm G, Muhr R, Jaenicke R (1992) *Protein Eng* 5:191–195
- Bradford MM (1976) *Anal Biochem* 72:248–254
- Christie PJ (1997) *J Bacteriol* 179:3085–3094
- Collaborative Computational Project Number 4 (1994) *Acta Crystallogr Sect D: Biol Crystallogr* 50:760–763
- Cooper CA, Zhang K, Andres SN, Fang Y, Kaniuk NA, Hannemann M, Brummell JH, Foster LJ, Junop MS, Coombes BK (2010) *PLoS Pathog* 6:e1000751
- Cornelis GR (2006) *Nat Rev Microbiol* 4:811–825
- Cornelis GR, Van Gijsegem F (2000) *Annu Rev Microbiol* 54:735–774
- da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LM, do Amaral AM, Bertolini MC, Camargo LE, Camarotte G, Cannavan F, Cardozo J, Chambergo F, Ciapina LP, Cicarelli RM, Coutinho LL, Cursino-Santos JR, El-Dorry H, Faria JB, Ferreira AJ, Ferreira RC, Ferro MI, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP, Lemos EG, Lemos MV, Locali EC, Machado MA, Madeira AM, Martinez-Rossi NM, Martins EC, Meidanis JC, Menck F, Miyaki CV, Moon DH, Moreira LM, Novo MT, Okura VK, Oliveira MC, Oliveira VR, Pereira HA, Rossi A, Sena JA,

- Silva C, de Souza RF, Spinola LA, Takita MA, Tamura RE, Teixeira EC, Tezza RI, Trindade dos Santos M, Truffi D, Tsai SM, White FF, Setubal JC, Kitajima JP (2002) *Nature* 417:459–463
15. Deng W, Chen L, Peng WT, Liang X, Sekiguchi S, Gordon MP, Comai L, Nester EW (1999) *Mol Microbiol* 31:1795–1807
16. Dodson EJ, Winn M, Ralph A (1997) *Methods Enzymol* 277:620–633
17. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) *Nucleic Acids Res* 35:W522–W525
18. Doniach S (2001) *Chem Rev* 101:1763–1778
19. Edelhoch H (1967) *Biochemistry* 6:1948–1954
20. Evdokimov AG, Phan J, Tropea JE, Routzahn KM, Peters HK III, Pokross M, Waugh DS (2003) *Nat Struct Biol* 10:789–793
21. Fasman GD (1989) *Practical handbook of biochemistry and molecular biology*. CRC Press, Cleveland, p 601
22. Fattori J, Prando A, Martins AM, Rodrigues FHS, Tasic L (2011) *Protein Pept Lett* 18:158–166
23. Fischer H, de Oliveira Neto M, Napolitano HB, Polikarpov I, Craievich AF (2010) *J Appl Cryst* 43:101–109
24. Galán JE, Collmer A (1999) *Science* 284:1322–1328
25. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) *Nucleic Acids Res* 31:3784–3788
26. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) *Protein identification and analysis tools on the expasy server*. In: Walker JM (ed) *The proteomics protocols handbook*. Human Press Inc, Totowa, pp 571–607
27. Glatter O, Kratky O (1982) *Small angle x-ray scattering*. Academic Press Inc, London, p 239
28. Hammersley AP, Svensson SO, Hanfland M, Fitch AN, Hausermann D (1996) *Two-dimensional detector software: from real detector to idealized image or two-theta scan*. *High Press Res* 14:235–248
29. Hueck CJ (1998) *Microbiol Mol Biol Rev* 62:379–433
30. Holm L, Rosenström P (2010) *Nucleic Acids Res* 38:W545–W549
31. Imada K, Minamino T, Kinoshita M, Furukawa Y, Namba K (2010) *Proc Natl Acad Sci USA* 107:8812–8817
32. Job V, Mattei PJ, Lemaire D, Attree I, Dessen A (2010) *J Biol Chem* 285:23224–23232
33. Khater L, Alegria MC, Borin PFL, Santos TM, Docena C, Tasic L, Farah CS, Ramos CHI (2007) *Arch Microbiol* 188:243–250
34. Kinoshita M, Yamane M, Matsunami H, Minamino T, Namba K, Imada K (2009) *Acta Crystallogr Sect F: Struct Biol Cryst Commun* F65:825–828
35. Koch MH, Vachette P, Svergun DI (2003) *Q Rev Biophys* 36:147–227
36. Konarev PV, Volkov VV, Petoukhov MV, Svergun DI (2006) *J Appl Crystallogr* 39:277–286
37. Konarev PV, Volkov VV, Sokolova AV, Koch MHK, Svergun DI (2003) *J Appl Cryst* 36:1277–1282
38. Kozin MB, Svergun DI (2001) *J Appl Crystallogr* 34:33–41
39. Krissinel E, Henrick K (2004) *Acta Crystallogr Sect D: Biol Crystallogr* 60:2256–2268
40. Lipfert J, Doniach S (2007) *Annu Rev Biophys Biomol Struct* 36:307–327
41. Lunelli M, Lokareddy RK, Zychlinsky A, Kolbe M (2009) *Proc Natl Acad Sci USA* 106:9661–9666
42. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) *J Mol Biol* 247:536–540
43. Mylonas E, Svergun DI (2007) *J Appl Cryst* 40:s245–s249
44. Oh C-S, Carpenter SCD, Hayes ML, Beer SV (2010) *Microbiology* 156:1211–1220
45. Parsot C, Hamiaux C, Page AL (2003) *Curr Opin Microbiol* 6:7–14
46. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) *Q Rev Biophys* 40:191–285
47. Ribeiro E Jr, Regis WCB, Tasic L, Ramos CHI (2003) *Protein Expression Purif* 28:202–208
48. Roy A, Kucukural A, Zhang Y (2010) *Nat Protoc* 5:725–738
49. Sayle RA, Milner-White EJ (1995) *Trends Biochem Sci* 20:374–376
50. Bernstein HJ (2000) *Trends Biochem Sci* 25:453–455
51. Semisotnov GV, Kihara H, Kotova NV, Kimura K, Amemiva Y, Wakabayashi K, Serdyuk IN, Timchenko AA, Chiba K, Nikaido K, Ikura T, Kuwajima K (1996) *J Mol Biol* 262:559–574
52. Stebbins CE, Galán JE (2003) *Nat Rev Mol Cell Biol* 4:738–743
53. Stebbins CE, Galán JE (2001) *Nature* 414:77–81
54. Svergun DI (1999) *Biophys J* 76:2879–2886
55. Svergun DI, Koch MHJ (2003) *Rep Prog Phys* 66:1735–1782
56. Tasic L, Borin PFL, Khater L, Ramos CHI (2007) *Protein Expression Purif* 53:363–369
57. Wattiau P, Cornelis GR (1993) *Mol Microbiol* 8:123–131
58. Yip CK, Finlay BB, Strynadka NC (2005) *Nat Struct Mol Biol* 12:75–81
59. Zhang Y (2008) *BMC Bioinform* 9:40–47
60. Zhang Y (2009) *Proteins Struct Funct Bioinf* 77:100–113

Bacterial Secretion Chaperones

Juliana Fattori, Alessandra Prando, Adriana Martini Martins, Fábio Henrique dos Santos Rodrigues and Ljubica Tasic*

Chemical Biology Laboratory, Department of Organic Chemistry, Chemistry Institute, State University of Campinas, P.O. Box 6154, Campinas, S.P., 13083-970, Brazil

Abstract: Many Gram-negative bacteria are able to invade hosts by translocation of effectors directly into target cells in processes usually mediated by two very complex secretion systems (SSs), named type III (T3) and type IV (T4) SSs. These syringe-needle injection devices work with intervention of specialized secretion chaperones that, unlike traditional molecular chaperones, do not assist in protein folding and are not energized by ATP. Controversy still surrounds secretion chaperones primary role, but we can say that these chaperones act as: (i) bodyguards to prevent premature aggregation, or as (ii) pilots to direct substrate secretion through the correct secretion system. This family of chaperones does not share primary structure similarity but amazingly equal 3D folds. This mini review has the intent to present updated structural and functional data for several important secretion chaperones, either alone or in complex with their cognate substrates, as well to report on the common features and roles of T3, T4 and flagellar chaperones.

Keywords: Secretion chaperones, protein and DNA transport, protein-protein interactions, spectrometry and spectroscopy.

INTRODUCTION

Gram-negative bacteria utilize at least six or seven distinct secretion pathways to transport proteins, single-strand DNA (ssDNA), and/or protein-DNA complexes across the inner and/or outer membranes of the cell envelope [1-3]. Among these, the most complex pathways are type III (T3) and type IV (T4) secretion systems (SSs) that share some common features [4, 5]. One of them is a mechanism for sensing environmental stimuli in contact with the bacterium of a eukaryotic target cell [6]. Upon such stimuli, these two sophisticated molecular syringes allow the bacteria to pump specific molecules, called effectors, directly into the host cell cytoplasm. However, it is still questionable how these syringes penetrate the host plasma membrane although they can interact with some associated protein complexes in the eukaryotic cell membrane and/or cell wall and thus provide a continuous conduit for the delivery of effectors [7-9]. The T3SS also enables bacteria to assemble cell surface flagella [10], while T4SS [11] promotes horizontal gene transfer between different species by transportation of DNA and its complexes. To date, it is believed that effectors (proteins, ssDNA, complexes) have to possess at least partially unfolded structures and structures complementary in size to the inner channel, generally, around 2.8 nm in diameter [12, 13], to be transported from one to another cell cytoplasm. In the process of unfolding, the effector protein N-terminal amphipathic domain can serve as a secretion signal and also as a binding domain to a specific secretion chaperone [14-16]. This family of chaperones is very versatile with low primary

structure identity but amazing 3D structural similarities [17-20].

Although the bacterial secretion processes involve various proteins, structural, functional, effectors, among many others, the key proteins belong to class of molecular chaperones [7]. Molecular chaperones bind and stabilize proteins at intermediate stages of folding, assembly, translocation across membranes and degradation. This large family of proteins is responsible for quality control of the cell proteome, and is divided in few subfamilies according to activity and common features; like the heat-shock proteins with ATPase activity (Hsps), or the small heat-shock proteins without ATPase activity (sHsps). Heat shock proteins have been classified by molecular weight, for example, Hsp70 for the 70 kDa heat shock protein and they are among the most well-conserved proteins known. The correct folding of newly formed proteins and maintenance of protein structure under stress are provided by chaperones from Hsps families and the most studied system is the Hsp70 in *E. coli* [21a], that consists from chaperone DnaK, co-chaperone DnaJ, and the nucleotide exchanging factor known as GrpE protein. Recently, the DnaK/DnaJ involvement in bacterial invasion of mice by *Salmonella* has been reported, and demonstrated that bacterial survival and proliferation at higher temperatures are related to DnaK/DnaJ proper function [21a-c]. Also, the chaperone DnaK in *E. coli* enables the secreted effectors stability in cytoplasm prior to secretion. Therefore, secretion chaperones cooperate with other molecular cell chaperones in order to provide efficient bacterial effectors delivery in host cells.

It is assumed that one of the first steps in competent bacterial secretion is a very conserved and characteristic target (effector) protein-chaperone interaction [22], having the chaperone-binding region of the target protein wound around the chaperone [23], and ends with the breakage of this interaction as chaperones must release their effectors because

*Address correspondence to this author at the Chemical Biology Laboratory, Department of Organic Chemistry, Chemistry Institute, State University of Campinas, P.O. Box 6154, Campinas, S.P., 13083-970, Brazil; Tel: (+55-19)3521-1106; Fax: (+55-19)3521-3023; E-mail: ljubica@iqm.unicamp.br

they remain within the bacterial cytoplasm [24, 25]. These very complex processes of chaperone-assisted secretion in assembled T3SS, T4SS and flagellum are illustrated in a model as represented in Fig. (1). All characterized T3 and T4SSs contain a cytoplasmic/inner membrane ATPase with highest ATP activity in dodecamer (T3SS) or hexadimeric form (T4SS) [26]. These ATPases usually are prone to multiple protein-protein interactions with some cytoplasmic and inner membrane components of the SSs, including interactions with chaperones and a global T3 secretion chaperone [27]. Therefore, ATPase activities could be associated with the release and unfolding of complexes between chaperone-effector. Another potential energy source for T3 secretion is the proton motive force, as reported for the assembly of bacterial flagella [25].

In this review we discuss principally the structure of some important bacterial flagellar, T3 and T4 secretion chaperones and their role in bacterial locomotion, secretion activity, and pathogenicity.

SECRETION CHAPERONES: SUBSTRATE-SPECIFIC PROTEINS

Although the bacterial T3 or T4 export membrane components are obvious homologues [28], secretion chaperones have low sequence identity (18% between CesT and SigE, both identified as secretion chaperones for Tir in *E. coli* and

SigD in *S. enterica*, respectively) [29], and thus are unlikely to be homologues [30]. Secretion chaperones are usually encoded by virulence operons that contain an essential but currently anonymous gene that lies between the genes encoding the export ATPase [31] and a protein known or suspected to control hook or needle length [32, 33]. These virulence genes encode, in each case, a protein of a similar (14–18 kDa, sometimes 12–20 kDa) size, without significant sequence similarity, high helicity and low pI. The genes encoding T3 chaperones are also often localized adjacent to the genes encoding their effectors; thereby, they are coexpressed *in vivo*. Mutation of the chaperone genes usually leads to rapid degradation, aggregation and significant loss of secretion of one or, in some cases, two cognate effectors. Several of these T3 chaperones have been suggested to participate in negative feedback regulation of virulence genes in *Yersinia* species and some of them have been directly associated with transcription regulation [33–35].

Analysis of the hydrophobic surface areas of the CesT and SigE T3 chaperones provided insights into how they selectively bind to effectors. Their hydrophobic surface area is significantly larger compared to a typical soluble protein and to the probable interaction area with the exposed hydrophobic amino acids in the effectors [28, 29]. The degree of unfolding of the effectors by the T3 chaperones is also unclear. T3 chaperones, including CesT, generally bind to a

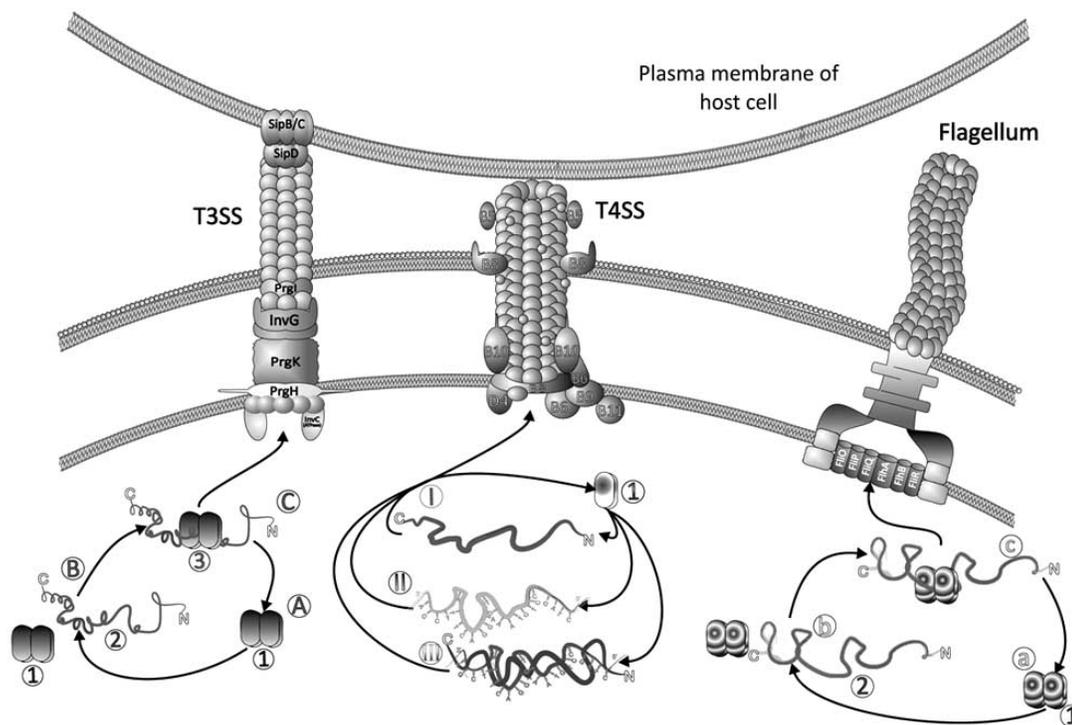


Figure 1. Illustration of bacterial secretion systems: T3SS (left), T4SS (middle), and flagellar (right). In T3 system secretion, three steps are involved: (A) the chaperone dimer (1) is free in the cytosol, (B) the dimer encounters its effector protein (2), and (C) the chaperone/effector complex (3) is guiding the delivery of effectors through the infection machinery. The T4SS is capable of translocating proteins (I), single strand DNA (II) or even single strand DNA/protein complexes (III). In this process, the chaperone (1), monomer or dimer, binds to its effectors, enabling their cytosolic stability, and finally, begins the translocation process. In the flagellar systems, the chaperone act as dimers (1), in such way that they bind to their effector protein, preventing aggregation and precipitation and, after this, begins further steps in the infectious mechanism (*i.e.* the building of the translocation machinery, or protein translocation itself). (Corel Draw 12, version 12.0.0.458, Corel Corporation; Adapted from [30, 71]).

limited region in the effector, often a <100-residue segment close to the N-terminus Fig. (1). There is as yet no evidence as to the degree of unfolding of this region or, more important, any resultant destabilization in the folding of the rest of the effector molecule as the T3 chaperone binds [36].

So far, the best described T4SS chaperone, VirE1, is even smaller than the cited T3 chaperones. It also has low pI and an amphipathic α -helix, the probable site for interaction with VirE2. The VirE1 T4 secretion chaperone forms a complex with the effector (VirE2). Gel filtration studies indicate that it has an apparent molar mass of ~70 to 80 kDa and a 2:1 molar ratio; these results are consistent with the dimerization proposition for VirE1 [37]. Formation of such a VirE1-VirE2 complex clearly seems to be important for preventing VirE2 aggregation. It is believed that the VirE1 dimer interacts dynamically with VirE2, and this complex associates with T4SS and then dissociates upon successful presentation of the substrate to the transfer channel. The minimum condition for this process to occur is that VirE1 disengages from VirE2 prior to interaction with ssDNA and the VirE2/ssDNA export [19, 20, 37].

In the flagellar system, the coordination of gene expression with secretion activity ensures that structural components of the flagellum are expressed only when required during the different stages of flagellar assembly. In this process, the flagellar proteins that polymerize to form the structure, including the hook-filament junction, filament and filament cap are exported by T3SS to the growing flagellum. Also, the components of the flagellum are secreted by a type III-like mechanism through the hook-basal body Fig. (1). Export of these structural components is strongly facilitated by cytosolic substrate-specific chaperones, which bind respectively to the C-terminal amphipathic domains of their substrates (chaperones: FlgN, FliT, and FliS) [38-40].

Structural and functional studies on secretion chaperones by crystallography, nuclear magnetic resonance (NMR) and mass spectrometry (MS) have shed light on the structural similarity of this family of proteins. These studies also provided valuable information about the molecular basis of the specific interactions between chaperones and effectors and their role on effector conformations. Further updated and important features are discussed in sequence.

X-RAY CRYSTALLOGRAPHY IN STRUCTURAL AND FUNCTIONAL STUDIES OF SECRETION CHAPERONES

Structural studies have revealed a remarkable degree of structural similarity among the different secretion chaperones despite a lack of significant sequence similarity and the wide variety of effectors that they chaperone and some of these are presented herein. The first to be discussed are T3SS chaperones that have been classified according to effector binding into four categories: (i) class IA chaperone: a single effector; (ii) class IB chaperone: multiple effectors; (iii) class II chaperone: the translocators; and (iv) class III: flagellar chaperones [7].

The crystallographic structures of the chaperones SicP and SigE from *Salmonella enterica* and SycE from *Yersinia pseudotuberculosis*, for example, confirmed that class IA

(proteins having ~130 residues) chaperones function as dimers [7, 41, 42]. These dimers exhibit similar 3D folds, in which each monomer contains five β strands and three α helices as illustrated in Fig. (2A).

The hydrophobic face of the C-terminal helix predicted from sequence analysis is buried into the core of the chaperone. In each case, the dimer surface exhibits two pairs of hydrophobic patches due to the internal two-fold symmetry. A binding region located within the substrate N-terminal has also been identified for most chaperones from this class [43]. The chaperone and effector complex structures, such as the SicP-SptP₃₅₋₁₃₉ and SycE-YopE₁₇₋₈₅ complexes from *Salmonella* and *Yersinia*, respectively, are remarkably similar, even though neither the chaperones nor their binding regions on the effectors share sequence similarities [42, 44]. In both cases, the chaperone-binding region is wrapped around the chaperone dimer and the interactions with the effector occur mainly between secondary structured elements of the effector and the two hydrophobic patches present on the chaperone dimer. Each of these patches confers little specificity to the interaction, since two identical hydrophobic patches of the chaperone dimer accommodate two different parts of the effector [7]. Furthermore, limited proteolysis of the chaperone SicP, associated with the effector SptP, and SycE with the effector YopE indicated that residues 35-139 of SptP and 17-85 of YopE are protected from degradation.

At the same time that some data on crystal structures of chaperones alone show that they form highly stable dimers in solution, the published cocrystal structures of chaperone-effector complexes provide information about the mutual interaction between these species. The cocrystal structures reveal, for example, that the N-terminal chaperone binding domains (CBD) of the effector interact with their chaperone in a highly unusual manner, forming extended, nonglobular polypeptides that wrap around both chaperones in the dimer through interactions with large hydrophobic patches [45].

Although most class I chaperones bind a single substrate (class IA) there are others such as InvB of *Salmonella* and Spa15 of *Shigella* Fig. (2, PDB code: 1RY9) that bind more than one substrate, and these chaperones belong to class IB. InvB, for example binds to at least four invasion *Salmonella* proteins: SipF, SopE, SopE2 and SipA. The crystal structure of the SipA(22-264)-InvB complex [46] presents similarities and differences from earlier chaperone-effector complexes in T3SS pathogenesis. The mode of interaction between them preserves the nonglobular character of the effector chaperone binding domain (CBD), but this interaction occurs with only one of the InvB molecules from the chaperone homodimer. This is in contrast to the interactions between *Salmonella* SptP with SicP, or *Yersinia* YopE with SycC, which involve a nonglobular polypeptide of the virulence factor that wraps around both molecules in the chaperone dimer. In the center of these interactions is the use of the same hydrophobic surface patch on each chaperone to bind different regions of the nonglobular effector. However, this nonglobular region of SipA interacts with the conserved hydrophobic patches of only one InvB molecule, leaving the other patches on the second InvB molecule exposed to solvent, contributing slightly more than one half to the total surface area buried upon complex formation. This observation suggests the pos-



Figure 2. Tertiary structure representation of (A) two T3SS (Protein Data Bank entries: 1RY9 and 1N5B, PyMol) chaperones, from *Shigella flexneri* (Spa15: 1RY9) and *Yersinia enterocolitica* (SycE: 1N5B), respectively. Secondary structure similarities can be clearly observed, since both show similar α -helical and β -sheet content. (B) Flagellar chaperone (FliS, black) with almost entire 2D α -helical structure in complex with its effector flagellin (white) (Protein Data Bank entry: 1ORY).

sibility that the InvB homodimer could bind two effector molecules at the same time. It is important to remember that these authors observed that the SipA₄₈₋₂₆₄ domain is not unfolded by InvB, but interacts through its N-terminal, hydrophobic portion and through proximal helices via globular protein-protein contacts with both molecules of the chaperone. Then, by structural analysis they [46] identified a common receptor-ligand type of interaction, involving a common binding motif in a short peptide of the effector and a conserved hydrophobic pocket in the chaperone.

The crystallographic studies can also be used to identify secretion chaperones, as in a recent study with *Salmonella enterica* serovar *Typhimurium* whose genome encodes T3SS proteins but the complement chaperone required for the virulence factor translocation was unknown [22]. Crystallographic analysis, after a reverse genetic approach, of SrcA at 2.5Å resolution allowed the identification of SrcA as a class I secretion chaperone. SrcA crystallized as a dimer and each monomer consisted of a small and a large domain. The larger domain mediates dimerization and is comprised of a twisted anti-parallel β -sheet flanked by α -helices (Fig. 2). The dimer interface occurs primarily through hydrophobic interactions and the total surface area buried at the dimer interface suggests that the chaperone would exist as a dimer in solution, which was confirmed by gel-filtration analysis. A comparison of the SrcA dimer interface with other class I chaperone family members indicated the overall similarity of quaternary structure shared between SrcA and class IA chaperones. This is in contrast to the class IB chaperone interface of Spa15, which, despite having similar tertiary structure to SrcA, adopts a distinct dimer interface. The interface for the SrcA homodimer is extensive and adopts a parallel configuration when comparing α -helices of opposing subunits similar to SicP (class IA) but different from Spa15 (class IB), which has a reduced dimer interface [22].

The class I chaperones seem to be amenable to crystallographic analysis, presenting a lot of solved structures, as already discussed. On the other hand, the structural data and information on the binding region(s) of class II chaperones

are scarce. Class II secretion chaperones are unique because they bind to at least two translocators that might associate to the eukaryotic membrane to form a pore that allows translocation of effectors. In a recent study [47], the crystal structure of class II IpgC (18-kDa protein) chaperone that binds two essential effectors, IpaB and IpaC, of *Shigella flexneri*, was reported. To understand the molecular basis of the chaperone IpgC function, these authors determined the crystal structure of functional apoIpgC by the multiwavelength anomalous diffraction (MAD) technique using selenomethionine (SeMet)-substituted protein and compared it to the crystal structure of IpgC complexed with the CBD of IpaB₅₁₋₇₂. Based on comparison, it has been proposed that the chaperone captures the translocator CBD in an extended conformation. They also identified the crucial role of N-terminal 21-amino acids in the dimerization of IpgC and defined the N termini of the CBD in IpaB. This information allowed predictions and manipulation for this class of chaperone-substrate interactions and could contribute in the development of specific antibacterial agents. Crystallographic studies have helped in establishing the concept that class I chaperones typically act as homodimers [42, 44] and, as recently demonstrated [48], despite putative structural divergence, some class II chaperones also form homodimers as, for example, SycD from *Yersinia enterocolitica* does. Therefore, dimerization must be a shared feature among class I and class II T3SS chaperones and important for recognition and downstream processing of the chaperone/effector complexes, since, despite the structural differences between these two classes of chaperones, their mode of interaction with effectors is conserved. The dimerization of SycD was later confirmed [49] by presenting the first experimental structure of a class II secretion chaperone, with two crystal structures at 1.95 and 2.6 Å resolution, showing the entire α -helical fold and revealing three tetratricopeptide repeat-like motifs predicted from the amino acid sequence. In both structures, the chaperone forms dimers utilizing residues from the first tetratricopeptide repeat-like motif, showing a deep dependency of dimerization on the same residues. It also was

verified that in solution SycD forms head-to-head homodimers. Therefore, for SycD two different assemblies of monomers into dimers seemed possible, based on the packing in the monoclinic crystals [49]. The finding that this chaperone can form two structurally different head-to-head homodimers with different arrangements of the monomers utilizing the same interface is unusual and seemed to be the first example of this type of dimerization that is almost exclusively mediated by hydrophobic contacts. Indeed, with this one can conclude that, in contrast to class I chaperones, where distinct features such as hydrophobic surface patches have been identified as essential for interaction with the specific secretion target, class II chaperones interact with multiple binding partners of different folds and functions and as a consequence, had to develop an arsenal of binding sites. In comparison with class I and II secretion chaperones, structural data about class III chaperones, or flagellar chaperones, are very poor, commonly encountered as additional information in studies focused on these chaperone effectors with almost no attention given to the chaperones themselves [50]. Class III chaperones bind to the extracellular filament proteins (or flagellin rod in the orthologous flagellar system) that polymerize into a helical structure following secretion from the bacterial cell. Secondary structure predictions suggest that class III chaperones adopt an extended alpha helical structure [32], which is confirmed by the crystal structure of the CesA chaperone in the enteropathogenic *E. coli* that binds the EspA filament protein [51]. The proteins FlgN, FliT and FliS (Fig. 2) of *Salmonella* are considered to belong to class III secretion chaperones due to such common features as low molar mass and a low pI, for example [50, 52, 53]. The interactions between these chaperones and their substrate(s) might be more labile in comparison to those involving the chaperones already discussed, because the complexes could not be recovered from the bacterium cytoplasm [7, 33, 53]. The size of complexes involving these chaperones suggests that a dimer of the chaperone is associated with one molecule of the substrate. Class III chaperones bind to the substrate C-terminal region that is proposed to mediate interactions between subunits in the flagellum [52, 53]. Therefore, flagellar chaperones seem to fine-tune the flagellar assembly process. FliT is a flagellar specific chaperone for the filament-capping protein FliD and is responsible for negative control of flagellar gene expression by binding to the FlhDC complex and is also known to interact with FliJ. FliT is a 14 kDa protein composed of 122 amino acid residues and facilitates the export of its cognate substrate on completion of hook-basal body assembly that, by binding to the effector, prevents its premature aggregation in the cytoplasm [23, 54]. The crystal structure of FliT was solved at 3.2 Å resolution and the asymmetric unit of the crystal contains two FliT molecules which form a dimer related by a pseudo-two fold local symmetry, producing a tetramer in the crystal, but after analytical centrifugation and size-exclusion chromatography experiments showing an equilibrium between monomer and dimer, this tetramer was considered as an artifact of the crystal packing. The chaperone consists of four α -helices and the core of the molecule is formed by an antiparallel α -helical bundle structure. The conformation of the C-terminal segment, which is an amphipathic α -helix, shows little distinction between the two molecules, and is flexible. The swapping of these domains produces the dimer.

These observations imply that the flexible C-terminal segment might play an important role in FliT function [23]. The antiparallel α -helix bundle core structure of FliT is similar to that of FliS, a flagellar specific chaperone for FliC, but the arrangement of the helices is quite different. FliS from *Salmonella typhimurium* has 135 amino acid residues and is predicted to be a predominantly α -helical protein [55]. However, the first crystallographic structure solved of a flagellar chaperone was the FliS from *Aquifex aeolicus* and indicated that the chaperone adopts a novel fold, distinct from the type III secretion chaperones earlier discussed. The structure of FliS solved at a 2.2 Å resolution [56] is an embellished, antiparallel four-helix bundle with a quasi-helical cap on one end formed by the 16 N-terminal residues. Therefore, although there are some superficial similarities between flagellar chaperones and the others T3 secretion chaperones, they do not share a common evolutionary origin. FliS from *Salmonella enterica* was reported to be a homodimer in solution [53], like the T3 secretion chaperones, but FliS from *A. aeolicus* seems to be a monomeric protein and, to support this observation the counterparts of all 40 amino acids that comprise the FliS binding site of *S. enterica* and FliC are bound by a single FliS polypeptide in the crystal structure of the *A. aeolicus* FliS-FliC complex [56]. In conclusion, it was demonstrated that FliS and the others type III secretion chaperones use different structural solutions to do the same thing: bind polypeptides in an extended, nonglobular conformation. Probably these two types of effector-chaperone complexes do not play analogous roles in T3 secretion and in flagellar biosynthesis. In contrast with class I and II chaperones FliS does not seem to be directly involved in the secretion of FliC, acting in the prevention of the premature polymerization of FliC in the bacterial cytosol [56]. In order to obtain more information about class III secretion chaperones and establish some class common features, it remains to be seen if FliS, FliT, and FlgN have similar tertiary structures like their counterparts in T3SS. However, although the amino acid sequences of FliT and FlgN are consistent with the α -helical folds, they seem to be unrelated to that of FliS.

Secretion chaperones are also required in type IV secretion systems (T4SS), which are similar to T3SS in direct effector transportation, but with one very important difference, since the first ones are also responsible for DNA exportation [6]. Other differences between these two systems are the secretion chaperones: the T3SS chaperones are extensively studied as already demonstrated; while the great number of papers on T4SS contains only data on activities of the effectors and their cognate chaperone. For example, in *Agrobacterium tumefaciens*, the prototype for T4SSs, protein VirE1 was identified as a secretion chaperone owing to some of its features, for example, it is a small (7 kDa) protein, with low pI and prevents the VirE2 from oligomerization and aggregation [19, 57, 58].

To date, there are no structural studies reported for VirE1, although structural information about this protein has been provided by structural studies on VirE1 complexed with its effector (VirE2) by electron microscopy (EM), small angle X-ray scattering (SAXS) [37, 58] and crystallography [59]. Despite earlier gel filtration assays that indicated that the VirE1-VirE2 complex has a 2:1 ratio, recent SAXS and EM studies indicate that, in co-expression, VirE1 and VirE2

form a heterodimer in a 1:1 molar ratio in solution, through the interaction of the VirE1 amphipathic α -helix C-terminal portion with a short N-terminal portion of effector (VirE2 protein, Fig. 3).

Recently the complex structure VirE1/VirE2 was solved utilizing crystallography [59] and the results obtained are in agreement with previously published SAXS and EM data. These results indicate a single fold in two structural domains of VirE2 that are connected by a single interdomain linker; while the VirE1 comprises a single helix situated between these two VirE2 domains with multiple interactions with both of them. All these observations imply that VirE2 possess a dynamic structure that can accommodate its different partners due to the flexibility of its interdomain linker, and indicates that VirE1 is a specific secretion chaperone that prevents premature interactions and oligomerization as a bodyguard to VirE2 until its secretion with bonded ssDNA.

FURTHER STRUCTURAL AND FUNCTIONAL STUDIES ON SECRETION CHAPERONES

Structural secretion chaperone research is limited to crystallographic studies on structured portions of protein fragments due to the aggregation-prone nature of free intact effectors and the intractability of intact chaperone-effector complexes to crystallize. This could be overcome by preparing the samples at concentrations and conditions suitable for NMR spectroscopy [60-62]. So far in T3SS, two theories have been proposed for the extended conformation of the effector CBD: (i) it is necessary to maintain or prime unfolding of the effector for transport through the narrow T3 secretion needle, and (ii) it forms a discrete three-dimensional signal targeting the chaperone-effector complex to a T3 secretion component required for translocation. Recent results demonstrating binding between a *Salmonella* chaperone-effector complex and the T3S ATPase, as well as ATP hydrolysis-dependent unfolding of the effector by the ATPase,

are consistent with both the unfolding and targeting models [63].

The structural and dynamic changes on an intact effector upon chaperone binding were studied by NMR [64]. Rather than maintaining an unfolded state in the effector, secretion chaperone SycE was found to promote structuring of the YopE CBD. Taken together with structural conservation of chaperone-effector complexes, these results were most consistent with a targeting model of action in which the CBD, in association with the chaperone constitutes a three-dimensional targeting signal [65]. The strongest experimental support for the targeting model comes from a *Yersinia* strain which expressed most of its effectors but had its HOPEM encoding DNA segment deleted (Δ HOPEM) [35]. In this strain, neither SycE nor the CBD region of YopE is required for translocation, whereas in wild-type strains both are required. Dispensability of SycE and the CBD region in the Δ HOPEM strain suggests that chaperones have a role in enhancing interaction between effectors and a T3 secretion component that is required for translocation. Other NMR results provided support for a targeting model of chaperone action and are inconsistent with unfolding models. As recently reported [41, 63, 64], the first 100 N-terminal residues of effector YopE protein are disordered and flexible as in natively unfolded proteins, but only in the absence of bound chaperone (SycE). By SycE binding, a pronounced disorder-to-order transition in the YopE CBD occurred but had no effect on other portions of this effector. No association between SycE-YopE and the *Yersinia* T3S ATPase YscN was detected, suggesting that the target of the putative signal in the SycE-YopE complex is some other component than YscN in *Yersinia*. Additionally, YopE residues form a solvent-exposed patch on the surface of the chaperone-binding domain (CBD) [64], and their effect on translocation is consistent with the structure of the CBD serving as a signal for translocation, and the effector being responsible for secretion but only when bonded to the chaperone.



Figure 3. Tertiary structure of VirE1/VirE2 complex from *Agrobacterium tumefaciens* T4SS, with small VirE1 chaperone (dark grey) preventing VirE2 (shade grey) aggregation and premature precipitation (Protein Data Bank entry: 3BTP).

Upon introduction with suitable ionization techniques, such as ESI (Electrospray Ionization) and MALDI (Matrix-Assisted Laser Desorption/Ionization) [66-68], Mass Spectrometry (MS) has been widely used in the study of proteins due to very precise information on mass, sequence and even proteins 3D structure. There are however only a few studies on secretion chaperones in which MS was applied. Among these, we can cite the MS sequencing experiments using tandem mass spectrometry applied in secretion chaperones functional studies and some on complexes among chaperones and effectors. Recently, functional studies on SsaE protein were conducted by nano-LC-ESI-MS/MS [69]. This protein is encoded by the T3SS operon within the *Salmonella* pathogenicity island 2 (SPI-2) *locus* and has 15 kDa and low pI, as do mostly known secretion chaperones. After coimmunoprecipitation assays, the bound proteins were eluted and subjected to MS analyses. Seven SPI-2 proteins: SsaH, SsaK, SsaN, SsaQ, SseA, SseB, and SseC, were identified as SsaE-interacting proteins. It also was verified that mutants without the *locus* for SsaE didn't secrete SPI-2 effectors, probably due to failure in assembly of the needle complex. The T3 apparatus is associated with an ATPase that presumably provides the energy for the secretion process and, interestingly, mass spectrometric analysis revealed that SsaE is coeluted with the predicted SPI-2 T3SS ATPase SsaN [69].

In another work [69], the study of the complexes formed between T3SS chaperones from *Yersinia pestis* and target proteins was reported. The interactions between the proteins were first identified by surface plasmon resonance (SPR), and then the complexes were submitted to MALDI-TOF-MS experiments, which confirmed even the formation of complexes identified as ones with weak interactions by the previous tests. The high affinity interactions exhibited very strong mass/charge signals. The SPR and MALDI-TOF-MS data were complementary, and provided a powerful method for rapidly identifying protein-protein interactions involved in complex macromolecular assemblies, such as T3SS are, on association of secretion chaperones and their effector proteins, thus enabling understanding some key steps in the assembly and function of these macromolecular systems. Besides, considering that unfolding is a prerequisite for T3 secretion, the interactions identified are likely to occur either before or after this stage in the process. Hence a hierarchical delivery of proteins through the injectosome may be partially controlled by the stability of chaperone-effector complexes. Therefore, the cited experiments could help in the elucidation of the transport of virulence factors across the bacterial and host cell membranes [70].

FINAL REMARKS

The central role in secretion strategy involving bacterial T3, flagellar and T4 secretion systems is occupied by specific secretion chaperones when secretion competencies are detected or triggered by a cell. The T3SS specific dimeric chaperones represent a large family of proteins with limited sequence identity ($\leq 20\%$), very well conserved folds and modes for effector binding in which the chaperone dimers provide rigid surfaces around which effectors wrap an ~25–100-residue chaperone-binding (CBD) region. T3SS chaperones are considered pilots for presenting the effectors to se-

cretion system, and also bodyguards since they enable effectors to remain soluble even when produced in higher concentrations. On the other hand, all known flagellar and T4 secretion chaperones are almost entirely composed from α -helices and act principally as bodyguards for their effectors by preventing premature association and aggregation of the effectors. Although structural studies on secretion chaperones, having crystallography as the leading technique in structural assignments so far, have led to better understanding of virulence mechanisms that many Gram-negative pathogens use, there are still many questions to be answered.

ACKNOWLEDGEMENTS

The financial support from FAPESP, CNPq, CAPES, and Prof. Carol H. Collins (IQ-UNICAMP) for helpful suggestions about English grammar and style are gratefully acknowledged.

REFERENCES

- [1] Rodou, A.; Ankrah, D.; Stathopoulos, C. Toxins and secretion systems of *Photobacterium luminescens*. *Toxins*, **2010**, *2*, 1250-1264.
- [2] Cornelis, G. The type III secretion injectisome. *Nat. Rev. Microbiol.*, **2006**, *4*, 811-825.
- [3] Baron, C.; O' Callaghan, D.; Lanka, E. Bacterial secrets of secretion: Euroconference on the biology of type IV secretion processes. *Mol. Microbiol.*, **2002**, *43*, 1359-1365.
- [4] Galán, J.; Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature*, **2006**, *444*, 567-573.
- [5] Lavigne, J.-P.; Botella, E.; O' Callaghan, D. Type IV secretion system and their effectors: An update. *Pathol. Biol.*, **2006**, *54*, 296-303.
- [6] Cascales, E.; Christie, P. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.*, **2003**, *1*, 137-149.
- [7] Parsot, C.; Hamiaux, C.; Page, A. The various and varying roles of specific chaperones in type III secretion systems. *Curr. Opin. Microbiol.*, **2003**, *6*, 7-14.
- [8] Burns, D. L. Biochemistry of type IV secretion. *Curr. Opin. Microbiol.*, **1999**, *2*, 25-29.
- [9] Bingle, L.E.H.; Bailey, C.M.; Pallen, M.J. Type VI secretion: A beginner's guide. *Curr. Opin. Microbiol.*, **2008**, *11*, 3-8.
- [10] Anderson, D.; Schneewind, O. Type III machines of Gram-negative pathogens: injecting virulence factors into host cells and more. *Curr. Opin. Microbiol.*, **1999**, *2*, 18-24.
- [11] Gelvin, S. Finding a way to the nucleus. *Curr. Opin. Microbiol.*, **2010**, *13*, 53-58.
- [12] Cornelis, G.; Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.*, **2000**, *54*, 735-774.
- [13] Stebbins, C.; Galán, J. Priming virulence factors for delivery into the host. *Nature Rev. Mol. Cell Biol.*, **2003**, *4*, 738-743.
- [14] Akeda, Y.; Galán, J. Chaperone release and unfolding of substrates in type III secretion. *Nature*, **2005**, *437*, 911-915.
- [15] Darwin, K.; Miller, V. Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in *Salmonella typhimurium*. *EMBO J.*, **2001**, *20*, 1850 - 1862.
- [16] Galán, J.; Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science*, **1999**, *284*, 1322-1328.
- [17] Bachert, S.; Selbach, M. Role of type IV secretion in *Helicobacter pylori* Pathogenesis. *Cell. Microbiol.*, **2008**, *10*, 1573-1581.
- [18] Llosa, M.; Roy, C.; Dehio, C. Bacterial type IV secretion systems in human disease. *Mol. Microbiol.*, **2009**, *73*, 141-151.
- [19] Zhao, Z.; Sagulenko, E.; Ding, Z.; Christie, P. Activities of virE1 and the VirE1 secretion chaperone in export of the multifunctional VirE2 effector via an *Agrobacterium* type IV secretion pathway. *J. Bacteriol.*, **2001**, *183*(13), 3855-3865.
- [20] Dumas, F.; Duckely, M.; Pelczar, P.; van Gelder, P.; Hohn, B. An *Agrobacterium* VirE2 channel for T-DNA transport into plant cells. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 485-490.
- [21] (a) Melnikov, E.E.; Rotanova, T.V. Molecular chaperones. *Russ. J. Bioorg. Chem.*, **2010**, *36*, 5-14; (b) Takaya, A.; Tomoyasu, T.; Ma-

- tsui, H.; Yamamoto, T. The DnaK/DnaJ chaperone machinery of *Salmonella enterica* Serovar Typhimurium is essential for invasion of epithelial cells and survival within macrophages, leading to systemic infection. *Infect. Immun.*, **2004**, *27*, 1364-1373; (c) Pérez-Rodríguez, R.; Fisher, A.C.; Perlmutter, J.D.; Hicks, M.G.; Chanal, A.; Santini, C.-L.; Wu, L.-F.; Palmer, T.; DeLisa, M.P. An essential role for the DnaK molecular chaperone in stabilizing over-expressed substrate proteins of the bacterial twin-arginine translocation pathway. *J. Mol. Biol.*, **2007**, *367*, 715-730.
- [22] Cooper, C.; Zhang, K.; Andres, S.; Fang, Y.; Kaniuk, N.; Hanne-mann, M.; Brumell, J.; Foster, L.; Junop, M.; Coombes, B. Structural and biochemical characterization of SrcA, a multi-Cargo type III secretion chaperone in *Salmonella* required for pathogenic association with a host. *PLoS Pathog.*, **2010**, *6*, e100075 (1-11).
- [23] Imada, K.; Minamino, T.; Kinoshita, M.; Furukawa, Y.; Namba, K. Structural insight into the regulatory mechanisms of interactions of the flagellar type III chaperone FliT with its binding partners. *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, 8812-8817.
- [24] Evans, L.; Hughes, C. Selective binding of virulence type III export chaperones by FliJ escort orthologues InvI and YscO. *FEMS Microbiol. Lett.*, **2009**, *293*, 292-297.
- [25] Bennett, J.; Hughes, C. From flagellum assembly to virulence: the extended family of type III export chaperones. *Trends Microbiol.*, **2000**, *8*, 202-204.
- [26] Büttner, D.; He, S. Type III protein secretion in plant pathogenic bacteria. *Plant Physiol.*, **2009**, *150*, 1656-1664.
- [27] Büttner, D.; Bonas, U. Who comes first? How plant pathogenic bacteria orchestrate type III secretion. *Curr. Opin. Microbiol.*, **2006**, *9*, 193-200.
- [28] (a) Thomas, J.; Stafford, G.; Hughes, C. Docking of cytosolic chaperone-substrate complexes at the membrane ATPase during flagellar type III protein export. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 3945-3950; (b) He, S.; Nomura, K.; Whittam, T. Type III protein secretion mechanism in mammalian and plant pathogens. *Biochim. Biophys. Acta*, **2004**, *1694*, 181-206.
- [29] (a) Elliot, S.J.; Hutcheson, S.W.; Dubois, M.S.; Mellies, J.L.; Wainwright, L.A.; Batchelor, M.; Frankel, G.; Knutton, S.; Kaper, J.B. Identification of CesT, a chaperone for the type III secretion of Tir in enteropathogenic *Escherichia coli*. *Mol. Microbiol.*, **1999**, *33*, 1176-1189; (b) Darwin, K.H.; Robinson, L.S.; Miller, V.L. SigE is a chaperone for the *Salmonella enterica* Serovar Typhimurium invasion protein SigD. *J. Bacteriol.*, **2001**, *183*, 1452-1454.
- [30] Minamino, T.; Imada, K.; Namba, K.; Mechanisms of type III protein export for bacterial flagellar assembly. *Mol. Biosyst.*, **2008**, *11*, 1105-1115.
- [31] Ghosh, P. Process of protein transport by the type III secretion system. *Microbiol. Mol. Biol. Rev.*, **2004**, *68*, 771-795.
- [32] Poyraz, O.; Schmidt, H.; Seidel, K.; Delissen, F.; Ader, C.; Tenenboim, H.; Goosmann, C.; Laube, B.; Thünemann, A.; Zychlinsky, A.; Baldus, M.; Lange, A.; Griesinger, C.; Kolbe, M. Protein refolding is required for assembly of the type three secretion needle. *Nat. Struct. Mol. Biol.*, **2010**, *17*, 788-792.
- [33] Tasic, L.; Borin, P.; Khater, L.; Ramos, C. Cloning and characterization of three hypothetical secretion chaperone proteins from *Xanthomonas axonopodis* pv. *citri*. *Prot. Expr. Purif.*, **2007**, *53*, 363-369.
- [34] Lloyd, S.; Norman, M.; Rosqvist R, Wolf-Watz H. Yersinia YopE is targeted for type III secretion by N-terminal, not mRNA, signals. *Mol. Microbiol.*, **2001**, *39*, 520-531.
- [35] Boyd, A.; Grosdent, N.; Töttemeyer, S.; Geuijen, C.; Blevess, S.; Iriarte, M.; Lambermont, I.; Octave, J.; Cornelis, G. *Yersinia enterocolitica* can deliver Yop proteins into a wide range of cell types: development of a delivery system for heterologous proteins. *Eur. J. Cell Biol.*, **2000**, *79*, 659-671.
- [36] Wattiau, P.; Woestyne, S.; Cornelis, G. Customized secretion chaperones in pathogenic bacteria. *Mol. Microbiol.*, **1996**, *20*, 255-262.
- [37] Frenkiel-Krispin, D.; Wolf, S.G.; Albeck, S.; Unger, T.; Peleg, Y.; Jacobovitch, J.; Michael, Y.; Daube, S.; Sharon, M.; Robinson, C.V.; Svergun, D.I.; Fass, D.; Tzfira, T.; Elbaum, M. Plant Transformation by *Agrobacterium tumefaciens*: Modulation of single-stranded DNA-VirE2 complex assembly by VirE1. *J. Biol. Chem.*, **2007**, *282*, 3458-3464.
- [38] Khater, L.; Alegria, M.; Borin, P.; Santos, T.; Docena, C.; Tasic, L.; Farah, C.; Ramos, C. Identification of the flagellar chaperone FliN in the phytopathogen *Xanthomonas axonopodis* pathovar *citri* by its interaction with hook-associated FlgK. *Arch. Microbiol.*, **2007**, *188*, 243-250.
- [39] Cornelis, G. The type III secretion injectisome, a complex nanomachine for intracellular 'toxin' delivery. *Biol. Chem.*, **2010**, *391*, 745-751.
- [40] Paul, K.; Erhardt, M.; Hirano, T.; Blair, D.; Hughes, K. Energy source of flagellar type III secretion. *Nature*, **2008**, *451*, 489-492.
- [41] Birtalan, S.; Ghosh, P. Structure of the *Yersinia* type III secretory system chaperone SycE. *Nat. Struct. Biol.*, **2001**, *8*, 974-978.
- [42] Stebbins, C.; Galán, J. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature*, **2001**, *414*, 77-81.
- [43] Sory, M.; Boland, A.; Lambermont, I.; Cornelis, G. Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the *cyaA* gene fusion approach. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 11998-12002.
- [44] Birtalan, S.; Phillips, R.; Ghosh, P. Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Mol. Cell*, **2002**, *9*, 971-980.
- [45] Phan, J.; Tropea, J.; Waugh, D. Structure of the *Yersinia pestis* type III secretion chaperone SycH in complex with a stable fragment of YscM2. *Acta Crystallogr. D Biol. Crystallogr.*, **2004**, *60*, 1591-1599.
- [46] Lilic, M.; Vujanac, M.; Stebbins, C. A common structural motif in the binding of virulence factors to bacterial secretion chaperones. *Mol. Cell*, **2006**, *21*, 653-664.
- [47] Lunelli, M.; Lokareddy, R.; Zychlinsky, A.; Kolbe, M. IpaB-IpgC interaction defines binding motif for type III secretion translocator. *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 9661-9666.
- [48] Schmid, A.; Dittmann, S.; Grimminger, V.; Walter, S.; Heesemann, J.; Wilharm, G. *Yersinia enterocolitica* type III secretion chaperone SycD: Recombinant expression, purification and characterization of a homodimer. *Prot. Expr. Purif.*, **2006**, *49*, 176-182.
- [49] Büttner, C.; Sorg, I.; Cornelis, G.; Heinz, D.; Niemann, H. Structure of the *Yersinia enterocolitica* type III secretion translocator chaperone SycD. *J. Mol. Biol.*, **2008**, *375*, 997-1012.
- [50] Bennet, A.; Thomas, J.; Fraser, G.; Hughes, C. Substrate complexes and domain organization of the *Salmonella* flagellar export chaperones FlgN and FliT. *Mol. Microbiol.*, **2001**, *39*, 781-791.
- [51] Yip, C.; Finlay, B.; Strynadka, N. Structural characterization of a type III secretion system filament protein in complex with its chaperone. *Nat. Struct. Mol. Biol.*, **2005**, *12*, 75-81.
- [52] Fraser, G.; Bennet, J.; Hughes, C. Substrate-specific binding of hook-associated proteins by FlgN and FliT, putative chaperones for flagellum assembly. *Mol. Microbiol.*, **1999**, *32*, 569-580.
- [53] Auvray, F.; Thomas, J.; Fraser, G.; Hughes, C. Flagellin polymerization control by a cytosolic export chaperone. *J. Mol. Biol.*, **2001**, *308*, 221-229.
- [54] Kinoshita, M.; Yamane, M.; Matsunami, H.; Minamino, T.; Namba, K.; Imada, K. Purification, crystallization and preliminary X-ray analysis of FliT, a bacterial flagellar substrate-specific export chaperone. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.*, **2009**, *F65*, 825-828.
- [55] Muskotál, A.; Király, R.; Sebestyén, A.; Gugolya, Z.; Végh, B.; Vonderviszt, F. Interaction of FliS flagellar chaperone with flagellin. *FEBS Lett.*, **2006**, *580*, 3916-3920.
- [56] Evdokimov, A.; Phan, J.; Tropea, J.; Routzahn, K.; Peters III, H.; Pokross, M.; Waugh, D. Similar modes of polypeptide recognition by export chaperones in flagellar biosynthesis and type III secretion. *Nat. Struct. Biol.*, **2003**, *10*, 789-793.
- [57] Sundberg, C.; Ream, W. The *Agrobacterium tumefaciens* chaperone-like protein, VirE1, interacts with VirE2 at domains required for Single-Stranded DNA binding and cooperative interaction. *J. Bacteriol.*, **1999**, *181*, 6850-6855.
- [58] Duckely, M.; Oomen, C.; Axthelm, F.; Van Gelder, P.; Waksman, G.; Engel, A. The VirE1-VirE2 complex of *Agrobacterium tumefaciens* interacts with single-stranded DNA and forms channels. *Mol. Microbiol.*, **2005**, *58*, 1130-1142.
- [59] Dym, O.; Albeck, S.; Unger, T.; Jacobovitch, J.; Branzburg, A.; Michael, Y.; Frenkiel-Krispin, D.; Wolf, S.G.; Elbaum, M. Crystal structure of the *Agrobacterium* virulence complex VirE1-VirE2 reveals a flexible protein that can accommodate different partners. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 11170-11175.
- [60] Takahashi, H.; Shimada, I. Production of isotopically labeled heterologous proteins in non-*E.coli* prokaryotic and eukaryotic cells. *J. Biomol. NMR*, **2010**, *46*, 3-10.

- [61] Chou, J.; Delaglio, F.; Bax, A. Measurement of one-bond¹⁵N-¹³C dipolar couplings in medium sized proteins. *J. Biomol. NMR*, **2000**, *18*, 101-105.
- [62] Rumpel, S.; Lakshmi, R.; Becker, S.; Zweckstetter, M. Assignment-free solution NMR method reveals CesT as an unswapped homodimer. *Prot. Sci.*, **2008**, *17*, 2015-2019.
- [63] Rodgers, L.; Mukerjea, R.; Birtalan, S.; Friedberg, D.; Ghosh, P. A solvent-exposed patch in chaperone-bound YopE is required for translocation by the type III secretion system. *J. Bacteriol.*, **2010**, *192*, 3114-3122.
- [64] Rodgers, L.; Gamez, A.; Riek, R.; Ghosh, P. The type III secretion chaperone SycE promotes a localized disorder-to-order transition in the natively unfolded effector YopE. *J. Biol. Chem.*, **2008**, *283*, 20857-20863.
- [65] Birtalan, S.; Phillips, R.; Ghosh, P. Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Mol. Cell*, **2002**, *9*, 971-80.
- [66] Whitehouse, C.; Dreyer, R.; Yamashita, M.; Fenn, J. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.*, **1985**, *57*, 675-679.
- [67] Karas, M.; Hillenkamp, E. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.*, **1988**, *60*, 2299-320.
- [68] Kinter, M.; Sherman, N. Protein Sequencing and Identification Using Tandem Mass Spectrometry. Wiley-Interscience, Inc, **2000**.
- [69] Swietnicki, W.; O'Brien, S.; Holman, K.; Cherry, S.; Brueggemann, E.; Tropea, J.; Hines, H.; Waugh, D.; Ulrich, R. Novel protein-protein interactions of the *Yersinia pestis* type III secretion system elucidated with a matrix analysis by surface plasmon resonance and mass spectrometry. *J. Biol. Chem.*, **2004**, *37*, 38693-38700.
- [70] Miki, T.; Shibagaki, Y.; Danbara, H.; Okada, N. Functional characterization of SsaE, a novel chaperone protein of the type III secretion system encoded by *Salmonella* pathogenicity island 2. *J. Bacteriol.*, **2009**, *191*, 6843-6854.
- [71] Buttner, D.; Bonas, U. Port of Entry the Type III Secretion Translocon. *Trends Microbiol.*, **2002**, *10*, 186-192.



Sugarcane Hsp101 is a hexameric chaperone that binds nucleotides

Thiago C. Cagliari^{a,b,1}, Viviane C.H. da Silva^{a,b,1}, Júlio C. Borges^c, Alessandra Prando^a, Ljubica Tasic^{a,d}, Carlos H.I. Ramos^{a,d,*}

^a Institute of Chemistry, University of Campinas-UNICAMP, P.O. Box 6154, 13083-970, Campinas, SP, Brazil

^b Institute of Biology, University of Campinas-UNICAMP, P.O. Box 6154, 13083-970, Campinas, SP, Brazil

^c Institute of Chemistry of São Carlos, University of São Paulo, São Carlos, SP, 13560-970, Brazil

^d Instituto Nacional de Ciência e Tecnologia em Biologia Estrutural e Bioimagem, Brazil

ARTICLE INFO

Article history:

Received 24 July 2011

Received in revised form 23 August 2011

Accepted 24 August 2011

Available online 31 August 2011

Keywords:

Hsp101

Sugarcane

Protein folding

Molecular chaperone

Heat shock protein

AAA+ protein

Analytical ultracentrifugation

STD-NMR

ABSTRACT

The Clp/Hsp100 AAA+ chaperone family is involved in recovering aggregated proteins and little is known about other orthologs of the well studied ClpB from *Escherichia coli* and Hsp104 from *Saccharomyces cerevisiae*. Plant Hsp101 is a good model for understanding the relationship between the structure and function of Hsp100 proteins and to investigate the role of these chaperones in disaggregation processes. Here, we present the cloning and purification of a sugarcane ortholog, SHsp101, which is expressed in sugarcane cells and is a folded hexamer that is capable of binding nucleotides. Thus SHsp101 has the structural and functional characteristics of the Clp/Hsp100 AAA+ family.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Protein misfolding and aggregation play important roles in many pathologies. Heterologous expression of genes sometimes generates inclusion bodies. Failure in proper folding results in a loss of biological function or even the formation of aggregated species, which may be harmful to the organism by developing into conformational diseases related to more than 40 pathologies [1,2]. To avoid either misfolding or its destructive consequences, cells have developed a complex quality control system involving molecular chaperones that is devoted to maintaining protein homeostasis [3,4].

Molecular chaperones are usually referred to as heat shock proteins (Hsps), although some molecular chaperones are not Hsps and *vice versa*, and are named according to the molecular mass of the monomer protein (for example, the 70 kDa

heat shock protein, Hsp70, and the 100 kDa heat shock protein, Hsp100) [4,5]. Chaperones from the Clp/Hsp100 family are ring-shaped AAA+ (ATPase associated with diverse cellular activities) proteins. ClpB from *Escherichia coli* and its eukaryotic homolog Hsp104 from yeast *Saccharomyces cerevisiae* are the most studied [5–7]. ClpB and Hsp104 have the extraordinary ability to solubilize aggregated proteins with the help of the Hsp70 system. Hsp100s are hexameric chaperones that are formed by an amino-terminal domain and two nucleotide-binding domains, NBD or AAA+ domains, which are connected by a linker named the M-domain in the case of ClpB and Hsp104, but not other Clp proteins (Fig. 1A). Orthologs of these proteins are present in bacteria, fungi, plants and mitochondria but are missing in metazoa.

Although not lethal when absent, the presence of Hsp104 and ClpB largely increases survival under stress conditions. Neither ClpB nor Hsp104 alone are capable of disaggregating proteins, but cells lacking *clpb* or *hsp104* are incompetent at eliminating aggregated proteins formed under heat shock conditions [8–10]. Hsp100 proteins cooperate with other chaperones to aid in protein folding. Efficient protein disaggregation and reactivation in *E. coli* is only achieved when ClpB, DnaK, DnaJ and GrpE are present [9,10]. Another partner, small Hsp, participates by binding partially unfolded proteins, thereby not allowing the formation of oligomeric aggregates. Hsp70 and Hsp100 facilitate the

Abbreviations: AUC, analytical ultracentrifugation; CD, circular dichroism; Hsp, heat shock protein; STD-NMR, saturation transfer difference NMR.

* Corresponding author at: Institute of Chemistry, University of Campinas-UNICAMP, P.O. Box 6154, 13083-970, Campinas, SP, Brazil. Tel.: +55 19 3521 3096; fax: +55 19 3521 3023.

E-mail address: cramos@iqm.unicamp.br (C.H.I. Ramos).

¹ Equally contributed to this work.

assay [21] using a kit (Bio-Rad). A model of SHsp101 based on the crystal structure of *Thermus thermophilus* ClpB (gi:38492937, PDB entry number 1QVR) [7], which shares 50% identity with sugarcane Hsp101, was generated using 3D-JIGSAW [22].

2.2. Spectroscopy

Circular dichroism (CD) spectra were recorded with a JASCO model J-810 CD spectropolarimeter equipped with a thermoelectric sample temperature controller (Peltier system) following standard procedures [23]. Data were collected from 260 to 200 nm, averaged at least three times, using cuvettes with a 1 mm path length and 2 μ M protein in 25 mM Tris–HCl, pH 7.5, 250 mM NaCl, and 1 mM β -mercaptoethanol. Data was analyzed with both Spectra Manager[®] (Jasco) and Origin[®] 7.5 (OriginLab). Emission fluorescence spectra were recorded on an Aminco Bowman[®] Series 2 (SLM-AMINCO) fluorimeter using quartz cells with 10 mm \times 10 mm optical path length. Protein concentrations of 2 μ M in the same buffer described for CD experiments were used. Emission fluorescence spectra of tryptophan were obtained with excitation at 295 nm (bandpass of 4 nm) and with emission from 305 to 420 nm (bandpass of 8 nm) and analyzed either by their maxima wavelength or by their center of spectral mass ($\langle \lambda \rangle$) as described by the equation below:

$$\langle \lambda \rangle = \left(\frac{\sum \lambda_i F_i}{\sum F_i} \right) \quad (1)$$

where λ_i is the wavelength and F_i is the fluorescence intensity at λ_i . Spectroscopy experiments were performed in the absence and presence of 200 μ M MgATP.

2.3. Protein extraction and immunoblotting

Plant material (leaf tissue) from *S. officinarum* cultivar IAC 93-3046 was homogenized in a mortar and pestle with two volumes (w/v) of 25 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA and protease inhibitors (1 μ M E64, 1 μ M leupeptin, 150 nM aprotinin, 500 μ M AEBSF (CALBIOCHEM[®]), and 1 mM phenylmethanesulfonyl fluoride). After centrifugation at 2500 g for 30 min, 15 ng of the supernatant was separated by 10% SDS-PAGE, electrotransferred onto a nitrocellulose membrane (Millipore), and stained with Ponceau S (Sigma–Aldrich). The membranes were blocked for 1 h with TBS-T (Tris-buffered saline containing 0.03% Tween-20) supplemented with 1% BSA and then incubated with polyclonal sugarcane Hsp101 antiserum (anti-SHsp101) for 2 h (1:500 dilution). After 3 washes with TBS-T, the membranes were incubated for 1 h with the secondary antibody (anti-rabbit IgG coupled to alkaline phosphatase, Bio-Rad) and detection was performed using an Alkaline Phosphatase Conjugate Substrate Kit (Bio-Rad). All experiments were performed at room temperature.

2.4. Dynamic light scattering, analytical ultracentrifugation and SEC-MALS

For the dynamic light scattering (DLS), analytical ultracentrifugation (AUC) and size exclusion chromatography with multi-angle light scattering (SEC-MALS) experiments, 300–2000 μ g/ml SHsp101 was tested in 25 mM Tris–Cl, pH 7.5, 250 mM NaCl and 1 mM Beta-mercaptoethanol in the absence or presence of 200 μ M nucleotides. SednTerp software (<http://www.jphilo.mailway.com/download.htm>) was used to calculate the density ($\rho = 1.00923$ g/mL) and viscosity ($\eta = 1.0321 \times 10^{-2}$ poise) of the buffer and the protein partial specific volume ($V_{bar} = 0.7384$ mL/g).

Dynamic light scattering (DLS) experiments were performed using a DynaPro-MS/X (Protein Solutions) instrument controlled by

DynamicsTM v.6.2.05 software. Samples were equilibrated at 20 °C for approximately 20 min and then submitted to approximately 300 measurements (10 s each). The DLS experiments were conducted to determine the experimental diffusion coefficient D at each protein concentration. Each value was extrapolated to standard conditions (temperature of 20 °C and in water, $D_{20,W}$). The values of $D_{20,W}$ were plotted versus protein concentration to determine the value of zero by extrapolation, $D_{20,W}^0$, which is a unique property of the macromolecule because it should avoid interference from the solvent and high protein concentrations.

Analytical ultracentrifugation experiments were performed using a Beckman Optima XL-A analytical ultracentrifuge using an AN-60Ti rotor at 20 °C and analyzed as described elsewhere [24–26]. Sedimentation experiments were conducted from 3000 to 16,000 rpm with data acquisition at 279 nm. Sedimentation velocity analysis was performed using the software Sedfit [27–29] to calculate the sedimentation coefficient, s . As described above for the DLS experiments, each value of s was extrapolated to standard conditions (temperature of 20 °C and in water, $s_{20,W}$). The values of $s_{20,W}$ were plotted versus protein concentration to determine the value of zero by extrapolation, $s_{20,W}^0$, which is a unique property of the macromolecule because it should avoid interference from the solvent and high protein concentrations. The molecular mass (M) of a protein can be estimated from s and D by the following equation:

$$M = \frac{sRT}{D(1 - V_{bar}\rho_f)} \quad (2)$$

where R is the gas constant, T is the absolute temperature and f is the frictional coefficient.

Equilibrium sedimentation analyses involved fitting an absorbance model versus the cell radius data using nonlinear regression. Distribution of the protein along the cell, which was obtained in the equilibrium sedimentation experiments, was fit with the following equation [30]:

$$C = C_0 e^{\left[\frac{M(1 - V_{bar}\rho)\omega^2(r^2 - r_0^2)}{2RT} \right]} \quad (3)$$

where C is the protein concentration at radial position r , C_0 is the protein concentration at radial position r_0 and ω is the centrifugal angular velocity.

Multi-angle laser light-scattering coupled with size exclusion chromatography (SEC–MALLS) was used to analyze the distribution of the masses of purified SHsp101. Data analyses were performed by chromatographic separation of 160 μ M protein using a 1 cm \times 60 cm Superdex 200 (GE Healthcare) column at room temperature with a flow rate of 0.5 mL/min. The experiments were performed continuously on the column eluate as it passed through a DAWN EOS System (Wyatt Technology Corporation, USA), and the data were analyzed using the Astra software package (version 5.3.4.14).

2.5. Saturation-transfer difference (STD) and NMR analyses

Purified SHsp101 (120 μ M) and nucleotides (81 mM ATP, 80 mM ADP and 46 mM ATP γ S) were prepared in 10 mM Tris–HCl buffer at pH 8.0 with a protein to ligand molar ratio of 1:100. The sample contained a total volume of 600 μ L and was completed with D₂O. All NMR spectra were recorded on a 500 MHz Varian INOVA spectrometer equipped with a 5 mm Tri-Res probe operating at a hydrogen frequency of 499.89 MHz at 298 K without sample spinning. The residual HDO signal was used as the internal reference (set at 4.70 ppm). The 1D STD NMR experiments employed the pulse sequence PRESAT, Water package from Varian to suppress the residual HDO signal. A spin-lock filter with a strength of 2 kHz and duration of 10 ms was applied to suppress the protein

Table 1
Purification summary of SHsp101 per 1 L of culture.

Purification step	Total protein ^{a,b} (mg)	SHsp101 ^b (mg)	Yield	Purity (%)
Soluble lysate	280	143	100	51
Affinity chromatography	113	96	67	85
Size exclusion chromatography	43	42	29	98

Errors are less than 4%.

^a 3 g of cells from 1 L of culture were lysed by sonication.

^b SHsp101 concentration was determined by the absorbance at 280 nm, and the total protein concentration was measured with the Bradford protein assay (see Section 2).

background. On-resonance irradiation of the protein was performed at a chemical shift of 0.0 ppm and the off-resonance frequency was set at 30.0 ppm where no protein signals are present. A train of selective Gaussian-shaped pulses (50 ms) with a power level of 32 Hz for the corresponding square shape was applied for saturation. The number of selective pulses (n) determined the pre-saturation period. The standard value was 58 pulses, leading to a total length of the saturation train of 3 s. Spectra were subtracted internally via phase cycling after every scan to minimize artifacts arising from temperature and magnetic field instability. The length of the spin-lock pulse was set to 40 ms. The 1D STD NMR spectra were acquired with 1024 scans and 2 s relaxation delay, and the free induction decay (FID) was acquired over 14,402 complex data points covering a spectral width of 12001.2 Hz. NMR spectra were multiplied by an exponential line-broadening function of 1 Hz and zero-filled by a factor of two prior to Fourier transformation. Spectral processing was performed on a Sun workstation using Vnmrj software (Varian package).

The $\{^1\text{H}, ^{13}\text{C}\}$ heteronuclear correlation (HETCOR) analysis was performed by dissolving 5.0 mg of ATP γ S in 40 μL of D_2O (as the control) or by dissolving 2.5 mg of ATP γ S into 20 μL of SHsp101 (stock solution: 120 μL) in 20 μL of D_2O . The HETCOR spectra were acquired with 1024 scans and 1.5 s relaxation delays, and the free induction decay (FID) was acquired over 4096 complex data points covering a spectral width of 26,324.4 Hz. A 500 MHz Varian INOVA spectrometer equipped with a nanoprobe was used, operating at a carbon frequency of 125.71 MHz at 298 K.

3. Results and discussion

3.1. Pure and folded sugarcane SHsp101 was produced

Members of the ring-shaped Clp/Hsp100 AAA+ family capable of reverting protein aggregation have been identified in several organisms such as bacteria (*E. coli* ClpB [11,31] and *Thermus thermophilus* ClpB [32]) fungi (*Sacharomicies cerevisiae* Hsp104 [33]) and plants (*Arabidopsis thaliana* [16,17]). The presence of a gene coding for an Hsp100 protein in sugarcane was determined by sequencing and annotating sugarcane ESTs [34]. To characterize the protein encoded by this gene, we cloned and sequenced the entire cDNA encoding for sugarcane Hsp101, SHsp101 (Fig. 1B). The cDNA was cloned into a pET28a vector for expression as a His-tagged construct. SHsp101 has 912 residues and a predicted molecular mass of 100.9 kDa for the monomer (Fig. 1B). This protein was named SHsp101 because it is 84% identical to AtHsp101 from *A. thaliana* (Supplementary material (Fig. S1)). Due to the His-tag tail, the recombinant monomeric construct has a predicted molecular mass of 103.1 kDa (Fig. 1B). A model for the structure of SHsp101 (Fig. 1C) was prepared based on the crystal structure of *T. thermophilus* ClpB (PDB entry number 1QVR) [7], which shares 50% identity with sugarcane Hsp101. The middle domain (M-domain) forms a coiled coil that is inserted between the NBD1 and NBD2 domains and points away from the body of the chaperone, forming propeller blades in the hexameric model [5,7]. The NBD domains bind nucleotides. The two tryptophans, whose emitted fluorescence was studied (see

below), are highlighted in red. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)

Approximately 50% of the expressed SHsp101 was soluble in the supernatant of the cell lysate, as determined by densitometric quantification of the bands in Coomassie Blue-stained SDS-PAGE gels (data not shown). The recombinant fused protein was purified using a HiTrap Chelating affinity column (Amersham Pharmacia Biotech) followed by a HiLoad Superdex 200 pg 26/60 molecular exclusion column (Amersham Pharmacia Biotech). In the first step of purification, SHsp101 was approximately 85% pure (Fig. 2, lane 2; Table 1). After gel filtration, the protein was approximately 98% pure (Fig. 2, lane 3). The yield of the recombinant SHsp101 was approximately 42 mg/L of culture (Table 1), as measured by both the Bradford method and the absorbance at 280 nm.

The circular dichroism (CD) spectrum (Fig. 3A), which had minima at 208 and 222 nm with signals of approximately $-17,000$ and $-15,000$ deg cm^2 dmol^{-1} , respectively, indicative of an α -helical protein (Table 2), revealed that the expressed sugarcane Hsp101 was folded. The amount of helix predicted from the signal at 222 nm [23] was approximately 40%, which is in good agreement with the ClpB homologs [7,35]. SHsp101 was also studied by CD in the presence and absence of both ADP and ATP. However, the spectra in the presence of nucleotides were undistinguishable from that in the absence of nucleotides (data not shown).

Emission fluorescence spectroscopy of Trp can be used to gain information about the environment surrounding this residue because of its sensitivity to the polarity of the environment. SHsp101 contains two Trp residues, which are both located in the second nucleotide-binding domain (Fig. 1B and C). The emitted fluorescence spectrum of the protein under native conditions exhibited maximum intensity at approximately 351 nm with a spectral center of mass at approximately 357 nm (Fig. 3B and Table 2). Under denaturing conditions (6 M Gdm-Cl), the

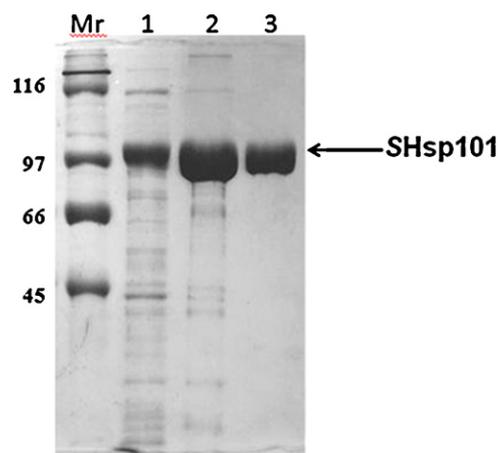


Fig. 2. SHsp101 purification followed by SDS-PAGE. Mr, molecular mass marker (masses in kDa on the left); lane 1, soluble fraction from bacterial lyses; lane 2, from the affinity chromatography step; lane 3, from the gel filtration chromatography step. The arrow indicates the purified SHsp101.

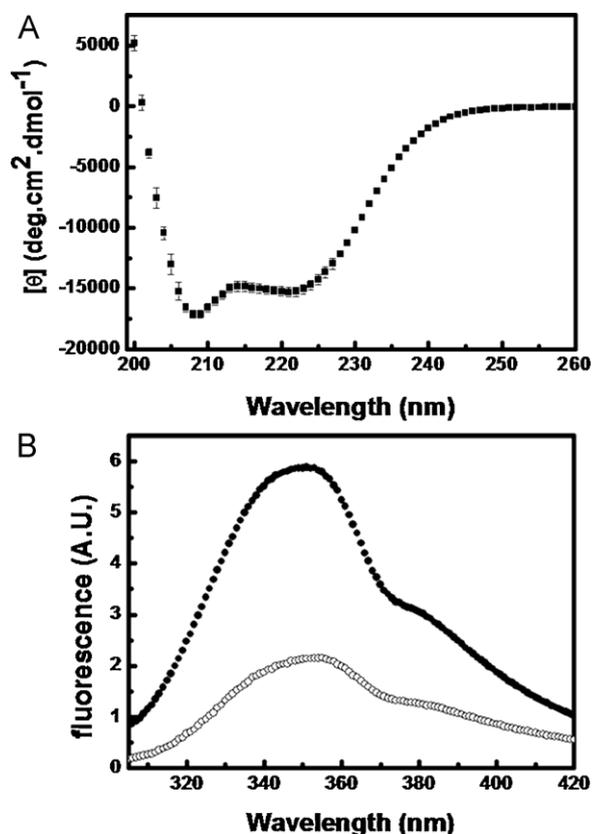


Fig. 3. Spectroscopic measurements indicate that the produced SHsp101 was folded. Experiments were performed in a buffer containing 25 mM Tris-HCl (pH 7.5), 500 mM NaCl and 1 mM 2- β -mercaptoethanol at 25 °C. (A) Circular dichroism. Residual molar ellipticity $[\theta]$ was measured from 200 to 260 nm, with minima observed at 208 and 222 nm with a signal of approximately $-17,000$ and $-15,000$ deg cm² dmol⁻¹, respectively, indicative of an alpha helical protein. The amount of helix predicted from the signal at 222 nm [24] is approximately 40%. (B) Emission fluorescence. Sugarcane Hsp101 (closed circle) and sugarcane Hsp101 in the presence of 6 M guanidinium chloride (open circle) emitted fluorescence spectra, which were measured with excitation at 295 nm and emission from 305 to 420 nm. A.U., arbitrary units.

emitted fluorescence spectrum of SHsp101 was of approximately three times less intense with a maximum intensity at approximately 356 nm and a center of mass at approximately 361 nm (Fig. 3B). Because both Trps contributed to the emitted spectrum, at least one Trp is at least partially buried in the interior of the protein. As shown by CD spectroscopy, the presence of either ADP or ATP

did not cause any relevant modification in the emitted fluorescence spectrum (data not shown).

3.2. Sugarcane Hsp101 is a hexamer that binds nucleotides

The oligomerization state is crucial for proteins from the Clp/Hsp100 AAA+ family. Oligomerization into hexamers is essential for function and to form the characteristic quaternary structure of a ring with an axial pore or channel [7,36–38]. To characterize the oligomeric state of SHsp101, a series of hydrodynamic experiments were conducted. First, analytical ultracentrifugation (AUC) sedimentation equilibrium of SHsp101 was carried out to obtain direct information on the oligomeric state of the protein. The molecular mass of a sedimenting particle is derived independently of the sedimentation and diffusion coefficients and is obtained by fitting the concentration distribution of the macromolecules at equilibrium. Fig. 4A shows that the best fit for the SHsp101 sedimentation equilibrium data was to a single species with a molecular mass of 620 ± 20 kDa (Fig. 4A and Table 2), which is in good agreement with an expected value of 618 kDa predicted for the hexameric species (Table 2). Values of the sedimentation coefficient $s_{20,W}^0$ (Fig. 4B, top) and diffusion coefficient $D_{20,G}^0$ (Fig. 4b, bottom) of 15.2 S and 2.4×10^{-7} cm²/s, respectively, were determined by extrapolation. These results agreed well with those of $s_{20,W} = 16$ –17 S for ClpB [35,39] and $s = \sim 16$ S and $D = 2.5 \times 10^{-7}$ cm²/s for Hsp104 [37]. The values of $s_{20,W}^0$ and $D_{20,W}^0$ were used in Eq. (2) (see Section 2) to estimate a molecular mass of 611 kDa for SHsp101 (Table 2). Multi-angle laser light-scattering in conjunction with size exclusion chromatography (SEC-MALLS) experiments were used to determine the molecular mass and oligomeric state of SHsp101. Fig. 4C shows a plot of the calculated molecular masses versus the elution time for the selected scattering peaks, which correspond to the UV-monitored elution peaks. An average molecular mass of 619 ± 6 kDa was obtained (Fig. 4C and Table 2), confirming that SHsp101 is a single species with the molecular mass of a hexamer. The masses measured by AUC and SEC-MALS experiments are equal to that predicted for a hexameric species, which is in good agreement with data obtained for other Clp/Hsp100 proteins such as *E. coli* [39] and *T. thermophilus* [7] ClpBs and yeast Hsp104 [36,37].

The AUC experiments were also performed in the presence of both ADP and ATP nucleotides without any measurable change in the oligomeric state of the protein. An oligomeric state that is independent of nucleotide binding has been shown for Hsp104 [40] but not for ClpB [36]. In addition to that, because both the sedimentation (s) and diffusion (D) coefficients are in very good agreement with those measured for other members of Clp/Hsp100 family [35,37,39], they likely share at least some global quaternary structure similarity. The values obtained for the hydrodynamic coefficients and molecular masses can be used to provide insight on the shape of the protein by the so-called Perrin or shape factor F [41]. The frictional coefficient for a rigid sphere having an identical volume as the molecule of interest (f_0) and the frictional coefficient (f) are obtained from measurements via the Stokes equation [42]. The closer the Perrin factor (f/f_0 ratio) is to 1, the more spherical the protein is. Our results indicate that SHsp101 has a Perrin factor of approximately 1.6, which is in good agreement with the values predicted for Hsp104 [37] and ClpB [43]. This analysis may have some limitations because high flexibility and mobility are characteristics of proteins belonging to the Clp/Hsp100 family [43].

Because ADP or ATP appear not to have measurable effects on the secondary structure, on the environment of the Trp residues or on the oligomeric state of SHsp101, we next investigated the interaction of the nucleotides with the protein with a saturation transfer

Table 2
SHsp101 biophysical parameters.

Circular dichroism at: -208 nm	$-17,000$ deg cm ² dmol ⁻¹
-222 nm	$-15,000$ deg cm ² dmol ⁻¹
Emitted Trp fluorescence spectrum:	
-maximum intensity	351 nm
-center of mass	357 nm
Sedimentation coefficient ($s_{20,W}^0$) from analytical ultracentrifugation sedimentation velocity	15.2 S
Diffusion coefficient ($D_{20,W}^0$) from dynamic light scattering	2.4×10^{-7} cm ² /s
Molecular mass based on: -Amino acid content for a hexamer	618.6 kDa
-AUC equilibrium sedimentation	616 kDa
$-s_{20,W}^0$ and $D_{20,W}^0$ using Eq. (2)	611 kDa
-SEC-MALLS	619 kDa

Errors are 4% or less (see text). $s_{20,W}^0$ and $D_{20,W}^0$ were measured by AUC velocity sedimentation and DLS, respectively.

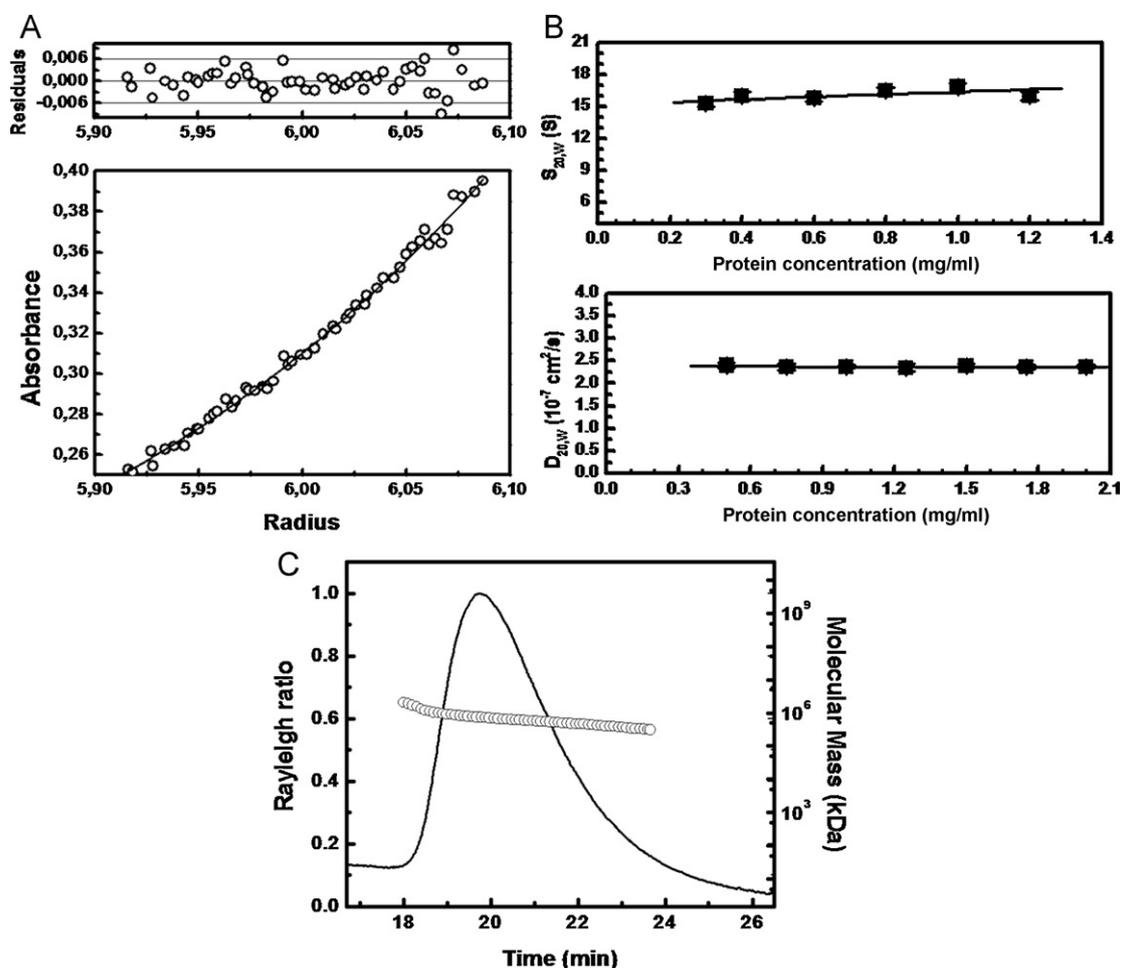


Fig. 4. SHsp101 is a hexamer. Experiments were performed in a buffer containing 25 mM Tris-HCl (pH 7.5), 500 mM NaCl and 1 mM 2- β -mercaptoethanol. (A) Analytical ultracentrifugation sedimentation equilibrium. The sedimentation equilibrium analyses were performed from 0.3 to 1 mg/mL and the best fit was determined by the randomness of the distribution of the residuals and by minimization of the variation of the variance. The figure shows the best fit of the experimental data at 3000 rpm at 20 °C to a molecular mass of 620 ± 20 kDa. Additional data and fitting are shown in the [Supplementary material](#) (Fig. S2). All data agree with a molecular mass of 620 ± 20 kDa and have a variance of approximately 10^{-5} . (B) Apparent sedimentation coefficient $s_{20,W}^0$ (top) and apparent diffusion coefficient $D_{20,W}^0$ (bottom) as a function of protein concentration. The values of $s_{20,W}^0$ and $D_{20,W}^0$ were determined by extrapolation to be 15.2 S and 2.4×10^{-7} cm²/s, respectively. A molecular mass of 611 kDa was calculated for SHsp101 using these values (from Eq. (2); see Section 2). (C) SEC-MALS. The protein was analyzed using a MALLS detector during elution with size exclusion chromatography, and the major peak presented a polydispersity of $1.1 \pm 0.2\%$ and an average molecular mass of 619 ± 6 kDa.

difference (STD) NMR experiments, which are capable of detecting weak binding with great sensitivity [44]. The STD experiments for protein–ligand systems are based on magnetization transfer from protein to ligands and rely upon the difference between two experiments: on-resonance and off-resonance. The signal of the protein is first saturated by a train of selective pulses in a range where only protein signals exist. When a protein becomes saturated, ligands that are in exchange between bound and free forms become saturated when bound to the protein. This mechanism happens via spin diffusion, a process where the saturation propagates across the hydrogens in a protein by H–H intramolecular dipolar interactions. The saturation transfer to the ligands occurs through intermolecular cross relaxation at the protein–ligand interface. Because of the long T_1 (relaxation time), these ligands remain in a saturated state when they are free. The off-resonance experiment consists of an identical pulse train but in a range where there are not protein or ligands signals. By subtraction of these two spectra, the obtained spectrum contains only the hydrogen signals belonging to molecules that bind to the protein. Resonance signals from nonbinders do not show up in the difference spectrum. Also, the degree of saturation of the individual hydrogens of a small ligand molecule reflects their proximity with the protein surface [45].

The STD experiments showed that nucleotides (ATP and ADP) bind to SHsp101. The STD NMR spectra for SHsp101 with ATP γ S, ATP, and ADP showed very similar profiles (data not shown). Fig. 5A shows the spectrum obtained for Hsp100/ATP γ S, which is similar to the other nucleotide interactions. The observed signals were attributed to the following hydrogen atoms: purine H-8 and H-2 and ribose H-1'. These experiments all allowed for group epitope mapping (GEM) with values around 100%. The STD NMR experiments provided evidence for the SHsp101–nucleotide interaction with only three interacting hydrogen atoms.

To specify the interactions between the adenosine nucleotides and SHsp101, the heteronuclear correlations between carbon and hydrogen atoms of the ligand in the absence and presence of the chaperone were compared. The NMR HETCOR experiments required the use of a nanoprobe due the low sample amount. Because of the possible ATPase activity of Hsp101, only non-hydrolyzable ATP γ S was monitored, as shown on Fig. 5B. All of the purine carbon atom signals, C₀5, C₈, C₀4, C₀6 and C₂, and only C1' and C2' of the ribofuranose showed different chemical shifts upon binding. The H-2 and H-8 hydrogen signals showed more pronounced chemical shift changes ([Supplementary material Table S1](#)). In conclusion, SHsp100 binds to ATP, ATP γ S and ADP by associating with their adenosine and partly interacting with the ribose

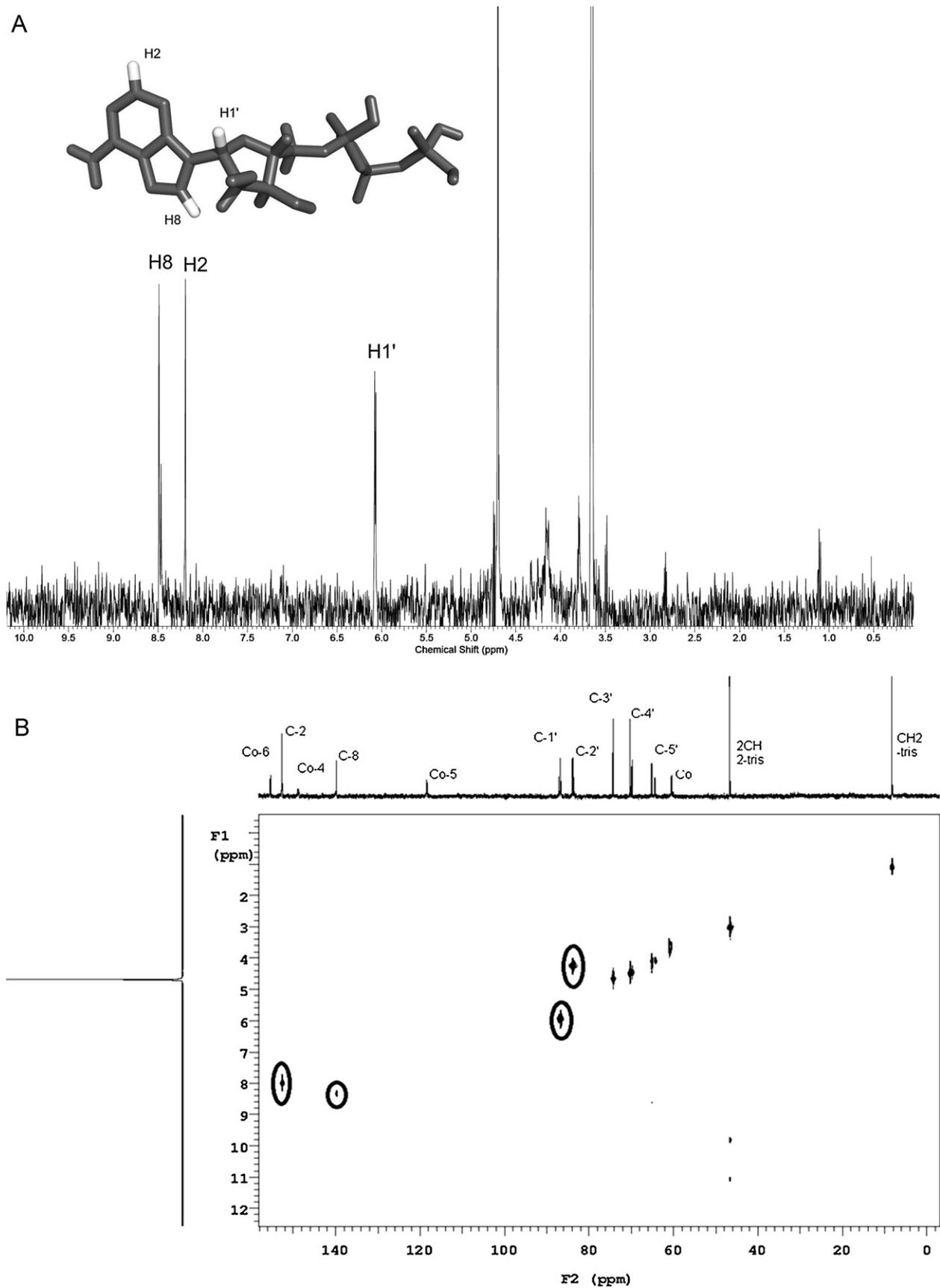


Fig. 5. SHsp101 binds nucleotides. (A) STD NMR spectrum. The STD NMR spectrum of a sample containing ATP- γ S (1 mM) and SHsp101 (10 μ M) showed three interacting hydrogen atoms and that the signals for H-1', H-2 and H-8 were \sim 100% saturated. The ligand-interacting hydrogen atoms of the ATP- γ S ribbon structure are shown in white. The same was observed in other two cases (SHsp101 with ATP or ADP). (B) HETCOR NMR spectra: a) ATP- γ S free and b) ATP- γ S in the presence of SHsp101. The ATP- γ S signals, which are correlations $\{^{13}\text{C}, ^1\text{H}\}$ that changed when in contact with the protein (SHsp101), are circled.

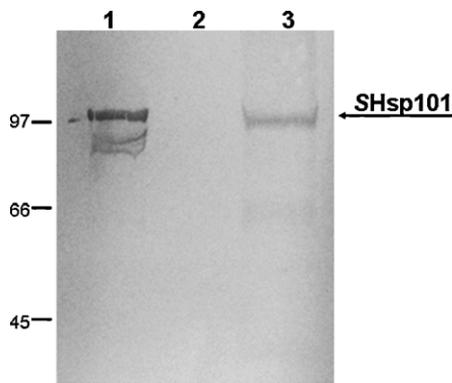


Fig. 6. SHsp101 is expressed in sugarcane. SHsp101 was detected in sugarcane extracts (leaf) by immunoblotting. Lane 1: purified recombinant sugarcane SHsp101 (35 ng). Lane 2: buffer alone, control. Lane 3: protein extract (15 µg) from sugarcane leaf. The proteins were subjected to SDS-PAGE and transferred to a nitrocellulose membrane. The blot was probed with an antiserum (1:500) anti-recombinant SHsp101, and the immunological complex was detected with AP-conjugated anti-rabbit IgG (BioRad).

moieties and three hydrogens H-2, H-8 and H-1' were most affected by binding to SHsp101. Changes in the chemical shifts of carbons C-2', C1', C2, C₀-4, C₀-5, C₀-6, and C-8 were also observed. Therefore, the phosphate ATPγS region does not appear to interact with the protein. Because the quaternary carbon atoms are not directly hydrogen bonded (because these carbon atoms are bound to four other carbons or atoms), their chemical shift changes were not observed in the HETCOR experiments. However, the $\delta^{13}\text{C}$ chemical shifts were changed, as shown in Table S1 for C₀-4, C₀-5 and C₀-6.

ATP is very important to the function of Clp/Hsp100 proteins as the ability to disrupt protein aggregates depends on nucleotide binding and hydrolysis [5]. Although both nucleotide binding sites, NBD1 and NBD2, are essential for function, they appear to play different roles. For example, NBD1 seems to be necessary for substrate binding whereas NBD2 is not [37]. Also, an asymmetric ATPase activity between the domains seems to be an essential feature for substrate remodeling [46]. Our results revealed that SHsp101 binds nucleotides and the effects caused by this binding will be further investigated.

4. Sugarcane expresses Hsp101

The presence of mRNA encoding for SHsp101 was confirmed by the EST sugarcane genome project [34,47]. Herein, we aimed to determine if the SHsp101 protein is normally expressed in sugarcane cells. Serum against the purified recombinant SHsp101 was used in western blotting analysis to assess the presence of SHsp101 in sugarcane cell extracts. SDS-PAGE was loaded with recombinant SHsp101 in one lane and protein extracts from leaf tissues of sugarcane without further treatment in another lane. After running, the gel was incubated with serum against recombinant SHsp101 (Fig. 6). The recombinant SHsp101 was labeled as expected; additionally, a protein with an identical molecular mass was observed in the sugarcane extract, strongly indicating that SHsp101 is expressed in sugarcane. The aforementioned findings open up the possibility of further investigations. For instance one should ask whether the expression of SHsp101 is highly increased when sugarcane is under stress conditions. Additionally, yeast Hsp104 is involved in the propagation of the prion factor under non-stresses conditions [48] but little is known about the function of Clp/Hsp100 proteins in plants when these organisms are not stressed.

5. Conclusions

Clp/Hsp100 AAA+ proteins are potentially key players for improving the expression of heterologous proteins in bacterial cells. Although these chaperones are important for biotechnological purposes, little is known on their mechanism of substrate binding and remodeling and why they are absent in the cytosol of metazoan cells. Studies of other orthologs may help to add information to what is already known about ClpB and Hsp104. Here, one of these orthologs, sugarcane Hsp101, was characterized, confirming that this protein belongs to the Clp/Hsp100 AAA+ family and will help provide information on the relationship between the structure and function for these proteins. The initial cloning and characterization of this protein reported provided basic information on the function of SHsp101, such as nucleotide binding and expression. This information may help to unravel the functions of Clp/Hsp100 proteins. Further functional studies are underway and will be published in the near future.

Acknowledgements

We thank Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), Ministério da Ciência e Tecnologia/Conselho Nacional de Pesquisa e Desenvolvimento (MCT/CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) for grants and fellowships. We thank Centro Nacional de Pesquisa em Energia e Materiais (CNPEM) for the use of the LEC-LNBio facility.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijbiomac.2011.08.027.

References

- [1] C.H.I. Ramos, S.T. Ferreira, *Protein Pept. Lett.* 12 (2005) 213–222.
- [2] L.M. Luheshi, D.C. Crowther, C.M. Dobson, *Curr. Opin. Chem. Biol.* 12 (2008) 25–31.
- [3] B. Bukau, J. Weissman, A. Horwich, *Cell* 125 (2006) 443–451.
- [4] A. Tiroli-Cepeda, C.H.I. Ramos, *Protein Pept. Lett.* 18 (2011) 101–109.
- [5] S.M. Doyle, S. Wickner, *Trends Biochem. Sci.* 34 (2008) 40–48.
- [6] J.R. Glover, S. Lindquist, *Cell* 94 (1998) 73–82.
- [7] S. Lee, M.E. Sowa, Y.H. Watanabe, P.B. Sigler, W. Chiu, M. Yoshida, et al., *Cell* 17 (2003) 229–240.
- [8] D.A. Parsell, A.S. Kowal, M.A. Singer, S. Lindquist, *Nature* 373 (1994) 475–478.
- [9] A. Mogk, T. Tomoyasu, P. Goloubinoff, S. Rüdiger, D. Röder, H. Langen, et al., *EMBO J.* 18 (1999) 6934–6949.
- [10] M. Zolkiewski, *J. Biol. Chem.* 274 (1999) 28083–28086.
- [11] P. Goloubinoff, A. Mogk, A.P.B. Zvi, T. Tomoyasu, B. Bukau, *Proc. Nat. Acad. Sci.* 96 (1999) 13732–13737.
- [12] A. Mogk, C. Schlieker, C. Strub, W. Rist, J. Weibezahn, B. Bukau, *J. Biol. Chem.* 278 (2003) 17615–17624.
- [13] M. Haslbeck, A. Miess, T. Stromer, S. Walter, J. Buchner, *J. Biol. Chem.* 280 (2005) 23861–23868.
- [14] K. Liberek, A. Lewandowska, S. Zietkiewicz, *EMBO J.* 27 (2008) 328–335.
- [15] A. Singh, A. Grover, *Plant Mol. Biol.* 74 (2010) 395–404.
- [16] Y.-R.J. Lee, R.T. Nagao, J.L. Key, *Plant Cell* 6 (1994) 1889–1897.
- [17] E.C. Schirmer, S.E. Lindquist, E. Vierling, *Plant Cell* 6 (1994) 1899–1909.
- [18] D.R. Gallie, D. Fortner, J. Peng, D. Puthoff, *J. Biol. Chem.* 277 (2002) 39617–39626.
- [19] A.L. Vettore, F.R. da Silva, E.L. Kemper, P. Arruda, *Genet. Mol. Biol.* 24 (2001) 1–7.
- [20] H. Edelhock, *Biochemistry* 6 (1967) 1948–1954.
- [21] M.M. Bradford, *Anal. Biochem.* 72 (1976) 248–254.
- [22] P.A. Bates, L.A. Kelley, R.M. McCallum, M.J. Stenberg, *Proteins* 5 (2001) 39–46.
- [23] D.H.A. Correa, C.H.I. Ramos, *African J. Biochem. Res.* 3 (2009) 164–173.
- [24] S.E. Harding, D.J. Winzor, in: S.E. Harding, B.Z. Chowdhry (Eds.), *Protein–Ligand Interactions: Hydrodynamics and Calorimetry: Practical Approach*, Oxford University Press, Oxford, 2001, pp. 105–135.
- [25] D.J. Winzor, S.E. Harding, in: S.E. Harding, B.Z. Chowdhry (Eds.), *Protein–Ligand Interactions: Hydrodynamics and Calorimetry: Practical Approach*, Oxford University Press, Oxford, 2001, pp. 74–104.
- [26] J.C. Borges, C.H.I. Ramos, *Curr. Med. Chem.* 18 (2011) 1276–1285.
- [27] P. Schuck, *Biophys. J.* 78 (2000) 1606–1619.

- [28] P. Schuck, M.A. Perugini, N.R. Gonzales, G.J. Howlett, D. Schubert, *Biophys. J.* 82 (2002) 1096–1111.
- [29] P. Schuck, *Anal. Biochem.* 320 (2003) 104–124.
- [30] M.L. Johnson, J.J. Correia, D.A. Yphantis, H.R. Halvorson, *Biophys. J.* 36 (1981) 575–588.
- [31] C.L. Squires, S. Pedersen, B.M. Ross, C. Squires, *J. Bacteriol.* 173 (1991) 4254–4262.
- [32] K. Motohashi, Y. Watanabe, M. Yohda, M. Yoshida, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 7184–7189.
- [33] Y. Sanchez, S. Lindquist, *Science* 248 (1990) 1112–1115.
- [34] J.C. Borges, M.C. Peroto, C.H.I. Ramos, *Gen. Mol. Biol.* 24 (2001) 85–92.
- [35] M.E. Barnett, A. Zolkiewska, M. Zolkiewski, *J. Biol. Chem.* 275 (2000) 37565–37571.
- [36] S. Lee, M.E. Sowa, Y.H. Watanabe, P.B. Sigler, W. Chiu, M. Yoshida, et al., *J. Struct. Biol.* 146 (2004) 99–105.
- [37] B. Bösl, V. Grimminger, S. Walter, *J. Biol. Chem.* 280 (2005) 38170–38176.
- [38] P. Wendler, J. Shorter, D. Snead, C. Plisson, D.K. Clare, S. Lindquist, et al., *Mol. Cell.* 34 (2009) 81–92.
- [39] M. Zolkiewski, M. Kessel, A. Ginsburg, M.R. Maurizi, *Protein Sci.* 8 (1999) 1899–1903.
- [40] D.A. Parsell, A.S. Kowal, S. Lindquist, *J. Biol. Chem.* 269 (1994) 4480–4487.
- [41] C.R. Cantor, P.R. Schimmel, *Size and shape of macromolecules*, in: L.W. McCombs (Ed.), *Biophysical Chemistry, Part II: Techniques for the Study of Biological Structure and Function*, W.H. Freeman and Company, New York, 1980, pp.539–590.
- [42] D.C. Teller, E. Swanson, C. De Haen, *Methods Enzymol.* 61 (1979) 103–124.
- [43] S. Zietkiewicz, M.J. Slusarz, R. Slusarz, K. Liberek, S. Rodziewicz-Motowidlo, *Biopolymers* 93 (2010) 47–60.
- [44] H. Huang, J. Milojevik, G. Melacini, *J. Phys. Chem. B.* 112 (2008) 5795–5802.
- [45] B. Meyer, T. Peters, *Angew. Chem. Int.* 42 (2003) 864–890.
- [46] S.M. Doyle, J. Shorter, M. Zolkiewski, J.R. Hoskins, S. Lindquist, S. Wickner, *Nat. Struct. Mol. Biol.* 14 (2007) 114–122.
- [47] J.C. Borges, T.C. Cagliari, C.H.I. Ramos, *J. Plant Physiol.* 164 (2007) 505–513.
- [48] Y.O. Chernoff, S.L. Lindquist, B. Ono, S.G. Inge-Vechtomov, S.W. Liebman, *Science* 268 (1995) 880–884.
- [49] W.L. DeLano, *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, 2002.