



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

DEPARTAMENTO DE QUÍMICA ORGÂNICA

**Resolução Estrutural de Proteínas Hipotéticas, Chaperonas de
Secreção, da Bactéria *Xanthomonas axonopodis* pv. *citri***

Tese de Doutorado

Aluna: Juliana Fattori

Orientadora: Profa. Dra. Ljubica Tasic

Laboratório de Química Biológica
<http://sites.google.com/site/lqb250/>

16 de setembro de 2011
Campinas – SP – Brasil

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

F269r Fattori, Juliana (1981-).
Resolução estrutural de proteínas hipotéticas,
chaperonas de secreção da bactéria *Xanthomonas
axonopodus* pv. *citri* / Juliana Fattori. – Campinas, SP:
[s.n.], 2011.

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

1. *Xanthomonas axonopodis* pv. *citri*. 2. Chaperonas
de secreção. 3. Caracterização estrutural.
4. Espalhamento de raios-X a baixo ângulo. I. Tasic,
Ljubica. II. Universidade Estadual de Campinas. Instituto
de Química. III. Título.

Informações para Biblioteca Digital

Título em inglês: Structural insights on hypothetical proteins, secretion chaperones from
Xanthomonas axonopodis pv. *citri*

Palavras-chave em inglês:

Xanthomonas axonopodis pv. *citri*
Secretion chaperones
Structural characterization
Small angle X ray scattering

Área de concentração: Química Orgânica

Titulação: Doutor em Ciências

Banca examinadora:

Ljubica Tasic [Orientador]
Letícia Khater Covesi
Mário Tyago Murakami
Anita Jocelyne Marsaioli
Wanda Pereira Almeida

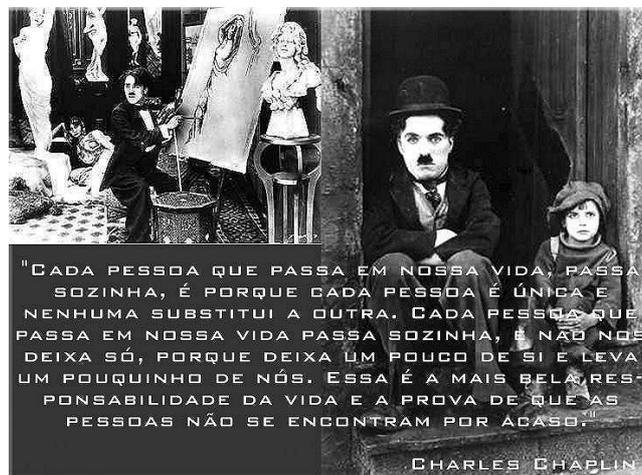
Data de defesa: 16/09/2011

Programa de pós-graduação: Química



A arte de sorrir cada vez que o mundo diz “NÃO”.

Jon Lucien/Guilherme Arantes



Aos meus pais,
Edson e Rosa,
sempre.

Agradecimentos

Antes e acima de tudo agradeço a Deus por tudo que me tem dado, por estar sempre a meu lado e por me conceder força e perseverança necessárias para que eu chegasse até aqui.

Em ordem cronológica, agradeço ao Thiago por, em 2001 e depois em 2003, ter me mostrado tantas portas que poderiam ser abertas neste caminho acadêmico.

Em seguida, na minha iniciação científica, agradeço ao professor Carlos Ramos pela oportunidade de fazer parte do seu grupo de pesquisas em 2003; por me fazer ver que saber inglês era extremamente importante; por me apresentar o imenso mundo dos artigos científicos e por nos levar a ler pelo menos um artigo por semana para as reuniões de grupo. Agradeço também a todos do antigo LBM que me ajudaram bastante; Willian, Veruska, Ana, Letícia, Júlio, Silvia, Kaliandra, Natália e Buba para quem desde aquela época eu dizia brincando “quando eu crescer quero ser igual a você”.

Durante o mestrado, tive oportunidade e experiência únicas de aprendizado. Agradeço imensamente ao professor Luiz Carlos Dias por me orientar neste período, pela oportunidade de fazer parte do seu grupo, pelos ensinamentos e pela contagiante paixão pela Química Orgânica e Síntese Orgânica. O que sei hoje de orgânica devo a ele e seu grupo. Obrigada por confiar em mim na apresentação do BMOS e por me deixar ‘cuidar’ das suas filhas, sou apaixonada pela ‘família Lu’. Agradeço também a todos do LQOS daquela época, que me ensinaram bastante, em especial Carol, Dimas, Airton, Anderson, Moita e Fer por serem para mim exemplos de disciplina, seriedade e comprometimento com a pesquisa.

Mais recentemente, durante o doutorado também tive ótimas experiências. Agradeço muito a professora Ljubica pela orientação e oportunidade, pelos ensinamentos, pela amizade, compreensão, pela sutileza nas broncas, por me dar espaço para aprender e crescer profissionalmente junto com seu grupo. Obrigada por tudo Buba, como você mesma disse nos meus primeiros dias no laboratório, você seria como uma segunda mãe. Agradeço a todos do laboratório de Química Biológica ou do I-250, não vou citar nomes para não correr o risco de esquecer alguém, mas agradeço a todos sem distinção, aos que já saíram e aos que acabaram de entrar, pela convivência, pelas ajudas e por aguentarem minhas chatices! Agradeço em especial a Junko por me divertir com sua sinceridade; ao Fabio pelas nossas valiosas discussões científicas, desde as teóricas às de ordem prática e pelas tantas figuras feitas no Corel; minha querida Almas, por estar sempre tentando me animar; ao Leandro ou pequena criança pela ajuda com o nosso artigo; ao professor Nelson por sempre se interessar pelo andamento do meu projeto, indicar artigos e ajudar nas questões práticas; a Iza pelas deliciosas tortas africanas; a Stephany pelo estojó de bailarina que

ganhei e aos que em algum dia me deram uma “caroninha” sempre bem vinda. E, à Marcelita por tudo, por às vezes ser um pouco minha mãe e em outras minha amiga!

Agradeço muito ao professor Ricardo Aparicio, pela colaboração, pelas idas ao LNLS para as medidas de SAXS; pelas discussões e auxílio sempre pronto na confecção de figuras. Agradeço também aos professores Cláudio Tormena e Alviclér Magalhães pelas ajudas no RMN; ao grupo do professor Carlos Ramos por sempre estarem prontos a ajudar; professor Fabio Gozzo e seu grupo, em especial Alana, Alexandre e Pilau pelo super auxílio com MS; ao pessoal do Laboratório de RMN, Anderson, Sonia e Paula por sempre me ajudarem e a todo pessoal do IQ, professores, técnicos, pessoal da limpeza e zeladoria.

Agradeço aos professores Luzia Koike, Reis, Paulo Miranda e Pili pela oportunidade de ser PED deles e ter aprendido um pouco sobre a arte de ensinar com eles. Agradeço aos professores Luciana Oliveira, Paulo Miranda e Fernando Coelho pelos conselhos no exame geral, sempre achei que escrevia melhor do que falava. Agradeço aos professores Wanda, Fábio Gozzo e Néelson Durán pelas valiosas sugestões no exame de qualificação; foram ótimas contribuições para a finalização do trabalho. Também gostaria de agradecer a todos os membros da banca da defesa, inclusive os suplentes, pela disponibilidade e por aceitarem o convite prontamente.

Agradeço a FAPESP pelo auxílio financeiro; ao Instituto de Química pela excelente infra-estrutura que possibilitou a realização do trabalho e que tem sido minha segunda casa há uns 10 anos; e agradeço o LNLS e LNBio também pela infra-estrutura.

Agradeço ao Gabri por me ajudar a ser uma pessoa menos nervosa e me ensinar que pra tudo a gente dá um jeitinho. Obrigada a todos os meus amigos que sempre me ajudaram: Quel, Robson, Marcelita, Dimas, Carol, Junkito, Almas, Sil, Ciça, Moita, Walmir, Gustavo e Paula. Valéria, Cláudia e PH pelo auxílio profissional. Luana e Luiza, Aninha, Alek, Juninho, Veri, Tatá, Bibica, Davi e todos os meus amores do Boldrini por existirem, fazerem ou terem feito parte da minha vida e por darem um colorido aos meus dias.

Por último, porém mais importante agradeço a minha família. Meus pais, dois anjos da guarda que Deus me deu e a minha irmã por sempre dar alegria a nossa casa. Agradeço todos os dias simplesmente pelo fato de vocês existirem, serem quem são, e fazerem parte da minha vida. Pai e mãe sempre me dando o suporte e auxílio mais que necessários, vou copiar uma frase da Dri que achei extremamente apropriada, vocês são co-autores desta conquista. Obrigada aos três por tudo.

Curriculum Vitae

Juliana Fattori

Formação Acadêmica

- 2008-2011 Instituto de Química – UNICAMP
Doutorado em Ciências
Projeto: Resolução estrutural de proteínas hipotéticas, chaperonas de secreção, da bactéria *Xanthomonas axonopodis* pv. *citri*
Orientadora: Profa. Dra. Ljubica Tasic
Agência Financiadora: FAPESP (08/50426-6)
- 2005-2007 Instituto de Química – UNICAMP
Mestrado em Química
Projeto: Síntese de 2-amino-3,5-dióis
Orientador: Prof. Dr. Luiz Carlos Dias
Agência Financiadora: CNPq
- 2000-2004 Universidade Estadual de Campinas – UNICAMP
Bacharelado em Química
Iniciação Científica: (Laboratório Nacional de Luz Síncrotron)
Projeto: Estudo de mutantes envolvidos na propensão de formar alfa-hélices em mioglobina de espermacete para entender sua diferença de estabilidade em relação ao seu homólogo em cavalo.
Orientador: Prof. Dr. Carlos Henrique Inácio Ramos
Agência Financiadora: FAPESP (03/04300-7)

Publicações em Periódicos

1. Fattori, J.; Prando, A.; Assis, L.H.P.; Aparício, R.; Tasic, L. “Structural Insights on Two Hypothetical Secretion Chaperones from *Xanthomonas axonopodis* pv. *citri*” *Protein J* **2011**, *30*, 324-333.
2. Fattori, J.; Prando, A.; Martins, A.M.; Rodrigues, F.H.S., Tasic, L. “Bacterial Secretion Chaperones” *Protein Pept Lett* **2011**, *18*, 158-166.
3. Azevedo, M.B.M.; Tasic, L.; Fattori, J.; Rodrigues, F.H.S.; Cantos, F.C.; Ribeiro, L.P.; Paula, V.; Ianzer, D.; Santos, R.A.S. “New formulation of an old drug in

hypertension treatment: the sustained release of captopril from cyclodextrin nanoparticles” *Int J Nanomedicine* **2011**, *6*, 1005-1016.

4. Dias, L.C.; Fattori, J.; Perez, C.C. “An Efficient Procedure for the Synthesis of 2-N-Boc-amino-3,5-Diols” *Tetrahedron* **2008**, *64*, 5891-5903.
5. Dias, L.C.; Fattori, J.; Perez, C.C. “Addition of allyltrichlorostannanes to aldehydes: application in the synthesis of 4-N-Boc-amino-3-hydroxy ketones” *Tetrahedron Lett.* **2007**, *49*, 557-561.
6. Regis, W.C.B.; Fattori, J.; Santoro, M.M.; Jamim, M.; Ramos, C.H.I. “On the difference in stability between horse and sperm whale myoglobins” *Arch Biochem Biophys* **2005**, *436*, 168-177.

Resumos em Congressos de maior relevância

1. Fattori, J.; Prando, A.; Aparicio, R.; Tasic, L. “SAXS and NMR Analyses of Hypothetical Secretion Chaperones” XXXIX Annual Meeting of Brazilian Biochemistry and Molecular Biology Society, 2010, Foz do Iguaçu, PR, Brasil.
2. Fattori, J.; Tasic, L. “Structural studies on secretion chaperones from *Xanthomonas axonopodis* pv. *citri*” I Congreso Iberoamericano de Química, Bioquímica e Ingeniería Química, 2009, Havana, Cuba.
3. Fattori, J.; Stevanatto, K.B.; Krogh, R.; Andricopulo, A.D.; Dias, L.C. “Synthesis and biological studies on sphingosine analogues” The 4th Brazilian Symposium on Medicinal Chemistry, 2008, Porto de Galinhas, PE, Brasil.
4. Fattori, J.; Dias, L.C. “A short and efficient synthesis of 2-N-Boc-amino-3,5-diols” 12th Brazilian Meeting on Organic Synthesis, 2007, Itapema, SC, Brasil.
5. Fattori, J.; dos Santos, T.L.; Dias, L.C. “Síntese de 2-amino-3,5-dióis” 30^a Reunião da Sociedade Brasileira de Química, 2007, Águas de Lindóia, SP, Brasil.
6. Fattori, J.; Dias, L.C. “Synthesis of 2-amino-3,5-diol derivatives” The 3rd Brazilian Symposium on Medicinal Chemistry, 2006, São Pedro, SP, Brasil.

Apresentação oral em Congressos

1. Fattori, J.; Perez, C.C.; Dias, L.C. “An efficient procedure for the synthesis of 2-N-Boc-amino-3,5-diols” 12th Brazilian Meeting on Organic Synthesis, 2007, Itapema, SC, Brasil.

RESUMO

RESOLUÇÃO ESTRUTURAL DE PROTEÍNAS HIPOTÉTICAS, CHAPERONAS DE SECREÇÃO, DA BACTÉRIA *Xanthomonas axonopodis* pv. *citri*.

O presente estudo teve por objetivo a caracterização estrutural de quatro possíveis chaperonas de secreção da bactéria *Xanthomonas axonopodis* pv. *citri* (*Xac*). Esta caracterização é importante para entender as funções destas proteínas e conseqüentemente compreender os mecanismos de virulência da *Xac* que é a causadora do cancro cítrico. Das quatro proteínas estudadas, a XACb0033 é considerada uma possível chaperona de secreção do sistema de secreção do tipo IV; a XAC0419 e XAC1346 são possíveis chaperonas de secreção do sistema de secreção do tipo III e a XAC1990, foi identificada em 2007 como uma proteína do sistema flagelar (FlgN). As quatro proteínas, clonadas em pET23a, foram expressas em *E. coli*, purificadas e obtidas enoveladas conforme as análises de dicroísmo circular (CD). Por CD também se observou um alto conteúdo helicoidal na estrutura secundária das proteínas. Elas foram caracterizadas por filtração em gel analítica, espectrometria de massas, técnicas de difusão de ressonância magnética nuclear (DOSY-NMR) e espalhamento de raios-X a baixo ângulo (SAXS). Por SAXS obteve-se um modelo do envelope molecular de duas proteínas que foi condizente com a estrutura tridimensional obtida por bioinformática. Baseando-se nos resultados obtidos foi possível especular a classificação da XAC0419 como provável chaperona secretória da classe I e a FlgN (XAC1990) foi classificada como uma chaperona flagelar (classe III); e para a XACb0033 observou-se similaridade estrutural com a VirE1, que é a única chaperona secretória do sistema de secreção do tipo IV conhecida até o momento.

ABSTRACT

STRUCTURAL INSIGHTS ON HYPOTHETICAL PROTEINS, SECRETION CHAPERONES, FROM THE BACTERIUM *Xanthomonas axonopodis* pv. *citri*.

The aim of this study was the structural characterization of four possible secretion chaperones from the bacterium *Xanthomonas axonopodis* pv. *citri* (*Xac*). This characterization is very important to better understand the function of such proteins and the virulence mechanisms from *Xac*, which is the bacterium responsible for provoking the citrus canker. Among the target proteins, XACb0033 is a possible secretion chaperone from type IV secretion system, XAC0419 and XAC1346 are possible secretion chaperones from type III secretion system and XAC1990 was earlier identified as a flagellar protein (FlgN). All four target proteins, cloned into pET23a, were expressed in *E. coli*, purified and found folded as observed in the circular dichroism analyses (CD). It was also verified (CD experiments) that these proteins have a high helical content. These proteins were further characterized by analytical gel filtration, mass spectrometry (MS), diffusion techniques in nuclear magnetic resonance (DOSY-NMR) and small angle X-ray scattering (SAXS). Molecular envelopes were built for two target proteins applying the SAXS data. These envelopes were in good agreement with the 3D structures predicted by bioinformatics. Finally, XAC0419 was considered a possible class I secretion chaperone based on the obtained results. XAC1990 (FlgN) was confirmed as a flagellar chaperone (class III), while obtained results for XACb0033 pointed structural similarity with VirE1, the unique type IV secretion system chaperone known so far.

Sumário

LISTA DE ABREVIATURAS E CONVENÇÕES.....	xvii
LISTA DE TABELAS	xxi
LISTA DE FIGURAS	xxiii
Capítulo 1 – Introdução	1
1.1 MOTIVAÇÃO – CONSIDERAÇÕES GERAIS	3
1.1.1. <i>Xanthomonas axonopodis</i> pv. <i>citri</i> (Xac).....	8
1.1.2. <i>Sistemas de Secreção</i>	11
1.1.3. <i>Chaperonas de Secreção</i>	17
1.2 ESTUDOS ESTRUTURAIIS DE PROTEÍNAS.....	29
1.2.1. <i>Técnicas auxiliares</i>	31
1.2.2. <i>NMR-DOSY</i>	45
1.2.3. <i>SAXS</i>	47
Capítulo 2 – Objetivos.....	61
Capítulo 3 – Parte Experimental.....	65
3.1 EXTRAÇÃO DO DNA PLASMIDIAL (LISE ALCALINA)	67
3.2 PREPARAÇÃO DE CÉLULAS COMPETENTES	68
3.3 TRANSFORMAÇÃO POR CHOQUE TÉRMICO	69
3.4 EXPRESSÃO PROTÉICA	70
3.5 ELETROFORESE EM GEL	71
3.6 LISE CELULAR	73
3.7 TRATAMENTO COM URÉIA (REENOVELAMENTO).....	74
3.8 PURIFICAÇÃO DAS PROTEÍNAS	75
3.9 DETERMINAÇÃO DA CONCENTRAÇÃO DE PROTEÍNAS (UV/VIS).....	78
3.10 MÉTODO DE BRADFORD	79
3.11 ESPECTROMETRIA DE MASSAS (MS).....	80
3.12 FLUORESCÊNCIA DE EMISSÃO	83
3.13 DICROÍSMO CIRCULAR (CD)	83
3.14 ESPALHAMENTO DE LUZ DINÂMICO (DLS).....	84

3.15 RESSONÂNCIA MAGNÉTICA NUCLEAR (NMR)	84
3.16 ESPALHAMENTO DE RAIOS-X A BAIXO ÂNGULO (SAXS)	86
3.17 BIOINFORMÁTICA	88
3.18 CRISTALOGRAFIA	90
Capítulo 4 – Resultados e Discussão	95
4.1 XAC1346	95
4.2 XAC0419	116
4.3 XACb0033	137
4.4 XAC1990: FlgN	162
Capítulo 5 – Conclusão e Perspectivas	185
5.1 XAC1346	185
5.2 XAC0419	187
5.3 XACb0033	189
5.4 XAC1990: FlgN	192
ANEXOS	197
TABELAS DAS CONDIÇÕES DE CRISTALIZAÇÃO	199
ARTIGO DE REVISÃO SOBRE CHAPERONAS DE SECREÇÃO	207
ARTIGO COM RESULTADOS OBTIDOS PARA XAC0419 E FlgN	217

Lista de Abreviaturas e Convenções

ATP: adenosina 5'-trifosfato

BLAST: ferramenta de busca por alinhamento (do inglês *Basic Local Alignment Search Tool*)

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

CCD: Dispositivo de Carga Acoplada (do inglês *Charge-Coupled Device*)

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

Da: Dalton

Dbppste: sequência de pulso de DOSY (do inglês *DOSY Bipolar Pulse Pair Stimulated Echo Experiment*)

DLS: Espalhamento de Luz Dinâmico (do inglês *Dinamic Light Scattering*)

DOSY: Espectroscopia orientada por Difusão (do inglês *Diffusion Ordered Spectroscopy*)

DTT: ditioneitol

EDTA: ácido etilenodiamino tetra-acético

FPLC: cromatografia líquida de proteínas por pressão (do inglês *Fast Protein Liquide Chromatography*)

HEPES: ácido 4-(2-hidroxietil)-1-piperazinaetanosulfônico

IPTG: isopropil β -D-1-tiogalactopiranosídeo

Jeffamina: polieteramina

k_{av} : coeficiente de partição

LB: Luria Bertrani

LC-MS/MS: Espectrometria de Massas acoplada à cromatografia líquida (do inglês *Liquid chromatography - Mass Spectrometry*)

LNLS: Laboratório Nacional de Luz Síncrotron

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

MES: ácido 2-(N-morfolino)etanosulfônico

MOPS: ácido 3-(N-morfolino)propanosulfônico

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

MS/MS: experimentos de espectrometria de massas em sequência

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

PDB: banco de dados de proteínas (do inglês *Protein Data Bank*)

PEG: polietilenoglicol

PFGSE: Gradiente de campo pulsado de eco de spin (do inglês *Pulse Field Gradient Spin Echo*)

pI: ponto isoelétrico

pv.: pathovar

R_f: fator de retenção

rpm: rotações por minuto

SAXS: Espalhamento de raios-X a baixo ângulo (do inglês *Small Angle X-Ray Scattering*)

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

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*)

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

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

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

UV/VIS: Espectroscopia no ultravioleta/visível

VC: volume de coluna

Xac: *Xanthomonas axonopodis* pv. *citri*

WATERGATE: sequência de pulso para supressão de água em NMR (do inglês *Water Suppression by Gradient-Tailored Excitation*)

Lista de Tabelas

Tabela III-1. Soluções utilizadas na lise alcalina.....	68
Tabela III-2. Soluções utilizadas na preparação de células competentes.....	69
Tabela III-3. Soluções utilizadas na transformação bacteriana.....	70
Tabela III-4. Solução estoque dos antibióticos e IPTG.....	71
Tabela III-5. Soluções tampão utilizadas no preparo de amostras.....	72
Tabela III-6. Solução de acrilamida utilizada na preparação dos géis.....	73
Tabela III-7. Géis de acrilamida (15%).....	73
Tabela III-8. Soluções corante e descorante.....	73
Tabela III-9. Soluções utilizadas na lise celular.....	74
Tabela III-10. Soluções utilizadas no processo de reenovelamento.....	75
Tabela III-11. Soluções para calibração da filtração em gel analítica.....	77
Tabela III-12. Soluções para calibração da filtração em gel analítica II.....	78
Tabela III-13. Reagente de Bradford preparado.....	80
Tabela III-14. Soluções utilizadas na preparação das amostras para MS.....	83
Tabela III-15. Soluções tampões utilizadas em SAXS.....	87
Tabela III-16. Soluções tampão para cristalização.....	91
Tabela III-17. Soluções precipitantes para cristalização.....	91
Tabela III-18. Soluções salinas para cristalização	92
Tabela IV-1. Proteínas em estudo, sequência de aminoácidos e massa molar (MM).....	95
Tabela IV-2. Dados para calibração da filtração em gel na Superdex 200 ($V_c=58,3$ mL).....	103
Tabela IV-3. Dados de dicroísmo circular da XAC1346.....	109
Tabela IV-4. Dados da calibração da filtração em gel analítica.....	121
Tabela IV-5. Dados de dicroísmo circular da XAC0419.....	125
Tabela IV-6. Concentração dos padrões utilizados nas medidas de SAXS.....	127
Tabela IV-7. Dados de dicroísmo circular da XACb0033.....	146
Tabela IV-8. Dados dos experimentos de NMR-DOSY.....	153
Tabela IV-9. Dados da calibração do experimento de DOSY.....	155
Tabela IV-10. Dados de dicroísmo circular da FlgN.....	169

Lista de Figuras

- Figura I-1.** Distribuição da produção de laranjas ao redor do mundo em 2005 e a representação de plantas contaminadas com o cancro cítrico.....**04**
- Figura I-2.** Sistemas de secreção de bactérias Gram-negativas (I-VI), onde MI, ME e MH correspondem respectivamente à membrana interna, externa e do hospedeiro. O Sistema de Secreção do tipo VII não está representado na figura, pois é encontrado apenas em bactérias Gram-positivas.....**11**
- Figura I-3.** Representação dos sistemas de secreção do tipo III (T3SS) e do tipo IV (T4SS); e do flagelo da *Xac*..... **13**
- Figura I-4.** Representação de um sistema flagelar (*Escherichia coli*).....**15**
- Figura I-5.** Estruturas de chaperonas de secreção do T3SS da classe I, diméricas. A) Estrutura da SycE de *Yersinia pseudotuberculosis* (PDB:1JYA). B) Estrutura da InvB de *Salmonella* sp. (PDB: 2FM8).....**21**
- Figura I-6.** Estruturas de chaperonas de secreção do T3SS da classe II e III. A) Estrutura da SycD de *Yersinia enterocolitica* (PDB:2VGX), dimérica. B) Estrutura da FliT de *Salmonella* sp. (PDB: 3A7M).....**22**
- Figura I-7.** Estrutura da chaperona de secreção do T4SS, VirE1 (em vermelho) complexada com sua proteína parceira, a VirE2 (em azul; PDB:3BTP).....**24**
- Figura I-8.** Ilustração de espectros de CD de diferentes tipos de estrutura secundária de proteínas, sendo em preto (1) exemplo de estrutura exclusivamente helicoidal, em vermelho (2) de folhas- β antiparalelas, em verde (3) de estrutura randômica e em azul (4) de estrutura desenovelada do colágeno.....**31**
- Figura I-9.** Ilustração do diagrama de energia (Diagrama de Jablonski) para o fenômeno de fluorescência, onde E corresponde a energia, S_0 ao estado fundamental, S_1' ao estado excitado singlete e S_1 ao estado singlete; 1, 2 e 3 referem-se aos estágios do processo de fluorescência descritos no texto.....**36**
- Figura I-10.** Ilustração de um cromatograma de filtração em gel e a relação linear entre o coeficiente de partição e o logaritmo da massa molecular para proteínas globulares.....**40**

Figura I-11. Ilustração de dados de SAXS. A) Curvas de espalhamento, tendo o logaritmo da intensidade em função do vetor de espalhamento. B) Região de Guinier, obtida a partir da curva de espalhamento. C) Curvas da distribuição de distâncias $\{P(r)\}$ em proteínas.....	48
Figura I-12. Ilustração da relação do formato das curvas de SAXS e a forma da molécula. À direita, curvas de espalhamento com logaritmo da intensidade em função do vetor espalhamento (s) e à esquerda as curvas de distribuição de distâncias, tendo as curvas a mesma cor da molécula referida.....	49
Figura I-13. Ilustração da relação da dimensão máxima obtida pela $P(r)$ e a massa molar da proteína. Sendo observada uma relação linear entre essas duas grandezas para moléculas esféricas (em preto) e os desvios que ocorrem em função da forma da molécula estão mostrados em cores diferentes.....	49
Figura I-14. Abordagem sinérgica para determinação estrutural de proteínas. Fluxograma descrevendo a utilização combinada das técnicas de cristalografia de raios-X, NMR e métodos computacionais na determinação da estrutura tridimensional de proteínas.....	56
Figura IV-1. Fluxograma geral da expressão das proteínas alvo.....	97
Figura IV-2. Fluxograma geral do processo de lise do precipitado bacteriano.....	98
Figura IV-3. Gel de eletroforese SDS-PAGE (15%) da expressão da XAC1346, tendo da esquerda para a direita: marcador de massa molar (P7709S-NEB); NI corresponde à amostra não induzida; 3H a 3 horas e 16H a 16 horas depois da adição de IPTG.....	99
Figura IV-4. Gel de eletroforese SDS-PAGE (15%) da lise bacteriana, tendo da esquerda para a direita: marcador de massa molar de proteína (P7709S-NEB); (L) lisado total; (SL) sobrenadante da lise; (PL) precipitado da lise e outro marcador de massa molar.....	99
Figura IV-5. Fluxograma da preparação do sobrenadante da lise para posterior purificação da XAC1346.....	100
Figura IV-6. Primeira purificação da XAC1346. (A) Cromatograma da purificação da XAC1346 (DEAE-Sepharose), com o pico correspondente indicado. B) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: marcador de massa molar (P7709S-NEB) e algumas frações coletadas.....	101
Figura IV-7. Segunda purificação da XAC1346, por filtração em gel. A) Cromatograma da filtração em gel, sendo o pico correspondente à proteína indicado. B) Gel de eletroforese	

SDS-PAGE (15%) referente ao cromatograma tendo na sequência: o marcador de massa molar (kit LMW – GE Healthcare) e diferentes frações da purificação.....	102
Figura IV-8. Ilustração dos dados da filtração em gel analítica. A) Cromatogramas utilizados para a construção da curva de calibração, onde: C corresponde a conalbumina (75 kDa), N a anidrase carbônica (29 kDa), R a ribonuclease A (13,7 kDa), O a ovalbumina (44 kDa) e A a aprotinina (6,5 kDa). B) Curva de calibração tendo o coeficiente de partição em função do logaritmo da massa molar.....	104
Figura IV-9. Gel de eletroforese SDS-PAGE (15%) da alíquota advinda da filtração em gel analítica, referente à XAC1346, (marcador de massa molar – P7709S/NEB – em destaque a esquerda).....	105
Figura IV-10. Curva do R_f do gel de eletroforese em função do logaritmo da massa molar.....	105
Figura IV-11. Curva de calibração de Bradford para quantificação da proteína. A) Curva para altas concentrações. B) Curva para baixas concentrações protéicas.....	106
Figura IV-12. Gel de eletroforese SDS-PAGE (15%) de alíquotas da XAC1346 em solução, tendo da esquerda para a direita: marcador de massa molar (kit LMW – GE Healthcare) e duas alíquotas contendo a proteína pura, utilizadas para quantificação. A tabela ao lado indica a concentração dessa proteína nas alíquotas 1 e 2.....	107
Figura IV-13. Ilustração da identificação da XAC1346 por MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 33% de cobertura na primeira e 36% na segunda amostra.....	108
Figura IV-14. Espectro de dicroísmo circular da proteína XAC1346 ($14,3 \mu\text{mol L}^{-1}$) em tampão fosfato de sódio (10 mmol L^{-1} , pH 8).....	109
Figura IV-15. Modelo da XAC1346. Modelo do I-TASSER (modelo de maior pontuação dentre os modelos gerados, $C\text{-score} = -3,14$) com N- e C-terminais marcados, e a hélice anfipática destacada em azul.....	112
Figura IV-16. Comparação das curvas de CD, em preto a curva experimental e em vermelho a curva teórica (200-240 nm), gerada pelo K2D2 para a estrutura calculada pelo I-TASSER.....	112

Figura IV-17. Sobreposição dos modelos tridimensionais da XAC1346, obtidos por dois programas diferentes. Em vermelho a estrutura calculada pelo QUARK e em azul a estrutura calculada pelo I-TASSER.....	113
Figura IV-18. Resultados comparativos para: (A) Acima a estrutura 3D da XAC1346 calculada pelo QUARK, ressaltando N- e C-terminais, mais a região anfipática em azul no C-terminal e embaixo a sequência de aminoácidos; (B) Acima a estrutura da FliJ (PDB 3AJW), e embaixo sua sequência de aminoácidos.....	115
Figura IV-19. Sobreposição das estruturas: da XAC1346 calculada pelo QUARK, em vermelho, ressaltando o N- e o C-terminal e FliJ (PDB 3AJW) em azul, com a região onde as hélices se sobrepõem destacada.....	115
Figura IV-20. Alinhamento das estruturas: da XAC1346 calculada pelo QUARK, em vermelho, e FliJ (PDB 3AJW) em azul, sendo alinhados 72 aminoácidos dos 108 da XAC1346 (<i>TM-Score</i> = 0,47; onde <i>TM-Score</i> >0,5 indica que as estruturas partilham o mesmo enovelamento e <i>TM-Score</i> <0,2 indica que não há similaridade estrutural).....	116
Figura IV-21. Gel de eletroforese SDS-PAGE (15%) da expressão da XAC0419, tendo na sequência: marcador de massa molar (P7709S – NEB); NI corresponde à amostra não induzida; 3H a 3 horas e 16H a 16 horas depois da adição de IPTG.....	117
Figura IV-22. Gel de eletroforese SDS-PAGE (15%) da lise bacteriana, tendo indicados: marcador de massa molar de proteína (P7709S – NEB); solução de proteínas com massa molar conhecida; (L) lisado total; (SL) sobrenadante da lise e (PL) precipitado da lise; com monômero e dímero indicados.....	117
Figura IV-23. Fluxograma da preparação do sobrenadante da lise para posterior purificação da XAC0419.....	118
Figura IV-24. Primeira purificação da XAC0419. (A) Cromatograma da purificação com o pico correspondente à proteína indicado. B) Géis de eletroforese SDS-PAGE (15%) da purificação, tendo da esquerda para a direita: marcador de massa molar (kit LMW/GE Healthcare e P7709S/NEB) e algumas frações coletadas, com as frações contendo dímero e monômero indicadas.....	119
Figura IV-25. Segunda purificação da XAC0419, por filtração em gel. A) Cromatograma da filtração em gel, sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) referente ao cromatograma tendo na sequência: marcador	

de massa molar (P7709S/NEB) e diferentes frações da purificação, com a fração utilizada na filtração em gel analítica indicada no gel.....	120
Figura IV-26. Gel de eletroforese SDS-PAGE (15%) da alíquota advinda da filtração em gel analítica referente à XAC0419 (marcador de massa molar kit LMW/GE Healthcare).....	120
Figura IV-27. Filtração em gel analítica. A) Cromatogramas utilizados para a construção da curva de calibração. B) Curva de calibração tendo o coeficiente de partição em função do raio de Stokes (R_s).....	122
Figura IV-28. Gel de eletroforese SDS-PAGE (15%) de alíquotas da XAC0419 em solução, tendo na sequência: marcador de massa molar (kit LMW – GE Healthcare) e alíquotas contendo a proteína pura, utilizadas para quantificação. A tabela ao lado indica a concentração dessa proteína nas alíquotas 1 e 2.....	123
Figura IV-29. Identificação da XAC0419 por MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 11% de cobertura para a amostra dimérica (superior) e 39% para a amostra monomérica (inferior).....	124
Figura IV-30. Espectros de Dicroísmo Circular da XAC0419 na forma monomérica ($7 \mu\text{mol L}^{-1}$) e dimérica ($18 \mu\text{mol L}^{-1}$) em tampão acetato de sódio (50mmol L^{-1} , pH 5). O branco foi feito com o mesmo tampão.....	124
Figura IV-31. Espectro de NMR de ^1H da XAC0419 ($0,1 \text{mmol L}^{-1}$) em tampão fosfato de sódio 5mmol L^{-1} (D_2O 10%) obtido no equipamento Varian INOVA 500 MHz (499,98 MHz), utilizando a sequência de pulso WATERGATE para supressão do sinal de água em 4,8 ppm.....	126
Figura IV-32. Resultados de SAXS para a XAC0419. Curva de espalhamento de raios-X da proteína, construída, utilizando o programa PRIMUS.....	128
Figura IV-33. Resultados de SAXS para a XAC0419. Região de Guinier da curva de espalhamento de raios-X da proteína.....	128
Figura IV-34. Curva de distribuição de distâncias da XAC0419, $P(r)$, obtida com GNOM/AUTOGNOM.....	129
Figura IV-35. Curva de Kratky da XAC0419, obtida pelo tratamento dos dados de SAXS.....	129

Figura IV-36. Estrutura calculada por bioinformática para a XAC0419 (<i>C-score</i> = -0,24), com N- e C-terminais marcados e a hélice anfipática em azul.....	131
Figura IV-37. Resultados de SAXS e bioinformática. À direita três vistas diferentes do modelo gerado por SAXS (em cinza) sobreposto com a estrutura calculada pelo I-TASSER (em vermelho). À esquerda a mesma estrutura calculada referente à vista superior da direita, com o potencial de carga superficial calculado sendo, em azul potencial positivo, em vermelho negativo e em cinza os resíduos hidrofóbicos.....	132
Figura IV-38. Comparação das curvas de CD, em vermelho a curva experimental e em preto a curva teórica, gerada pelo K2D2 para a estrutura obtida pelo I-TASSER.....	133
Figura IV-39. Resultados de bioinformática. A) Duas vistas diferentes da estrutura 3D da XAC0419 calculada pelo QUARK, ressaltando o N- e o C-terminal, mais a região anfipática em rosa no C-terminal. B) Sobreposição dos modelos tridimensionais da XAC0419, em vermelho a estrutura calculada pelo I-TASSER e em azul a estrutura calculada pelo QUARK.....	135
Figura IV-40. Alinhamento das estruturas: da XAC0419 calculada pelo QUARK, em vermelho, e SycE de <i>Yersinia pseudotuberculosis</i> dimérica em azul (PDB:1JYA), sendo alinhados 67 aminoácidos dos 103 da XAC0419 (<i>TM-Score</i> = 0,46).....	136
Figura IV-41. Géis de eletroforese SDS-PAGE (15%). A) Expressão da XACb0033, tendo da esquerda para a direita: marcador de massa molar de proteína (kit LMW/GE Healthcare); NI corresponde à amostra não induzida e 16h a 16 horas depois da adição de IPTG. B) Lise da XACb0033, tendo da esquerda para a direita: marcador de massa molar de proteína (P7702S/NEB); L corresponde ao lisado total, SL ao sobrenadante da lise e PL precipitado da lise.....	137
Figura IV-42. Géis de eletroforese SDS-PAGE (15%) da lavagem do precipitado (A) e do tratamento com uréia (B), tendo: marcador de massa molar de proteína (kit LMW/GE Healthcare e P7703/NEB); amostras do lisado (L); sobrenadante da lise (S); precipitado da lise (P); água das lavagens do precipitado (L1 e L2); reenovelamento (R); sobrenadante do primeiro reenovelamento (SR1); precipitado do primeiro reenovelamento (PR1); sobrenadante do segundo reenovelamento (SR2) e precipitado do segundo reenovelamento (PR2) da XACb0033.....	138

Figura IV-43. Fluxograma da obtenção da XACb0033 solúvel para posterior purificação.....	139
Figura IV-44. Gel de eletroforese SDS-PAGE (15%) da lavagem dos precipitados do reenovelamento da XACb0033, tendo na sequência: marcador de massa molar de proteína (kit LMW/GE Healthcare); água das lavagens do precipitado do primeiro reenovelamento (L1 e L2); precipitado do primeiro reenovelamento (PR1); água das lavagens do precipitado do segundo reenovelamento (L21 e L22) e precipitado do segundo reenovelamento (PR2).....	140
Figura IV-45. Cromatograma da purificação do sobrenadante do reenovelamento da XACb0033 (A), sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) da purificação da XACb0033, tendo na sequência: marcador de massa molar (P7703/NEB), EC correspondendo a amostra que foi aplicada na coluna seguida de diferentes frações da purificação.....	141
Figura IV-46. Gel de eletroforese SDS-PAGE (15%) de alíquotas da XACb0033 em solução, tendo na sequência: o marcador de massa molar (kit LMW/GE Healthcare) e as diferentes frações de uma purificação, destacando as bandas cortadas para sequenciamento por Espectrometria de Massas. Ao lado tabela de medidas das concentrações das soluções marcadas.....	142
Figura IV-47. Filtração em gel da XACb0033. A) Cromatograma, sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) referente ao cromatograma tendo na sequência: marcador de massa molar (P7709S/NEB) e a amostra referente ao pico indicado no cromatograma.....	143
Figura IV-48. Filtração em gel analítica da XACb0033. Cromatograma, sendo o pico correspondente à proteína e seu volume de eluição indicados na figura.....	144
Figura IV-49. Espectro de massas da XACb0033 intacta, com massa do monômero e dímero indicadas.....	144
Figura IV-50. Identificação da XACb0033 por MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 91% de cobertura da sequência.....	145
Figura IV-51. Dicroísmo Circular da XACb0033. Curva de CD experimental com a proteína 9,1 $\mu\text{mol L}^{-1}$ em tampão fosfato de sódio 10 mM pH 7,4.....	146

Figura IV-52. Fluorescência de emissão da XACb0033. Espectros de (máximo 349 nm), da proteína 4 $\mu\text{mol L}^{-1}$ em tampão fosfato de sódio 10 mM pH 7,4. O branco foi feito com este tampão.....	147
Figura IV-53. Resultados de DLS para a XACb0033. Ao lado gel de eletroforese da amostra submetida ao ensaio, tendo o marcador de massa molar (kit LMW/GE Heathcare) e a amostra da proteína.....	148
Figura IV-54. Espectro de NMR-DOSY da XACb0033 0,12 mmol L^{-1} em 10% de D_2O obtido no equipamento Varian INOVA 500 MHz, utilizando a seqüência de pulso Dbppste, onde em F1 tem-se o coeficiente de difusão (D) multiplicado por 10^{-10} e a projeção do espectro de ^1H em F2.....	149
Figura IV-55. Espectros de NMR de ^1H das proteínas padrão (indicadas no espectro com as respectivas concentrações) em tampão fosfato de sódio 50 mmol L^{-1} (D_2O 10%) obtido no equipamento Varian INOVA 500 MHz (499,98 MHz), utilizando a seqüência de pulso WATERGATE para supressão do sinal de água em 4,8 ppm.....	151
Figura IV-56. Espectro de NMR de ^1H da BSA 0,9 mmol L^{-1} em D_2O obtido no equipamento de 500 MHz, utilizando a seqüência WATERGATE para supressão do sinal de água em 4,8 ppm.....	152
Figura IV-57. Espectro de NMR de ^1H da lisozima 4,1 mmol L^{-1} em D_2O obtido no equipamento de 500 MHz, utilizando a seqüência WATERGATE para supressão do sinal de água em 4,8 ppm.....	152
Figura IV-58. Curva com os dados de NMR tratados da lisozima 4 mmol L^{-1} em D_2O (99%) obtido no equipamento de 500 MHz, utilizando a seqüência de pulso Dbppste. Sinal em -7,8 (log D) referente ao sinal da água residual (HDO).....	153
Figura IV-59. Curvas relacionando o logaritmo do coeficiente de difusão com o logaritmo da massa molar (A) o coeficiente de difusão com o raio de Stokes (B) das proteínas padrão.....	154
Figura IV-60. Curva relacionando o coeficiente de difusão com a massa molar, em escala logarítmica.....	155
Figura IV-61. Estrutura calculada para a XACb0033. Modelo do I-TASSER (modelo de maior pontuação dentre os modelos gerados, <i>C-score</i> = -2,28) com N- e C-terminal	

marcados; em azul a hélice anfipática e os dois resíduos de Trp (W26 e W70) em regiões desordenadas.....	157
Figura IV-62. Espectros de CD da XACb0033, curva calculada em preto (200-240 nm) e curva experimental em azul.....	158
Figura IV-63. Resultados de bioinformática. A) Duas formas diferentes de representação da estrutura 3D da XACb0033 calculada pelo QUARK, ressaltando N- e C-terminais, a região anfipática em vermelho no N-terminal e os resíduos W26 e W70 em vermelho. B) Sobreposição dos modelos tridimensionais, em vermelho a estrutura calculada pelo I-TASSER e em azul a estrutura calculada pelo QUARK.....	161
Figura IV-64. Alinhamento das estruturas: da XACb0033 calculada pelo QUARK em vermelho, e VirE1 de <i>Agrobacterium tumefaciens</i> em azul, sendo alinhados 48 aminoácidos dos 65 da VirE1 (<i>TM-Score</i> = 0,46).....	161
Figura IV-65. Géis de eletroforese SDS-PAGE (15%). A) Expressão da FlgN, tendo na sequência: marcador de massa molar de proteína (kit LMW/GE Healthcare); amostras não induzidas (NI) e amostras 3 horas depois da adição de IPTG (3H); mistura de proteínas com massa molar conhecida. B) Lise bacteriana, tendo na sequência: proteínas com massa molar conhecida (<i>Gel filtration Calibration kit</i> LMW); 3 horas depois da adição de IPTG (3H); lisado total (L); sobrenadante da lise (SL) e precipitado da lise (PL).....	163
Figura IV-66. Fluxograma da obtenção da FlgN solúvel para posterior purificação.....	164
Figura IV-67. Géis de eletroforese SDS-PAGE (15%) do reenovelamento da FlgN, tendo na sequência: marcador de massa molar de proteína (P7709S/NEB); lisado total (L); sobrenadante da lise (SL) e precipitado da lise (PL), água das lavagens do precipitado da lise (L1,2 e 3), precipitado final (Pf); reenovelamento (R); sobrenadante do reenovelamento (SR) e precipitado do reenovelamento (PR).....	164
Figura IV-68. Cromatograma da purificação do sobrenadante do reenovelamento da FlgN (A), sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: proteínas com massa molar conhecida (<i>Gel filtration Calibration kit</i> LMW), EC correspondendo à amostra que foi aplicada na coluna seguida de diferentes frações da purificação.....	165
Figura IV-69. Filtração em gel da FlgN. A) Cromatograma, sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) referente ao	

cromatograma tendo na sequência: proteínas com massa molar conhecida (<i>Gel filtration Calibration kit</i> LMW), a amostra que aplicada na coluna (EC) e a amostra referente ao pico indicado no cromatograma.....	166
Figura IV-70. Filtração em gel analítica da FlgN. Cromatograma, sendo o pico correspondente à proteína e seu volume de eluição indicados na figura.....	167
Figura IV-71. Gel de eletroforese SDS-PAGE (15%) de alíquotas da FlgN em solução, tendo na sequência: marcador de massa molar (P7709S/NEB) e diferentes frações de uma purificação. Ao lado tabela de medidas das concentrações das soluções marcada.....	167
Figura IV-72. Identificação da XAC1990 por MS/MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 96% de cobertura da sequência.....	168
Figura IV-73. Dicroísmo Circular da FlgN. Curva de CD experimental com a proteína (24 $\mu\text{mol L}^{-1}$) em tampão fosfato de sódio (50 mmol L^{-1} , pH 8).....	169
Figura IV-74. Fluorescência de emissão da FlgN. Espectros de emissão (máximo 351 nm), da proteína 2,6 $\mu\text{mol L}^{-1}$ em tampão fosfato de sódio 10 mmol L^{-1} , pH 7,4. O branco foi feito com este tampão.....	170
Figura IV-75. Dados de DLS da FlgN. Distribuição de tamanhos da proteína. Ao lado gel de eletroforese da amostra submetida ao ensaio, tendo o marcador de massa molar (P7709S/NEB) e amostras da proteína.....	171
Figura IV-76. Espectro de NMR de ^1H da FlgN (0,22 mmol L^{-1}) em tampão fosfato de sódio 5 mmol L^{-1} (D_2O 10%) obtido no equipamento Varian INOVA 500 MHz (499,98 MHz), utilizando a sequência de pulso WATERGATE para supressão do sinal de água em 4,8 ppm.....	172
Figura IV-77. Dados de SAXS da FlgN. A) Curva de espalhamento de SAXS. B) Região de Guinier da curva de espalhamento.....	173
Figura IV-78. Curva de distribuição de distâncias da XAC1990, $P(r)$	174
Figura IV-79. Curva de Kratky para a FlgN, obtida pelo tratamento dos dados de SAXS.....	174
Figura IV-80. Representação em <i>cartoon</i> da estrutura da FlgN com N- e C- terminais marcados (I-TASSER <i>C-score</i> = -2,66), a hélice anfipática destacada em laranja e o resíduo de Trp (W78) em azul (exposto ao solvente) (A). A mesma estrutura calculada, representada por superfície, com o potencial de carga superficial correspondente (potencial	

positivo em azul, negativo em vermelho e em cinza as regiões hidrofóbicas), destacando o W78 exposto ao solvente (B). Em (C) três vistas do modelo gerado pelos dados de SAXS da FlgN (em cinza) sobreposto com a estrutura calculada pelo I-TASSER (em vermelho).....176

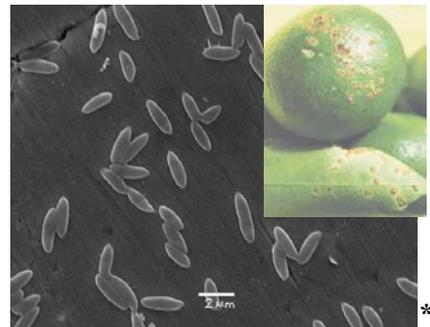
Figura IV-81. Comparação das curvas de CD da FlgN, em vermelho a curva experimental e em preto a curva teórica (200-240 nm), gerada pelo K2D2 para a estrutura calculada pelo I-TASSER.....177

Figura IV-82. Resultados de bioinformática. A) Estrutura da FlgN calculada pelo QUARK, ressaltando N- e C-terminais, e o resíduo de triptofano em azul. B) Sobreposição dos modelos tridimensionais da FlgN, em azul a estrutura obtida pelo I-TASSER e em vermelho a estrutura calculada pelo QUARK.....179

Figura IV-83. Resultados de SAXS e bioinformática. A) Estrutura da FlgN da *Xac*, calculada pelo I-TASSER. B) Estrutura cristalográfica da FlgN de *Pseudomonas aeruginosa* (PDB 2FUP, depositada em 2007, dados não publicados).....180

Figura IV-84. Alinhamento das estruturas: da FlgN da *Xac* calculada pelo QUARK, em vermelho, e FlgN de *Pseudomonas aeruginosa* em azul (PDB:2FUP), sendo alinhados 72 aminoácidos dos 110 da FlgN da *Xac* (*TM-Score* = 0,45).....180

Capítulo 1



1.1 MOTIVAÇÃO

*Adaptado de <http://karantina.deptan.go.id/old/index.php?link=news&id=59> e Martins, A.M. 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.

1. Introdução

1.1 Motivação – Considerações Gerais

O cancro cítrico é uma doença da citricultura caracterizada pelo aparecimento de pequenas manchas amarelas circulares nas folhas e nos frutos, sendo que estas manchas se tornam marrons com o avanço da doença (Figura I-1). As manchas nos frutos são salientes e nos ramos podem ser identificadas lesões pardas em forma de crostas.¹

Essa doença está presente na agricultura brasileira desde 1957, proveniente da Ásia e atinge todas as variedades de citros.² Como o cancro cítrico é uma doença bastante contagiosa que pode levar a queda dos frutos e perdas na produção agrícola, ele representa um fator de grande impacto na economia brasileira assim como na de outros países produtores de citros (Figura I-1).³

As frutas cítricas são produzidas no mundo todo, sendo que em 2004 foi observada a existência de 140 países produtores.⁴ Cerca de 70% da produção mundial está concentrada no hemisfério norte, em particular nos países mediterrâneos e nos Estados Unidos da América, entretanto, o Brasil é o maior produtor mundial. De acordo com a Organização *FAO (Food and Agricultural Organization of United Nations)* das Nações Unidas, em 2007 os maiores

¹ (a) Página da Gestora de Projetos Agromundo. <http://www.agromundo.com.br/?p=19495> (acessado em Junho de 2011). (b) Lopes, T.P. Caracterização Estrutural do Complexo de Proteínas Hipotéticas – XACb0032/XACb0033 da Bactéria *Xanthomonas axonopodis* pv. *citri*. Dissertação de Mestrado, Instituto de Química, Universidade Estadual de Campinas, SP, Novembro 2007.

² (a) Bitancourt, A.A. O Cancro Cítrico. *O Biológico* **1957**, *23*, 101-111. (b) Borin, P.F.L. Caracterização estrutural da proteína hipotética XACb0033 da bactéria *Xanthomonas axonopodis* pv. *citri*. Dissertação de Mestrado, Instituto de Química, Universidade Estadual de Campinas, SP, Janeiro de 2010.

³ Gottwald, T.R.; Graham, J.H.; Schubert, T.S. Citrus Canker: the pathogen and its impact. *Plant Health Progress* [Online] **2002**, doi10.1094/PHP-2002-0812-01-RV.

⁴ http://www.ask.com/wiki/Citrus_production?qsrc=3044 (acessado em Junho de 2011).

produtores mundiais de citros eram: Brasil, Estados Unidos, México, Índia, China, Espanha, Itália, Irã, Egito e Paquistão (Figura I-1), sendo esta ordem descrita, mais especificamente em relação à produção de laranjas, porém estes também são os maiores produtores de citros no geral.^{4,5}

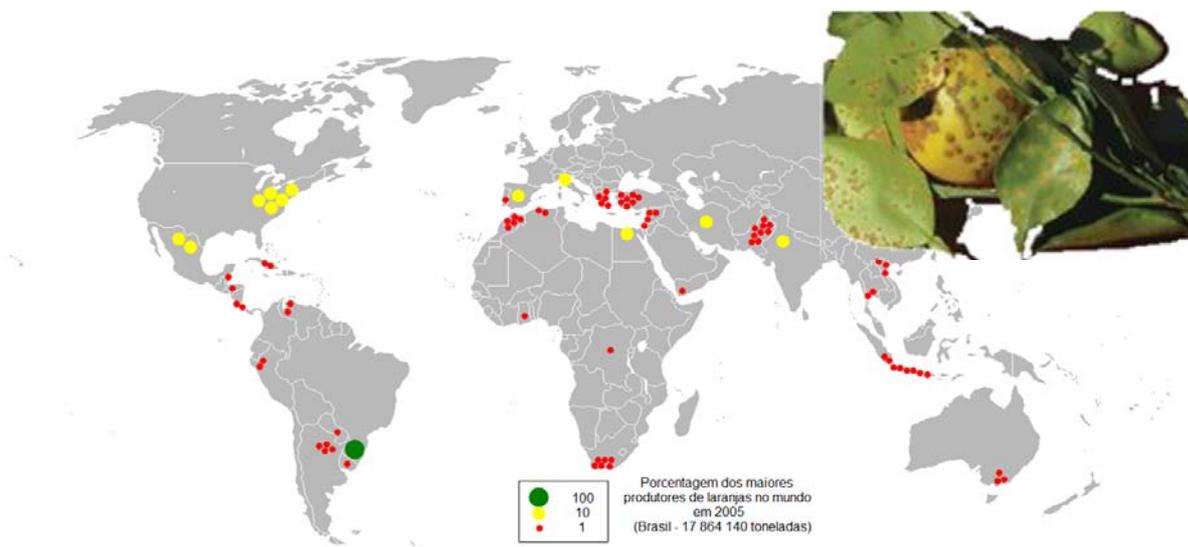


Figura I-1. Distribuição da produção de laranjas ao redor do mundo em 2005 e a representação de plantas contaminadas com o cancro cítrico no canto superior direito.⁵

Dentro do Brasil, o parque citrícola paulista é o maior do mundo, composto por mais de 200 milhões de pés de laranja em 628 mil hectares, responsáveis por 53% da produção mundial de suco, empregando mais de 400 mil pessoas e gerando cerca de US\$1,5 bilhões anuais.^{1,5} Considerando tais números pode-se avaliar as perdas econômicas advindas do aparecimento do cancro cítrico.

Acredita-se que a doença, observada pela primeira vez no estado de São Paulo em 1957, tenha se originado no sudeste da Ásia (Índia), mas atualmente

⁵ (a) Belasque Jr., J.; Fernandes, N.G.; Massari, C.A. O Sucesso da Campanha de Erradicação do Cancro Cítrico no Estado de São Paulo, Brasil. *Summa Phytopathol.* **2009**, *35*, 91-92. (b) Adaptado de <http://www.xn--gew.net/articles/123535/Production>, <http://www.freshfromflorida.com/pi/canker/> (acessado em Junho de 2011).

está presente em vários países como Japão, países da região central da África, no estado da Flórida nos Estados Unidos, entre outros.

Algumas dessas áreas erradicaram o cancro cítrico e outras, como o Brasil, tem programas de erradicação em andamento. Entretanto o cancro parece ser endêmico na maioria das áreas em que aparece e devido a sua rápida propagação, aos severos danos causados e ao grande impacto no comércio e exportações, a doença é uma ameaça significativa para os produtores de citros.^{4,6}

Segundo levantamento do Fundo de Defesa da Citricultura (Fundecitrus), a incidência da doença entre 2009 e 2010 em São Paulo foi de 0,14 para 0,44%, representando um crescimento de 214% neste período. Entretanto, este aumento já era esperado, uma vez que em junho de 2009 o governo do estado atenuou a legislação de controle da doença.^{1a} Em 1999, após observação da incidência de 0,7% de cancro nos pomares, a legislação passou a ser bastante rígida, sendo obrigatória a erradicação de todas as árvores com cancro, além de uma área em um raio de 30 metros de distância de uma árvore contaminada e ainda todo um pomar (talhão). Isso levou a uma diminuição na incidência para 0,27% em 2000, variando desde então de 0,10 a 0,22% até 2009.^{1a} Em 2009, por pressão dos agricultores, a Secretaria da Agricultura de São Paulo revogou a determinação referente aos talhões, o que fez com que o índice voltasse a disparar em 2010. Após seis meses desta mudança na legislação, o número de novos casos de cancro cítrico cresceu quase 80% em relação ao primeiro semestre de 2009. Com isso o produtor precisa intensificar as medidas de controle do cancro, principalmente nos

⁶ Facincani, A.P. Análise Proteômica do Fitopatógeno *Xanthomonas axonopodis* pv. *citri*. Tese de Doutorado, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista/Campus Jaboticabal, 2007.

períodos de chuva, dado que o clima úmido favorece a proliferação da bactéria causadora da doença.^{1a}

Além da erradicação, podem ser empregados outros métodos de controle do cancro cítrico, como o estabelecimento de “quebra ventos” nas áreas de produção⁶, porém, em 2008, foi mostrado que o uso destes “quebra-ventos” contribui pouco no controle da doença.⁷ Outra medida é o controle da larva minadora dos citros, uma vez que as lesões provocadas pela larva intensificam a infecção pelo cancro cítrico.⁸

Uma terceira medida que também é bastante empregada é o uso de pulverizadores de cobre, sendo que alguns estudos mostraram que a utilização destes bactericidas contendo cobre diminui a incidência inicial da doença e o crescimento epidêmico.⁹ Além de que, o cobre diminui a incidência e a severidade das lesões nas folhas e frutos provocadas pelo cancro cítrico.⁷ Entretanto, apesar de alguns estudos mostrarem uma relação custo-benefício positiva do efeito do cobre no controle da doença, a eliminação das plantas contaminadas e adjacentes ainda é o método mais eficaz de controle do cancro cítrico.^{5b,10}

⁷ (a) Behlau, F.; Belasque Jr., J.; Bergamin Filho, A.; Graham, J.H.; Leite Jr., R.P.; Gottwald, T.R. Cooper sprays and windbreaks for control of citrus canker on young orange trees in southern Brazil. *Crop Prot.* **2008**, *27*, 807-813. (b) Graham, J.H.; Gottwald, T.R.; Cubero, J.; Achor, D.S. *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradication of citrus canker. *Mol. Plant Pathol.* **2004**, *5*, 1-15.

⁸ Jesus Jr., W.C.; Belasque Jr., J.; Amorin, L.; Christiano, R.S.C.; Parra, J.R.P.; Bergamin Filho, A. Injuries Caused by Citrus Leafminer (*Phyllocnistis citrella*) Exacerbate Citrus Canker (*Xanthomonas axonopodis* pv. *citri*) Infection. *Fitopatol. Bras.* **2006**, *31*, 277-283.

⁹ (a) Behlau, F.; Amorin, L.; Belasque Jr., J.; Bergamin Filho, A.; Leite Jr., R.P.; Graham, J.H.; Gottwald, T.R. Annual and polyetic progression of citrus canker on trees protected with copper sprays. *Plant Pathol* **2010**, *59*, 1031-1036. (b) McGuire, R.G. Evaluation of Bactericidal Chemicals for Control of *Xanthomonas* on Citrus. *Plant Dis.* **1988**, *72*, 1016-1020.

¹⁰ Behlau, F.; Belasque Jr., J.; Graham, J.H.; Leite Jr., R.P. Effect of frequency of copper applications on control of citrus canker and the yield of young bearing sweet orange trees. *Crop Prot.* **2010**, *29*, 300-305.

Assim, como a erradicação é uma medida bastante utilizada, porém que gera prejuízos;¹¹ existe a necessidade de se entender melhor o mecanismo de interação entre a planta hospedeira e o fitopatógeno, para o desenvolvimento de novas e melhores estratégias de controle e combate ao cancro cítrico.¹² Atualmente um estudo¹³ mostrou, por exemplo, que a regulação da expressão de terpenos na casca das frutas cítricas pode conferir resistência ao fitopatógeno causador da doença. Este estudo indicou que o acúmulo de limoneno na casca do fruto parece estar envolvido na interação da planta com o fitopatógeno, e que a diminuição desta quantidade de limoneno confere resistência à planta. Podendo ser esta uma nova estratégia para a produção de frutos transgênicos resistentes ao cancro cítrico.

Na literatura existem três tipos de cancro cítrico descritos: os tipos A, B e C, além das subdivisões. Os tipos B e C foram encontrados apenas na América do Sul, enquanto o tipo A, originário da Ásia, está espalhado atualmente pelo mundo todo, inclusive na América do Sul e é o tipo mais severo e que gera maiores perdas econômicas.¹²

¹¹ (a) del Campo, R.; Russi, P.; Mara, P.; Mara, H.; Peyrou, M.; León, I.P.; Gaggero, C. *Xanthomonas axonopodis* pv. *citri* enters the VBNC state after copper treatment and retains its virulence. *FEMS Microbiol. Lett.* **2009**, *298*, 143-148. (b) Belasque Jr., J.; Bergamin Filho, A.; Massari, C.A. Prováveis conseqüências do abrandamento da metodologia de erradicação do cancro cítrico no Estado de São Paulo. *Trop. Plant Pathol.* **2010**, *35*, 314-317.

¹² Moreira, L.M.; Almeida, N.F.; Potnis, N.; Digiampietri, L.A.; Adi, S.S.; Bortolossi, J.C.; da Silva, A.C.; da Silva, A.M.; Moraes, F.E.; Oliveira, J.C.; Souza, R.F.; Facincani, A.P.; Ferraz, A.L.; Ferro, M.I.; Furlan, L.R.; Gimenez, D.F.; Jones, J.B.; Kitajima, E.W.; Laia, M.L.; Leite Jr, R.P.; Nishiyama, M.Y.; Neto, J.R.; Nociti, L.A.; Norman, D.J.; Ostroki, E.H.; Pereira Jr, H.A.; Staskawicz, B.J.; Tezza, R.I.; Ferro, J.A.; Vinatzer, B.A.; Setubal, J.C. Novel insights into genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifolli*. *BMC Genomics* [Online] **2010**, *11*, 238.

¹³ Rodríguez, A.; Andrés, V.S.; Cervera, M.; Redondo, A.; Alquézar, B.; Shimada, T.; Gadea, J.; Rodrigo, M.J.; Zacarías, L.; Palou, L.; López, M.; Castañera, P.; Peña, L. Terpene Down-Regulation in Orange Reveals the Role of Fruit Aromas in Mediating Interactions with Insect Herbivores and Pathogens. *Plant Physiol.* **2011**, *156*, 793-802.

O cancro cítrico do tipo A é causado pela bactéria *Xanthomonas axonopodis* pv. *citri*, denominada atualmente de *Xanthomonas citri* subsp. *citri*.⁸ Como ainda existem controvérsias a respeito desta denominação, no presente texto a bactéria está referida como *Xanthomonas axonopodis* pv. *citri*.

1.1.1. *Xanthomonas axonopodis* pv. *citri* (*Xac*)

A *Xanthomonas axonopodis* pv. *citri* é uma bactéria Gram-negativa fitopatogênica, causadora do cancro cítrico do tipo A.¹⁴ O gênero *Xanthomonas* em si, representa o maior grupo de bactérias fitopatogênicas, e apresenta grande importância econômica devido às doenças que provoca.⁶ A *Xac*, em especial, por não requerer um vetor específico de infecção, é de fácil disseminação e altamente contagiosa, sendo transferida entre os hospedeiros pela ação do vento, chuva ou manuseio.¹⁵

Esta bactéria apresenta respiração aeróbica, formato de bastonetes e possui um flagelo polar responsável por sua motilidade, sendo que recentemente foi observado que este flagelo é requerido para a formação do biofilme maduro e, conseqüentemente, está envolvido no desenvolvimento do cancro cítrico.¹⁶ O crescimento da *Xac* é endofítico, sendo bem adaptada ao ambiente de plantas, porém dificilmente encontrada livre no solo; se adapta bem em temperaturas de 20 a 30 °C, o que associado a altos teores de umidade do ar é ótimo para o crescimento e desenvolvimento da infecção. Além disso,

¹⁴ Brunings, A.M.; Gabriel, D.W. *Xanthomonas citri*: breaking the surface. *Mol. Plant Pathol.* **2003**, *4*, 141-157.

¹⁵ Khater, L. Identificação e caracterização de chaperones de secreção no genoma do fitopatógeno *Xanthomonas axonopodis* patovar *citri*. Tese de Doutorado, Instituto de Biologia, Universidade Estadual de Campinas, SP, Fevereiro de 2006.

¹⁶ Malamud, F.; Torres, P.S.; Roeschlin, R.; Rigano, L.A.; Enrique, R.; Bonomi, H.R.; Castagnaro, A.P.; Marano, M.R.; Vojnov, A.A. The *Xanthomonas axonopodis* pv. *citri* flagellum is required for mature biofilm and canker development. *Microbiol.* **2011**, *157*, 819-829.

produz polissacarídeos extracelulares como a goma xantana que auxiliam na sua dispersão e sobrevivência.^{2b,6,15}

Além da fácil disseminação, a *Xac* é bastante resistente e adaptável a diferentes ambientes, tendo como hospedeiro todas as variedades de citros, diferindo no grau das lesões. E, recentemente foi mostrado que a mutação de um único aminoácido no gene de patogenicidade *pthA* da *Xac* pode alterar a formação do cancro cítrico.¹⁷

A *Xac* penetra na planta hospedeira pelos estômatos e pelos ferimentos, como os causados pela larva minadora de citros e se multiplica nos espaços intercelulares, onde também secreta os polissacarídeos que formam um biofilme e provocam o encharcamento no tecido infectado. O espaço intercelular, no qual a *Xac* se encontra é pobre em nutrientes, contendo substâncias antimicrobianas e ácidas e, portanto, para sobreviver e se multiplicar a bactéria recruta um grande número de genes e seus produtos transcritos, ligados a virulência e patogenicidade. Em resposta a essa infecção a planta, preparada para detectar a presença dos patógenos, produz uma série de substâncias antimicrobianas e outros tipos de resposta de estresse. Porém, o encharcamento e aumento da permeabilidade das células leva à perda de nutrientes, fazendo com que a planta fique vulnerável à multiplicação da *Xac*, que acaba suprimindo os mecanismos de defesa do hospedeiro ou levando a uma resposta associada a uma reação de hipersensibilidade das células atacadas.^{6,14,15,18}

¹⁷ Lin, H.-C.; Chu, M.-K.; Lin, Y.-C.; Deng, W.-L.; Chang, H.; Hsu, S.-T.; Tzeng, K.-C. A single aminoacid substitution in PthaA of *Xanthomonas axonopodis* pv. *citri* altering canker formation on grapefruit leaves. *Eur. J. Plant Pathol.* **2011**, *130*, 143-154.

¹⁸ Enrique, R.; Siciliano, F.; Favaro, M.A.; Gerhardt, N.; Roeschlin, R.; Rigano, L.A.; Sendin, L.; Castagnaro, A.P.; Vojnov, A.A.; Marano, M.R. Novel demonstration of RNAi in citrus reveals importance of citrus callose synthase defense against *Xanthomonas citri* subsp. *citri*. *Plant Biotechnol. J.* **2011**, *9*, 394-407.

Vale mencionar, a observação de que a parede celular da planta é a primeira defesa dela, sendo a principal barreira contra a infecção da *Xac*.¹⁸ Durante o processo de proliferação na célula hospedeira, a *Xac*, como as demais bactérias Gram-negativas patogênicas, secreta fatores de virulência macromoleculares para a célula hospedeira durante o processo de infecção.¹⁹

Assim, a infecção começa com a atuação dos sistemas de secreção (tipos I-VI), nos quais fatores macromoleculares, em geral proteínas ou complexos de ácidos nucléicos e proteínas, são excretados para a célula da planta. No caso dos sistemas de secreção III e IV, por exemplo, esses fatores de virulência são injetados diretamente no citosol da célula hospedeira, onde irão interferir nos processos celulares.⁶ Portanto, o início da infecção no cancro cítrico se dá com a atuação dos sistemas de secreção e no caso da *Xac*, em particular, sistemas secretórios do tipo II, III e IV (T2SS, T3SS e T4SS)¹⁴ identificados no seu genoma, sequenciado em 2002.^{20,21}

¹⁹ (a) Papanikou, E.; Karamanou, S.; Economou, A. Bacterial protein secretion through the translocase nanomachine. *Nature Rev. Microbiol.* **2007**, *5*, 839-851. (b) Holland, I.B. Translocation of bacterial proteins—an overview. *Biochim. Biophys. Acta* **2004**, *1694*, 5-16.

²⁰ da Silva, A.C.; Ferro, J.A.; Reinach, F.C.; Farah, C.S.; Furlan, L.R.; Quaggio, R.B.; Monteiro-Vitorello, C.B.; Van Sluys, M.A.; Almeida, N.F.; Alves, L.M.; do Amaral, A.M.; Bertolini, M.C.; Camargo, L.E.; Camarotte, G.; Cannavan, F.; Cardozo, J.; Chambergo, F.; Ciapina, L.P.; Cicarelli, R.M.; Coutinho, L.L.; Cursino-Santos, J.R.; El-Dorry, H.; Faria, J.B.; Ferreira, A.J.; Ferreira, R.C.; Ferro, M.I.; Formighieri, E.F.; Franco, M.C.; Greggio, C.C.; Gruber, A.; Katsuyama, A.M.; Kishi, L.T.; Leite, R.P.; Lemos, E.G.; Lemos, M.V.; Locali, E.C.; Machado, M.A.; Madeira, A.M.; Martinez-Rossi, N.M.; Martins, E.C.; Meidanis, J.C.; Menck, F.; Miyaki, C.V.; Moon, D.H.; Moreira, L.M.; Novo, M.T.; Okura, V.K.; Oliveira, M.C.; Oliveira, V.R.; Pereira, H.A.; Rossi, A.; Sena, J.A.; Silva, C.; de Souza, R.F.; Spinola, L.A.; Takita, M.A.; Tamura, R.E.; Teixeira, E.C.; Tezza, R.I.; Trindade dos Santos, M.; Truffi, D.; Tsai, S.M.; White, F.F.; Setubal, J.C.; Kitajima, J.P. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **2002**, *417*, 459-463.

²¹ (a) Alegria, M.C.; Docena, C.; Khater, L.; Ramos, C.H.I.; Silva, A.C.R.; Farah, S.C. New Protein-Protein Interactions Identified for the Regulatory and Structural Components and Substrates of the Type III Secretion System of the Phytopathogen *Xanthomonas axonopodis* Pathovar *citri*. *J. Bacteriol.* **2004**, *186*, 6186-6197. (b) Alegria, M.C.; Souza, D.P.; Andrade, M.O.; Docena, C.; Khater, L.; Ramos, C.H.I.; Silva, A.C.R.; Farah, S.C. New Protein-Protein Interactions Involving the Products of the Chromosome- and Plasmid-Encoded Type IV Secretion Loci of the Phytopathogen *Xanthomonas axonopodis* pv. *citri*. *J. Bacteriol.* **2005**, *187*, 2315-2325.

1.1.2. Sistemas de secreção

Dentre os sete tipos de sistemas de secreção (I-VII) presentes em bactérias (Figura I-2), cuja denominação foi feita de acordo com a ordem de descoberta dos mesmos, quatro utilizam um mecanismo de secreção geral.

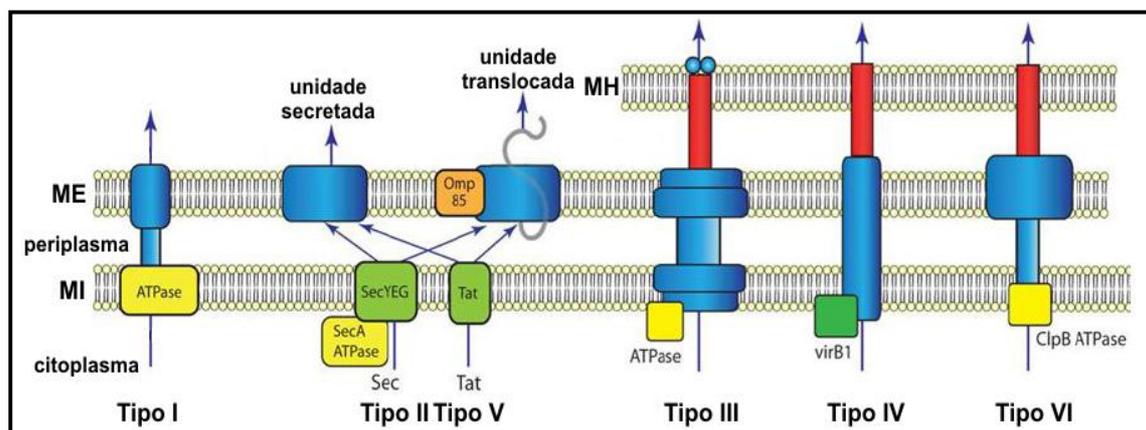


Figura I-2. Sistemas de secreção de bactérias Gram-negativas (I-VI), onde MI, ME e MH correspondem respectivamente à membrana interna, externa e do hospedeiro.²² O Sistema de Secreção do tipo VII não está representado na figura, pois é encontrado apenas em bactérias Gram-positivas.

No mecanismo de secreção geral,²³ o transporte das proteínas através da membrana interna é realizado pelo sistema *sec*-dependente (sistemas de secreção do tipo II e V, por exemplo), onde as proteínas a serem secretadas possuem uma sequência sinal na extremidade N-terminal que será clivada no periplasma; por isso, esses sistemas são chamados de *sec*-dependentes e eles

²² Adaptado de Tseng, T.-T.; Tyler, B.M.; Setubal, J.C. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol.* **2009**, *9*(Supl I): S2 doi:10.1186/1471-2180-9-S1-S2.

²³ (a) Economou, A. Following the leader: bacterial protein export through the Sec pathway. *Trends Microbiol.* **1999**, *7*, 315-320. (b) Henderson, I.R.; Navarro-Garcia, F.; Desvaux, M.; Fernandez, R.C.; Ala'Aldeen, D. Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Bio. Rev.* **2004**, *68*, 692-744.

são classificados de acordo com a forma em que as proteínas a serem secretadas atravessam a membrana externa.²⁴

Diferente dos sistemas de secreção *sec*-dependentes, os sistemas de secreção capazes de exportar o substrato diretamente do citoplasma para fora da membrana externa da bactéria são chamados de *sec*-independentes e compreendem os sistemas do tipo I, III, IV e VI.²⁵

Os sistemas de secreção do tipo III e IV possuem algumas características em comum, como por exemplo, ambos necessitam de contato físico com a célula hospedeira, requerem que uma chaperona secretória se ligue à proteína a ser secretada e acredita-se que a exportação desta seja feita em somente uma etapa, via um canal de secreção. Entretanto, além do sistema de secreção do tipo IV também funcionar como *sec*-dependente,²⁶ ele pode exportar DNA fita simples e complexo DNA/proteína até a célula hospedeira e não há nenhuma evidência de que o sistema do tipo III possa fazer o mesmo, sendo esta a diferença fundamental entre eles.²⁷

O sistema de secreção do tipo III (T3SS – Figura I-3) possui um formato cilíndrico, se assemelhando a uma agulha de injeção, e é composto por aproximadamente 20 a 30 proteínas que formam um complexo que atravessa as membranas interna e externa do envelope bacteriano e se projeta para fora da bactéria, sendo considerado o mais complexo dentre os sistemas de secreção.^{28,29} Durante o processo de secreção, as proteínas a serem

²⁴ Thanassi, D.G.; Hultgren, S.J. Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr. Opin. Cell Biol.* **2000**, *12*, 420-430.

²⁵ Büttner, D.; Bonas, U. Port of entry – the type III secretion translocon. *Trends Microbiol.* **2002**, *10*, 186-192.

²⁶ Burns, D.L. Type IV transporters of pathogenic bacteria. *Curr. Opin. Microbiol.* **2003**, *1*, 29-34.

²⁷ Christie, P.J.; Vogel, J.P. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* **2000**, *8*, 354-360.

²⁸ Smith, C.L.; Hultgren, S.J. Bacteria thread the needle. *Nature* **2001**, *414*, 29-31.

secretadas devem ser reconhecidas e, como o canal de secreção é estreito (3 nm de diâmetro) devem estar desenoveladas antes da secreção para passar pelo canal. Para atender a esse requisito, o mecanismo de reconhecimento do substrato envolve vários meios de sinalização, além do auxílio de proteínas acessórias, como as chaperonas de secreção supracitadas.³⁰

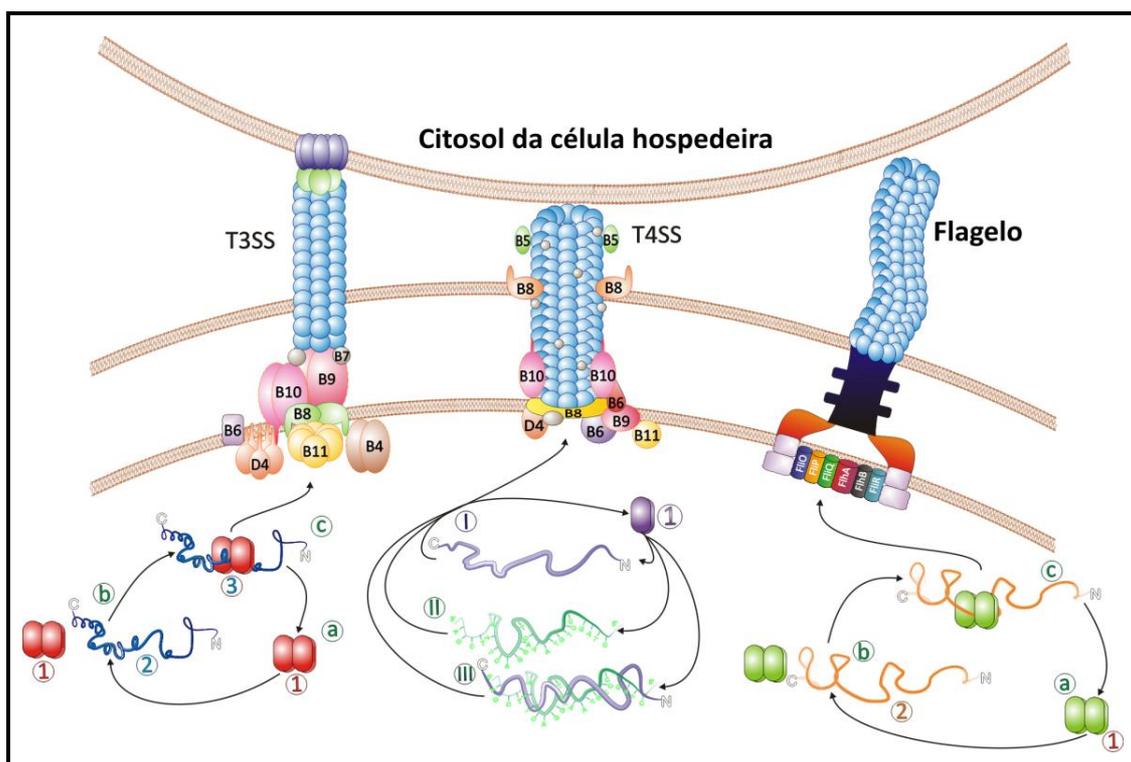


Figura I-3. Representação dos sistemas de secreção do tipo III (T3SS) e do tipo IV (T4SS); e do flagelo da *Xac*.³¹

Assim, foi sugerido que as chaperonas de secreção citosólicas se ligam especificamente ao substrato e auxiliam na translocação.³²

²⁹ Galán, J.E.; Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **1999**, *284*, 1322-1328.

³⁰ Galán, J.E.; Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **2006**, *444*, 567-573.

³¹ Adaptado de Fattori, J.; Prando, A.; Martins, A.M.; Rodrigues, F.H.S.; Tasic, L. Bacterial Secretion Chaperones. *Protein Pept. Lett.* **2011**, *18*, 158-166.

O sistema de secreção do tipo IV (T4SS – Figura I-3), por sua vez, está relacionado com os componentes da maquinaria de conjugação bacteriana, no qual a transferência através do envelope bacteriano é mediada por uma estrutura supramolecular denominada formação do par conjugado. Este complexo formado é responsável por estabelecer o contato entre as células levando à formação de um canal de conjugação pelo qual se dá a translocação do substrato a ser transferido.³³

Este sistema de conjugação bacteriana, além de exportar *t*-DNA, exporta também proteínas e proteínas associadas ao DNA fita simples.³⁴ E no geral, a translocação é realizada em uma única etapa, não havendo necessidade de um intermediário periplasmático, porém, observou-se que deve haver uma chaperona de secreção mediando a transferência.³⁵ Desta forma, a exportação de proteínas, independente de DNA, através do sistema de secreção do tipo IV deve ser realizada de maneira análoga à do sistema de secreção do tipo III, primeiramente com a formação do complexo entre a proteína a ser secretada e a sua chaperona secretória.

Vale comentar também, sobre o sistema flagelar (Figura I-3), devido às similaridades com o T3SS, além de que é sabido que a *Xac* possui um flagelo polar.

O sistema flagelar é um sistema complexo que além de ser responsável pela motilidade, desempenha um papel fundamental na adesão, formação de

³² Wattiau, P.; Cornelis, G.R. SycE, a chaperone-like protein of *Yersinia enterocolitica* involved in the secretion of YopE. *Mol. Microbiol.* **1993**, *8*, 123-131.

³³ Cascales, E.; Christie, P.J. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **2003**, *1*, 137-149.

³⁴ Christie, P.J. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol. Microbiol.* **2001**, *40*, 294-305.

³⁵ Deng, W.; Chen, L.; Peng, W.-T.; Liang, X.; Sekiguchi, S.; Gordon, M.P.; Comai, L.; Nester, E.W. VirE1 is a specific molecular chaperone for the exported single-stranded-DNA-binding protein VirE2 in *Agrobacterium*. *Mol. Microbiol.* **1999**, *31*, 1795-1807.

biofilme e invasão do hospedeiro. Um sistema flagelar típico consiste de seis componentes (Figura I-4): um corpo basal que inclui os anéis MS, P e L; um motor e um ‘interruptor’; um gancho; um filamento e um aparato de exportação (T3SS).³⁶

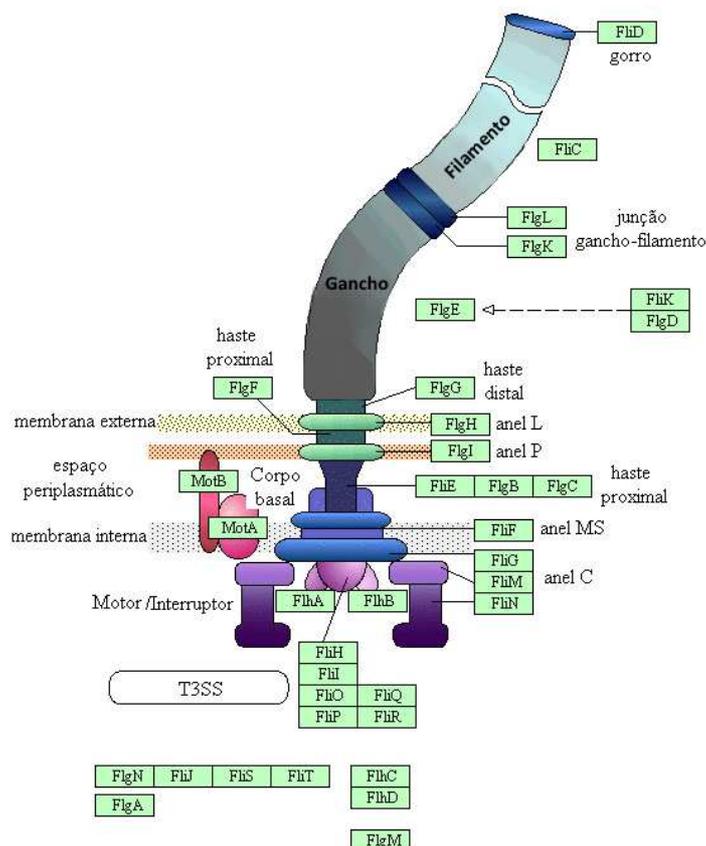


Figura I-4. Representação de um sistema flagelar (*Escherichia coli*).³⁷

Nos sistemas flagelares mais estudados (*E. coli* e *Salmonella entérica* sp. *Typhimurium*) observou-se que 50 genes estão envolvidos na biossíntese e função do flagelo; dos quais metade codificam os componentes estruturais do

³⁶ Macnab, R.M. Action at a Distance – Bacterial Flagella Assembly. *Science* **2000**, *290*, 2086-2087.

³⁷ Adaptado de www.genome.jp/kegg/pathway/eco/eco02040.html (acessado em Setembro de 2011).

flagelo e os outros são responsáveis pela regulação da montagem flagelar e detecção/processamento dos sinais externos aos quais o flagelo responde.³⁸

As características estruturais do flagelo, juntamente com a evidência de homologia entre algumas proteínas como a FliI e subunidades ATP sintase, por exemplo, sugerem que ele se originou como um sistema de secreção primitivo, envolvendo primeiramente uma ATPase e depois adicionaram-se o gancho e os componentes do filamento por duplicação de genes e diversificação. Esta afirmação de sua origem também é suportada pela clara correlação entre o sistema flagelar e o T3SS, cuja arquitetura genética é semelhante e derivada de um complexo de genes flagelares.³⁸ Por isso considera-se que o sistema flagelar é o ancestral do T3SS.

As estruturas do corpo basal tanto do flagelo quanto do T3SS são bastante similares e podem estar relacionadas funcionalmente. Assim, pode se propor que, em bactérias que possuem um flagelo e um T3SS, como a *Xac*, deve haver algum grau de coordenação na montagem e função destes dois sistemas. Em um estudo com *P. aeruginosa*, por exemplo, explorando a relação entre a motilidade e a secreção de fatores de virulência pelo T3SS, foi sugerido que existe uma regulação negativa direta entre a motilidade flagelar e a eficiência do T3SS.³⁹ A produção do flagelo requer um investimento considerável das reservas do sistema flagelar e esta operação é dispendiosa em termos energéticos; sendo que estas condições desfavoráveis podem metabolicamente regular negativamente o T3SS.

³⁸ (a) Liu, R.; Ochman, H. Stepwise formation of the bacterial flagellar system. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 7116-7121. (b) Khan, I.H.; Reese, T.S.; Khan, S. The cytoplasmic component of the bacterial flagellar motor. *Proc. Nat. Acad. Sci. USA* **1992**, *89*, 5956-5960.

³⁹ Soscia, C.; Hachani, A.; Bernadac, A.; Filloux, A.; Bleves, S. Cross Talk between Type III Secretion and Flagellar Assembly Systems in *Pseudomonas aeruginosa*. *J. Bacteriol.* **2007**, *189*, 3124-3132.

Neste estudo com *P. aeruginosa*, os autores observaram que existe uma oposição entre a motilidade e o T3SS, de tal forma que o estado móvel da bactéria, correspondendo a uma fase inicial do seu ciclo de infecção, pode ser reprimido quando o T3SS é requerido.³⁹

Apesar da necessidade de elucidar as bases moleculares da relação entre estes dois sistemas, sabe-se que o sistema flagelar, assim como o T3SS, recruta chaperonas de secreção citosólicas, como as proteínas FlgN, FliJ, FliT e FliS (Figura I-4), na montagem do flagelo.⁴⁰

Considerando que estas chaperonas de secreção estão relacionadas aos sistemas de secreção do tipo III e IV e à montagem do flagelo, que são estruturas cujos genes foram identificados no genoma da *Xac*, uma discussão mais aprofundada sobre elas encontra-se na sequência.

1.1.3. Chaperonas de secreção

As chaperonas de secreção representam uma família peculiar de chaperonas moleculares, que não apresentam similaridade ou homologia de sequência primária, mas possuem algumas características comuns.

Chaperonas moleculares, por sua vez, são consideradas uma grande família de proteínas que têm por função assistir outras proteínas no seu enovelamento ou desenovelamento, reenovelamento das proteínas desnaturadas após stress, bem como sua posterior organização em uma estrutura oligomérica biologicamente ativa;⁴¹ sendo, portanto responsáveis

⁴⁰ Blocker, A.; Komoriya, K.; Aizawa, S.-I. Type III secretion systems and bacterial flagella: Insights into their function from structural similarities. *Proc. Nat. Acad. Sci. USA* **2003**, *100*, 3027-3030.

⁴¹ Ellis, R.J. The molecular chaperone concept. *Semin. Cell Biol.* **1990**, *1*, 1-9.

pelo controle de qualidade do proteoma da célula.⁴² Existem muitas famílias diferentes de chaperonas moleculares e cada uma age de maneira peculiar no enovelamento/desenovelamento de proteínas, sendo que novas funções continuam sendo descobertas.

No caso específico das chaperonas de secreção, algumas de suas características comuns são: têm massa molar menor que 20.000 Da, possuem baixo ponto isoelétrico (pI), 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.⁴³ Normalmente, são codificadas por genes localizados próximos aos genes que codificam suas proteínas parceiras, o que ajuda na sua identificação.⁴⁴

Apesar de não apresentarem homologia de estrutura primária, as estruturas cristalográficas de algumas delas mostram que suas estruturas terciárias são bastante similares, dependendo de sua subclasse.⁴⁵

As chaperonas de secreção dos T3SSs são subdivididas em classes, de acordo com suas proteínas parceiras: as chaperonas da classe I (com cerca de 130 resíduos de aminoácidos) se associam a uma (classe IA) ou várias (classe

⁴² (a) Ellis, R.J.; van der Vies, S.M. Molecular chaperones. *Annu. Rev. Biochem.* **1991**, *60*, 321-347. (b) Ali, M.M.; Roe, S.M.; Vaughan, C.K.; Meyer, P.; Panaretou, B.; Piper, P.W.; Prodromou, C.; Pearl, L.H. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **2006**, *440*, 1013-1017.

⁴³ (a) Bennett, J.C.; Hughes, C. From flagellum assembly to virulence: the extended family of type III export chaperones. *Trends Microbiol.* **2000**, *8*, 202. (b) Wattiau, P.; Woestyn, S.; Cornelis, G.R. Customized secretion chaperones in pathogenic bacteria. *Mol. Microbiol.* **1996**, *20*, 255-262.

⁴⁴ Cornelis, G.R.; Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **2000**, *54*, 735-774.

⁴⁵ (a) Page, A.L.; Parsot, C. Chaperones of the type III secretion pathway: jacks of all trades. *Mol. Microbiol.* **2002**, *46*, 1-11. (b) Stebbins, C.E.; Galán, J.E. Priming virulence factors for delivery into the host. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 738-743.

IB) proteínas de virulência, enquanto que as da classe II (com cerca de 160 resíduos de aminoácidos) se associam com duas proteínas translocadoras (proteínas que formam o poro na membrana externa do hospedeiro). As chaperonas das classes IA e II são codificadas por genes próximos aos dos seus substratos e as da classe IB, por outro lado, são codificadas dentro do operon dos componentes do aparato do sistema de secreção.⁴⁶ E, por fim, como o sistema flagelar é considerado o ancestral do T3SS,⁴⁷ esta classificação é estendida e as chaperonas de secreção flagelares constituem a classe III.^{43a} As chaperonas desta última classe apresentam as mesmas características comuns já citadas, entretanto se ligam à região C-terminal de sua(s) proteína(s) parceira(s),⁴⁸ ao contrário das demais que se ligam na região N-terminal das parceiras.

Apesar de não serem classificadas conjuntamente, as chaperonas de secreção dos sistemas de secreção do tipo IV (T4SSs) também partilham das características supracitadas, sendo que a chaperona mais estudada deste tipo é a VirE1, uma proteína pequena, com baixo pI e que previne a oligomerização e agregação da sua parceira, a VirE2.⁴⁹

⁴⁶ Parsot, C.; Hamiaux, C.; Page, A.L. The various and varying roles of specific chaperones in type III secretion systems. *Curr. Opin. Microbiol.* **2003**, *6*, 7-14.

⁴⁷ Macnab, R.M. The bacterial flagellum: reversible rotary propeller and type III export apparatus. *J. Bacteriol.* **1999**, *181*, 7149-7153.

⁴⁸ Fattori, J.; Prando, A.; Martins, A.M.; Rodrigues, F.H.S.; Tasic, L. Bacterial Secretion Chaperones. *Protein Pept. Lett.* **2011**, *18*, 158-166.

⁴⁹ (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.

No geral, as chaperonas de secreção se ligam aos aminoácidos 50-100 da sua proteína parceira, o que significa, portanto, que o seu sítio de ligação é logo após a sequência sinal na extremidade N-terminal. Através das estruturas tridimensionais, observou-se que estas chaperonas mantêm o domínio de ligação na proteína parceira em um estado desenovelado.^{45,50,51,52}

Além da função de proporcionar um estado parcialmente desenovelado à proteína parceira, também foi proposto que a chaperona de secreção tem a importante função de dirigir o encaminhamento do complexo chaperona/proteína parceira para a maquinaria de secreção, dado que a remoção do domínio de ligação da chaperona, em alguns complexos, previne a secreção da proteína parceira.^{48,53} Portanto, a sequência sinal existente na proteína a ser secretada deve ser um sinal geral para a secreção e as chaperonas devem conferir especificidade ao processo.⁵⁴ Por fim, uma terceira função também discutida para estas chaperonas é a de prevenir agregação ou interações prematuras da sua proteína parceira antes da secreção.⁴⁸

⁵⁰ (a) Schesser, K.; Frithz-Lindsten, 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. (b) Sory, M.P.; Boland, A.; Lambermont, I.; Cornelis, G.R. 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.

⁵¹ Woestyn, S.; Sory, M.P.; Boland, M.; Leguenne, O.; Cornelis, G.R. 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.

⁵² Stebbins, C.; Galán, J. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **2001**, *414*, 77-81.

⁵³ Lee, S.H.; Galán, J.E. Salmonella type III secretion-associated chaperones confer secretion-pathway specificity. *Mol. Microbiol.* **2004**, *51*, 483-495.

⁵⁴ Luo, Y.; Betero, M.G.; Frey, E.A.; Pfuetzner, R.A.; Wenk, M.R.; Creagh, L.; Marcus, S.L.; Lim, D.; Sicheri, F.; Kay, C.; Hatnes, C.; Finlay, B.B.; Strynadka, N.C.J. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nat. Struct. Biol.* **2001**, *8*, 1031-1036.

A estrutura cristalográfica das chaperonas de secreção SicP e SigE de *Salmonella entérica* e SycE de *Yersinia pseudotuberculosis* (Figura I-5) confirmaram que as chaperonas da classe IA atuam como dímeros que exibem estrutura tridimensional similar, sendo que cada monômero contém cinco folhas- β e três hélices- α segregadas.^{46,48,55}

Estudos estruturais, por cristalografia de raios-X com as chaperonas de secreção InvB para *Salmonella sp.* (Figura I-5) e Spa15 para *Shigella sp.*, que pertencem a classe IB, também revelaram a atuação das mesmas como dímeros, com apenas algumas diferenças em relação às chaperonas da classe IA, no modo de interação com a parceira, dado que se ligam a mais de um substrato.⁴⁸

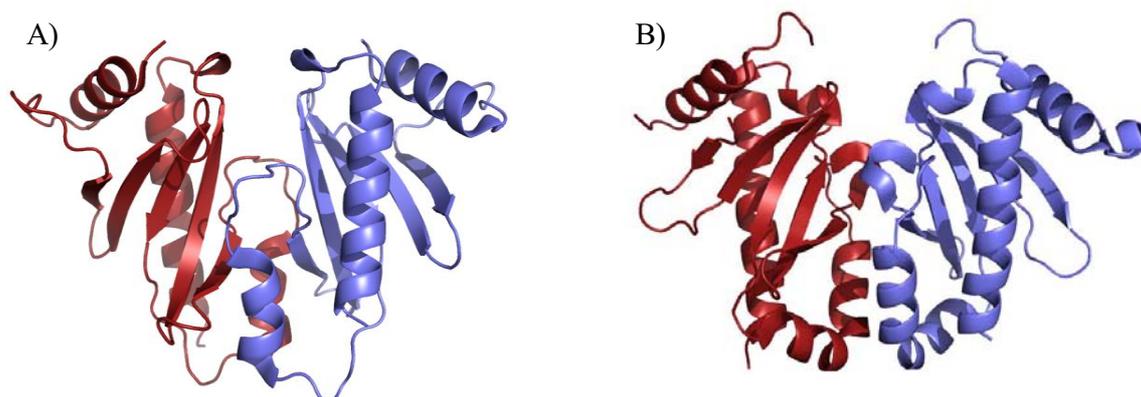


Figura I-5. Estruturas de chaperonas de secreção do T3SS da classe I, diméricas. A) Estrutura da SycE de *Yersinia pseudotuberculosis* (PDB:1JYA).⁵⁶ B) Estrutura da InvB de *Salmonella sp.* (PDB: 2FM8).⁵⁷

⁵⁵ Evdokimov, A.G.; Tropea, J.E.; Routzahn, K.M.; Waugh, D.S. Three-dimensional structure of the type III secretion chaperone SycE from *Yersinia pestis*. *Acta Cryst.* **2002**, D58, 398-406.

⁵⁶ Birtalan, S.; Ghosh, P. Structure of the *Yersinia* type III secretory system chaperone SycE. *Nat. Struct. Biol.* **2001**, 8, 974-978.

⁵⁷ Lilić, M.; Vujanac, M.; Stebbins, C.E. A Common Structural Motif in the Binding of Virulence Factors to Bacterial Secretion Chaperones. *Molecular Cell* **2006**, 21, 653-664.

Conforme se observa na Figura I-5, as proteínas da classe I apresentam estruturas tridimensionais bastante semelhantes, com regiões de hélices- α e folhas- β segregadas ($\alpha+\beta$), num enovelamento do tipo ($\alpha-\beta\beta-\alpha-\beta\beta-\alpha$).

Por outro lado, a estrutura cristalográfica da SycD de *Yersinia enterocolitica* (Figura I-6), chaperona de secreção da classe II, mostrou um enovelamento do tipo helicoidal, diferentemente das chaperonas da classe I.⁵⁸ Entretanto, a dimerização é uma característica comum entre as chaperonas de secreção das classes I e II e apesar de algumas diferenças estruturais destas duas classes de chaperonas de secreção, o modo de interação com a proteína parceira é conservado.^{48,59}

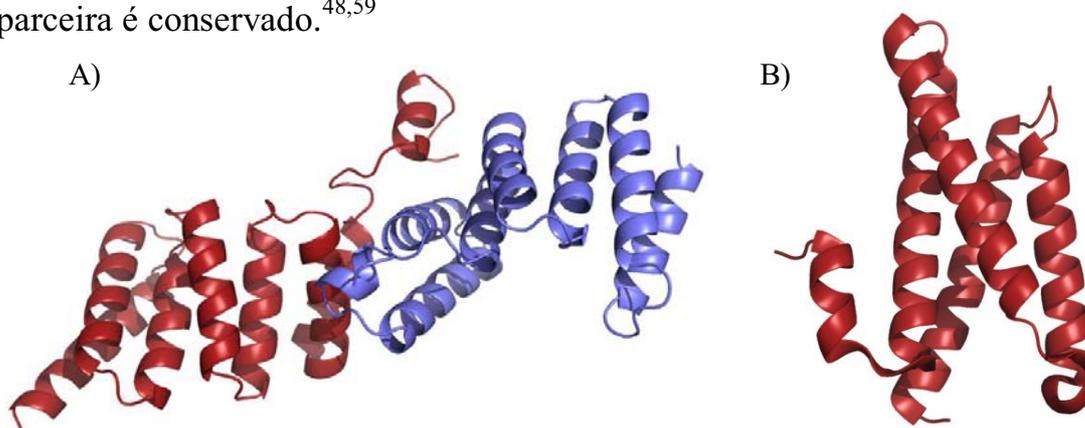


Figura I-6. Estruturas de chaperonas de secreção do T3SS da classe II e III. A) Estrutura da SycD de *Yersinia enterocolitica* (PDB:2VGX), dimérica.⁵⁸ B) Estrutura da FliT de *Salmonella* sp. (PDB: 3A7M).⁶⁰

As chaperonas de secreção da classe III por sua vez apresentam menos dados estruturais, porém cálculos de estrutura secundária sugerem que elas apresentam uma estrutura helicoidal alongada. A estrutura cristalográfica da FliT (Figura I-6), por exemplo, mostrou que esta chaperona consiste de quatro hélices- α e a região central da molécula é formada por hélices antiparalelas.

⁵⁸ 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.

⁵⁹ 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.

Além de que, foi observado que a mesma atua como dímero.⁶⁰ Já no caso da FliS, existem algumas controvérsias. A FliS de *Aquifex aeolicus*, segundo sua estrutura cristalográfica, é formada por quatro hélices- α antiparalelas e atua como monômero; enquanto que a FliS de *Salmonella entérica* atua como dímero, como as demais chaperonas de secreção.⁶¹

No caso das chaperonas de secreção do T4SS, tem-se estudos estruturais da VirE1 de *Agrobacterium tumefaciens*, no complexo com sua parceira, a VirE2.⁶² Apesar de resultados por filtração em gel indicarem a formação de um complexo VirE1-VirE2 na razão 2:1, dados de microscopia eletrônica, SAXS e cristalografia de raios-X mostraram a formação de um heterodímero na razão 1:1 (Figura I-7).^{48,49}

⁶⁰ 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.

⁶¹ (a) 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. (b) 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.

⁶² 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.

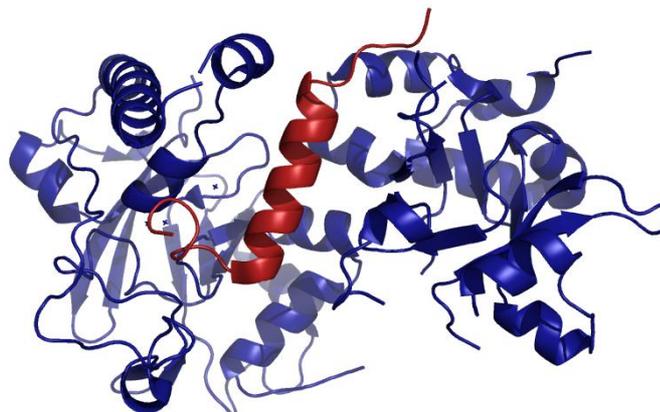


Figura I-7. Estrutura da chaperona de secreção do T4SS, VirE1 (em vermelho) complexada com sua proteína parceira, a VirE2 (em azul; PDB:3BTP).⁶²

A estrutura cristalográfica ilustrada na Figura I-7 mostrou que a VirE1 consiste de uma única hélice- α entre os dois domínios estruturais da VirE2.

Portanto, conforme apresentado, as chaperonas de secreção são proteínas peculiares dos sistemas de secreção dos tipos III e IV e do sistema flagelar, de acordo com os dados até a presente data. E, no genoma da *Xac*, que possui um cromossomo e dois plasmídeos (pXAC33 e pXAC64) foram encontrados domínios que codificam estes dois sistemas de secreção (T3SS e T4SS) e também um flagelo polar.²⁰ Entretanto, não foram encontrados genes que codificam as chaperonas secretórias características destes sistemas.

Porém, dentre as 1294 proteínas hipotéticas preditas no genoma da *Xac*, quarenta foram identificadas como possíveis chaperonas de secreção.^{15,63}

Assim, acredita-se que o conhecimento da estrutura destas chaperonas e, conseqüentemente, o entendimento do papel desenvolvido por elas e pelas

⁶³ Khater, L.; Santos, T.M.; Alegria, M.C.; Docena, C.; da Silva, A.C.R.; Ramos, C.H.I. 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.

demais proteínas que constituem os sistemas secretórios durante a infecção da *Xac* pode auxiliar no estabelecimento das bases moleculares do cancro cítrico e, posteriormente, na prevenção ou eliminação da doença.

1.2 ESTUDOS ESTRUTURAIS DE PROTEÍNAS

1.2 Estudos Estruturais de Proteínas

O conhecimento do arranjo espacial dos aminoácidos da cadeia polipeptídica é extremamente importante, pois a estrutura de uma proteína está diretamente relacionada com sua função. E esta função determina o interesse científico e econômico pelas proteínas, em diversas aplicações, como no entendimento das vias bioquímicas nos processos biológicos para aplicações industriais, na agricultura e também na medicina.⁶⁴

A estrutura primária de uma proteína, ou seja, a sequência dos aminoácidos é que determina o arranjo espacial deles através de impedimento estérico e das interações entre eles, que podem ser: eletrostáticas, hidrofóbicas, interações de Van der Waals e ligações de hidrogênio.⁶⁵

De acordo com estas interações, principalmente ligações de hidrogênio são formados os elementos da estrutura secundária das proteínas: hélices- α , folhas- β , voltas e regiões randômicas. Estes elementos de estrutura secundária, por sua vez, interagem entre si formando a estrutura terciária da proteína. Muitas vezes para desempenharem suas funções, algumas proteínas atuam como oligômeros ou requerem a ligação de um cofator gerando sua estrutura quaternária.⁶⁴

Para determinar a estrutura tridimensional das proteínas, as duas técnicas empregadas, que fornecem detalhes em nível atômico, são a espectroscopia de Ressonância Magnética Nuclear (NMR) e a cristalografia de raios-X. Ressaltando que o número de estruturas depositadas no banco de dados de proteínas pela técnica de cristalografia é maior que o número de estruturas depositadas por NMR.

⁶⁴ Nelson, D.L.; Cox, M.M. *Lehninger Principles of Biochemistry*, 4th Ed.; W.H. Freeman & Company: New York, 2004; Capítulo 3.

⁶⁵ Melnikov, E.E.; Rotanova, T.V. *Molecular Chaperones. Russ. J. Bioorg. Chem.* **2010**, *36*, 1-10.

Essa diferença no número de estruturas resolvidas por cristalografia de raios-X e NMR ocorre por razões como o fato do poder de resolução desta última ser menor que o da cristalografia de raios-X, uma vez que a informação obtida para uma mesma proteína é muito mais complexa quando se utiliza NMR. Outro fato importante é o limite de massa molecular encontrado na NMR, pois o complexo de maior massa molecular cuja estrutura foi resolvida com sucesso por NMR, apresenta 64 kDa. Existe ainda o fato de que para alguns conjuntos de dados obtidos podem-se determinar duas ou mais conformações possíveis para uma proteína. Além de que o custo do aparato experimental da NMR aumenta com a complexidade do sistema em estudo.⁶⁶

Outras técnicas como a Microscopia Eletrônica Criogênica (*Cryo Electron-Microscopy* – Cryo-EM) e Espalhamento de raios-X a baixo ângulo (*Small Angle X-ray Scattering* – SAXS) vêm sendo bastante utilizadas atualmente na determinação da estrutura de proteínas. Informações estruturais complementares e de grande importância no entendimento dos modelos obtidos pelas técnicas supracitadas podem ser fornecidas por diferentes técnicas como: espectroscopia no Infravermelho (IR), no Ultravioleta/Visível (UV/VIS), Raman, Dicroísmo Circular (CD), Ressonância de Spin Eletrônico, Fluorescência de emissão, Espectrometria de Massas (MS), Ultracentrifugação Analítica (AUC), Filtração em gel Analítica e técnicas calorimétricas (Calorimetria de Titulação Isotérmica – ITC e Calorimetria Diferencial de Varredura – DSC).

As técnicas de caracterização estrutural utilizadas neste trabalho estão descritas mais detalhadamente na sequência.

⁶⁶ Página da *University Medical School of Debrecen*. <http://www.cryst.bbk.ac.uk/pps97/assignments/projects/ambrus/html.htm> (acessado em Setembro de 2011).

1.2.1. Técnicas auxiliares

Na análise de estrutura secundária de proteínas, uma técnica utilizada frequentemente é a **espectroscopia de dicroísmo circular (CD)**, na região do ultravioleta distante (*far-UV*), 190-240 nm.⁶⁷ Esta técnica baseia-se na diferença de absorção da luz circularmente polarizada a esquerda e a direita, sendo que em proteínas a ligação peptídica é a responsável por essa diferença na absorção. Os diferentes arranjos da cadeia primária das proteínas (hélices- α e/ou folhas- β) interagem de modo peculiar com a radiação polarizada gerando espectros de CD com a elipticidade (θ em graus) em função do comprimento de onda (Figura I-8).

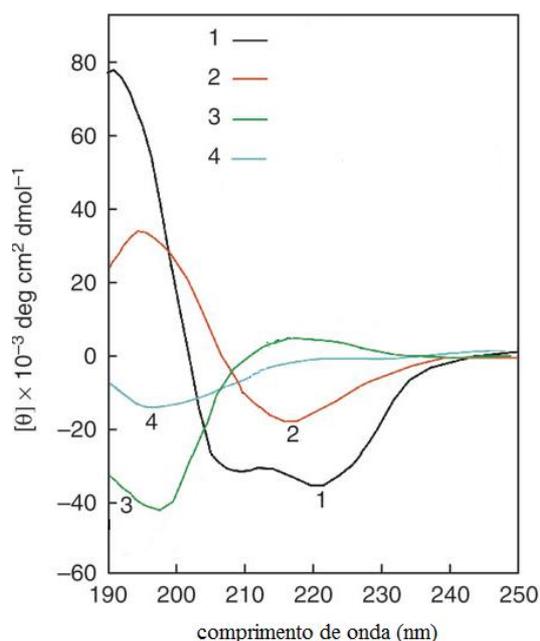


Figura I-8. Ilustração de espectros de CD para diferentes tipos de estrutura secundária de proteínas, sendo em preto (1) exemplo de estrutura exclusivamente helicoidal, em vermelho (2) de folhas- β antiparalelas, em verde (3) de estrutura randômica e em azul (4) da estrutura desenovelada do colágeno.⁶⁸

⁶⁷ Walla, P.J. *Modern Biophysical Chemistry – Detection and Analysis of Biomolecules*; Willey-VCH: Weinheim, 2009; Capítulo 4.

⁶⁸ Adaptado de Greenfield, N.J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **2007**, *1*, 2527-2535.

Essa diferença na absorção da luz circularmente polarizada se deve à assimetria da ligação peptídica, sendo que a função amida das proteínas possui duas transições eletrônicas de baixa energia, bem caracterizadas pelos sinais negativos de CD na região de 215-230 nm, referente à transição $n \rightarrow \pi^*$ e positivos na região de 190-200 nm referente à transição $\pi_0 \rightarrow \pi^*$.⁶⁹

Conforme se observa na Figura I-8, uma proteína composta apenas por hélices- α apresenta dois mínimos de magnitude semelhantes em 208 e 222 nm, sendo que o sinal em 222 nm está relacionado com as ligações de hidrogênio presentes neste tipo de estrutura (número 1 na Figura I-8). Já as proteínas compostas apenas por folhas- β apresentam um sinal negativo entre 210-220 nm e um sinal positivo entre 195-200 nm, porém os espectros deste tipo de proteínas são mais diversos, pois podem existir folhas- β antiparalelas, paralelas ou uma mistura das duas (número 2 na Figura I-8). E as proteínas com estrutura randômica ou desordenada apresentam sinal negativo intenso na região de 200 nm (número 3 na Figura I-8). Enquanto as proteínas desenoveladas apresentam sinal negativo na região de 200 nm (número 4 na Figura I-8).⁷⁰

A técnica de CD pode ser utilizada em estudos conformacionais de proteínas, informando sobre o tipo de enovelamento, *i.e.*, estrutura secundária, e a predominância de estruturas randômicas, ou hélices- α , ou folhas- β ; e em estudos de mudanças conformacionais com variações temperatura e pH, por exemplo.^{69,70}

Quando se utiliza a espectroscopia de dicroísmo circular na região do UV de vácuo (abaixo de 190 nm) podem ser obtidas informações adicionais

⁶⁹ Corrêa, D.H.A.; Ramos, C.H.I. The use of circular dichroism spectroscopy to study protein folding, form and function. *African J. Biochem. Res.* **2009**, 3, 164-173.

⁷⁰ Kelly, S.M.; Price, N.C. The application of circular dichroism to studies of protein folding and unfolding. *Biochim. Biophys. Acta* **1997**, 1338, 161-185.

sobre a estrutura das proteínas. Entretanto, medidas de CD nesta região de comprimento de onda são limitadas em um espectropolarímetro convencional devido à absorção alta da amostra e a baixa intensidade da fonte de radiação. A utilização da luz síncrotron como fonte de radiação, desenvolvida em 1980, acompanhada da evolução da instrumentação resultaram no desenvolvimento recente de espectrômetros de CD com radiação síncrotron que são adequados para analisar conformações e enovelamento de proteínas em uma ampla gama de condições. Fato que possibilitou, pela primeira vez, a obtenção de espectros de proteínas em soluções aquosas em regiões de comprimentos de onda tão baixos quanto 160 nm.⁷¹

Esta técnica melhora a utilidade da técnica de CD em biologia estrutural, pois o alto fluxo do feixe de radiação tem importantes consequências espectrais nos estudos de proteínas. A principal consequência é a possibilidade da aquisição de dados em comprimentos de onda mais baixos, que são mais sensíveis às diferenças na estrutura secundária; e em comprimentos de onda muito baixos, cujos sinais referentes às transições com transferências de carga contêm informações sobre a estrutura terciária.⁷²

⁷¹ Wallace, B.A.; Janes, R.W. Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics. *Curr. Opin. Chem. Biol.* **2001**, *5*, 567-571.

⁷² Wallace, B.A.; Janes, R.W. Circular dichroism and synchrotron radiation circular dichroism spectroscopy: tools for drug discovery. *Biochem. Soc. Trans.* **2003**, *31*, 631-633.

Outra técnica espectroscópica que já foi bastante utilizada em estudos estruturais de proteínas, principalmente na caracterização da estrutura secundária, é a espectroscopia no infravermelho com transformada de Fourier (**FTIR** – do inglês *Fourier transform infrared*).

A espectroscopia na região do infravermelho é baseada nas vibrações moleculares e a energia destas vibrações corresponde à região do infravermelho do espectro eletromagnético.

A utilização desta técnica permite identificar grupos funcionais como O–H e C=O, por exemplo, que apresentam frequências de vibrações características. Os grupos com frequências de vibração características, interessantes na caracterização estrutural de proteínas incluem C=O, –COOH, O–H e N–H, porém espectros de infravermelho de proteínas são mais conhecidos pela presença de frequências de vibração do grupo amida.

As bandas largas nos espectros de infravermelho de proteínas referente ao grupo amida (ligação peptídica) são resultado da sobreposição da absorção de vários componentes que representam diferentes elementos estruturais como hélices- α , folhas- β , voltas ou estruturas desordenadas. A questão da sobreposição de bandas pode ser resolvida no processo de deconvolução, aplicando a transformada de Fourier e calculando a segunda derivada da curva no espectro. Estes procedimentos permitem identificar diferentes estruturas presentes em uma proteína e alterações conformacionais da mesma.⁷³ E, para obtenção da informação a respeito da fração de cada elemento de estrutura secundária em uma proteína foram desenvolvidos diferentes métodos, dentre os quais os mais populares são o de ajuste das curvas e reconhecimento de padrão.

⁷³ Haris, P.I.; Severcan, F. FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media. *J. Mol. Catal. B: Enzym.* **1999**, *7*, 207-221.

Apesar de algumas facilidades da espectroscopia no infravermelho, como a obtenção de informação estrutural de proteínas em diferentes ambientes, por exemplo, em solução aquosa, em solventes orgânicos ou na forma de cristais; e o fato de não haver efeitos fluorescentes ou de espalhamento de luz na aquisição dos dados; ela apresenta uma grande desvantagem que é a banda larga de absorção de água que se sobrepõe as bandas dos grupos amida.⁷⁴

Assim, com o grande desenvolvimento instrumental de outras técnicas bem mais práticas como CD que provem o mesmo tipo de informação, porém com espectros de mais fácil interpretação, a espectroscopia no infravermelho não tem sido muito utilizada atualmente nos estudos estruturais de proteínas.

Uma técnica que pode ser utilizada como sonda de estrutura terciária, porém fornecendo dados não tão detalhados quanto os fornecidos por NMR e cristalografia, é a **fluorescência de emissão**. A fluorescência é resultado de um processo de três estágios que ocorre com algumas moléculas chamadas de fluoróforos. Em um primeiro estágio, ocorre a excitação da molécula, quando um fóton de energia $h\nu_{\text{ex}}$ fornecido, por exemplo, por um laser, é absorvido pelo fluoróforo criando um estado eletrônico singlete excitado (S_1' , Figura I-8). Este estado excitado existe por um tempo de tipicamente 1-10 ns e durante este tempo o fluoróforo sofre mudanças conformacionais e interage com o ambiente, sendo a energia do estado excitado dissipada levando a um estado singlete intermediário (S_1 , Figura I-8) a partir do qual se origina a fluorescência de emissão; porém nem todas as moléculas excitadas retornam ao estado fundamental por emissão de

⁷⁴ Zuber, G.; Prestrelski, S.J.; Benedek, K. Application of Fourier transform infrared spectroscopy to studies of aqueous protein solutions. *Anal. Biochem.* **1992**, *207*, 150-156.

fluorescência. O terceiro estágio consiste na fluorescência de emissão, quando um fóton de energia $h\nu_{em}$ é emitido, retornando o fluoróforo ao seu estado fundamental (S_0 , Figura I-9). Ressaltando que como ocorre a dissipação de energia durante o tempo de vida do estado excitado, a energia do fóton emitido é menor que a do fóton absorvido; sendo esta diferença de energia ($h\nu_{ex} - h\nu_{em}$) chamada de deslocamento de Stokes.⁷⁵

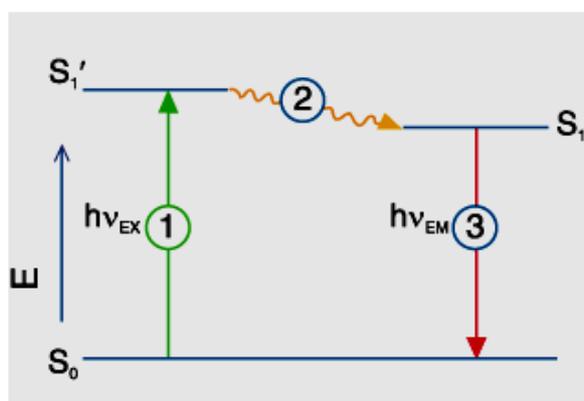


Figura I-9. Ilustração do diagrama de energia (Diagrama de Jablonski) para o fenômeno de fluorescência, onde E corresponde a energia, S_0 ao estado fundamental, S_1' ao estado excitado singlete e S_1 ao estado singlete; 1, 2 e 3 referem-se aos estágios do processo de fluorescência descritos no texto.⁷⁶

A espectroscopia de fluorescência de emissão pode ser utilizada em estudos conformacionais de proteínas, dado que os aminoácidos fenilalanina, tirosina e triptofano apresentam propriedades fluorescentes, devido às duplas ligações conjugadas em suas estruturas e conseqüentemente a deslocalização eletrônica. As cistinas, formadas pela ligação de duas cisteínas também apresentam esta propriedade. Entretanto apenas o triptofano é utilizado experimentalmente como sonda de fluorescência intrínseca devido ao seu rendimento quântico ser alto o suficiente para a obtenção de um sinal visível

⁷⁵ Lackovicz, J.R. *Principles of Fluorescence Spectroscopy*; Kluwer Academy/Plenum Publisher: New York, 1983.

⁷⁶ Adaptado de: <http://www.physics.nus.edu.sg/~Biophysics/pc3267/Fluorescence%20Spectroscopy%202007.pdf> (Acessado em agosto de 2011).

de emissão. Sendo que por rendimento quântico entende-se a razão do número de fótons emitidos pelo número de fótons absorvidos.⁷⁵ O triptofano (Trp-W) apresenta uma absorção por volta de 295 nm (sem interferência da tirosina) e emissão em torno de 340 nm e é altamente sensível ao ambiente, com características de emissão de fluorescência bem diferentes, de acordo com o grau de exposição ao meio hidrofílico ou hidrofóbico.

Assim, quando o resíduo de triptofano encontra-se mais exposto ao solvente apresenta um máximo de emissão ao redor de 350-355 nm, enquanto o Trp (W) em um ambiente hidrofóbico, no estado nativo da proteína possui um máximo em ~335 nm, fazendo com que esta técnica possa ser utilizada em estudos de mudanças conformacionais de proteínas.

A espectrometria de massas (MS) também é uma técnica poderosa e muito utilizada para estudos de proteínas quanto à pureza, massa molecular, formação de oligômeros e sequenciamento. Com o aumento do desempenho e versatilidade da instrumentação de MS, rotineiramente surgem novas estratégias de análise de proteínas tendo a espectrometria de massas como elemento central.⁷⁷ O uso da espectrometria de massas para identificar, sequenciar e investigar a estrutura covalente das proteínas foi acelerado com o projeto Genoma, sendo a MS aplicada principalmente no lugar da degradação de Edman. O desenvolvimento de fontes de ionização mais brandas, capazes de gerar íons moleculares intactos das proteínas como *electrospray* (ESI) e desorção e ionização a laser assistida por matriz (MALDI) foi o maior responsável pela grande utilização da MS na análise de proteínas. Isto associado ao desenvolvimento de novos analisadores de massas e analisadores multiestágios, como o híbrido quadrupolo-tempo de vôo ou o tempo de vôo

⁷⁷ Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **2006**, *312*, 212-217.

sequencial (Q-ToF e ToF-ToF) fez da espectrometria de massas a técnica central no surgimento da Proteômica.

No geral, a análise proteômica baseada em MS é feita em três estágios: primeiro a proteína é isolada da sua fonte biológica podendo ser fracionada ou não e digerida; os peptídeos resultantes da digestão são então submetidos à análise de espectrometria de massas qualitativa e quantitativa e por fim o grande grupo de dados gerado é analisado em programas apropriados para dedução da sequência de aminoácidos e/ou quantificação da proteína na amostra.⁷⁷ A identidade dos peptídeos é assinalada aos espectros de MS/MS através de buscas em bancos de dados, realizada de maneira a gerar resultados consistentes e a análise estatística dos resultados é crítica para garantir confiabilidade nas identificações.⁷⁸

As amostras para análise por MALDI, por exemplo, podem ser preparadas de várias maneiras, sendo mais utilizado o método de evaporação de solvente em matriz, na qual a matriz é um composto orgânico pequeno, facilmente ionizável e que permite a ionização branda da proteína. E 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.

Portanto a análise de proteínas, especialmente a proteômica, dirigiu o desenvolvimento da MS na última década, visando o melhoramento na exatidão das massas e no poder de resolução, entre outros. Entretanto, as

⁷⁸ (a) Baldwin, M.A. Protein identification by mass spectrometry: Issues to be considered. *Mol. Cell Proteomics* **2004**, *3*, 1-9. (b) Carr, S. The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell Proteomics* **2004**, *3*, 531-533. (c) Keller, A.; Nesvizhskii, A.I.; Koller, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383-53-92.

novas tecnologias em MS advindas da demanda proteômica podem ser aplicadas a outros tipos de biomoléculas como metabólitos, lipídeos e carboidratos.⁷⁷ Ressaltando que recentemente a espectrometria de massas vem sendo utilizada em estudos estruturais de proteínas, através de experimentos de ligação cruzada, onde podem ser determinadas algumas restrições espaciais entre os pares ligados pelo agente de ligação cruzada. E, utilizando modelagem molecular podem ser obtidas estruturas tridimensionais que satisfaçam tais restrições.⁷⁹

A filtração em gel analítica é uma técnica cromatográfica onde a separação é feita baseada no volume hidrodinâmico das moléculas. A fase estacionária utilizada é uma matriz polimérica com poros de diferentes tamanhos responsáveis pela separação das moléculas. Sendo que as moléculas de menor volume entram nos poros da matriz e, portanto levam um tempo maior para saírem da coluna. Porém esta técnica pode ser utilizada para se obter informações estruturais de proteínas, uma vez que a partir do coeficiente de partição pode se estimar o raio hidrodinâmico ou raio de Stokes da proteína.

Esta técnica também é bastante utilizada para estimar a massa molecular de proteínas, entretanto faz-se necessária a utilização de curvas de calibração que expressam a dependência do volume de eluição em relação ao logaritmo da massa molecular. De fato, o parâmetro que governa o volume de eluição ou o coeficiente de partição da molécula não é sua massa molecular, mas sim o seu raio de Stokes.⁸⁰ Assim, a massa molecular só pode ser calculada por

⁷⁹ Singh, P.; Panchaud, A.; Goodlett, D.R. Chemical Cross-Linking and Mass Spectrometry As a Low Resolution Protein Structure Determination Technique. *Anal. Chem.* **2010**, *82*, 2636-2642.

⁸⁰ Winzor, D.J. Analytical exclusion chromatography. *J. Biochem. Biophys. Methods* **2003**, *56*, 15-52.

filtração em gel quando se utiliza um segundo experimento, como por exemplo, a ultracentrifugação analítica, também baseado no raio de Stokes. Porém, como na construção da curva de calibração geralmente são utilizadas proteínas globulares, pode se obter uma estimativa da massa molecular a partir desta técnica pois para proteínas globulares a relação entre o volume de eluição (coeficiente de partição - k_{av}) e o logaritmo da massa molecular é linear (Figura I-10).

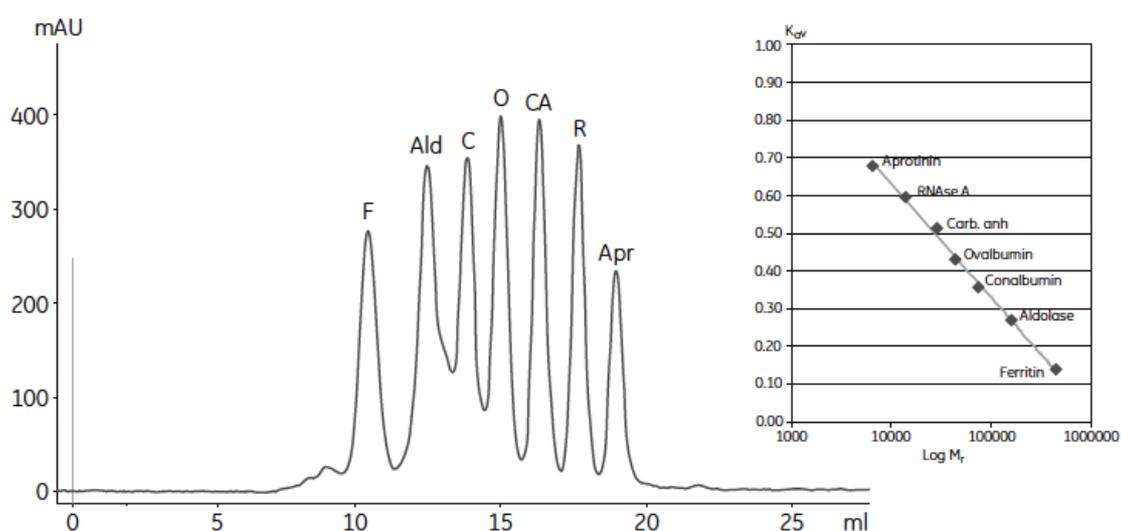


Figura I-10. Ilustração de um cromatograma de filtração em gel e a relação linear entre o coeficiente de partição (k_{av}) e o logaritmo da massa molecular para proteínas globulares.⁸¹

O mesmo é válido para estimativas do raio de Stokes (R_s) de uma proteína, sendo que normalmente o raio de Stokes das proteínas utilizadas na calibração é determinado por experimentos que medem o coeficiente de difusão (D) da molécula e através da Equação (Eq. 1) de Stokes-Einsten se obtém o R_s .⁸²

$$D = kT/6\pi\eta R_s \quad (\text{Eq.1})$$

⁸¹ Adaptado de: [http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/0548AC9B0B2AA8DAC1257628001D30DB/\\$file/28951560AD.pdf](http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/0548AC9B0B2AA8DAC1257628001D30DB/$file/28951560AD.pdf) (Acessado em Agosto de 2011).

⁸² Erickson, H.P. Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proced. Online* **2009**, *11*, 32-51.

onde: k é a constante de Boltzman, T a temperatura e η é a viscosidade da solução.

O coeficiente de difusão pode ser determinado experimentalmente por técnicas como **espalhamento de luz dinâmico (DLS)**, por exemplo, ou técnicas de difusão de Ressonância Magnética Nuclear (DOSY-NMR), discutidas na sequência.

Além do coeficiente de difusão, a técnica de espalhamento de luz dinâmico (DLS) também permite obter informações do estado oligomérico da proteína em solução, tendência em agregação, raio hidrodinâmico da mesma e conseqüentemente sua massa molecular, além de uma estimativa da forma da molécula. De maneira simplista esta técnica mede o movimento Browniano e o relaciona com o tamanho das partículas.

Partículas dispersas em um meio líquido movem-se ao acaso (movimento Browniano). Para uma dispersão de partículas esféricas, o coeficiente de difusão, como visto na Eq. 1, é inversamente proporcional ao raio hidrodinâmico ou raio de Stokes. Devido a este movimento, a intensidade da luz espalhada por um conjunto de partículas em solução sofre flutuações ao longo do tempo. E, portanto, a intensidade da luz que atinge o detector sofre flutuações devido ao movimento das partículas.

Desse modo, existe uma relação entre a velocidade de flutuação da luz espalhada e o coeficiente de difusão das partículas.⁸³

Na prática incide-se radiação laser sobre a solução e analisa-se a intensidade das flutuações da luz espalhada em determinado ângulo

⁸³ (a) Chu, B. *Laser Light scattering: Basic Principles and Practice*, 2nd ed.; Dover Publications: USA, 2007. (b) Lomakin, A.; Teplow, D.B.; Benedek, G.B. *Quasielastic light scattering for protein assembly studies*, Amyloid proteins: Methods and Protocols. Methods in molecular biology, vol. 299. Sigurdsson, E.M., Ed.; Humana Press Inc.: Totowa, NJ, 2005; pp 153-174.

fornecendo uma curva com a distribuição de tamanho de partículas em solução.

A partir destes dados de raio hidrodinâmico e coeficiente de difusão também é possível ter-se ideia da forma da molécula, uma vez que, moléculas alongadas e volumosas apresentam maior resistência ao movimento de difusão e, portanto, um menor coeficiente de difusão (D) que as moléculas compactas ou esféricas. Esta resistência ao movimento é devido a uma maior fricção ou atrito com o sistema. Tal força de resistência citada, denominada de força friccional, depende do coeficiente friccional (f) que é uma constante da molécula que relaciona seu raio hidrodinâmico (R_H) e a viscosidade da solução (η).

Assim, partindo do princípio de que uma molécula esférica e compacta sofre a menor fricção de todas as formas possíveis para uma molécula, e conseqüentemente apresenta o maior coeficiente de difusão possível (D), pode-se calcular seu coeficiente friccional (f_0) em função do raio de uma esfera compacta e lisa (R_0) através da seguinte equação (Eq. 2):

$$R_0 = (3 \cdot MM \cdot V_{\text{bar}} / 4\pi N)^{1/3} \quad (\text{Eq. 2, sendo } f_0 = 6\pi\eta R_0)$$

onde: MM é a massa molar da proteína, V_{bar} o volume parcial específico de uma molécula e N é número de Avogrado.

Portanto, considerando que a maioria das proteínas e biomoléculas não são esféricas e sim com formato elipsóide, tem-se que o coeficiente de difusão do elipsóide é menor que o da esfera, para o mesmo volume. Assim, sendo R_0 o raio de uma esfera não hidratada de mesma massa molar, tem-se que o raio hidrodinâmico (R_H) obtido por experimentos de filtração em gel ou DLS, por exemplo, será sempre maior que R_0 . Conhecendo então o raio hidrodinâmico

experimental (R_H) de uma proteína, e conseqüentemente seu coeficiente friccional (f), é possível avaliar se a mesma é esférica ou globular a partir da razão friccional (f/f_0).

Sendo f_0 o coeficiente friccional de uma esfera lisa com certa massa molar, proteínas globulares sempre apresentarão valores de f/f_0 próximos de um (dificilmente será igual a um, devido a solvatação das moléculas). Entretanto proteínas de mesma massa molar, porém com formatos diferentes de uma esfera (alongadas ou achatadas), apresentaram valores de f/f_0 maiores que um.⁸⁴ Fazendo com que através de experimentos como DLS, filtração em gel ou AUC possam ser úteis na especulação do tamanho e forma das proteínas.

Ressaltando que R_H e conseqüentemente f também pode ser obtido através de experimentos de difusão por NMR, aplicando a Eq.1.

Conforme citado anteriormente, a ultracentrifugação analítica (AUC) é uma técnica que pode ser utilizada para complementar resultados da filtração em gel analítica.⁸⁵

A AUC é utilizada para determinar a massa molar de proteínas, bem como estados de oligomerização e agregação. Ela é considerada uma técnica absoluta, pois, ao contrário da filtração em gel analítica, não há necessidade de calibração no processo de análise, o que diminui os erros nas medidas.⁸⁵

Existem basicamente duas metodologias principais utilizadas na AUC, velocidade de sedimentação e sedimentação em equilíbrio. A primeira se baseia na sedimentação controlada de macropartículas através do uso de altas

⁸⁴ Kaszynska, K.; Banachowicz, E.; Slósarek, G.; Morawiec, A.; Gawronska, I.; Barciszewski, J. Dynamic Light Scattering and NMR Studies of Napin. *J. of Solution Chem.* **2002**, *13*, 987-993.

⁸⁵ Carvalho, F.A.O. Estudo da hemoglobina extracelular gigante de *Glossoscolex paulistus* (HbGp) por ultracentrifugação analítica e fluorescência em função do pH. Dissertação de Mestrado, Instituto de Química, Universidade de São Paulo, São Carlos, SP, 2009.

velocidades de centrifugação e a sedimentação em equilíbrio se fundamenta em princípios termodinâmicos, sendo a sedimentação sensível a massa molar da proteína e não a sua forma. Por isso, o uso desta última torna possível a determinação da massa molar da proteína, independente de sua forma.⁸⁶

Independente da metodologia utilizada, na AUC o sistema de detecção utilizado é um sistema ótico, que permite acompanhar o gradiente de concentração da macromolécula na solução, sendo este gradiente causado pela aplicação de uma força gravitacional. O deslocamento (sedimentação) da proteína durante a ultracentrifugação é analisado através de medidas de absorvância ou índices de refração, em comprimentos de onda específicos em função do tempo. Ressaltando que os comprimentos de onda mais utilizados nas análises são na região do espectro em que as ligações peptídicas absorvem radiação (210-230 nm) e 280 nm para as proteínas contendo aminoácidos aromáticos em sua estrutura primária.^{85,86}

Assim, à medida que a proteína sedimenta, ocorre uma variação na concentração da amostra ao longo do tubo. E, esta diferença de concentração em função do tempo é medida como uma diferença de absorção.

Utilizando programas computacionais adequados é possível obter informação sobre a forma da partícula através das relações já discutidas entre o coeficiente de sedimentação (S), o coeficiente de difusão (D) e o coeficiente friccional (f).⁸⁶

⁸⁶ Borges, J.C.; Ramos, C.H.I. Analysis of Molecular Targets of *Mycobacterium tuberculosis* by Analytical Ultracentrifugation. *Curr. Med. Chem.* **2011**, *18*, 1276-1285.

1.2.2. NMR-DOSY

A técnica de Ressonância Magnética Nuclear (NMR) pode ser utilizada para estudar os movimentos moleculares, em especial a difusão molecular de líquidos, através de medidas de tempo de relaxação de spin nuclear, dado que, estes dependem do tempo de reorientação molecular e, conseqüentemente dos seus processos de difusão.⁸⁷

As medidas de difusão por NMR são baseadas em experimentos que envolvem a refocalização da magnetização em campos não homogêneos, chamados de experimentos de eco de spin, onde se aplica um pulso de gradiente de campo na amostra, cuja função é produzir a força de um campo magnético linearmente dependente da posição. Assim, após a aplicação de um pulso de gradiente, os spins ficam com a fase codificada espacialmente.

Devido a algumas dificuldades apresentadas, o experimento de eco de spin foi melhorado com a aplicação de eco de spin com gradiente de campo pulsado, que permite determinar a posição dos núcleos (spin) com o primeiro gradiente e medir a extensão de seus movimentos no intervalo Δ entre os dois gradientes de campo pulsados.^{87b} Esta metodologia (PFGSE) pode então, ser estendida de maneira multidimensional através do incremento da amplitude do gradiente de campo (g) e /ou de sua duração (δ). Sendo que, o conjunto de dados obtidos, após processamento, pode ser apresentado como uma matriz bi ou tridimensional, com os coeficientes de difusão (D) ao longo de um dos eixos.^{87b,88}

⁸⁷ (a) Antalek, B. Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results. *Concepts Magn. Reson.* **2002**, *14*, 225-258. (b) Laverde Jr., A. Estudos de RMN Aplicados à Química Supramolecular e Discriminação Enantiomérica. Tese de Doutorado, Instituto de Química, Universidade Estadual de Campinas, Campinas, SP, 2001.

⁸⁸ Altieri, A.S.; Hinton, D.P.; Byrd, R.A. Association of Biomolecular Systems *via* Pulsed Field Gradient NMR Self-Diffusion Measurements. *J. Am. Chem. Soc.* **1995**, *117*, 7566-7567.

A espectroscopia de Ressonância Magnética Nuclear orientada por difusão (DOSY) foi desenvolvida por Morris e Johnson,⁸⁹ sendo que o método combina os ecos de spin com esquemas de inversão de dados gerando uma visão bidimensional, na qual o espectro de NMR de ^1H é mostrado em uma dimensão enquanto na outra dimensão tem-se o ‘espectro’ dos coeficientes de difusão.^{87,90} Ao contrário da DLS a técnica de DOSY detecta todos os componentes que apresentam núcleos ativos em NMR (normalmente ^1H), indiferente de seus tipos, utilizando-se da seletividade da NMR para distinguir espécies moleculares.⁹¹

As aplicações mais comuns dos experimentos de DOSY envolvem a determinação da distribuição de espécies, de acordo com suas massas moleculares em um meio, ou estudos da interação de solutos com compostos macromoleculares, entretanto, utilizando a Equação de Stokes-Einstein (Eq. 1) estes experimentos podem ser aplicados em estudos de biomoléculas em solução,⁹² como, por exemplo, no estudo de mudanças conformacionais de proteínas em solução⁹³ e na determinação do estado oligomérico de proteínas em solução,⁹⁴ como é o caso apresentado no presente estudo.

⁸⁹ Morris, K.F.; Johnson Jr., C.S. Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **1992**, *114*, 3139-3141.

⁹⁰ Macchioni, A.; Ciancaleoni, G.; Zuccaccia, C.; Zuccaccia, D. Determining accurate molecular sizes in solution through NMR diffusion spectroscopy. *Chem. Soc. Rev.* **2008**, *37*, 479-489.

⁹¹ Lin, M., Shapiro, M., Wareing, J. Diffusion-Edited NMR–Affinity NMR for Direct Observation of Molecular Interactions. *J. Am. Chem. Soc.* **1997**, *119*, 5249-5250.

⁹² De Souza, A.A.; Laverde Jr., A. Aplicação da Espectroscopia de Ressonância Magnética Nuclear para Estudos de Difusão Molecular em Líquidos: A Técnica DOSY. *Quim. Nova* **2002**, *25*, 1022-1026.

⁹³ Tillet, M.L.; Lian, L.-Y.; Norwood, T.J. Practical Aspects of the Measurement of Diffusion of Proteins in Aqueous Solution. *J. Magn. Reson.* **1998**, *133*, 379-384.

⁹⁴ Krishnan, V.V. Determination of Oligomeric State of Proteins in Solution from Pulsed-Field-Gradient Self-Diffusion Coefficient Measurements. A Comparison of Experimental, Theoretical, and a Hard-Sphere Aproximated Values. *J. Magn. Reson.* **1997**, *124*, 468-473.

1.2.3. SAXS

A técnica de espalhamento de raios-X a baixo ângulo (ângulos menores que 3-5°) também tem sido utilizada no estudo de proteínas em solução. O tipo de espalhamento em SAXS é um processo elástico que ocorre quando um feixe de raios-X atravessa a amostra e interage com seus elétrons, onde a radiação reemitida por esses elétrons é espalhada isotropicamente e as ondas espalhadas interferem entre si, cancelando-se totalmente em algumas direções.

A técnica de SAXS fornece informações estruturais sobre a não homogeneidade da partícula com dimensões características de dez a algumas centenas de angstroms (Å).⁹⁵ As não homogeneidades ou flutuações na densidade eletrônica são tidas como centros espalhadores e, supondo que os centros espalhadores estão no vácuo, a amplitude do espalhamento é proporcional ao número de mols de elétrons por unidade de volume, ou seja, à densidade eletrônica. Se os centros espalhadores estão imersos em outro meio, como é o caso de partículas em solução, o espalhamento surge do contraste de densidade eletrônica entre a partícula e o solvente.⁹⁶ Em proteínas, este contraste pode ser pequeno, resultando em um espalhamento muito fraco exigindo maior tempo de coleta ou a utilização de soluções protéicas mais concentradas (o que poderia prejudicar o padrão de espalhamento, uma vez que as proteínas poderiam se agregar). Desta forma, feixes de raios-X produzidos pela radiação síncrotron são empregados, pois o fluxo de fótons do feixe que incide sobre a amostra é muito maior em comparação aos raios-X obtidos de forma convencional, propiciando um melhor padrão de

⁹⁵ da Silva, J.C. Estudo de mudanças conformacionais de macromoléculas em solução usando espalhamento de raio-X. Dissertação de Mestrado, Instituto de Física, Universidade Estadual de Campinas, SP, Fevereiro de 2009.

⁹⁶ (a) Glatter, O.; Kratky, O., Eds. *Small Angle X Ray Scattering*; Academic Press Inc.: Londres, 1982; Capítulo 5. (b) Glatter, O. X-Ray Small Angle Scattering of Molecules Composed of Subunits. *Acta Phys. Austriaca* **1972**, *36*, 307-315.

espalhamento em condições experimentais adequadas, tornando possível registrar esses casos de contraste fraco com maior eficiência.⁹⁷

A partir das curvas de espalhamento obtidas e do tratamento matemático dos dados é possível obter, com base no formato das curvas, parâmetros dimensionais (Figura I-11 e I-12).⁹⁸

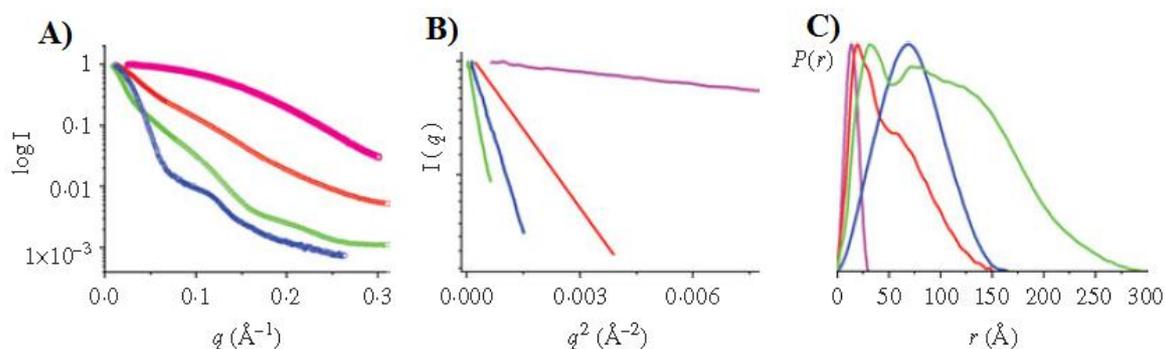


Figura I-11. Ilustração de dados de SAXS. A) Curvas de espalhamento, tendo o logaritmo da intensidade em função do vetor de espalhamento. B) Região de Guinier, obtida a partir da curva de espalhamento. C) Curvas da distribuição de distâncias $\{P(r)\}$ em proteínas.⁹⁹

Como citado, a partir do tipo das curvas obtidas é possível inferir a forma da proteína (Figura I-12) e relacionar com a massa molecular (Figura I-13).

⁹⁷ Svergun, D.I.; Koch, M.H.J. Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **2003**, *66*, 1735-1782.

⁹⁸ Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **1999**, *76*, 2879-2886.

⁹⁹ Adaptado de Putnam, C.D.; Hammel, M.; Hura, G.L.; Tainer, J.A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **2007**, *40*, 191-285.

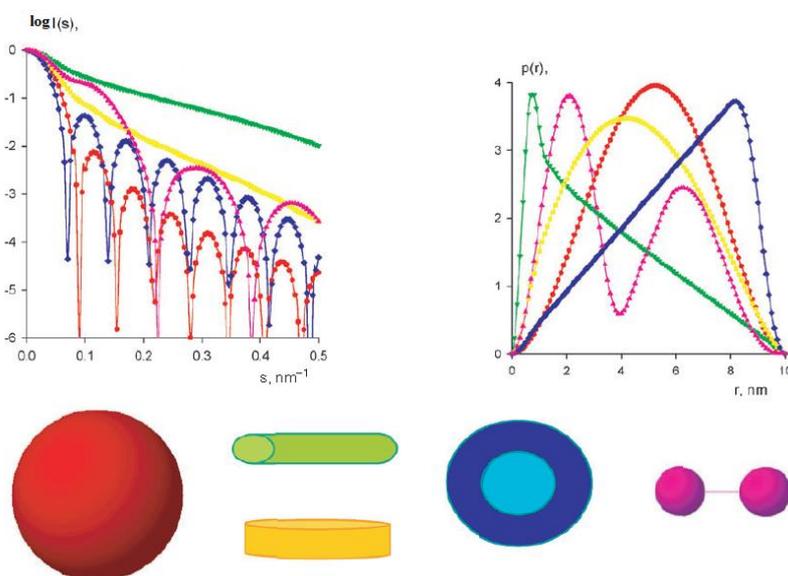


Figura I-12. Ilustração da relação do formato das curvas de SAXS e a forma da molécula. À direita, curvas de espalhamento com logaritmo da intensidade em função do vetor espalhamento (s) e à esquerda as curvas de distribuição de distâncias, tendo as curvas a mesma cor da molécula referida.¹⁰⁰

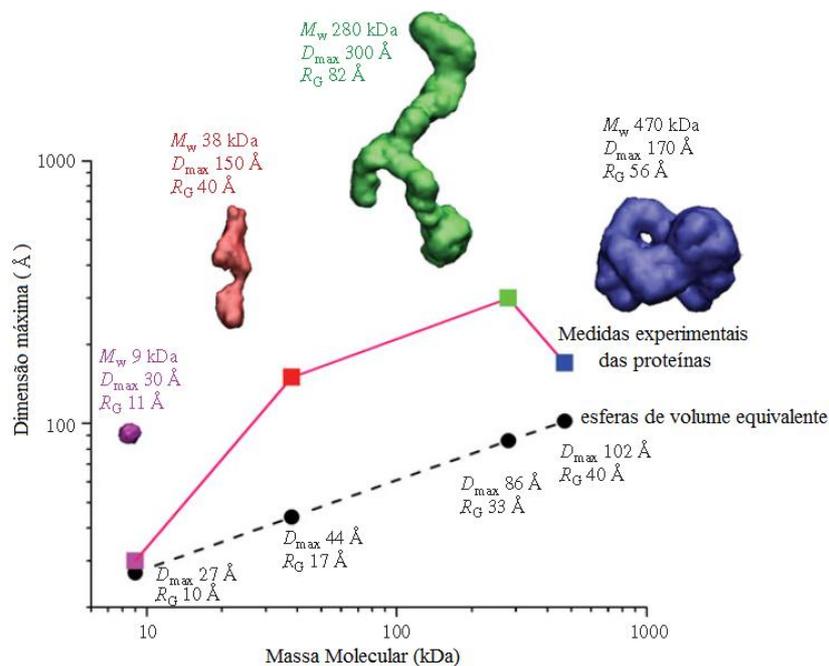


Figura I-13. Ilustração da relação da dimensão máxima obtida pela $P(r)$ e a massa molar da proteína. Sendo observada uma relação linear entre essas duas grandezas para moléculas esféricas (em preto) e os desvios que ocorrem em função da forma da molécula estão mostrados em cores diferentes.⁹⁹

¹⁰⁰ Adaptado da referência 97.

Estes parâmetros dimensionais utilizados em programas apropriados permitem a construção de modelos do envelope molecular, que podem gerar informações sobre estruturas terciárias e quaternárias da macromolécula, porém sem detalhes em nível atômico.

Por fim, para obtenção de detalhes em nível atômico e resolução da estrutura de uma proteína utiliza-se a **cristalografia de raios-X** ou Ressonância Magnética Nuclear, que são técnicas que se complementam em estudos de estrutura e dinâmica de proteínas.¹⁰¹ A cristalografia é o exame da estrutura interna de um cristal através de técnicas de difração de raios-X. Para distinguir átomos individuais, o comprimento de onda da radiação utilizada deve ser igual ou menor que a distância entre os átomos, sendo que estas distâncias são da ordem de angstroms e a radiação eletromagnética com comprimento de onda nesta faixa pertence aos raios-X. Quando um feixe de raios-X passa por um cristal, alguns raios interagem com os átomos e são espalhados. O espalhamento pode resultar em um fenômeno chamado de difração, que depende da forma de interação entre a radiação e os átomos. Analisando a difração é possível, então, obter informações a respeito da estrutura do cristal e de seus componentes. Ressaltando que, na determinação da estrutura 3D de uma proteína pela cristalografia de raios-X o processo de cristalização da amostra, onde se determina a condição ótima de formação de cristais é crucial. E após a obtenção dos dados de difração do cristal, estes são utilizados na construção e refinamento do modelo molecular com coordenadas atômicas da proteína.

¹⁰¹ (a) Rhodes, G. *Crystallography made crystal clear: a guide for users of macromolecular models*; Academic Press: San Diego, 1993. (b) Evans, J. *Biomolecular NMR Spectroscopy*; Oxford University Press: Nova Iorque, 1995.

Entretanto, como a cristalização de proteínas pode ser, em alguns casos, um processo complicado, experimentos de SAXS, Cryo-EM e as técnicas auxiliares, como DLS e filtração em gel analítica são cada vez mais utilizadas na investigação de estruturas de proteínas.

Tanto nos estudos por cristalografia de raios-X como em NMR, os **métodos computacionais** são bastante utilizados como ferramenta auxiliar na determinação da estrutura protéica. De fato, como é sabido, a sequência de aminoácidos é que determina a estrutura terciária de uma proteína, portanto na teoria, é possível calcular a estrutura tridimensional a partir de sua sequência de aminoácidos. Entretanto, apesar dos avanços computacionais e tecnológicos atuais, ainda não é possível prever corretamente a estrutura protéica tridimensional a partir da estrutura primária. Fato que pode ser constatado pela maior quantidade de sequências de proteínas disponíveis em relação às estruturas protéicas.¹⁰²

A forma de uma proteína corresponde à conformação de mais baixa energia da molécula e reflete a combinação das diferentes propriedades de seus aminoácidos constituintes. Estas conformações de baixa energia correspondem a um enovelamento energeticamente favorável da cadeia polipeptídica através de interações hidrofóbicas e ligações de hidrogênio entre os aminoácidos da cadeia.

Apesar de vários tipos de interações entre a cadeia lateral serem característicos de alguns elementos estruturais, no geral não há uma simples correlação entre a sequência de aminoácidos e a estrutura da proteína. Além de que, diferentes sequências podem adotar enovelamentos similares.

¹⁰² Dodson, E.J. Protein Predictions. *Nature* **2007**, *450*, 176-177.

A partir de uma estrutura conhecida é relativamente fácil entender as interações atômicas que contribuem para tal enovelamento. Entretanto, conhecendo a sequência de aminoácidos, buscar interações atômicas que determinem o enovelamento é uma tarefa bem mais complicada. Primeiramente é necessário quantificar as contribuições energéticas de cada tipo de interação e em seguida observar o efeito das conformações moleculares nestas contribuições individuais. Lembrando que, mesmo uma proteína pequena pode adotar várias conformações diferentes.¹⁰²

Os métodos computacionais utilizados no cálculo de estruturas protéicas podem ser: *ab initio*, Rosetta, simulações de dinâmica molecular e outros programas disponíveis eletronicamente.

O método *ab initio* corresponde a um grupo de métodos nos quais as estruturas moleculares podem ser calculadas basicamente aplicando a Equação de Schrödinger, constantes fundamentais e número atômico. Esta metodologia permite que as diversas propriedades moleculares que dependam da energia total do sistema ou de diferenças de energia (por exemplo, parâmetro de rede, compressibilidade, distâncias interatômicas, energia de formação e outros) possam ser determinadas com extrema precisão. O método Rosetta é uma variação do *ab initio* que estuda os núcleos das proteínas em um momento de equilíbrio termodinâmico (estado mínimo de energia). Através dele calcula-se qual das formas potenciais de uma proteína apresenta energia mais baixa e identifica regiões flexíveis, sendo um dos melhores métodos para calcular estruturas tridimensionais de proteínas.¹⁰³

No caso da dinâmica molecular, o objetivo é calcular a energia associada a uma dada conformação molecular, estudando a interação entre as

¹⁰³ Petrey, D.; Honig, B. Protein Structure Prediction: Inroads to Biology. *Molecular Cell* **2005**, *20*, 811-819.

moléculas, em relação à estrutura, função e movimento. Ela pode ser usada para descrever uma possível estrutura de uma sequência a partir da minimização da energia potencial. São simulações que modelam a proteína como sendo um sistema newtoniano de átomos ligados por molas. Por isso, além das interações eletrostáticas entre os átomos, há ainda que se considerar o cálculo dos parâmetros relativos às energias de alongamento, torção e de dobramento da ligação química. Os parâmetros usados na dinâmica molecular são obtidos por cálculos de mecânica quântica em fragmentos de peptídeos e, além da determinação da estrutura das proteínas, pode ser utilizada também para entender o mecanismo de enovelamento/desenovelamento.¹⁰³

Este campo de cálculo de estrutura a partir da sequência de aminoácidos foi estimulado por uma rede criada em 1994 (CASP - *Critical Assessment of Protein Structure Prediction*) que visa a avaliação das estimativas, comparando a estrutura calculada com a estrutura cristalográfica obtida experimentalmente, utilizando sistemas rigorosos de pontuação (*scores*).¹⁰²

Assim, os métodos computacionais para predição de estrutura de proteínas podem ser classificados em quatro grupos: modelagem por comparação; reconhecimento do enovelamento; métodos de primeiros princípios com informação de base de dados, e, métodos de primeiros princípios sem informação de base de dados. Sendo que, os dois últimos baseiam-se na hipótese termodinâmica de Anfinsen, que considera o processo de enovelamento de uma proteína em sua estrutura nativa um processo puramente físico que depende apenas da sequência de aminoácidos e do ambiente em que se encontra (solvente).¹⁰⁴

¹⁰⁴ Floudas, C.A. Computational Methods in Protein Structure Prediction. *Biotechnol. Bioeng.* **2007**, *97*, 207-213.

Os métodos de modelagem por comparação se baseiam na premissa que sequências relacionadas evolutivamente exibem estrutura tridimensional similar, ou seja, que a similaridade de sequência sugere similaridade estrutural. Já os métodos de reconhecimento de enovelamento se fundamentam no fato de que o número de diferentes estruturas protéicas enoveladas é bem mais limitado que o vasto número de diferentes sequências obtidas pelo genoma.¹⁰⁴

Os métodos de primeiros princípios com informação de base de dados podem ser subdivididos em: métodos de recombinação baseados em fragmentos; métodos híbridos que combinam a comparação de múltiplas sequências, otimização de Monte Carlo com pontuação das funções e agrupamento; e, os métodos que combinam informação de estrutura secundária com restrições terciárias. Nos métodos híbridos utiliza-se, por exemplo, um modelo de átomos unidos com três átomos por resíduo e obtém-se informação de interações a longa distância através da identificação de contatos das cadeias laterais. É feito, então, o agrupamento e seleção dos grupos de maior população que resulta em confôrmeros próximos do confôrmero nativo de energia mais baixa. Um exemplo bem sucedido desta metodologia é a abordagem hierárquica TASSER que combina a identificação de um modelo através da amostragem de estruturas e construção do modelo via um rearranjo contínuo de fragmentos utilizando Monte Carlo e faz uma análise posterior utilizando pontuações TM (do inglês *Template Modeling scores*). E, esta abordagem foi introduzida e aplicada no CASP6 e CASP7.^{103,104}

Por último, o método de primeiros princípios sem informação de base de dados, tenta identificar estruturas da proteína em seu meio com energia livre mais baixa, utilizando somente as leis da física e a sequência de

aminoácidos. Devido à vasta gama de alvos e à inabilidade de aplicar informações de base de dados, este é o método mais difícil dentre os métodos de predição de estrutura terciária de proteínas. Entretanto, é bem apropriado no cálculo de estrutura de proteínas que, a princípio, não apresentam homólogas.^{102,104}

Finalmente, com base nos dados da rede CASP (CASP5, 6 e 7) ficou evidente que o método de primeiros princípios com informação de base de dados, mais especificamente os métodos baseados em fragmentos e métodos híbridos, são os que apresentam melhores resultados na predição de estrutura tridimensional de proteínas.¹⁰⁴

Vale destacar que a combinação das técnicas utilizadas na determinação da estrutura de proteínas, principalmente NMR, cristalografia de raios-X e métodos computacionais têm auxiliado bastante na obtenção das estruturas terciárias das proteínas com alto grau de confiabilidade (Figura I-14).¹⁰⁵

Considerando que o objetivo da biologia estrutural é a determinação de estruturas representativas das proteínas no meio biológico de uma maneira rápida e eficiente, a tendência atual é uma aplicação combinada da cristalografia de raios-X, NMR e dos métodos computacionais na determinação de estruturas protéicas. Isto se deve ao fato de que cada técnica individualmente apresenta suas limitações, por exemplo, a determinação estrutural por cristalografia de raios-X é uma abordagem rápida, entretanto existem dois grandes obstáculos que são o processo de cristalização e o problema das fases.

¹⁰⁵ Szymczyna, B.R.; Taurog, R.E.; Young, M.J.; Snyder, J.C.; Johnson, J.E.; Williamson, J.R. Synergy of NMR, Computation, and X-Ray Crystallography for Structural Biology. *Structure* **2009**, *17*, 499-507.

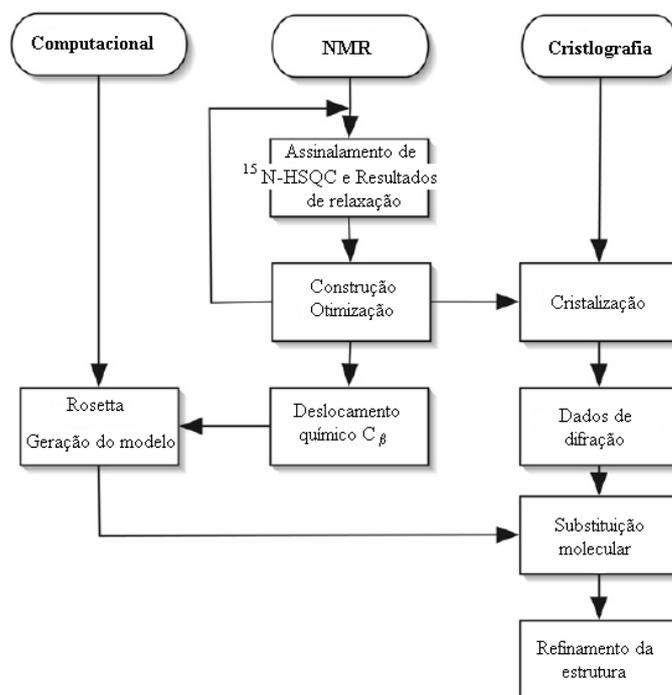


Figura I-14. Abordagem sinérgica para determinação estrutural de proteínas. Fluxograma descrevendo a utilização combinada das técnicas de cristalografia de raios-X, NMR e métodos computacionais na determinação da estrutura tridimensional de proteínas.¹⁰⁶

A Ressonância Magnética Nuclear pode ser uma ferramenta extremamente poderosa para superar estes obstáculos, pois fornece informações que auxiliam na identificação de regiões da proteína mais apropriadas para cristalização. Desde que se utilize uma amostra solúvel e monodispersa de uma proteína de tamanho adequado, pela NMR pode-se obter informação de regiões flexíveis da proteína que atrapalham a cristalização, regiões onde há interação com ligantes e regiões propensas à cristalização, através de experimentos de ¹H, ¹⁵N-HSQC e relaxamento de ¹⁵N.

¹⁰⁶ Adaptado da referência 105.

Por outro lado, na NMR, existe a limitação do tamanho da proteína, além de que a atribuição dos deslocamentos químicos da cadeia lateral dos aminoácidos e as correlações de distâncias podem ser difíceis e demoradas, representando os gargalos desta técnica na determinação estrutural de proteínas.¹⁰⁵

Entretanto, a identificação de regiões da proteína passíveis de se estudar estruturalmente dificulta o trabalho tanto na cristalografia de raios-X quanto na NMR. E, nesse sentido, os cálculos *ab initio* discutidos anteriormente representam um meio alternativo para a obtenção de estruturas, porém tais previsões são dependentes da localização de um mínimo global em um vasto espaço conformacional.

Assim, a partir de dados obtidos por NMR têm-se informações sobre restrições estruturais que podem melhorar os modelos obtidos pelos métodos computacionais, além das informações que são úteis para os ensaios de cristalização. As estruturas obtidas pelos métodos computacionais ou por NMR podem ser utilizadas como modelos para substituição molecular em cristalografia de raios-X. Os métodos computacionais por sua vez têm auxiliado na rápida atribuição dos resíduos e determinação de estruturas por NMR. E, dados de cristalografia também são úteis na construção de modelos pelos métodos computacionais.¹⁰⁵

Portanto, a estratégia de se utilizar conjuntamente estas três técnicas é extremamente interessante e útil, uma vez que acelera o processo de determinação estrutural e garante uma maior confiabilidade nos resultados obtidos, dado que, de maneira global, aproveitam-se os pontos fortes de cada técnica.

Capítulo 2



OBJETIVOS

*Adaptado de <http://www.plantmanagementnetwork.org/elements/view.aspx?ID=411>; ref. 3.

2. Objetivos

O objetivo deste estudo foi a caracterização estrutural de quatro proteínas hipotéticas da bactéria *Xanthomonas axonopodis* pv. *citri*, sendo que estas quatro proteínas são classificadas como possíveis chaperonas de secreção.

Para alcançar tal objetivo foram propostos os seguintes objetivos específicos:

- Obtenção das quatro proteínas alvo (XACb0033, XAC1346, XAC0419 e XAC1990) puras e enoveladas.
- Caracterização das mesmas por métodos espectroscópicos como Dicroísmo Circular, Fluorescência de emissão, Ressonância Magnética Nuclear, Espalhamento de Luz Dinâmico e Espectrometria de Massas.
- Realização de experimentos de SAXS com as quatro proteínas para propor tamanho, forma e estado oligomérico das mesmas em solução.
- Obtenção de dados estruturais através de ferramentas de bioinformática e comparação com os obtidos experimentalmente.

Capítulo 3



PARTE EXPERIMENTAL

*Adaptado de http://zju.molmicro-foodsafety.net/Structure_en.asp e Nelson, D.L.; Cox, M.M. Lehninger Principles of Biochemistry, 4th Ed.; W.H. Freeman & Company: New York, 2004; Capítulo 3.

3. Parte Experimental

As etapas de clonagem, seleção de vetores de expressão, linhagens de bactérias e condições de expressão já haviam sido padronizadas em trabalhos anteriores de membros do grupo junto ao Prof. Dr. Carlos Ramos.^{15,107,108}

O vetor utilizado para a expressão foi o pET23a. Assim, na construção dos vetores de expressão, os genes que codificam as proteínas de interesse foram clonados no plasmídeo pET23a entre os sítios de restrição das enzimas *NdeI* e *Sall*, originando os seguintes vetores: pET23a-XAC1346, pET23a-XAC0419, pET23a-XACb0033 e pET23a-XAC1990. A amplificação do DNA foi feita pela reação em cadeia da polimerase (PCR, do inglês *polymerase chain reaction*) e a sequência correta dos genes clonados nos plasmídeos foi confirmada pelo sequenciamento de DNA.¹⁰⁸

Em seguida, estes vetores de expressão foram introduzidos em células de *E. coli* para produção das proteínas recombinantes. As células utilizadas para a expressão das proteínas alvo foram células da linhagem BL21(DE3)pLysS (Invitrogen).¹⁰⁹

A linhagem BL21(DE3) é utilizada para expressão de proteínas recombinantes, pois apresenta deleção da protease *lon* e da protease de membrana externa *omT*. Estas últimas são enzimas envolvidas na degradação

¹⁰⁷ Tasic, L. Caracterização estrutural de proteínas (chaperones) de secreção da bactéria *Xanthomonas axonopodis* pv. *citri*. Relatório de pós-doutorado, Laboratório Nacional de Luz Síncrotron, Campinas, SP, Junho de 2004.

¹⁰⁸ Tasic, L.; Borin, P.; Khater, L.; Ramos, C.; Cloning and characterization of three hypothetical secretion chaperone proteins from *Xanthomonas axonopodis* pv. *citri*. *Protein Expression Purif.* **2007**, *53*, 363-369.

¹⁰⁹ a) Mandel, M.; Higa, A. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **1970**, *53*, 159-162. b) Cohen, S.N.; Chang, A.C.Y.; Hsu, L. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *E. coli* by R-factor DNA. *Proc. Nat. Acad. Sci. USA* **1972**, *69*, 2110-2114.

das proteínas durante a etapa de purificação.¹¹⁰ Além disso, esta linhagem possui o gene que codifica a enzima T7 RNA polimerase, controlado pelo promotor lacUV5, que é expresso em função da adição de IPTG. Assim, estes sistemas permitem a obtenção elevada da proteína cujo DNA está clonado no vetor pET23a. A linhagem BL21(DE3)pLysS possui ainda a expressão aumentada da lisozima T7 e é resistente a cloranfenicol, sendo utilizada para a expressão de proteínas tóxicas à bactéria. A lisozima T7 ajuda na lise bacteriana e é um inibidor natural da T7 RNA polimerase, portanto nesta linhagem a expressão basal de genes sob o controle do promotor T7 é eliminada.¹¹¹

Para a inserção do DNA das proteínas, em pET23a (pET23a-DNA), nas células de *E. coli* foi utilizado o método de choque térmico, por ser um método mais barato e eficiente.

E, a expressão das proteínas pela bactéria foi induzida com a adição de IPTG (isopropil β -D-1-tiogalactopiranosídeo), um composto mimético da alolactose. Este último, por sua vez, é um metabólito da lactose que desencadeia a transcrição do operon *lac*, conforme citado acima. Porém, diferente da alolactose, o IPTG apresenta um átomo de enxofre (S), que leva à formação de uma ligação química não hidrolisável pela célula, evitando sua degradação e, portanto, deixando sua concentração constante durante a expressão.

¹¹⁰ Seraphim, T.V. Estudos bioquímicos e biofísicos de proteínas de choque térmico da família HSP40 de cana-de-açúcar e de levedura. Dissertação de Mestrado, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Fevereiro d 2010.

¹¹¹ Studier, W.F.; Rosenberg, A.H.; Dunn, J.J.; Dubendorff, J.W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **1990**, *185*, 60-89.

3.1. Extração do DNA plasmidial (lise alcalina)¹¹²

Foi feito um inóculo de uma colônia de *E. coli*, cepa BL21(DE3)plysS transformada com o DNA referente à proteína de interesse (pET23a-DNA, gentilmente cedido pelo Prof. Dr. Carlos Ramos) em 1,5 mL de meio LB (Tabela III-1) contendo $50 \mu\text{g mL}^{-1}$ de ampicilina e $34 \mu\text{g mL}^{-1}$ de cloranfenicol, incubando por 16 horas a 37°C e 200 rpm (Incubadora Tecnal TE-420). Transcorrido este tempo, a suspensão foi transferida para um microtubo e centrifugada (Centrífuga Eppendorf 5804R) a 13200 rpm durante 5 minutos. O sobrenadante foi descartado, sendo removido por aspiração a vácuo.

O precipitado foi ressuspendido em 100 μL da solução I (Tabela III-1) gelada, agitando em vortex até completa dissolução. Adicionou-se em seguida 200 μL da solução II recém-preparada (Tabela III-1), misturando rapidamente por inversão e o tubo foi mantido em banho de gelo por 5 minutos. A esta suspensão foram adicionados 150 μL da solução III gelada (Tabela III-1) e homogeneizou-se a mistura por agitação em vortex com o microtubo invertido. Deixou-se o microtubo em banho de gelo por mais 5 minutos e a mistura foi centrifugada a 13200 rpm por 5 minutos, transferindo o sobrenadante para um novo microtubo. Ao sobrenadante adicionou-se 200 μL de fenol e 200 μL de clorofórmio, agitando em vortex. Novamente foi feita uma centrifugação a 13200 rpm por 5 minutos e o sobrenadante foi transferido cuidadosamente para um novo microtubo. A este foram adicionados 800 μL de etanol absoluto gelado para precipitação do DNA, centrifugando em seguida, nas mesmas condições já citadas. Removeu-se o sobrenadante por aspiração a vácuo e o precipitado foi lavado com 200 μL de etanol 70%

¹¹² Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; 1989, 2nd ed.

gelado. O sobrenadante foi novamente removido por aspiração e deixou-se o precipitado secar ao ar por 10 minutos. O DNA foi então ressuscitado em 50 μL de TE/RNase (Tabela III-1) e armazenado em freezer -80°C .

Tabela III-1. Soluções utilizadas na lise alcalina

Meio de cultura LB		Solução I		Solução II	
peptona	1,0% (m/v)	glicose	50 mmol L ⁻¹	NaOH	200 mmol 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 imediatamente antes do uso.	
pH 7,0; autoclavado		pH 8,0; autoclavada e estocada a 4 °C		Tampão TE/RNase	
		Solução III		Tris	10 mmol L ⁻¹
		CH ₃ COOH	2 mol L ⁻¹	EDTA	1 mmol L ⁻¹
		CH ₃ COONa	3 mol L ⁻¹	RNase	20 $\mu\text{g mL}^{-1}$
		estocada a 4 °C		pH 8; estocado a -4°C	

3.2. Preparação de células competentes

Foi feito o inóculo, a partir de uma placa de cultura fresca da cepa BL21(DE3)plysS de *E. coli* (Invitrogen – C6060-03), em 20 mL de meio de cultura PSI (Tabela III-2) contendo 34 $\mu\text{g mL}^{-1}$ de cloranfenicol e deixado por 16 horas sob agitação (200 rpm – Incubadora Tecnal TE-420) a 37 °C. Transcorrido este tempo foi feita a diluição destes 20 mL em 200 mL de meio PSI contendo 34 $\mu\text{g mL}^{-1}$ de cloranfenicol e deixado a 37 °C e 200 rpm, monitorando a absorbância em 600 nm (Espectrofotômetro Bel Photonics SP1105). Quando se atingiu a absorbância de 0,6 (0,5–0,7), o crescimento foi interrompido colocando a cultura em banho de gelo e centrifugou-se (Centrífuga Beckman Coulter Allegra X-22R) a suspensão por dez minutos a 4000 rpm e 4 °C. O precipitado foi então ressuscitado em 25 mL do tampão

de transformação (TBI – Tabela III-2), deixado quinze minutos no gelo e, posteriormente, centrifugado por quinze minutos a 4000 rpm, 4 °C. Descartou-se o sobrenadante e o precipitado foi novamente ressuscitado em 8 mL do tampão TBII (Tabela III-2). Foram feitas alíquotas de 150 µL em microtubos resfriados e estas células foram acondicionadas em -80 °C.

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

Meio de cultura PSI		TBI		TBII	
peptona	1,6% (m/v)	NaOH	30 mmol L ⁻¹	MOPS	10 mmol 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		
		glicerol	13,5%		
pH 7,2; autoclavado		pH 6,0; autoclavado e acondicionado em 4 °C		pH 7,0; autoclavado e acondicionado em 4 °C	

3.3. Transformação por choque térmico

Aos 150 µL das células competentes BL21(DE3)plysS foram adicionados 1,5 µL do DNA da proteína de interesse, pET23a-proteína alvo, com agitação sutil. Esta suspensão foi deixada por trinta minutos em banho de gelo, e um minuto e trinta segundos a 42 °C, sem agitação. Em seguida foi colocada rapidamente em banho de gelo e adicionou-se 1 mL de meio SOC (Tabela III-3) recém-preparado sem antibióticos e incubou-se por uma hora a 37 °C, 200 rpm (Incubadora Tecnal TE-420). Transcorrido este tempo, 250 µL desta suspensão foram plaqueados em placas de Petri contendo meio LB sólido (Tabela III-3), 50 µg mL⁻¹ de ampicilina e 34 µg mL⁻¹ de cloranfenicol.

Ressaltando que foram feitos controles positivo e negativo para testar a viabilidade das células competentes preparadas. As placas foram deixadas por 16 horas em estufa (Nova Ética – 410/1 ND) a 37 °C para crescimento bacteriano, e foram selecionadas algumas das colônias que cresceram, para testes de mini-indução. 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 quinze dias.

Tabela III-3. Soluções utilizadas na transformação bacteriana

Meio de cultura LB sólido		Meio de cultura SOC		
peptona	1,0% (m/v)	peptona	2,0% (m/v)	Em fluxo laminar, antes do uso, adicionou-se ao meio SOC: MgCl ₂ 10 mmol L ⁻¹ glicose 20 mmol L ⁻¹ (gelada)
extrato de levedura	0,5% (m/v)	extrato de levedura	0,5% (m/v)	
NaCl	1,0% (m/v)	NaCl	10 mmol L ⁻¹	
ágar	1,5% (m/v)	KCl	2,5 mmol L ⁻¹	
pH 7,0; autoclavado		pH 7,0; autoclavado		

3.4. Expressão protéica

Foi feito o pré-inóculo, a partir de uma placa de cultura fresca da cepa BL21(DE3)plyS de *E. coli* transformada com o DNA referente à proteína de interesse, em 5 mL de meio LB (Tabela III-1) contendo 50 µg mL⁻¹ de ampicilina e 34 µg mL⁻¹ de cloranfenicol (antibióticos apropriados), incubando por 16 horas a 37 °C e 200 rpm (Incubadora Tecnal TE-420). Em seguida, estes 5 mL foram diluídos em 500 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. Atingida a absorbância de aproximadamente 0,8, em 600 nm

(Espectrofotômetro Bel Photonics SP1105), a expressão protéica 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 $37 \text{ }^\circ\text{C}$ por 3 ou 16 horas – dependendo da proteína em questão. Transcorrido este tempo, a suspensão foi centrifugada a 3500 rpm (Centrífuga Cientec CT-6000R), $4 \text{ }^\circ\text{C}$ por 15 minutos e o precipitado foi acondicionado a $-20 \text{ }^\circ\text{C}$.

Tabela III-4. Solução estoque dos antibióticos e IPTG

Ampicilina	Cloranfenicol	IPTG
sal de sódio de ampicilina 50 mg mL^{-1} em água:etanol 50:50	34 mg mL^{-1} em etanol	1 mol L^{-1} em água estéril

As três soluções foram filtradas em filtros de $0,22 \text{ }\mu\text{m}$ e estocadas a $-4 \text{ }^\circ\text{C}$.

3.5. Eletroforese em gel

A análise qualitativa das proteínas foi feita utilizando eletroforese em gel de acrilamida (SDS-PAGE 15% - Tabela III-7). As amostras de $5\text{-}20 \text{ }\mu\text{L}$ (dependendo do tipo de amostra) foram aplicadas no gel de empacotamento e separadas no gel de separação (géis previamente preparados) pela aplicação de uma diferença de potencial. A eletroforese foi conduzida utilizando o aparato MiniVE Vertical Electrophoresis System (Amershan Biosciences), utilizando 180 V, 50 mA durante aproximadamente 100 minutos, sendo que neste aparato o gel fica imerso no tampão de corrida (Tabela III-7). Transcorrido este tempo o gel foi incubado em uma solução corante (Tabela III-8) por cerca de 20 minutos sob agitação e em seguida lavado abundantemente com água corrente. Por fim o gel foi deixado imerso no descorante (Tabela III-8) sob agitação por 20 minutos, trocando a solução descorante de 3 a 4 vezes. Todas

as corridas eletroforéticas foram feitas aplicando um marcador de massa molar para posterior identificação das proteínas alvo. No caso foram utilizados os marcadores: P7709S (New England Biolabs-NEB) contendo proteínas de 7 a 175 kDa e o kit LMW (Low Molecular Weight Calibration Kit for SDS Electrophoresis – GE Healthcare), contendo proteínas de 14,4 a 97 kDa, diluídos em tampão para padrão (Tabela III-5).

As amostras provenientes da indução foram preparadas da seguinte maneira: alíquotas de 1 mL da indução foram centrifugadas a 13200 rpm (centrífuga Eppendorf 5424) por um minuto, 4 °C e o sobrenadante foi descartado. Ao precipitado, adicionou-se 100 µL de água e 100 µL de tampão de amostra e agitou-se a mistura em vortex. Em seguida, estas amostras foram sonicadas (sonicador Sonic VCX-750) utilizando amplitude de pulso de 10% por 15 segundos. Antes da aplicação no gel estas amostras foram colocadas em banho de ultrassom por 15 minutos e fervidas por 5 minutos. As demais amostras foram apenas diluídas pela metade com tampão de amostra antes da aplicação no gel.

Tabela III-5. Soluções tampão utilizadas no preparo de amostras

Tampão de amostra		Tampão de padrão	
Tris	50 mmol L ⁻¹	Tris	62,5 mmol L ⁻¹
DTT	100 mmol L ⁻¹	DTT	100 mmol L ⁻¹
SDS	2% (m/v)	SDS	2% (m/v)
azul de bromofenol	0,1% (m/v)	azul de bromofenol	0,01% (m/v)
glicerol	10% (m/v)	glicerol	10% (m/v)
pH 6,8		pH 6,8	

Tabela III-6. Solução de acrilamida utilizada na preparação dos géis

Mix 30%	
Acrilamida - CH ₂ CHCONH ₂	29% (m/v)
Bis-acrilamida - (CH ₂ CHCONH ₂)CH ₂	1% (m/v)

Tabela III-7. Géis de acrilamida (15%)

Gel de empacotamento		Gel de separação	
Mix 30%	5% (v/v)	Mix 30%	15% (v/v)
Tris-HCl 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)
(NH ₄) ₂ S ₂ O ₈ (APS)	0,1% (m/v)	(NH ₄) ₂ S ₂ O ₈ (APS)	0,1% (m/v)
C ₆ H ₁₆ N ₂ (TEMED)	0,04% (v/v)	C ₆ H ₁₆ N ₂ (TEMED)	0,04% (v/v)
Tampão de corrida			
Tris	25 mmol L ⁻¹	glicerol	2% (m/v)
SDS	0,2% (m/v)	pH	8

*ajustar pH da solução tampão

Tabela III-8. Soluções corante e descorante

Corante		Descorante	
Comassie brilliant blue R-250	0,25% (m/v)	CH ₃ CH ₂ OH	30% (v/v)
CH ₃ CH ₂ OH	30% (v/v)	CH ₃ COOH	10% (v/v)
CH ₃ COOH	10% (v/v)		

3.6. Lise celular

O precipitado referente a 1 L da expressão protéica foi ressuspenso em 20 mL de tampão de lise (20 mL L⁻¹ – Tabela III-9) e as células foram lisadas por sonicação (sonicador de ponta Sonic VCX-750) utilizando amplitude de pulso de 30% por 2 minutos (pulsando 10 s e descansando 40 s)

em banho de gelo. Esta suspensão foi então centrifugada a 15000 rpm (Centrífuga Beckman Coulter Allegra X-22R), 4 °C por 30 minutos. Após análise de alíquotas do sobrenadante e do precipitado por eletroforese, o precipitado foi estocado a -20 °C e o sobrenadante foi submetido à diálise em tampão de purificação (Tabela III-9), sendo utilizado normalmente para 15 mL da suspensão 300 mL de tampão por 30 minutos a 4 °C sob agitação, com cerca de 6 trocas do tampão para posterior purificação. Ressaltando que a diálise foi feita utilizando membranas de celulose regenerada com porosidade de 3,5 kDa (Dialysis Tubing – Fisher Scientific) fervidas antes do uso.

Tabela III-9. Soluções utilizadas na lise celular

Tampão de lise		Tampão diálise pH8		Tampão diálise pH5	
Tris	100 mmolL ⁻¹	NaH ₂ PO ₄	3,6 mmolL ⁻¹	CH ₃ COOH	18 mmolL ⁻¹
NaCl	100 mmolL ⁻¹	Na ₂ HPO ₄	46,4 mmolL ⁻¹	CH ₃ COONa	32 mmolL ⁻¹
EDTA	1 mmolL ⁻¹	pH 8,0		pH 5,0	
pH 8,0; autoclavado					

3.7. Tratamento com uréia (reenovelamento)¹¹³

O precipitado, contendo a proteína de interesse, obtido a partir da lise, foi ressuspendido em 10 mL de tampão de solubilização (Tabela III-10) por litro de indução e esta suspensão foi mantida sob agitação contínua por 30 minutos em banho de gelo. Transcorrido este tempo, a suspensão foi submetida à diálise com tampão de purificação (Tabela III-10) por 30 minutos, a 4 °C sob agitação, com cerca de 6-7 trocas do tampão. Após a diálise, a suspensão foi centrifugada (Centrífuga Beckman Coulter Allegra X-22R) a

¹¹³ Ribeiro Jr, E.; Regis, W.C.B.; Tasic, L.; Ramos, C.H.I. Fast purification of the Apo form and of a non-binding heme mutant of recombinant sperm whale myoglobin. *Protein Expression Purif.* **2003**, 28, 202-208.

15000 rpm, a 4 °C por 30 minutos e as alíquotas do sobrenadante e do precipitado foram analisadas por eletroforese em gel. O precipitado foi estocado a -20 °C e o sobrenadante foi submetido à purificação por cromatografia. O precipitado que ainda continha proteína foi lavado duas vezes com água destilada, centrifugado como descrito anteriormente e submetido a novo tratamento com uréia. Em alguns casos este tratamento foi efetuado três vezes. As águas das lavagens apresentaram a proteína pura em baixa concentração e foram posteriormente dialisadas com tampão adequado e concentradas.

Tabela III-10. Soluções utilizadas no processo de reenovelamento

Tampão de solubilização		Tampão diálise pH8	
NaH ₂ PO ₄	3,6 mmol L ⁻¹	NaH ₂ PO ₄	3,6 mmol L ⁻¹
Na ₂ HPO ₄	46,4 mmolL ⁻¹	Na ₂ HPO ₄	46,4 mmol L ⁻¹
NaCl	100 mmolL ⁻¹	NaCl	100 mmol L ⁻¹
ureia (NH ₂) ₂ CO	8 mol L ⁻¹		pH 8,0

3.8. Purificação das proteínas

As proteínas contidas nos sobrenadantes, tanto obtidos a partir da lise ou a partir do tratamento com uréia, foram purificadas por cromatografia líquida (FPLC ÄKTA Purifier UPC900 – GE Healthcare). Como o sobrenadante da lise apresenta muitos contaminantes, a purificação foi efetuada em duas etapas. Primeiramente por troca iônica, utilizando as colunas DEAE SepharoseTM Fast Flow (17 0709-01) ou CM SepharoseTM Fast Flow (17 0719-01), tendo como fase móvel as mesmas soluções tampões das diálises (Tabelas III-9 e 10). Sendo a proteína eluída com gradiente de solução de NaCl 1 mol L⁻¹. Em seguida, efetuou-se a segunda etapa da purificação, utilizando a cromatografia por filtração em gel. O sobrenadante obtido através

do tratamento com uréia, como não apresenta muitos contaminantes, foi acondicionado, por diálise, no tampão de eluição (Tabela III-10) e aplicado diretamente em uma coluna de filtração em gel. A cromatografia por filtração em gel foi feita utilizando uma coluna Superdex™ 200 Prep Grade (17 1043-01) ou Superdex™ 75 Prep Grade (17 1044-01) 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.

A concentração das amostras foi feita utilizando filtro AMICON de 3 kDa (Millipore), centrifugando a 4000 rpm e 4 °C (Centrífuga Eppendorf 5804R).

Os experimentos de filtração em gel analítica foram realizados em uma coluna de 15 mm de diâmetro e 300 mm de comprimento, com resina Superdex™ 200, tendo 58,2 mL de volume (mesma coluna utilizada na purificação). Os experimentos foram realizados no FPLC UPC900, utilizando um fluxo de 0,6 mL/min, com pressão máxima de 0,6 MPa na coluna e um detector de UV, monitorando a absorbância em 280 nm. A coluna foi equilibrada com 1 VC (60 mL), injetou-se 500 µL de amostra, correspondendo a ~0,9% do volume da coluna (2%, volume máximo permitido para uma boa resolução) e as corridas foram realizadas com cerca de 2 VC (100 mL). Para determinação do volume vazio da coluna (23,18 mL), utilizou-se uma solução de Azul de Dextran (~2 MDa), 2 mg mL⁻¹ em tampão fosfato de sódio (50 mmol L⁻¹, pH 7,2) e cloreto de sódio (150 mmol L⁻¹).

Para a construção da curva de calibração do coeficiente de partição em função da massa molecular das proteínas, utilizou-se o kit de calibração para filtração em gel da GE Healthcare (Gel Filtration Calibration kit LMW), contendo as proteínas: conalbumina (75 kDa), ovalbumina (45 kDa), anidrase carbônica (29 kDa), ribonuclease A (13,7 Da) e aprotinina (6,5 kDa). Foram

preparadas soluções de 3 mg mL^{-1} de cada uma das proteínas, com exceção da ovalbumina (4 mg mL^{-1}) na mesma solução tampão da Azul de Dextran. A partir destas soluções foram preparadas duas misturas (mix A e B – Tabela III-11), na mesma solução tampão, aplicadas separadamente na coluna, em um fluxo de $0,7 \text{ mL/min}$. A partir dos dados destas duas corridas cromatográficas foi construída a curva de calibração do coeficiente de partição (k_{av}) em função da massa molar das proteínas. Amostras de diferentes concentrações das proteínas em estudo, quase puras, foram aplicadas na coluna, utilizando os mesmos parâmetros da calibração, para estimar a massa molar aproximada das mesmas.

Tabela III-11. Soluções para calibração da filtração em gel analítica

Mix A		Mix B	
Aprotinina	$3,0 \text{ mmol L}^{-1}$	Ribonuclease A	$1,5 \text{ mmol L}^{-1}$
Ribonuclease A	$1,5 \text{ mmol L}^{-1}$	Anidrase Carbônica	$0,7 \text{ mmol L}^{-1}$
Ovalbumina	$0,5 \text{ mmol L}^{-1}$	Conalbumina	$0,3 \text{ mmol L}^{-1}$
Solução Tampão = Tampão de eluição pH 7,2			
NaH ₂ PO ₄		$15,8 \text{ mmol L}^{-1}$	
Na ₂ HPO ₄		$34,2 \text{ mmol L}^{-1}$	
NaCl		150 mmol L^{-1}	

Para a construção da curva de calibração do coeficiente de partição em função do raio de Stokes das proteínas, foram utilizadas diferentes proteínas com raio de Stokes conhecidos. Foram preparadas soluções de cada uma das proteínas na solução tampão fosfato de sódio 50 mmol L^{-1} pH 8, contendo 150 mmol L^{-1} de NaCl. Para obter o coeficiente de partição das proteínas padrão, nas condições utilizadas, foram aplicados $500 \mu\text{L}$ das seguintes soluções protéicas (Tabela III-12) em um fluxo de $0,6 \text{ mL/min}$: BSA ($35,3 \text{ \AA}$ - CAS 9048-46-8), ovalbumina ($30,5 \text{ \AA}$ - CAS 9006-59-1), α -quimiotripsin-

ogênio (20,9 Å - CAS 9035-75-0), ribonuclease A (16,4 Å - CAS 9001-99-4) e aprotinina (9,83 Å - CAS 9087-70-1), obtidas da Sigma-Aldrich. A partir dos volumes de eluição encontrados e coeficientes de partição calculados foi construída a curva de calibração do coeficiente de partição em função do raio de Stokes das proteínas. Amostras de diferentes concentrações das proteínas em estudo, puras, foram aplicadas na coluna, utilizando os mesmos parâmetros da calibração para o cálculo do raio de Stokes das mesmas. Ressaltando que as corridas cromatográficas, tanto para a calibração quanto para determinação do raio de Stokes das proteínas, foram feitas em triplicata.

Tabela III-12. Soluções para calibração da filtração em gel analítica II

Soluções protéicas			
Aprotinina	0,77 mmol L ⁻¹	Ovalbumina	0,18 mmol L ⁻¹
Ribonuclease A	0,30 mmol L ⁻¹	BSA	0,13 mmol L ⁻¹
α -quimiotripsinogênio			0,40 mmol L ⁻¹
Solução Tampão = Tampão de eluição pH 8,0			
	NaH ₂ PO ₄		3,4 mmol L ⁻¹
	Na ₂ HPO ₄		46,6 mmol L ⁻¹
	NaCl		150 mmol L ⁻¹

3.9. Determinação da concentração de proteínas (UV/VIS)

Para as proteínas contendo resíduos de triptofano, a concentração foi determinada por espectroscopia UV-VIS (espectrofotômetros Femto CIRRUS 80MB e Hitach U-200), verificando a absorvância em 280 nm,¹¹⁴ utilizando o

¹¹⁴ (a) Edelho, H. Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry* **1967**, *6*, 1948-1954. (b) Nozaki, Y. Determination of the concentration of protein by dry weight—A comparison with spectrophotometric methods. *Archives of Biochem. and Biophys.* **1986**, *249*, 437-446.

coeficiente de absorvidade molar calculado pelos programas sednterp®¹¹⁵ e Expasy,¹¹⁶ com base no número de resíduos de triptofano, tirosina e fenilalanina. Também foi feita a determinação da concentração utilizando a proteína em uma solução de cloreto de guanidina 6 mol L⁻¹, para desenovelamento da mesma, verificando a absorbância em 280 nm.¹¹⁷

3.10. Método de Bradford¹¹⁸

Para as proteínas que não apresentam resíduos aromáticos ou cistinas em sua estrutura primária, a quantificação foi feita pelo método de Bradford, que faz uso das propriedades espectrais do corante Comassie Brilliant Blue G-250 para estimar a quantidade de proteína em solução.

Este corante, um trifenilmetano dissulfonato de sódio (C₄₇H₄₈N₃NaO₇S₂), que em solução ácida apresenta coloração marrom, ao se ligar à proteína tem sua absorbância alterada de 465 para 595 nm, sendo que esta alteração é perceptível visualmente devido à mudança de coloração para tons de azul de acordo com a concentração das proteínas. A coloração marrom, em meio ácido, é referente à forma catiônica da molécula, com os três nitrogênios carregados positivamente e os dois grupos sulfônicos carregados negativamente. A ligação da proteína estabiliza a forma aniônica, negativamente carregada do corante, mesmo em soluções ácidas, produzindo a coloração azul; sendo a formação deste complexo corante-proteína utilizada para quantificar a concentração da proteína em solução.¹¹⁸

¹¹⁵ Philo, J.S. Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal. Biochem.* **2006**, *354*, 238-246.

¹¹⁶ Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784-3788.

¹¹⁷ Pace, C.N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **1995**, *11*, 2411-2423.

¹¹⁸ Bradford, M.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.

Para esta quantificação foi construída inicialmente uma curva de calibração utilizando soluções com diferentes concentrações de BSA (Sigma-Aldrich) no mesmo tampão que a proteína de interesse. Para concentrações menores de $50 \mu\text{g mL}^{-1}$, adicionou-se 200 μL do reagente de Bradford estoque (Tabela III-13) a 800 μL da solução de BSA, misturou-se 10 vezes por inversão, esperou-se 5 minutos e foi efetuada a leitura da absorbância em 595 nm. Utilizando sempre cubetas de plástico no espectrofotômetro Agilent 8453 ou Femto CIRRUS 80MB. Para concentrações mais altas, seguiu-se o mesmo procedimento, adicionando 40 μL da solução de BSA a 2 mL do reagente de Bradford diluído (Tabela III-13), misturou-se 10 vezes por inversão, esperou-se 5 minutos e foi efetuada a leitura da absorbância em 595 nm. Ressaltando que em alguns ensaios foi utilizado o reagente de Bradford comercial (B6916–Sigma).

Tabela III-13. Reagente de Bradford preparado

Solução estoque		Solução diluída	
Comassie brilliant blue G-250	$0,5 \text{ mg mL}^{-1}$	Solução estoque	20% (v/v)
CH ₃ OH	25% (v/v)	diluída em água e filtrada antes do uso	
H ₃ PO ₄	42,5% (v/v)		
solução estocada no escuro a 4 °C e filtrada antes do uso			

3.11. Espectrometria de Massas (MS)

Os experimentos de Espectrometria de Massas foram realizados com auxílio do grupo do Prof. Dr. Fabio C. Gozzo (IQ-UNICAMP), utilizando o equipamento Waters Synapt HDMS (ESI/MALDI-QToF, Waters). Para as análises de confirmação da sequência protéica e formação das ligações

dissulfeto utilizou-se a técnica de evaporação da gota em um MALDI-ToF/ToF, tendo como matriz uma solução recém-preparada do ácido α -ciano-4-hidróxi cinâmico (CHCA- 10 mg mL⁻¹) em água/acetonitrila (CH₃CN)/ácido trifluoracético (CF₃COOH) (4/5/1 em volume). As amostras protéicas desta etapa foram obtidas pela digestão de amostras de géis de eletroforese em uma dimensão (1D), conforme metodologia descrita a seguir.

As bandas cortadas dos géis de eletroforese foram colocadas em microtubos lavados com metanol e água deionizada. As bandas foram então lavadas por pelo menos três vezes com 200 μ L da solução B (Tabela III-14), sob agitação (vortex) por 10 minutos. Em seguida, lavou-se com 200 μ L de água deionizada e descartou-se o sobrenadante. Adicionou-se 25 μ L da solução D (Tabela III-14), para cobrir o gel, agitando rapidamente no vortex e deixou-se por uma hora a 56 °C. Transcorrido este tempo, o sobrenadante foi descartado, adicionou-se 25 μ L da solução E (Tabela III-14) ao gel, agitando rapidamente e deixou-se por 45 minutos, no escuro, à temperatura ambiente. Após este tempo descartou-se o sobrenadante e o gel foi lavado três vezes com 200 μ L da solução B (Tabela III-14) e depois com 200 μ L de água deionizada. Adicionou-se então 100 μ L de acetonitrila (acetonitrila grau HPLC - Merk KGaA) ao gel para desidratar, deixando cerca de 5 minutos à temperatura ambiente e depois se retirou o excesso de solvente. O gel foi seco no liofilizador para posterior digestão. Depois de seco, o gel foi reidratado com 15 μ L da solução de tripsina (Tripsin Gold-Mass Spec Grade, V5280 - Promega ou T-8642 - Sigma Chemical_Co) e deixado por 10 minutos a 4 °C. Em seguida adicionou-se a solução A (Tabela III-14) até cobrir o gel e este foi incubado a 37 °C por 16 horas. Transcorrido este tempo o sobrenadante foi transferido para um microtubo limpo e o gel foi lavado por pelo menos duas vezes com 30 μ L da solução C (Tabela III-14), agitando suavemente por

30 minutos e o sobrenadante foi juntado ao anterior. O sobrenadante foi então passado por uma coluna (zip-tip) C18 (Millipore) e adicionado à matriz para análise no espectrômetro de massas. Amostras do sobrenadante, concentradas em centrífuga do tipo *SpeedVac*, também foram analisadas diretamente por ESI-Q-ToF.

Amostras das proteínas puras intactas em solução tampão ou em água foram analisadas no ESI-Q-ToF para obtenção da massa exata das proteínas. Algumas amostras contendo proteína pura, porém de massa molar aparentemente semelhante à massa esperada, de acordo com análise de gel de eletroforese, foram submetidas a estudos por MS para identificação da proteína. Para tanto, as amostras protéicas foram digeridas de acordo com um procedimento semelhante ao anterior. Adicionou-se 100 μL de bicarbonato de amônio (NH_4HCO_3) 50 mmol L^{-1} a amostra protéica (cerca de 10 μL). Em seguida foi feita a redução, adicionando 5 μL de DTT 200 mmol L^{-1} , e fervendo a mistura por 10 minutos. Agitou-se a solução em vortex e a reação foi deixada por 45 min à temperatura ambiente. Na sequência foi feita a alquilação, adicionando-se 4 μL de iodoacetamida ($\text{CH}_2\text{I CONH}_2$) 1 mol L^{-1} . A solução foi agitada em vortex e deixada por mais 45 min à temperatura ambiente. A alquilação foi interrompida pela adição de 20 μL de DTT 200 mmol L^{-1} incubando a mistura reacional por 45 min à temperatura ambiente. Para a digestão adicionou-se 15 μL da solução de tripsina e incubou-se esta solução por 16 h a 37 °C. Transcorrido esse tempo, as amostras de proteínas digeridas foram aplicadas no espectrômetro de massas.

Ressaltando-se que as buscas para identificação das proteínas foram feitas utilizando o programa Mascot.¹¹⁹

¹¹⁹Matrix Science - Mascot (<http://dalton-server.iqm.unicamp.br/mascot>)

Tabela III-14. Soluções utilizadas na preparação das amostras para MS

Solução A		Solução C		Solução E	
NH ₄ HCO ₃	25 mmol L ⁻¹	CF ₃ COOH	5% (v/v)	CH ₂ ICONH ₂	55 mmol L ⁻¹
	pH 8	CH ₃ CN	50% (v/v)	NH ₄ HCO ₃	25 mmol L ⁻¹
Solução B		Solução D		Solução tripsina	
NH ₄ HCO ₃	25 mmol L ⁻¹	DTT	10 mmol L ⁻¹	tripsina	20 ngmL ⁻¹
CH ₃ CN	50% (v/v)	NH ₄ HCO ₃	25 mmol L ⁻¹	NH ₄ HCO ₃	25 mmol L ⁻¹
	pH 8		pH 8		

3.12. Fluorescência de Emissão

As medidas de fluorescência de emissão foram obtidas em um espectrofluorímetro Carey Eclipse (Varian), utilizando uma cubeta de quartzo com 1 cm de caminho óptico e as amostras de proteínas em concentrações de 1-5 $\mu\text{mol L}^{-1}$ em tampão fosfato de sódio (pH 8, 10 mmol L^{-1}). A excitação foi feita na região de 280 nm, fazendo uma varredura da emissão de 300 a 400 nm, utilizando a abertura da fenda de 5 nm em ambos os casos.

Ressaltando que todas as medidas de fluorescência de emissão tiveram a linha de base corrigida, utilizando como branco o tampão correspondente de cada amostra.

3.13. Dicroísmo Circular (CD)

As medidas de dicroísmo circular foram feitas em um espectropolarímetro JASCO J-720 (Tóquio, Japão), equipado com um controlador de temperatura (Peltier Type Control System PFD 4255, JASCO). Utilizando cubetas circulares de quartzo de 0,05 cm de caminho óptico ou cubetas de quartzo retangulares com 1 cm de caminho óptico e as amostras em concentrações de até 10 $\mu\text{mol L}^{-1}$ em tampão fosfato (pH 8) ou acetato (pH 5)

de sódio 10 mmol L^{-1} . As medidas foram feitas com uma velocidade de varredura de 20 nm/min de 198 a 260 nm , 16 ou 32 acumulações, a $20 \text{ }^\circ\text{C}$. Vale lembrar que as medidas foram feitas até atingir a voltagem máxima de 700 V na fotomultiplicadora. Para o registro e tratamento dos dados foram utilizados os programas Spectra Manager[®] (JASCO), CDNN Deconvolution^{®120} e K2D2.¹²¹

3.14. Espalhamento de Luz Dinâmico (DLS)

As medidas de DLS foram feitas em cubetas quartzo (ZS2112, QS3,00 mm, Hellma), em um equipamento Zetasizer nano series ZS (Malvern Instruments) e DLS-Wyat DynaPro-MS800 (Protein Solutions, LNBio/LNLS), com a amostra protéica em diferentes concentrações, em tampão na concentração de 10 mmol L^{-1} , a 20 ou $25 \text{ }^\circ\text{C}$. Foram feitas cerca de 50 medidas para cada amostra, com tempo de coleta de 10 segundos.

3.15. Ressonância Magnética Nuclear (NMR)

Preparo de amostras:

As amostras de $600 \text{ } \mu\text{L}$ foram preparadas com 10% de D_2O (Cambridge Isotope Laboratories), com as proteínas em concentrações de $\sim 0,5 \text{ mmol L}^{-1}$ em tampão fosfato ou acetato de sódio 10 mmol L^{-1} . As amostras do padrão de massa molar (kit LMW- GE Healthcare) foram preparadas em concentrações

¹²⁰ Böhm, G.; Muhr, R.; Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **1992**, *5*, 191-195.

¹²¹ (a) Andrade, M.A., P. Chacón, J.J. Merelo and F. Morán. Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Protein Eng.* **1993**, *6*, 383-390. (b) Merelo, J.J., M.A. Andrade, A. Prieto and F. Morán. Proteinotopic Feature Maps. *Neurocomputing* **1994**, *6*, 443-454.

na faixa de 0,5 mmol L⁻¹ em D₂O (99,9%), assim como as amostras de BSA, ovalbumina, α -quimiotripsinogênio, ribonuclease A e aprotinina.

Experimentos:

Os experimentos de NMR de ¹H e DOSY foram feitos no equipamento Varian INOVA 500 MHz utilizando a sonda de tripla ressonância (tri_res) de 5 mm a 25 °C ou a sonda penta, também de detecção indireta. Para a aquisição e processamento dos dados utilizou-se o software VNMRJ da Varian.

Os experimentos de ¹H foram obtidos com o equipamento operando na frequência de 499,98 MHz e os espectros foram referenciados pelo sinal residual da água (HDO) em 4,8 ppm, sendo que para supressão de água utilizou-se a sequência de pulso WATERGATE W5.¹²² Nesta sequência de pulso, os spins experimentam uma rotação de 180° através de pulsos sucessivos de 7,8, 18,5, 37,2, 70, 134,2, 134,2, 70, 37,2, 18,5, 7,8°. Foram executadas 128 repetições para cada amostra (número de transientes = 128).

Os experimentos de DOSY, por sua vez, foram feitos utilizando a sequência de pulso Dbppste. Inicialmente foi calibrada a largura do pulso de 90° (6-7 μ s), em seguida foram selecionados os parâmetros de aquisição: duração do gradiente de pulso (δ) de 3 ms e um tempo de difusão (Δ) de 100 ms, apropriados para macromoléculas. Foi feita então a otimização destes parâmetros variando a força do gradiente (G), normalmente de 500 a 28000 (G/s), até a obtenção do segundo espectro com 5% da intensidade do primeiro. Determinada então a melhor combinação entre estes três parâmetros, foi

¹²² (a) Liu, M.; Mao, X.; Ye, C.; Huang, H.; Nicholson, J.K.; Lindon, J.C. *J Magn. Reson.* **1998**, *132*, 125. (b) Sklenar, V.; Piotta, M.; Leppik, R.; Saudek, V. *J. Magn. Reson.* **1993**, *102A*, 241. (c) Piotta, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661. (d) Zheng, G.; Stait-Gardner, T.; Kumar, P.G.A.; Torres, A.M.; Price, W.S. PGSTE-WATERGATE: An STE-based PGSE NMR sequence with excellent solvent suppression. *J. Magn. Reson.* **2008**, *191*, 159-163.

selecionado o número de gradientes (32), tendo a força de gradiente mínima de 0 e máxima de 28000 (G/s). Foram acumulados 32 ou 64 espectros com tempo de aquisição de 120 minutos, sendo o último espectro com 2-5% da intensidade do primeiro.

3.16. Espalhamento de raios-X a baixo ângulo (SAXS)

As medidas de SAXS foram realizadas nas linhas de luz do LNLS: SAXS1, com detector Pilatus e SAXS2, que possui um detector do tipo CCD. Estas linhas operam na região de raios-X, com energia na faixa de 6-12 keV (1-2 Å). Para as análises foi utilizado um porta-amostra de mica, no qual se injetou ao redor de 250 µL de amostra. As medidas foram feitas a 25 °C, com comprimento de onda de 1,488 Å (CH₃(CH₂)₂₀COOAg - behenato de prata), a distância do detector foi de 1079,7 mm, correspondendo a um intervalo de q entre $0,0153 < q < 0,3056 \text{ \AA}^{-1}$ (sendo $q = 4\pi \sin\Theta / \lambda$, onde 2Θ é o ângulo de espalhamento). Os padrões de espalhamento das amostras protéicas e soluções tampão (Tabela III-15) foram coletados alternadamente, com um tempo de exposição de 600 s, tempo este, otimizado para reduzir os danos da radiação. Para cada amostra protéica foram coletadas cerca de seis medidas sucessivas. Como padrões, para posterior cálculo de massa molar, foram utilizados BSA e lisozima (Sigma-Aldrich), nas mesmas soluções tampões das proteínas e em diferentes concentrações (10 ou 3 mg/mL).

As imagens coletadas foram processadas utilizando o programa Fit2D¹²³ para normalização das intensidades, subtração dos sinais do tampão e integração das imagens, após construção de uma máscara gerando a curva de espalhamento. Sendo que a redução dos dados consiste na integração radial

¹²³ Hammersley, A.P.; Svensson, .O.; Hanfland, M.; Fitch, A.N.; Hausermann, D. Two-dimensional detector software: from real detector to idealized image or two-theta scan. *High Press Res.* **1996**, *14*, 235-248.

dos padrões de espalhamento 2D-SAXS para perfis de espalhamento 1D no programa Fit2D, seguido da média das curvas individuais, feita com o programa PRIMUS.¹²⁴

Tabela III-15. Soluções tampões utilizadas em SAXS

Tampão A		Tampão B		Tampão C	
Tris	5 mmolL ⁻¹	Tris-HCl	5 mmolL ⁻¹	Tris-HCl	5 mmolL ⁻¹
	pH 8		pH 5		pH 7

O raio de giro, R_g , e a intensidade de espalhamento no ângulo zero, $I(0)$, foram estimados utilizando a aproximação de Guinier, $I(q)=I(0)\exp(-q^2 R_g^{2/3})$, válida para ângulos pequenos ($q < 1,3/R_g$).^{96a} Os programas GNOM e AUTOGNOM^{124b} foram utilizados para obter a função de distribuição de distâncias, $P(r)$, a anisotropia da molécula e a distância máxima intramolecular, além de estimativas de R_g e $I(0)$ mais precisas. As curvas de Kratky, ($q^2 I(q) \times q$), calculadas a partir dos dados de espalhamento foram utilizadas para inferir o estado conformacional das proteínas em solução.¹²⁵ A massa molar das proteínas foi estimada, com margem de erro de cerca de 10%, utilizando um padrão (BSA ou lisozima) coletado nas mesmas condições das proteínas em estudo.¹²⁶ Também foi utilizado um método alternativo, para esta estimativa, baseado nas curvas isoladas, medidas em uma escala relativa

¹²⁴ (a) Koch, M.H.; Vachette, P.; Svergun, D.I. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **2003**, *36*, 147-227. (b) Konarev, P.V.; Volkov, V.V.; Petoukhov, M.V.; Svergun, D.I. ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2006**, *39*, 277-286.

¹²⁵ (a) Doniach, S. Changes in biomolecular conformations seen by small-angle X-ray scattering. *Chem. Rev.* **2001**, *101*, 1763-1778. (b) Putnam, C.D.; Hammel, M.; Hura, G.L.; Tainer, J.A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **2007**, *40*, 191-285.

¹²⁶ Mylonas, E.; Svergun, D.I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* **2007**, *40*, s245-s249.

(SaxsMow).¹²⁷ Foi obtido um modelo do envelope molecular, de baixa resolução, para duas proteínas estudadas a partir das curvas experimentais, utilizando o programa DAMMIF para gerar vinte modelos independentes. Em seguida a média destes modelos foi feita com os programas DAMAVER/DAMFILT.^{98b} Para uma melhor visualização foi feita uma máscara no arquivo de coordenadas obtido pelo procedimento anterior, utilizando o NCSMASK,¹²⁸ resultando em um modelo tridimensional final, de baixa resolução da estrutura das proteínas em estudo. Ressaltando que todas as análises foram feitas com a colaboração do Prof. Dr. Ricardo Aparicio.

3.17. Bioinformática

Foram realizadas buscas por homologia e similaridade de sequência primária através do alinhamento de sequências (BLAST), com o algoritmo blastp, utilizando como banco de dados o SWISSPROT e o PDB. As buscas foram feitas com as sequências das quatro proteínas em estudo, porém não resultou em nenhum alinhamento significativo (*E-values* ~1).¹²⁹ Na realidade, nenhum homólogo com estrutura tridimensional conhecida está disponível para as proteínas alvo. Assim, na ausência de modelos de alta resolução determinados por técnicas experimentais, a estrutura atômica das proteínas foi calculada por métodos computacionais usando o I-TASSER e o QUARK. As

¹²⁷ Fisher, H.; de Oliveira Neto, M.; Napolitano, H.B.; Polikarpov, I.; Craievich, A.F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **2010**, *43*, 101-109.

¹²⁸ (a) Collaborative Computational Project Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. Sect D: Biol. Crystallogr.* **1994**, *50*, 760-763. (b) Dodson, E.J.; Winn, M.; Ralph, A. Collaborative computational project, number 4: Providing programs for protein crystallography. *Methods Enzymol.* **1997**, *277*, 620-633.

¹²⁹ Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389-3402.

estruturas calculadas pelo I-TASSER¹³⁰ e QUARK¹³¹ auxiliaram na avaliação dos modelos de baixa resolução obtidos pelas curvas experimentais de SAXS. As análises dos modelos gerados pelas buscas de similaridade e alinhamento de estrutura foram feitas com o PDBeFold¹³², para classificação das mesmas na base de dados SCOP¹³³ (arquivo 1.73). As sobreposições dos modelos teóricos com os obtidos experimentalmente pelos dados de SAXS foram feitas utilizando o SUPCOMB.^{124,134} Para análise dos dados utilizou-se o GNUPLOT (gnuplot, versão gnuplot 4.4 patchlevel CVS) e para visualização e análises gráficas foram utilizados os programas RASMOL¹³⁵ e PyMOL (The PyMOL Molecular Graphics System, verso 1.3, Schrödinger, LCC).¹³⁶ Os dados de distribuição de cargas na superfície da estrutura protéica foram obtidos com o

¹³⁰ (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.

¹³¹ (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).

¹³² Krissinel, E.; Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect D: Biol. Crystallogr.* **2004**, *60*, 2256-2268.

¹³³ (a) Andreeva, A.; Howorth, D.; Chandonia, J.M.; Brenner, S.E.; Hubbard, T.J.; Chothia, C.; Murzin, A.G. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* **2008**, *36*, D419-D425. (b) Murzin, A.G.; Brenner, S.E.; Hubbard, T.; Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **1995**, *247*, 536-540.

¹³⁴ Kozin, M.B.; Svergun, D.I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **2001**, *34*, 33-41.

¹³⁵ (a) Sayle, R.A.; Milner-White, E.J. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **1995**, *20*, 374-376. (b) Bernstein, H.J. Recent changes to RasMol, recombining the variants. *Trends Biochem. Sci.* **2000**, *25*, 453-455.

¹³⁶ Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037-10041.

PDB2PQR¹³⁷ utilizando o campo de força AMBER e as propriedades eletrostáticas foram calculadas com a ferramenta APBS do PyMOL.

3.18. Cristalografia

As lamínulas, onde se deposita a gota da solução protéica, foram silanizadas utilizando Aquasil 20% ou solução de clorotrimetilsilano em etanol (10%, v/v).

As soluções de tampões (Tabela III-16), sais (Tabela III-18) e precipitantes (Tabela III-17) que constituem os kits de cristalização foram preparadas no laboratório de Biologia Estrutural e Cristalografia (IQ-UNICAMP). Todas as soluções foram preparadas em água deionizada, as soluções dos sais foram filtradas em filtros 0,22 µm e algumas soluções de precipitantes foram aquecidas durante a preparação. Ressaltando que foram preparados cerca de 50 mL de cada solução.

De posse das lamínulas silanizadas, preparou-se as placas de cristalização (já com silicone para selagem posterior), adicionando-se 500 µL de solução (correspondentes às 98 condições de cristalização dos kits comercializados HR2-110 e HR2-112 – Crystal Screen e Crystal Screen2 – da Hampton Research e algumas condições exóticas¹³⁸) em cada poço de cristalização. A amostra protéica em tampão Tris-HCl (5 mmol L⁻¹, pH 8), na concentração de 3,6 mg mL⁻¹ foi então colocada para cristalizar utilizando o método de *Hanging Drop*, *i.e.*, difusão de vapor em gota suspensa. Para tanto, se adicionou na lamínula uma gota no centro, contendo 1,5 µL da solução do

¹³⁷ Dolinsky, T.J.; Czodrowski, P.; Li, H.; Nielsen, J.E.; Jensen, J.H.; Klebe, G.; Baker, N.A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522-W525.

¹³⁸ Brzozowski, A.M.; Walton, J. Clear strategy screens for macromolecular crystallization. *J. Appl. Cryst.* **2001**, *34*, 97-101.

poço e 1,5 μL da solução protéica. Com uma pinça, a lamínula contendo a gota foi colocada de forma a fechar o poço, com a gota para baixo e, feito isso, para as 98 condições, a placa foi selada e acondicionada a 18 °C para crescimento dos cristais.

Estas placas foram observadas em microscópio ótico e não foi observado crescimento de nenhum cristal, exceto inorgânicos.

Tabela III-16. Soluções tampão para cristalização

Acetato de sódio 2 mol L ⁻¹ pH 4.6	Citrato de sódio tribásico 1 mol L ⁻¹ pH 5.6
Tris-HCl 1 mol L ⁻¹ pH 8.5	Imidazol 1 mol L ⁻¹ pH 6.5
HEPES 1 mol L ⁻¹ pH 7.5	MES 0,5 mol L ⁻¹ pH 6.5
Cacodilato de sódio 0,5 mol L ⁻¹ pH 6.5	Bicina 1 mol L ⁻¹ pH 9.0

Tabela III-17. Soluções precipitantes para cristalização

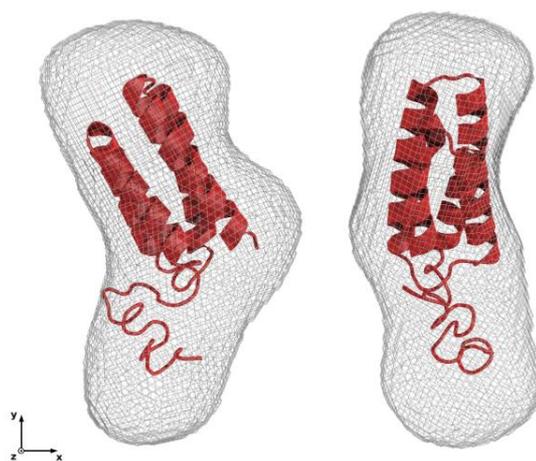
PEG 4000 50% (m/v)	PEG 8000 44,6%
Formato de sódio 3 mol L ⁻¹	PEG 1500 50%
PEG 6000 50%	Formato de magnésio 1 mol L ⁻¹
Etilenoglicol 50%	Brometo de hexadeciltrimetilamônio 1 mol L ⁻¹
Imidazol 1 mol L ⁻¹ pH 7	PEG 1000 50%
Sulfato de magnésio 2 mol L ⁻¹	PEG monometiléter 2000 50%
PEG 20000 50%	Jeffamina M-600 50% (v/v)
Formato de amônio 3 mol L ⁻¹	PEG monometiléter 5000 50%
PEG 10000 50%	Citrato de sódio tribásico 1,6 mol L ⁻¹ pH 6.5
Cloreto de magnésio 5 mol L ⁻¹	

Tabela III-18. Soluções salinas para cristalização

Cloreto de cálcio 1 mol L ⁻¹	Fosfato de potássio monobásico 1,5 mol L ⁻¹
Citrato de sódio tribásico 1,6 mol L ⁻¹	Acetato de zinco 0,5 mol L ⁻¹
Cloreto de magnésio 2 mol L ⁻¹	Cloreto de sódio 5 mol L ⁻¹
Acetato de amônio 1 mol L ⁻¹	Cloreto de cobalto (II) 0,5 mol L ⁻¹
Sulfato de amônio 3,5 mol L ⁻¹	Cloreto de cádmio 1 mol L ⁻¹
Sulfato de lítio 2 mol L ⁻¹	Tartarato de sódio e potássio 1,5 mol L ⁻¹
Acetato de sódio 2 mol L ⁻¹	Cloreto de ferro (III) 0,5 mol L ⁻¹
Cloreto de céσιο 0,5 mol L ⁻¹	Sulfato de zinco 0,5 mol L ⁻¹
Acetato de magnésio 1 mol L ⁻¹	Fosfato de sódio monobásico 1,5 mol L ⁻¹
Sulfato de cádmio 0,5 mol L ⁻¹	Cloreto de níquel (II) 0,5 mol L ⁻¹
Fosfato de amônio monobásico 2,5 mol L ⁻¹	

As 98 condições utilizadas estão apresentadas em anexo (Tabelas A1-A3).

Capítulo 4



RESULTADOS E DISCUSSÃO

4. Resultados e Discussão

Após sequenciamento do genoma da *Xac*,²⁰ quarenta, dentre as proteínas hipotéticas identificadas, foram consideradas prováveis chaperonas de secreção.^{15,63} Proteínas hipotéticas são proteínas cuja existência é predita pela análise genômica, entretanto não há evidências experimentais de que elas sejam expressas *in vivo*; e suas funções podem ser especuladas, através de buscas por homologia.

Estas quarenta proteínas hipotéticas foram consideradas possíveis chaperonas de secreção após uma análise mais pontual do genoma da *Xac* utilizando ferramentas de bioinformática; onde dentre as proteínas hipotéticas foram buscadas proteínas com características de chaperonas de secreção dos sistemas flagelar, T3SS e T4SS.⁶³ Algumas dessas quarenta proteínas hipotéticas da *Xac* foram clonadas por membros do grupo^{15,107,108} e dentre elas foram escolhidas quatro (Tabela IV-1) para a realização dos estudos estruturais aqui propostos. Ressaltando que a confirmação da sequência de aminoácidos (estrutura primária) delas foi feita por sequenciamento de DNA.

Tabela IV-1. Proteínas em estudo, sequência de aminoácidos e massa molar (MM)

Proteína	Sequência de aminoácidos	MM (Da)
XAC1346 (pI 5,6)	MKTIHMLSLSAIAALTLAACQGPDAEQARRDAASA AK HAAAATEQAADRAAVEAKAATADASVAARELAA DASRATANAADATADKTREMANKAEQNADKHDADHAQH	10945,9
XAC0419 (pI 7,0)	MPHDPSVVAHVEKTCTAQLAHFASDITRVDVHLRD ENGQRGGAADRTCSIEAHVAGLPPIAATNSAE TTASAVTGAARKLRRTAIEHARGKQHSKATAPPPDML	10767,0
XACb0033 (pI 5,6)	MAHVNSRVQKHRDALRMAGLRPVQIWVPDTRRPD FAEECRRQCRLAAQADMADTDMQRFMDEALADM DGWTE	8418,5
XAC1990 FlgN (pI 6,7)	MNVNEFLQRLSDALAGERQALLENIDGLMRHTQDKLS ALQALEAAMPAGEEERLRELAEANRANGALLARRRRE VNWALRHLGRTESAPSYDAKGQSSVLRGGRSLAVA	12121,6

A escolha foi feita baseada em resultados anteriores, que indicaram alto nível de expressão para duas delas, o que facilita o trabalho, principalmente em se tratando de estudos estruturais, nos quais se necessitam amostras protéicas concentradas. Além disso, estas quatro proteínas já tinham as condições de expressão padronizadas.^{15,107,108} E, como elas foram consideradas possíveis chaperonas de secreção da *Xac*, acredita-se que podem estar envolvidas no mecanismo de virulência da bactéria.

É importante mencionar que em um trabalho recente sobre o proteoma da *Xac*, que identificou 30% das proteínas preditas pelo genoma, através de estudos por MS, as proteínas XAC1346 e XAC0419 foram identificadas.⁶ Portanto estas duas proteínas hipotéticas são realmente expressas pela *Xac*.

Depois de confirmada – por mini-indução – a inserção do DNA de interesse nas células pela transformação, deu-se início a produção das proteínas em grande quantidade para os ensaios de caracterização. A expressão por tempo variado, dependendo da proteína em questão. Um esquema geral da expressão das proteínas em larga escala está mostrado no fluxograma da Figura IV-1.

Transcorrido o tempo de indução de cada proteína, o crescimento bacteriano foi interrompido e a suspensão protéica foi centrifugada. O precipitado bacteriano foi estocado a -20 °C para posterior lise, sendo observado que após cerca de um ano em estoque o precipitado ainda pôde ser utilizado normalmente.

Vale ressaltar que as colônias contendo o DNA das proteínas de interesse foram mantidas em estoque, em placas de Petri a 4 °C, sendo feita a manutenção a cada quinze dias.

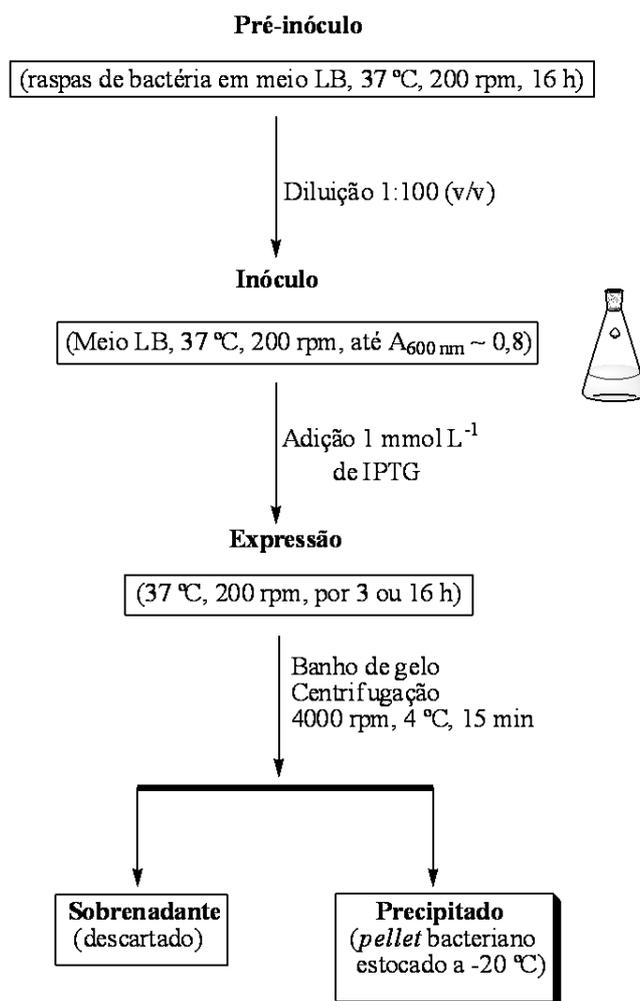


Figura IV-1. Fluxograma geral da expressão das proteínas alvo.

No procedimento de lise utilizou-se ultrassom de ponta (sonicação) para quebra da membrana celular e liberação do conteúdo da célula; procedimento que também é útil para fragmentar moléculas de DNA.

A lise, assim como a expressão protéica, foi acompanhada por eletroforese em gel, coletando alíquotas do lisado total, sobrenadante e precipitado obtidos após centrifugação. O processo de lise foi o mesmo para as quatro proteínas alvo (Figura IV-2).

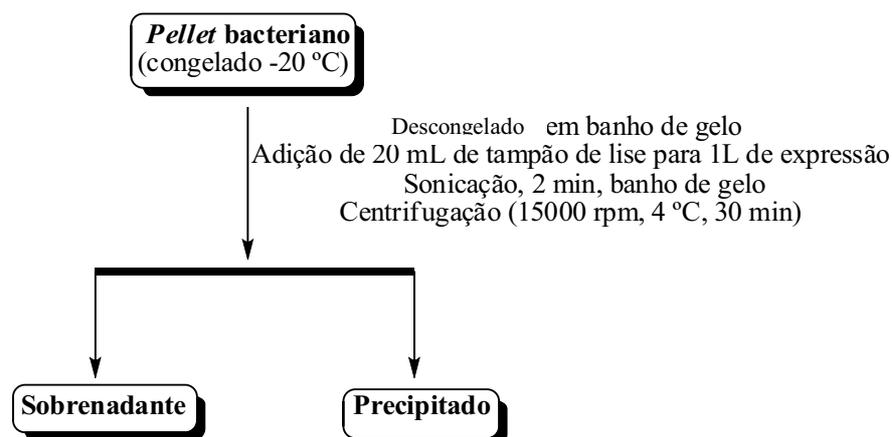


Figura IV-2. Fluxograma geral do processo de lise do precipitado bacteriano.

Após a lise, como duas proteínas alvo (XAC1346 e XAC0419) foram encontradas no sobrenadante e as outras duas no precipitado, cada uma recebeu um tratamento específico até a obtenção da proteína pura para os estudos estruturais. Portanto, na sequência a obtenção e caracterização de cada proteína será discutida individualmente.

Estas quatro proteínas vêm sendo estudadas pelo grupo da Profª. Dra. Ljubica desde 2005, portanto alguns ensaios como estabilidade térmica por dicroísmo circular e tempo de vida de fluorescência de triptofano, por exemplo, já haviam sido realizados anteriormente.^{1b,2b,108}

4.1. XAC1346

A expressão da proteína XAC1346 é baixa, tanto que não é bem visualizada em gel de eletroforese conforme se observa na Figura IV-3, entretanto, nas etapas subsequentes de lise e purificação, as bandas desta proteína começam a aparecer, ainda que fracas. É importante ressaltar que os géis de eletroforese apresentados no trabalho são géis qualitativos, sendo que não foram tomados cuidados de aplicação sempre da mesma quantidade de amostra para efeitos de uma análise quantitativa.

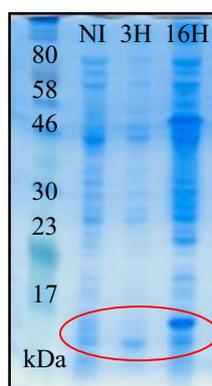


Figura IV-3. Gel de eletroforese SDS-PAGE (15%) da expressão da XAC1346, tendo da esquerda para a direita: marcador de massa molar (P7709S-NEB); NI corresponde à amostra não induzida; 3H a 3 horas e 16H a 16 horas depois da adição de IPTG.

Como pode ser observado na Figura IV-4, a XAC1346, mesmo não sendo muito bem visualizada nos géis de eletroforese, após a lise se encontra na porção sobrenadante, sendo o precipitado, portanto, descartado.

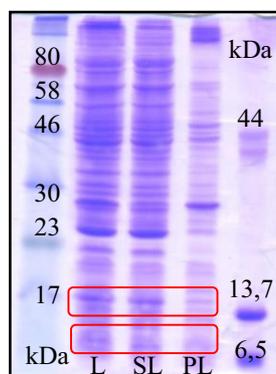


Figura IV-4. Gel de eletroforese SDS-PAGE (15%) da lise bacteriana, tendo da esquerda para a direita: marcador de massa molar de proteína (P7709S-NEB); (L) lisado total; (SL) sobrenadante da lise; (PL) precipitado da lise e outro marcador de massa molar.

Como a proteína foi encontrada no sobrenadante da lise, efetuou-se diálise em tampão fosfato de sódio (50 mmol L^{-1} , pH 8). Esse processo teve o intuito de retirar do sobrenadante todo e qualquer sal que estava nele pela

adição do tampão de lise e fazer com que a proteína se encontrasse no mesmo tampão a ser usado na purificação (Figura IV-5).

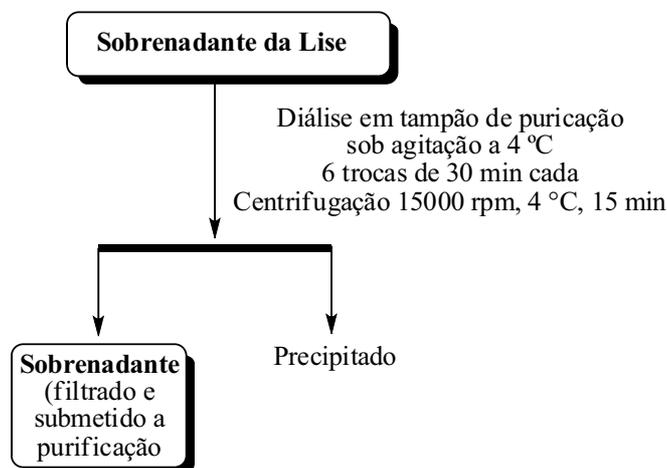


Figura IV-5. Fluxograma da preparação do sobrenadante da lise para posterior purificação da XAC1346.

No passo seguinte, a amostra foi filtrada e em seguida, foi realizada primeiramente a purificação por cromatografia líquida de troca iônica e depois por filtração em gel, pois as alíquotas do sobrenadante da lise apresentam muitos contaminantes.

A XAC1346 apresenta pI baixo (5,64), portanto, o sobrenadante da lise foi purificado utilizando uma coluna de troca aniônica (DEAE-Sepharose) com as amostras em tampão fosfato de sódio (50 mmol L^{-1} , pH 8). Neste pH, a proteína de interesse se encontra na forma de ânions e, portanto, ao passar pela coluna, interage com a resina por interação eletrostática. Em seguida, ao se adicionar gradativamente uma solução de cloreto de sódio, ocorre a troca aniônica, pois o ânion cloreto com maior afinidade se liga à resina e a proteína de interesse na sua forma aniônica é liberada (Figura IV-6).

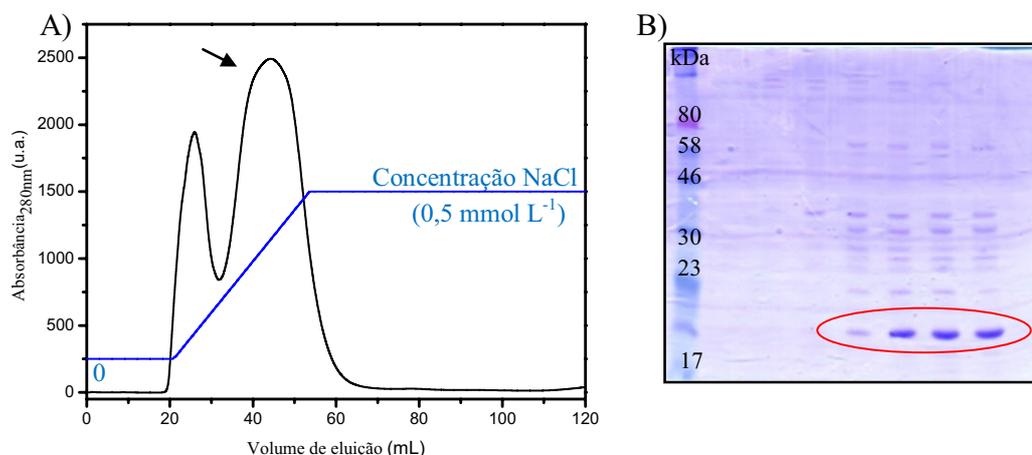


Figura IV-6. Primeira purificação da XAC1346. (A) Cromatograma da purificação da XAC1346 (DEAE-Sepharose), com o pico correspondente indicado. B) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: marcador de massa molar (P7709S-NEB) e algumas frações coletadas.

Na sequência, as alíquotas contendo a proteína de interesse, eluídas com cerca de 30% da solução de cloreto de sódio (1 mol L^{-1}), foram agrupadas e submetidas a etapa subsequente de purificação, cromatografia por filtração em gel, (Figura IV-7).

A técnica de filtração em gel também foi utilizada de maneira analítica, a fim de estimar a massa molar aproximada da proteína e determinar seu estado oligomérico. Para tanto, primeiramente foi feita uma calibração do experimento, utilizando o kit de calibração para filtração em gel, de baixa massa molar da GE Healthcare (*Gel filtration calibration kit LMW*) contendo as seguintes proteínas: aprotinina, ribonuclease A, anidrase carbônica, ovalbumina e conalbumina (Tabela IV-2).

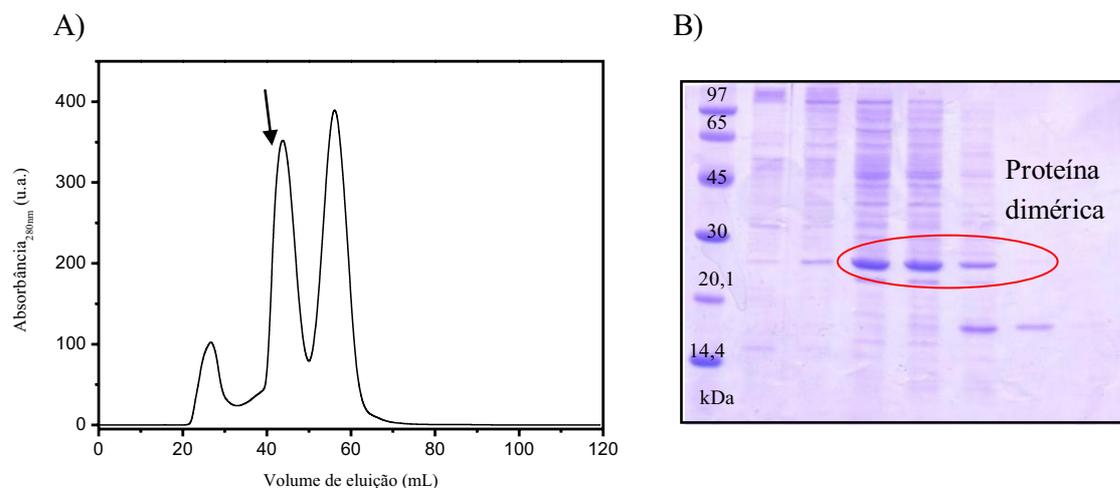


Figura IV-7. Segunda purificação da XAC1346, por filtração em gel. A) Cromatograma da filtração em gel, sendo o pico correspondente à proteína indicado. B) Gel de eletroforese SDS-PAGE (15%) referente ao cromatograma tendo na sequência: o marcador de massa molar (kit LMW – GE Healthcare) e diferentes frações da purificação.

Inicialmente foi aplicada na coluna Superdex 200 uma solução de Azul de Dextran (2 MDa) em tampão fosfato de sódio (50 mmol L⁻¹, pH 7,2), contendo cloreto de sódio (150 mmol L⁻¹). Aplicou-se 1 mL da solução correspondendo a menos de 2% do volume da coluna – máximo permitido para uma boa resolução – num fluxo de 0,7 mL/minuto para determinação do volume vazio da coluna que foi de 23,2 mL.¹³⁹

Em seguida, durante a calibração, foram feitas duas corridas com diferentes misturas para calibração. A primeira mistura continha aprotinina, ribonuclease A e ovalbumina (Mix A) no mesmo tampão da Azul de Dextran e a corrida cromatográfica foi realizada nas condições supracitadas. O mesmo procedimento foi seguido para a segunda mistura contendo ribonuclease A,

¹³⁹ (a) Erickson, H.P. Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proced. Online* **2009**, *11*, 32-51. (b) Winzor, D.J. Analytical exclusion chromatography. *J. Biochem. Biophys. Methods* **2003**, *56*, 15-52.

anidrase carbônica e conalbumina (Mix B). Para a construção da curva de calibração (Figura IV-8, Tabela IV-2) e determinação do coeficiente de partição das proteínas foi utilizada a fórmula mostrada a seguir:

$$k_{av} = \frac{V_e - V_0}{V_c - V_0}$$

onde: k_{av} - coeficiente de partição; V_e - volume de eluição; V_0 - volume vazio e V_c - volume da coluna.

Tabela IV-2. Dados para calibração da filtração em gel na Superdex 200 ($V_c = 58,3$ mL)

Proteína	Massa Molar (kDa)	Concentração aplicada	V_e (mL)	K_{av}
Aprotinina	6,5	69,2 $\mu\text{mol L}^{-1}$	47,7	0,70
Ribonuclease A	13,7	32,8 $\mu\text{mol L}^{-1}$	44,7	0,61
Anidrase Carbônica	29,0	15,5 $\mu\text{mol L}^{-1}$	41,6	0,52
Ovalbumina	44,0	18,2 $\mu\text{mol L}^{-1}$	38,7	0,44
Conalbumina	75,0	6 $\mu\text{mol L}^{-1}$	36,0	0,37

Para a construção da curva de calibração (Figura IV-8B) foi retirado o ponto referente à anidrase carbônica, pois esta proteína foi de difícil dissolução na solução tampão, formando um pouco de precipitado, necessitando centrifugação e filtração para retirada deste último. Excluindo este ponto da curva de calibração, foi obtido um melhor ajuste ($r^2 = 0,998$) em comparação com o valor antes da exclusão deste ponto ($r^2 = 0,992$).

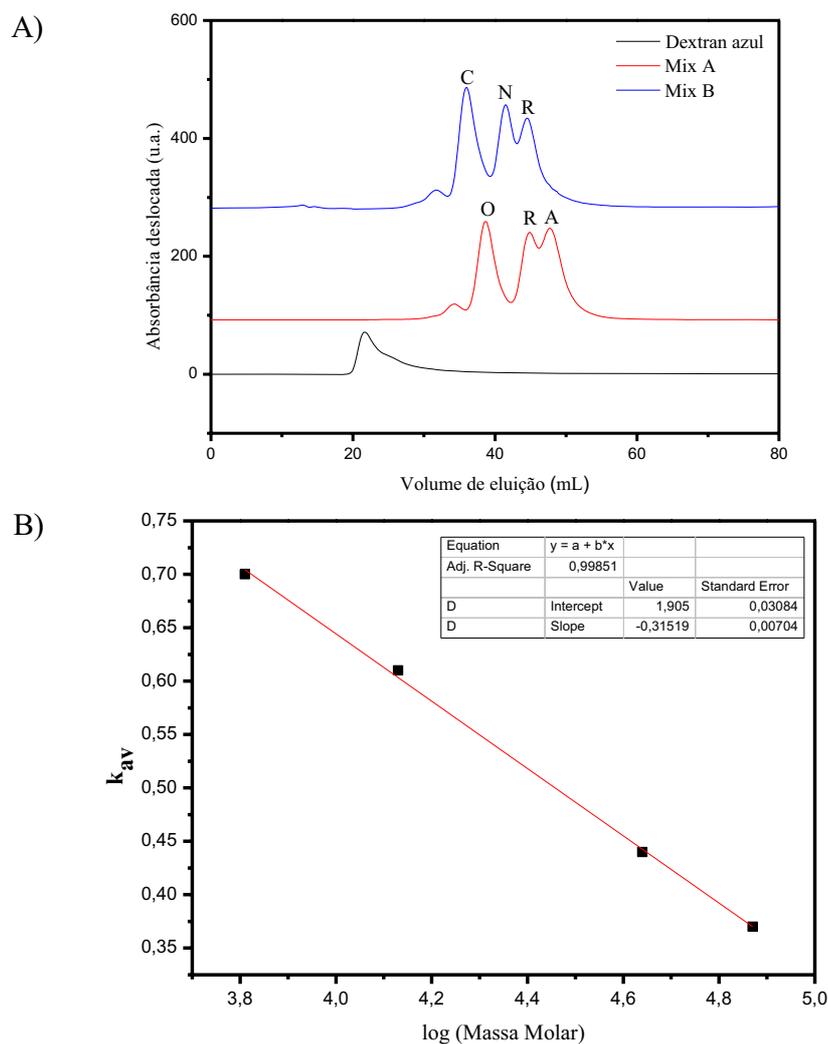


Figura IV-8. Ilustração dos dados da filtração em gel analítica. A) Cromatogramas utilizados para a construção da curva de calibração, onde: C corresponde a conalbumina (75 kDa), N a anidrase carbônica (29 kDa), R a ribonuclease A (13,7 kDa), O a ovalbumina (44 kDa) e A a aprotinina (6,5 kDa). B) Curva de calibração tendo o coeficiente de partição (k_{av}) em função do logaritmo da massa molar (log Massa Molar).

Tendo sido realizada a calibração da filtração em gel na coluna Superdex 200, foram aplicadas amostras contendo a XAC1346 relativamente pura, utilizando as mesmas condições da calibração. O volume de eluição da XAC1346 foi 43,5 mL (Figura IV-9), correspondendo a um coeficiente de

partição (k_{av}) de 0,58. Plotando este valor na curva de calibração, obtemos uma massa molar de 17688 Da; sendo a massa teórica 10945,9 Da (Tabela IV-1), pode-se inferir que provavelmente a proteína não é globular e/ou que a mesma se encontra em solução, nestas condições em forma dimérica.

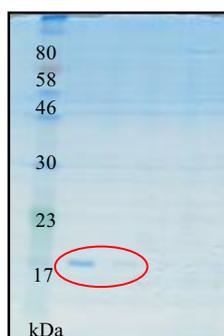


Figura IV-9. Gel de eletroforese SDS-PAGE (15%) da alíquota advinda da filtração em gel analítica, referente à XAC1346, (marcador de massa molar – P7709S/NEB – à esquerda).

Lembrando que, teoricamente o gel de eletroforese também pode ser utilizado como um primeiro resultado para estimar a massa molar, foi construída uma curva do fator de retenção (R_f) no gel *versus* o logaritmo da massa molar para um gel SDS-PAGE (Figura IV-10) e plotando o valor do R_f da XAC1356 nesta curva, obteve-se 17260 Da.

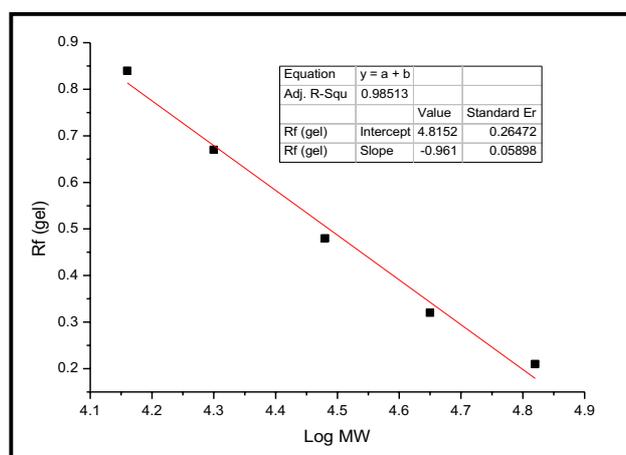


Figura IV-10. Curva do R_f do gel de eletroforese em função do logaritmo da massa molar.

Assim, concluiu-se que ambas estimativas devem indicar presença da proteína na forma dimérica em solução.

As amostras coletadas, contendo a proteína pura, foram concentradas e antes dos ensaios de caracterização estrutural, foi necessária a quantificação da proteína na amostra. Como a XAC1346 não apresenta resíduos de triptofano, tirosina e fenilalanina, que são responsáveis pela absorção de radiação em 280 nm, para a quantificação desta proteína utilizou-se o método de Bradford.¹¹⁸ Para tanto foi construída a curva de calibração utilizando como padrão BSA (Figura IV-11).

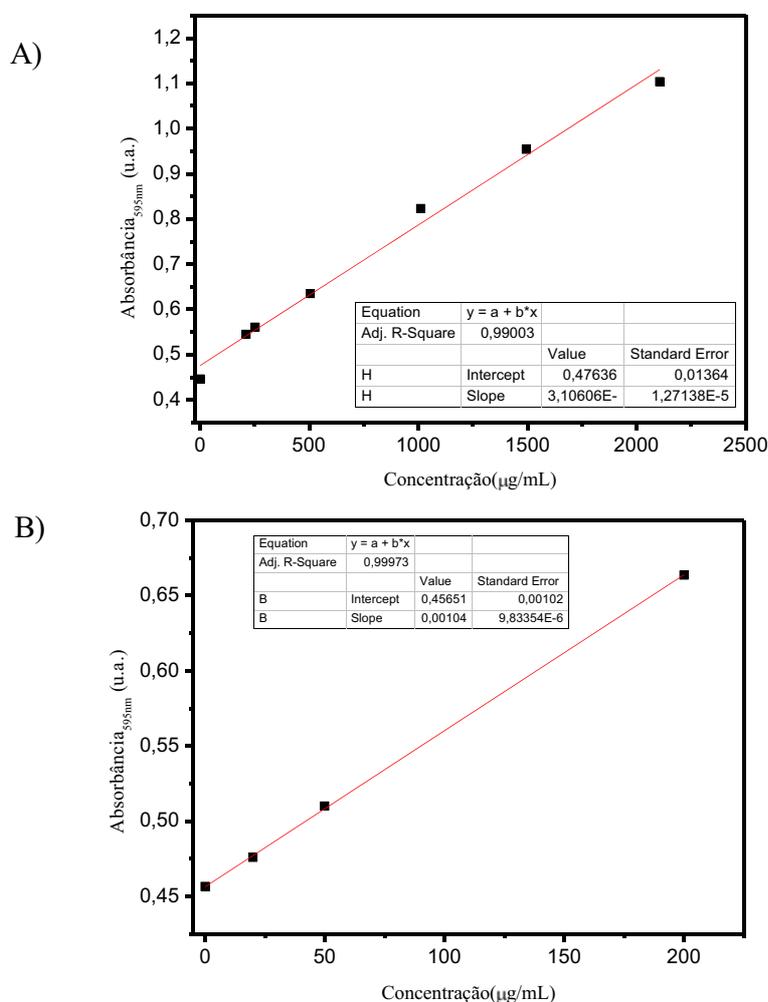


Figura IV-11. Curva de calibração de Bradford para quantificação da proteína. A) Curva para altas concentrações. B) Curva para baixas concentrações protéicas.

Em seguida foi feita a reação da amostra contendo a proteína e o reagente de Bradford e quantificou-se a mesma na amostra (Figura IV-12).

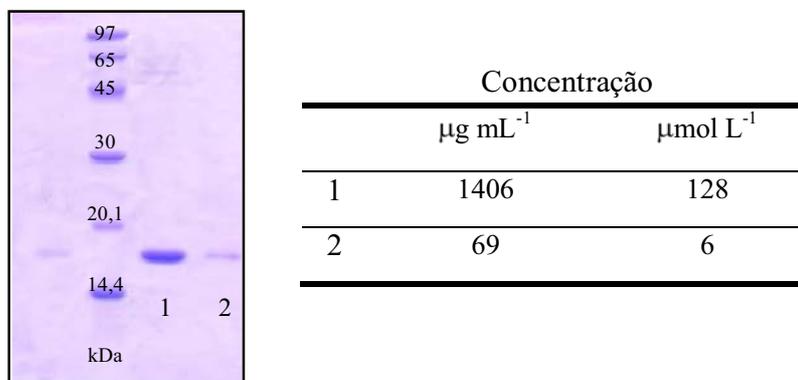


Figura IV-12. Gel de eletroforese SDS-PAGE (15%) de alíquotas da XAC1346 em solução, tendo da esquerda para a direita: marcador de massa molar (kit LMW – GE Healthcare) e duas alíquotas contendo a proteína pura, utilizadas para quantificação. A tabela ao lado indica a concentração dessa proteína nas alíquotas 1 e 2.

Como a massa molar estimada para a XAC1346 (~17,7 kDa) difere da massa molar teórica (21,9 kDa para a forma dimérica), as alíquotas representadas na Figura IV-2 foram submetidas a análises por espectrometria de massas para a identificação da proteína. Para tanto, foi efetuado o processo de redução seguido de alquilação e, por fim, foi feita a digestão com tripsina. Amostras da proteína digerida foram então levadas ao espectrômetro de massas para análise por LC-MS/MS. Para a identificação das proteínas, foi feita uma busca no programa de busca Mascot¹¹⁹, utilizando os espectros de MS/MS e a XAC1346 foi identificada nas duas amostras (Figura IV-13).

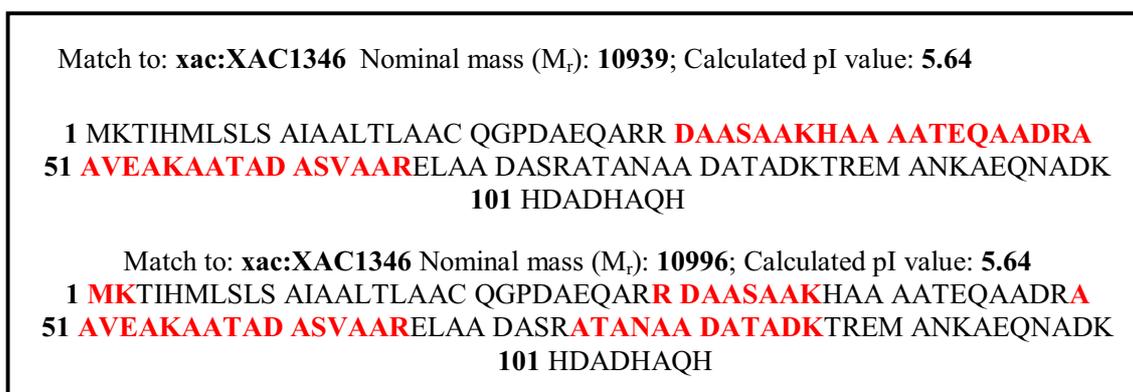


Figura IV-13. Ilustração da identificação da XAC1346 por MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 33% de cobertura na primeira e 36% na segunda amostra.

Apesar da baixa cobertura obtida pelos experimentos de espectrometria de massas, como a estrutura primária da proteína já havia sido obtida pelo sequenciamento de DNA e estes experimentos de MS foram realizados apenas para certificar que as alíquotas da purificação realmente correspondiam à proteína de interesse, os resultados obtidos foram considerados satisfatórios. O problema da baixa cobertura poderia ser resolvido com diferentes condições de digestão, por exemplo. Entretanto, não foi necessário realizar tais experimentos.

Tendo identificado a proteína, que como visto apresenta um comportamento anômalo no gel de eletroforese, primeiramente foram realizados experimentos de dicroísmo circular para avaliar se a proteína obtida estava enovelada e obter informações sobre a estrutura secundária da mesma (Figura IV-14).

Os dados de dicroísmo circular foram tratados por diferentes programas para calcular a quantidade de estrutura secundária. Como uma das características das chaperonas de secreção é o alto conteúdo helicoidal, na Tabela IV-3 estão apresentados somente os dados referentes à quantidade de

hélice- α . Também, utilizando programas de bioinformática foi calculada a estrutura secundária da XAC1346, com base na sua sequência de aminoácidos.

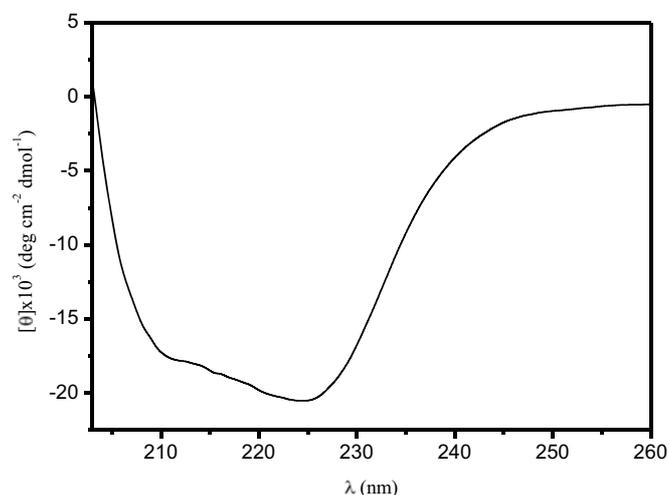


Figura IV-14. Espectro de dicroísmo circular da proteína XAC1346 ($14,3 \mu\text{mol L}^{-1}$) em tampão fosfato de sódio (10 mmol L^{-1} , pH 8).

Tabela IV-3. Dados de dicroísmo circular para a XAC1346

Programas utilizados	Porcentagem de hélice- α predita	Porcentagem de hélice- α calculada
GOR ¹⁴⁰	82%	-
NetSurfP ¹⁴¹	75%	-
K2D2 ¹²¹	-	67%
	-	$[\theta]_{222\text{nm}} = 60\%$ ⁶⁹

De acordo com a predição (Tabela IV-3) a proteína deve apresentar de 70 a 80% dos aminoácidos em estrutura helicoidal. Pelos dados experimentais de dicroísmo circular tem-se que a proteína apresenta cerca de 70% de

¹⁴⁰ Combet, C.; Blanchet, C.; Geourjon, C.; Deléage, G. *NPS@: Network Protein Sequence Analysis TIBS* **2000**, 25, 147.

¹⁴¹ Petersen, B.; Petersen, T.N.; Andersen, P.; Nielsen, M.; Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* **2009**, 9, 51 (doi:10.1186/1472-6807-9-51).

estrutura helicoidal. Assim, pode se dizer que os dados preditos são condizentes com os dados experimentais e que a XAC1346 tem um alto conteúdo de estrutura helicoidal, apresentando uma das características comuns das chaperonas de secreção conhecidas.

Foram realizadas medidas de DLS para obter uma estimativa do tamanho da proteína, através do raio de Stokes. O raio de Stokes ou raio hidrodinâmico corresponde ao raio de uma molécula solvatada, e refere-se ao caso idealizado de uma molécula esférica. Como muitas biomoléculas não são esféricas, o raio de Stokes aproximado representa o raio correspondente a uma esfera com volume equivalente.

Obteve-se um raio de Stokes de 29 Å, correspondendo a 93% em massa das partículas em solução, com uma polidispersidade de 7,3%, indicativos de uma amostra monodispersa ($59 \mu\text{mol L}^{-1}$). A massa molar estimada pelas medidas de DLS foi de cerca de 40 kDa, indicando, conforme discutido anteriormente, que a proteína não deve ser esférica, pois para uma molécula alongada, a massa molar estimada a partir do raio hidrodinâmico é sempre superior à massa molar real. A forma alongada da proteína é corroborada pela razão R_H/R_0 igual a 1,9 que indica uma proteína levemente alongada.

Foram feitos ensaios de NMR com intuito de obter o coeficiente de difusão da molécula e de posse deste dado, utilizando o raio de Stokes obtido por DLS poder inferir a forma da molécula. Porém, como a proteína agrega em altas concentrações, não foi possível realizar experimentos de difusão. E, as tentativas de concentrar a amostra para posterior análise de SAXS também levaram a agregação. Para solucionar este problema posteriormente podem ser feitos estudos de estabilidade da XAC1346, variando pH ou força iônica da solução, ou utilizando co-fatores por exemplo.

Buscas por homologia e similaridade de sequência primária no BLAST,^{129,142} utilizando os parâmetros padrão, não resultaram em alinhamentos significativos ($E\ value < 1$). Assim como as buscas usando o PDB como banco de dados. Portanto, como não foram encontradas estruturas com similaridade determinadas por técnicas experimentais, utilizou-se o programa I-TASSER¹³⁰ para cálculo da estrutura tridimensional da XAC1346. Este programa utiliza o alinhamento com diferentes estruturas conhecidas para calcular uma estrutura tridimensional com base na sequência de aminoácidos. No caso desta proteína, os cinco principais modelos de estruturas gerados pelo I-TASSER apresentaram um $TM = 0,36 \pm 0,12$ ($TM > 0,5$ indicam modelos de topologia correta e $TM < 0,17$ indicam modelos de topologia randômica). Este valor (TM) é uma forma de avaliação quantitativa da qualidade das estruturas geradas; é um algoritmo que calcula a similaridade da topologia de duas estruturas protéicas, sendo que valores inferiores a 0,17 indicam um modelo selecionado aleatoriamente da biblioteca do PDB. Sendo importante ressaltar que os modelos gerados podem não corresponder à estrutura real da molécula, por isso é preciso cuidado ao analisar os resultados, sempre os confrontando com dados experimentais.

De acordo com o modelo (Figura IV-15), os 13 resíduos que correspondem ao domínio anfipático no C-terminal⁶³ estão estruturados na forma de uma hélice- α , outra característica estrutural das chaperonas de secreção.

¹⁴²Altschul, S.F.; Wootton, J.C.; Gertz, E.M.; Agarwala, R.; Morgulis, A.; Schäffer, A.A.; Yu, Y.-K. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **2005**, *272*, 5101-5109.

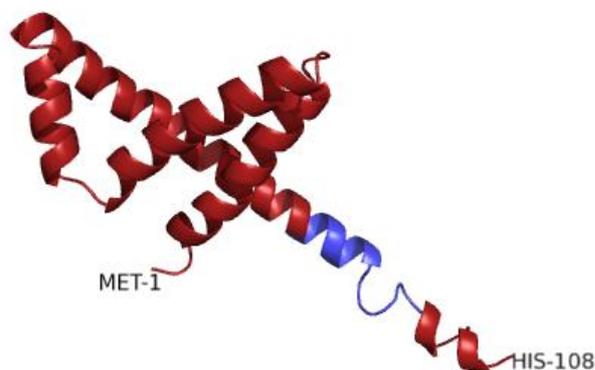


Figura IV-15. Modelo da XAC1346. Modelo do I-TASSER (modelo de maior pontuação dentre os modelos gerados, $C\text{-score} = -3,14$) com N- e C-terminais marcados, e a hélice anfipática destacada em azul.

De acordo com o modelo obtido pelo I-TASSER a proteína apresenta 79% de estrutura helicoidal. A partir do programa k2D2¹²¹ foi gerado um espectro de dicroísmo circular correspondente a uma proteína com 79% de hélice- α e feita uma comparação com o espectro obtido experimentalmente (Figura IV-16).

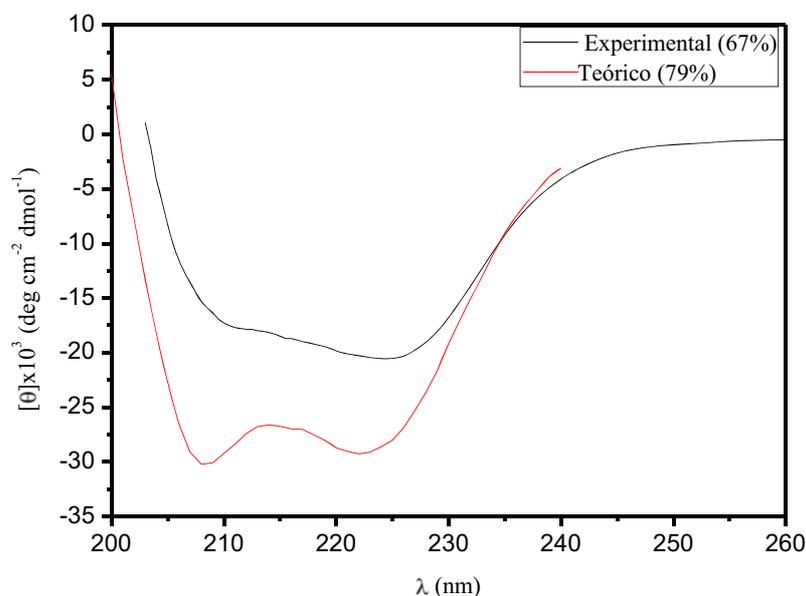


Figura IV-16. Comparação das curvas de CD, em preto a curva experimental e em vermelho a curva teórica (200-240 nm), gerada pelo K2D2 para a estrutura calculada pelo I-TASSER.

Fazendo uma comparação, a distância máxima observada no modelo calculado é de cerca de 70 Å, que corresponderia a uma esfera de 35 Å. Este valor é próximo ao obtido como raio de Stokes por DLS (29 Å). Como este dado experimental é menor que o calculado, provavelmente indica que a proteína em solução é mais compacta do que o modelo. A estrutura calculada apresenta forma alongada, corroborando também com os dados de DLS, onde foi suposto que a proteína é alongada.

Atualmente vem sendo desenvolvido um novo programa (QUARK) para cálculo de estrutura tridimensional, sem a utilização de estruturas conhecidas como moldes, baseando-se apenas na sequência de aminoácidos. O QUARK é um algoritmo computacional para o cálculo *ab initio* da estrutura de proteínas, que visa construir um modelo tridimensional baseando se apenas na sequência de aminoácidos da proteína. Os modelos são construídos a partir de pequenos fragmentos (1-20 aminoácidos) por simulações de Monte Carlo tendo como base o conhecimento do campo de força em nível atômico.

Através deste programa foi possível calcular mais um modelo tridimensional para a XAC1346 que é bastante semelhante ao obtido pelo I-TASSER, diferindo principalmente nas regiões randômicas (Figura IV-17). Sendo este modelo mais compacto, com uma distância máxima em torno de 60 Å e com 82% de estrutura helicoidal.

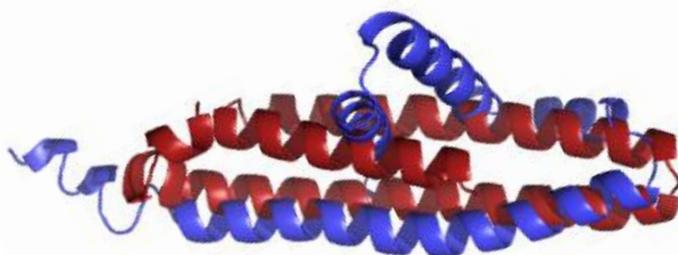


Figura IV-17. Sobreposição dos modelos tridimensionais da XAC1346, obtidos por dois programas diferentes. Em vermelho a estrutura calculada pelo QUARK e em azul a estrutura calculada pelo I-TASSER.

A XAC1346 é considerada uma possível chaperona de secreção do T3SS de acordo com buscas por homologia de sequência com chaperonas do T3SS. Tais buscas foram realizadas em trabalhos anteriores, utilizando algoritmos específicos. O gene que codifica a XAC1346 se encontra no cromossomo da *Xac* e observou-se 51% de similaridade com a proteína FliJ de *Rhodobacter sphaeroides*, cuja estrutura tridimensional não foi resolvida (gene gi1518878 – Figura IV-18).⁶³ Como foi discutido, é uma proteína pequena, com massa molar teórica de 10,9 kDa, pI de 5,6 e os resíduos envolvidos no domínio anfipático são os resíduos de 88 a 100, segundo a predição.⁶³

Por fim, analisando esta última estrutura calculada pelo QUARK frente à estrutura de uma FliJ com estrutura conhecida, no caso a FliJ de *Salmonella enterica* serovar Typhimurium¹⁴³ no PyMol, (Figura IV-19) pode se dizer que as duas proteínas além de apresentarem 51% de similaridade de estrutura primária, também apresentam uma certa similaridade estrutural.

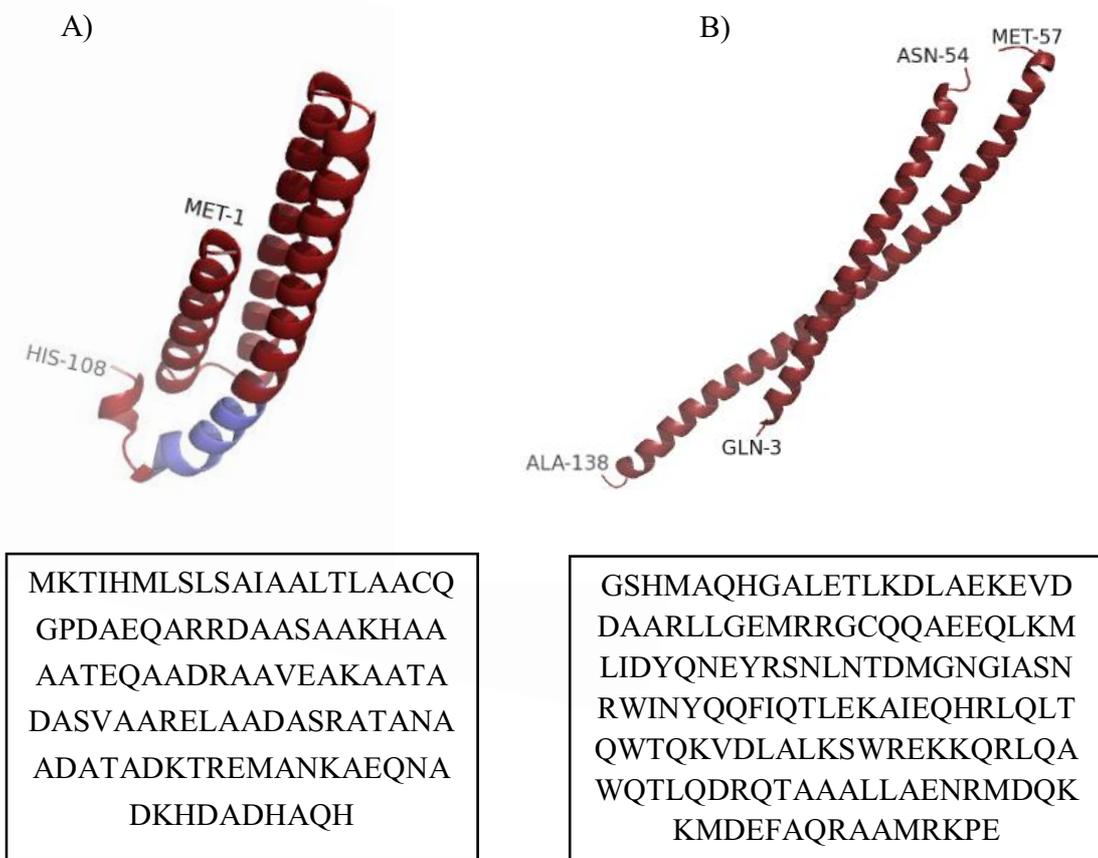


Figura IV-18. Resultados comparativos para: (A) Acima a estrutura 3D da XAC1346 calculada pelo QUARK, ressaltando N- e C-terminais, mais a região anfipática em azul no C-terminal e embaixo a sequência de aminoácidos; (B) Acima a estrutura da FliJ (PDB 3AJW),¹⁴³ e embaixo sua sequência de aminoácidos.

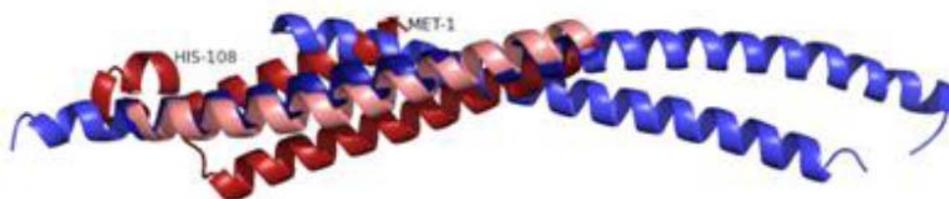


Figura IV-19. Sobreposição das estruturas: da XAC1346 calculada pelo QUARK, em vermelho, ressaltando o N- e o C-terminal e FliJ (PDB 3AJW)¹⁴³ em azul, com a região onde as hélices se sobrepõem destacada.

¹⁴³ Ibuki, T.; Imada, K.; Minamino, T.; Kato, T.; Miyata, T.; Namba, K. Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. *Nat. Struct. Mol. Biol.* **2011**, *18*, 277-282.

Fazendo o alinhamento das estruturas em programas apropriados¹⁴⁴ observa-se quantitativamente esta semelhança estrutural (Figura IV-20).

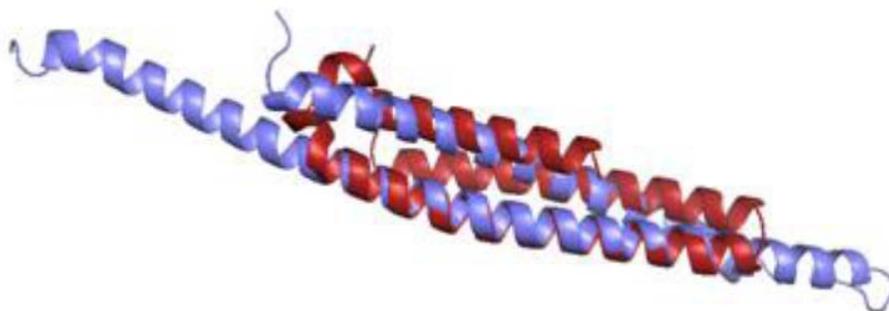


Figura IV-20. Alinhamento das estruturas: da XAC1346 calculada pelo QUARK, em vermelho, e FliJ (PDB 3AJW)¹⁴³ em azul, sendo alinhados 72 aminoácidos dos 108 da XAC1346 ($TM\text{-Score} = 0,47$; onde $TM\text{-Score} > 0,5$ indica que as estruturas partilham o mesmo enovelamento e $TM\text{-Score} < 0,2$ indica que não há similaridade estrutural).

Portanto, pode-se dizer que, apesar da ausência de dados estruturais para comprovar que a XAC1346 é uma chaperona de secreção, com os resultados apresentados observa-se um forte indício de que ela seja uma chaperona de secreção da *Xac*. Pode-se especular ainda que, como ela apresenta similaridade estrutural e de sequência com chaperonas flagelares, é expressa pela *Xac* que apresenta um flagelo, porém nenhuma chaperona flagelar foi predita no seu genoma, a XAC1346 tem forte indício de ser uma chaperona flagelar.

4.2. XAC0419

Assim como para a proteína discutida na seção anterior, a expressão da XAC0419 também é baixa (Figura IV-21), sendo mais bem visualizada apenas nos géis de eletroforese das purificações.

¹⁴⁴ Zhang, Y.; Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score *Nucleic Acids Res.* **2005**, *33*, 2302-2309.

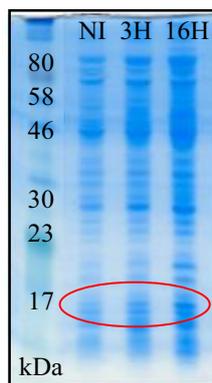


Figura IV-21. Gel de eletroforese SDS-PAGE (15%) da expressão da XAC0419, tendo na sequência: marcador de massa molar (P7709S – NEB); NI corresponde à amostra não induzida; 3H a 3 horas e 16H a 16 horas depois da adição de IPTG.

Da mesma maneira que a XAC1346, após a lise a XAC0419 se encontra na porção sobrenadante (Figura IV-22).

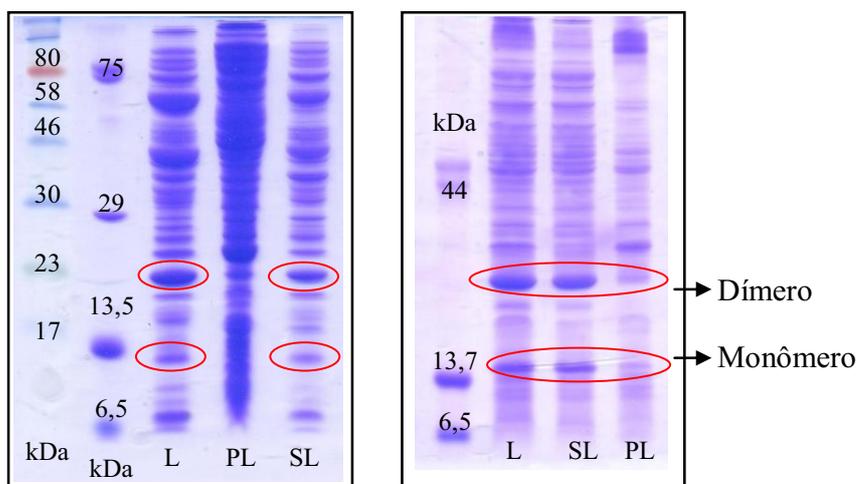


Figura IV-22. Gel de eletroforese SDS-PAGE (15%) da lise bacteriana, tendo indicados: marcador de massa molar de proteína (P7709S – NEB); solução de proteínas com massa molar conhecida; (L) lisado total; (SL) sobrenadante da lise e (PL) precipitado da lise; com monômero e dímero indicados.

A solução contendo a XAC0419 foi dialisada em tampão acetato de sódio (50 mmol L^{-1} pH 5), para acondicionar a proteína no tampão utilizado na purificação (Figura IV-23).

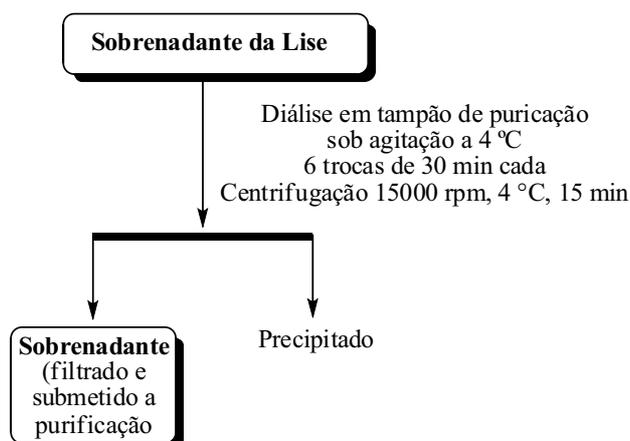


Figura IV-23. Fluxograma da preparação do sobrenadante da lise para posterior purificação da XAC0419.

Após centrifugar e filtrar a amostra, foi realizada a etapa de purificação por cromatografia líquida de troca catiônica, e em seguida a purificação por filtração em gel.

Esta proteína, cujo pI é 7,06, em tampão acetato de sódio (50 mmol L^{-1} , pH 5), encontra-se na forma de cátions e liga-se à resina aniônica por interações eletrostáticas, sendo retirada após aplicação de gradiente de cloreto de sódio (1 mol L^{-1}), com ~20% de sal (Figura IV-24).

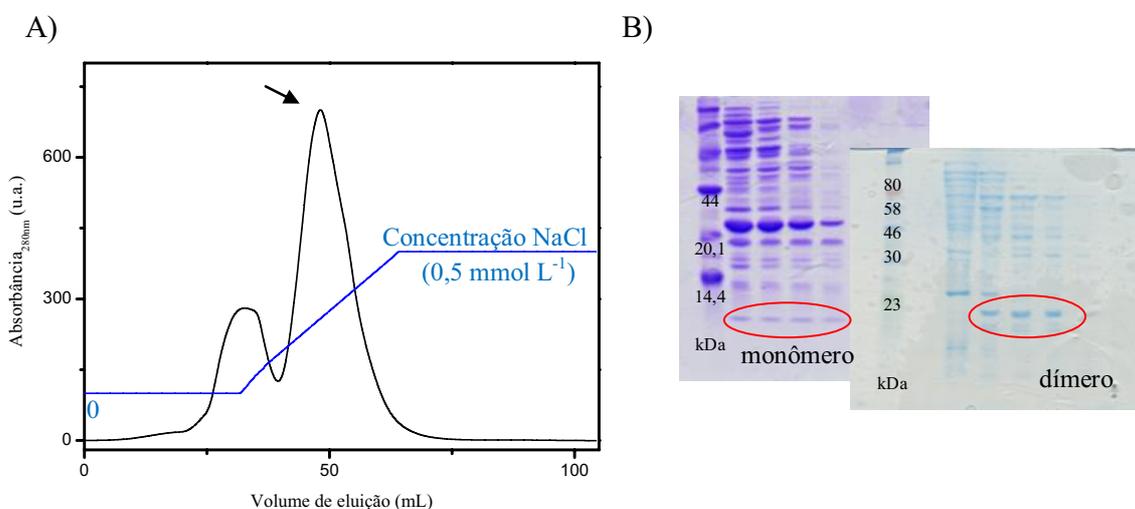


Figura IV-24. Primeira purificação da XAC0419 (CM-Sepharose). (A) Cromatograma da purificação com o pico correspondente à proteína indicado. B) Géis de eletroforese SDS-PAGE (15%) da purificação, tendo da esquerda para a direita: marcador de massa molar (kit LMW/GE Healthcare e P7709S/NEB) e algumas frações coletadas, com as frações contendo dímero e monômero indicadas.

As frações da CM-Sepharose contendo a proteína foram submetidas à nova purificação, utilizando cromatografia por filtração em gel (Superdex 200) para obtenção da proteína pura (Figura IV-25).

A técnica de filtração em gel também foi utilizada de maneira analítica, para se ter uma estimativa da massa molar da proteína e determinar o estado oligomérico da mesma. Para tanto, foi utilizada a curva de calibração construída anteriormente (Figura IV-8), cuja equação está demonstrada abaixo:

$$k_{av} = 1,905 - 0,31519 \cdot \log \text{ Massa Molar}$$

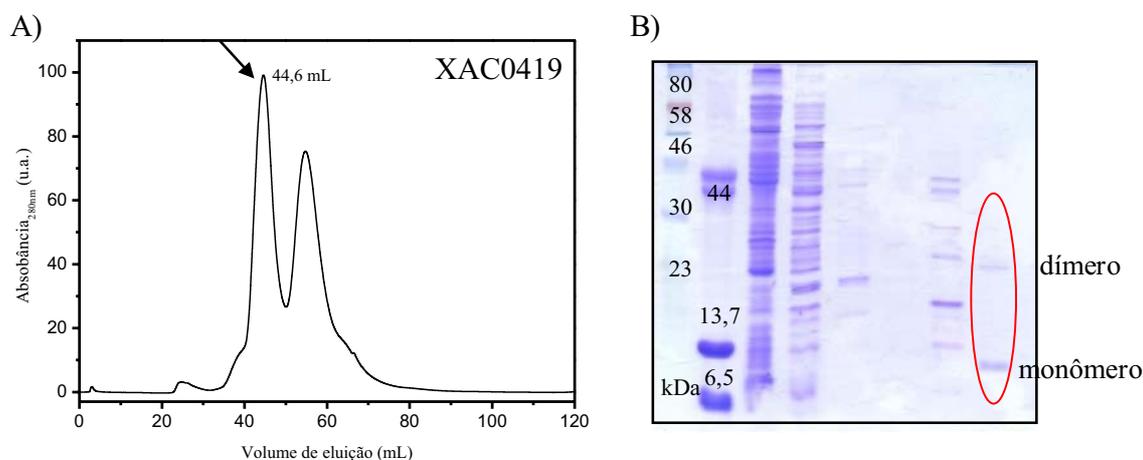


Figura IV-25. Segunda purificação da XAC0419, por filtração em gel. A) Cromatograma da filtração em gel, sendo o pico correspondente a proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) referente ao cromatograma tendo na sequência: marcador de massa molar (P7709S/NEB) e diferentes frações da purificação, com a fração utilizada na filtração em gel analítica indicada no gel.

A amostra quase pura da XAC0419 foi aplicada na coluna Superdex, utilizando as mesmas condições da calibração. O volume de eluição da XAC0419 foi 44,6 mL (Figura IV-26), correspondendo a um coeficiente de partição (k_{av}) de 0,61 e massa molar de 12818 Da. Sendo a massa teórica 10767,0 Da (Tabela IV-1), concluiu-se, portanto, que a proteína se encontra em solução na forma monomérica, nas condições utilizadas.

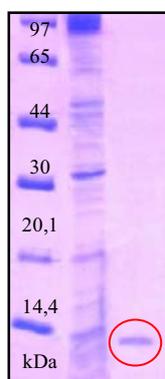


Figura IV-26. Gel de eletroforese SDS-PAGE (15%) da alíquota advinda da filtração em gel analítica referente à XAC0419 (marcador de massa molar kit LMW/GE Healthcare).

Utilizando novamente o gel de eletroforese para estimar a massa molar para a XAC0419, construiu-se uma curva de calibração para o gel da Figura IV-24B e plotando o valor do R_f tanto para a forma dimérica quanto monomérica da proteína nesta curva, obteve-se a massa molecular de 10715 Da para o monômero e 21379 Da para o dímero.

Para estimar o tamanho e a forma da XACb0419 através da filtração em gel analítica,^{139,145} novamente foi realizada uma calibração do experimento com proteínas de raio de Stokes conhecidos.

A nova calibração foi feita utilizando as seguintes proteínas padrão: aprotinina, ribonuclease A, α -quimiotripsinogênio, ovalbumina e BSA. Aplicou-se novamente na coluna Superdex 200 uma solução da dextran azul. Foram aplicados 500 μ L da solução, correspondendo a menos de 1% do volume da coluna. As cinco proteínas padrão foram aplicadas separadamente na coluna, em triplicata. Para a construção da curva de calibração (Tabela IV-4, Figura IV-27) e determinação do coeficiente de partição das proteínas utilizou-se a fórmula mostrada anteriormente.

Tabela IV-4. Dados da calibração da filtração em gel analítica

Proteína	R_s (Å)	Concentração aplicada	V_e (mL)	K_{av}
Aprotinina	9,83	770 μ mol L ⁻¹	47,93 \pm 0,04	0,70
Ribonuclease A	16,4	299 μ mol L ⁻¹	44,4 \pm 0,2	0,60
α -quimiotripsinogênio	20,9	400 μ mol L ⁻¹	42,69 \pm 0,04	0,56
Ovalbumina	30,5	182 μ mol L ⁻¹	38,52 \pm 0,07	0,44
BSA	35,3	123 μ mol L ⁻¹	35,84 \pm 0,05	0,36

¹⁴⁵ Job, V.; Mattei, P.J.; Lemaire, D.; Attree, I.; Dessen, A. Structural basis of chaperone recognition of type III secretion system minor translocator proteins. *J. Biol. Chem.* **2010**, 285, 23224-23232.

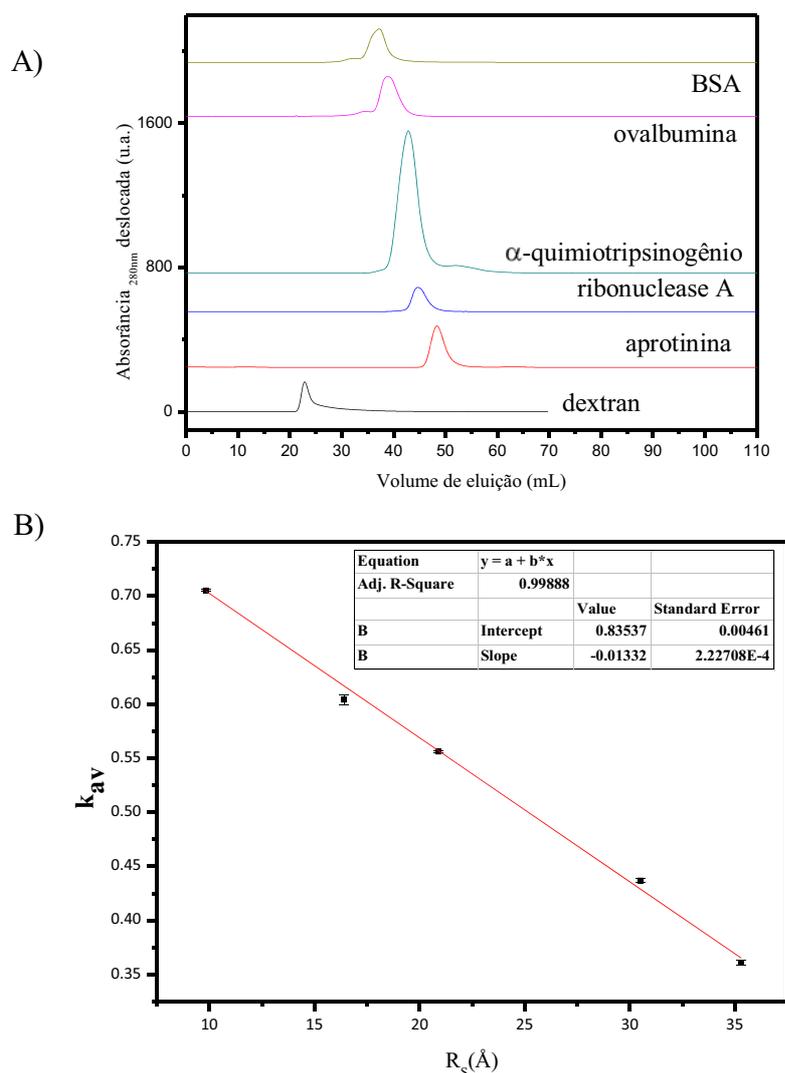


Figura IV-27. Filtração em gel analítica. A) Cromatogramas utilizados para a construção da curva de calibração. B) Curva de calibração tendo o coeficiente de partição (k_{av}) em função do raio de Stokes (R_s).

Utilizando o coeficiente de partição da XAC0419 na curva de calibração da Figura IV-27B, obteve-se uma estimativa do raio de Stokes da proteína, em torno de 17 Å. Comparando este valor com o raio de Stokes das proteínas usadas na calibração, pode-se dizer que provavelmente a proteína apresenta uma forma alongada.

Para a determinação da concentração da proteína em solução utilizou-se novamente o método de Bradford porque a XAC0419 também não apresenta resíduos de triptofano, tirosina e cistina. Utilizando as curvas de calibração da Figura IV-11 determinou-se a concentração da XAC0419 nas amostras (Figura IV-28).

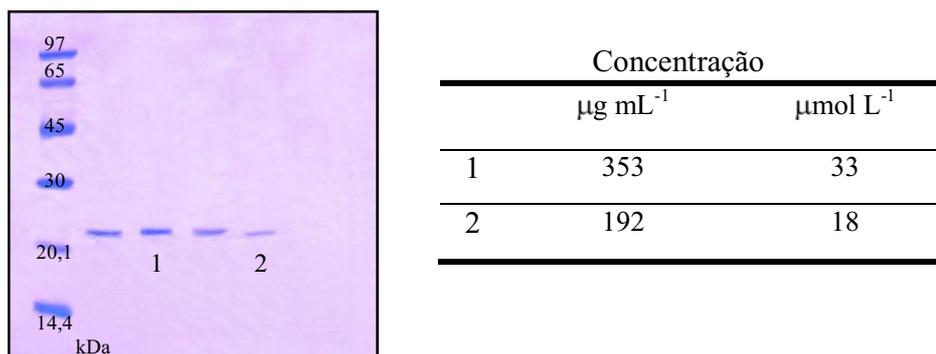


Figura IV-28. Gel de eletroforese SDS-PAGE (15%) de alíquotas da XAC0419 em solução, tendo na sequência: marcador de massa molar (kit LMW – GE Healthcare) e alíquotas contendo a proteína pura, utilizadas para quantificação. A tabela ao lado indica a concentração dessa proteína nas alíquotas 1 e 2.

Amostras contendo tanto a proteína dimérica quanto monomérica, também foram digeridas para identificação por espectrometria de massas. Novamente, para a identificação da proteína, foi feita uma busca no programa de busca Mascot¹¹⁹, utilizando os espectros de MS/MS e a XAC0419 foi identificada tanto na amostra monomérica quanto dimérica (Figura IV-29). Apesar da baixa cobertura de sequência, como a sequência de DNA já havia sido confirmada, estes resultados permitem certificar que a proteína isolada era a XAC0419.

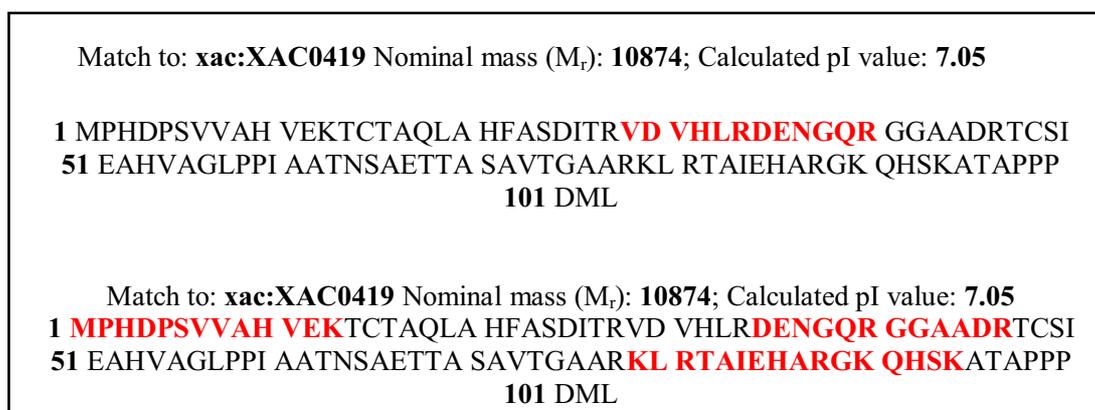


Figura IV-29. Identificação da XAC0419 por MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 11% de cobertura para a amostra dimerica (superior) e 39% para a amostra monomérica (inferior).

Na sequência foram realizados experimentos de dicroísmo circular para avaliar se a proteína obtida estava enovelada e obter informações sobre a estrutura secundária da mesma (Figura IV-30).

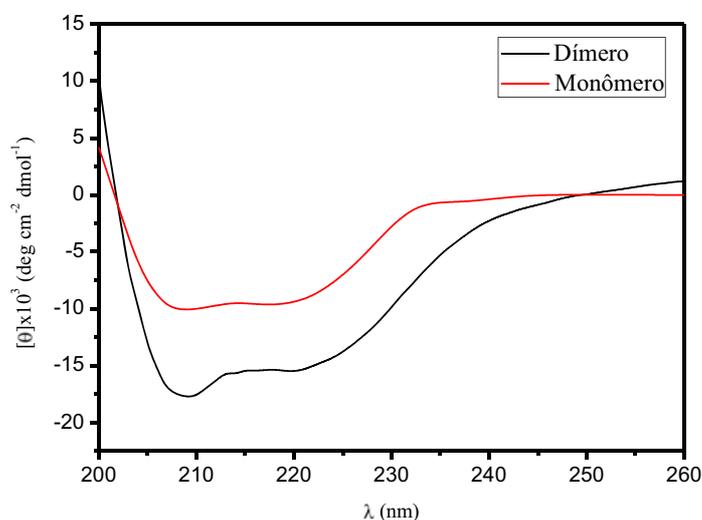


Figura IV-30. Espectros de Dicroísmo Circular da XAC0419 na forma monomérica ($7 \mu\text{mol L}^{-1}$, em vermelho) e dimerica ($18 \mu\text{mol L}^{-1}$, em preto) em tampão acetato de sódio (50mmol L^{-1} , pH 5). O branco foi feito com o mesmo tampão.

Os dados dos espectros de dicroísmo circular da XAC0419 foram utilizados em diferentes programas para calcular sua quantidade de estrutura secundária. Os valores calculados utilizando dados teóricos e experimentais da porção helicoidal desta proteína estão sumarizados na Tabela IV-5.

Tabela IV-5. Dados de dicroísmo circular para a XAC0419

Programas utilizados	Porcentagem de hélice- α teórica	Porcentagem de hélice- α experimental
GOR ¹⁴⁰	34%	-
NetSurfP ¹⁴¹	39%	-
Jpred ¹⁴⁶	41%	-
CDNN ¹⁴⁰	-	30% (54%)*
K2D2 ¹²¹	-	26% (60%)
		$[\theta]_{222\text{nm}} = 30\% (46\%)^{69}$
		$[\theta]_{208\text{nm}} = 39\% (59\%)$

*entre parêntesis está a porcentagem de hélice- α para a forma dimérica.

De acordo com a predição (Tabela IV-5) a proteína deve apresentar em torno de 38% dos resíduos de aminoácidos em estrutura helicoidal. Pelos dados experimentais de dicroísmo circular tem-se que a proteína apresenta cerca de 30% de estrutura helicoidal quando na sua forma monomérica, enquanto que na forma dimérica, apresenta de 50 a 60%. Assim, pode se dizer que os dados experimentais são condizentes com os calculados.

Também foram realizadas medidas de DLS para obter uma estimativa do tamanho desta proteína. O raio de Stokes foi estimado em 8 Å, correspondendo a 94% em massa das partículas em solução, com uma

¹⁴⁶ Cole, C.; Barber, J.D.; Barton, G.J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **2008**, *35*, W197-W201.

polidispersidade de 11%. Portanto uma amostra monodispersa ($55 \mu\text{mol L}^{-1}$). Entretanto, se compararmos com o raio obtido a partir da filtração em gel analítica, observa-se uma discrepância nestes valores. Porém, como a amostra submetida às medidas de DLS a princípio se apresentou agregada, sendo necessário centrifugá-la e filtrá-la novamente antes de efetuar a medida, acredita-se que o resultado da filtração em gel analítica seja mais coerente e mais confiável, descartando-se o resultado de DLS.

Também foram feitos ensaios de NMR para obtenção do coeficiente de difusão da molécula. Porém em uma amostra, pelo espectro de ^1H observou-se que a proteína estava desenovelada (Figura IV-31) e em outra amostra a proteína estava muito diluída, não sendo possível a realização dos experimentos de difusão. O que novamente demanda estudos posteriores adicionais da estabilidade da proteína.

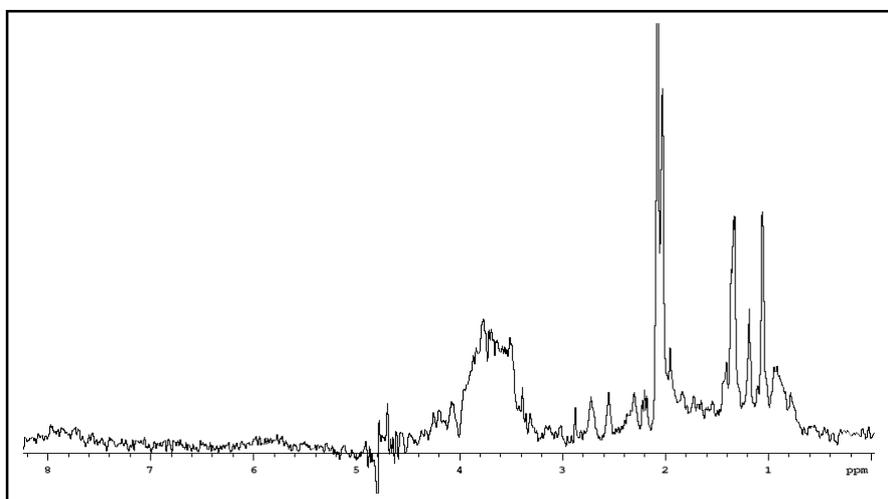


Figura IV-31. Espectro de NMR de ^1H da XAC0419 ($0,1 \text{ mmol L}^{-1}$) em tampão fosfato de sódio 5 mmol L^{-1} (D_2O 10%) obtido no equipamento Varian INOVA 500 MHz (499,98 MHz), utilizando a sequência de pulso WATERGATE para supressão do sinal de água em 4,8 ppm.

A ausência de sinais na região de 6-9 ppm indicam a presença da proteína desenovelada em solução (–NH– da ligação peptídica troca com D₂O e não é observado) .

Amostras da XAC0419 foram submetidas à análise de SAXS na linha SAXS2 do LNLS, equipada com um detector bidimensional do tipo CCD. Foram utilizadas, como amostras padrão, BSA e lisozima (Tabela IV-6), nos mesmos tampões das proteínas e em diferentes concentrações. A amostra analisada continha a XAC0419, (7,9 mg mL⁻¹), no tampão tris-HCl 5 mmol L⁻¹, pH 5.

Tabela IV-6. Concentração dos padrões utilizados nas medidas de SAXS

Padrão	Massa Molar	Concentração
BSA	66 kDa	10,43 mg mL ⁻¹
		3,0 mg mL ⁻¹
lisozima	14,4 kDa	10,21 mg mL ⁻¹
		3,2 mg mL ⁻¹

Os dados obtidos foram processados com a ajuda do professor colaborador Ricardo Aparicio (IQ-Unicamp) utilizando o programa fit2D,¹²³ com correção da atenuação da amostra e normalização pela intensidade transmitida do feixe, além de subtração do tampão, para construir uma curva média de espalhamento usando o PRIMUS^{124b} (Figura IV-32). Em seguida, construiu-se a curva de Guinier, ou região linear da curva de espalhamento (Figura IV-33).

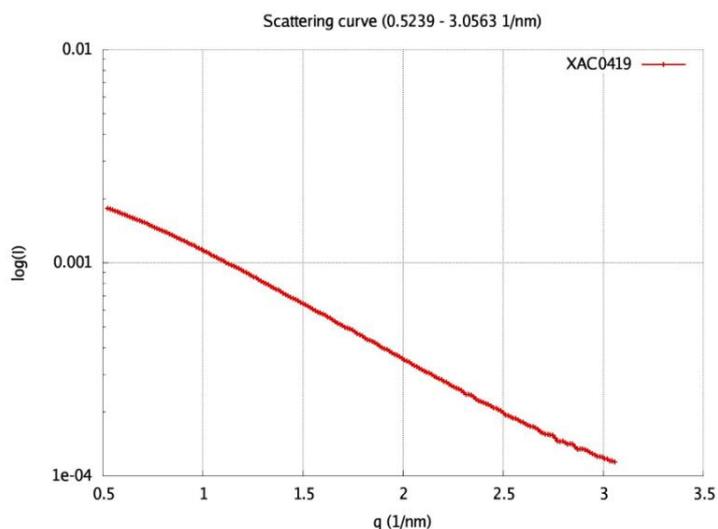


Figura IV-32. Resultados de SAXS para a XAC0419. Curva de espalhamento de raios-X da proteína, construídas utilizando o programa PRIMUS.

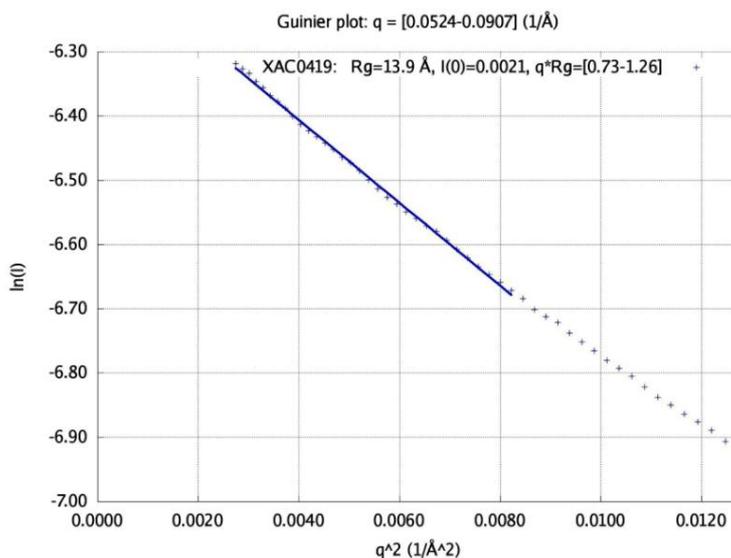


Figura IV-33. Resultados de SAXS para a XAC0419. Região de Guinier da curva de espalhamento de raios-X da proteína.

As análises subsequentes, de determinação de raio de giro e distribuição de distâncias (Figura IV-34) foram feitas com os programas GNOM/AUTOGNOM, DAMMIN, DAMAVER e outros do pacote ATSAS,

já descritos no Capítulo 3. A partir dos dados experimentais também foi construída a curva de Kratky (*Kratky plot*, sendo $I(q) \cdot q^2$ em função de q – Figura IV-35). Sendo que a curva de Kratky exibiu um máximo bem definido, característico de proteínas compactas com estrutura terciária estável, corroborando com a análise anterior de CD que confirmou a presença de elementos de estrutura secundária.

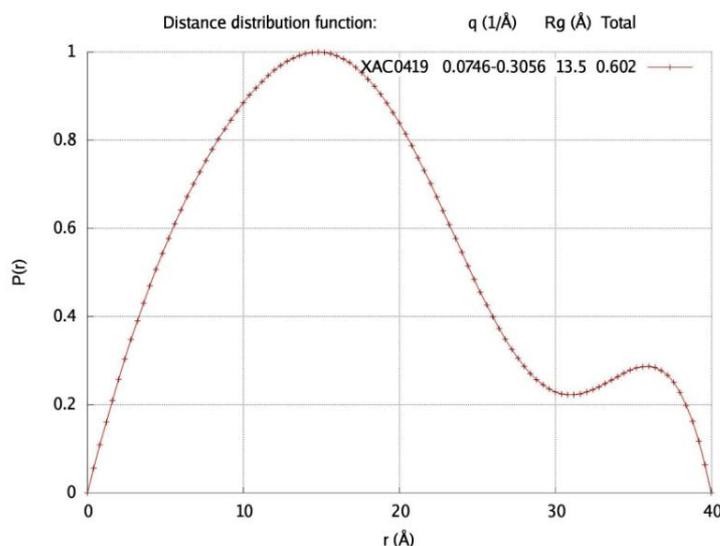


Figura IV-34. Curva de distribuição de distâncias da XAC0419, $P(r)$, obtida com GNOM/AUTOGNOM.

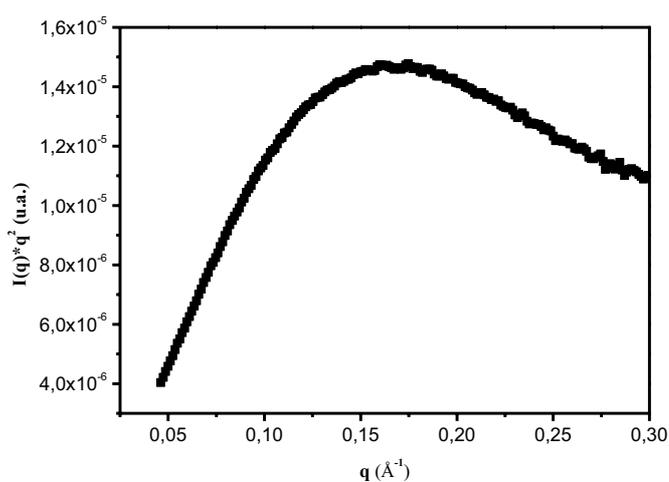


Figura IV-35. Curva de Kratky da XAC0419, obtida pelo tratamento dos dados de SAXS.

A partir destes processamentos obteve-se um raio de giro para a XAC0419 de 13,9 Å, que é condizente, como será visto a seguir, com o tamanho da proteína e concorda com resultados anteriores que indicaram a presença da proteína na forma monomérica em solução.

A partir da distribuição das distâncias na proteína obteve-se um raio de giro para a XAC0419 de ~13,5 Å, concordante com o obtido no processamento anterior. A massa molecular da proteína foi estimada de duas maneiras, utilizando um padrão para calibração¹²⁶ (no caso a BSA 10,43 mg/mL) e um método, relativamente novo, baseado em uma única curva medida em escala relativa.¹²⁷ Para a XAC0419 obteve-se uma massa molar de 8,6 kDa (21% de discrepância), relativa à proteína monomérica em solução, nas condições analisadas.

Todos os passos do processamento foram feitos utilizando Linux, o programa GNUPLOT¹⁴⁷ foi empregado para análise dos dados e o *plot* e os programas RASMOL¹³⁵ e PyMOL¹⁴⁸ para visualização dos modelos.

A partir destes dados de SAXS foi construído um modelo do envelope molecular desta proteína. O envelope obtido a partir das curvas experimentais é alongado conforme esperado. E, devido à baixa resolução intrínseca da técnica de SAXS, modelos de coordenadas atômicas são normalmente utilizados para interpretar estes envelopes^{124a,125b,149}. Como modelos experimentais de alta resolução não são disponíveis para as proteínas alvo, empregou-se a predição de estrutura terciária pelo servidor I-TASSER e foram gerados os modelos de coordenadas (Figura IV-36).¹³⁰

¹⁴⁷ Página do programa <http://www.gnuplot.info/>

¹⁴⁸ Página do programa <http://www.pymol.org/>

¹⁴⁹ Lipfert, J.; Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 307-327.

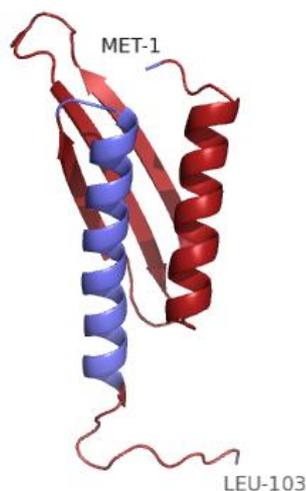


Figura IV-36. Estrutura calculada por bioinformática para a XAC0419 (C -score = -0,24), com N- e C-terminais marcados e a hélice anfipática em azul.

Os melhores modelos apresentaram exatidão estimada de (TM-score) $0,75 \pm 0,10$ para a XAC0419 sugerindo uma topologia correta (Figura IV-36).

Foram calculados outros quatro modelos, que compartilham as mesmas características estruturais gerais, diferindo principalmente nas regiões desordenadas. Os melhores resultados para alinhamentos de estrutura da XAC0419 pertencem à família d.204.1.1 (proteína Y de ligação de ribossomo) do SCOP,¹³³ cujo enovelamento consiste principalmente de folhas- β antiparalelas, com regiões α e β segregadas. Buscas no servidor DALI¹⁵⁰ levaram a resultados similares ao do SCOP. É importante mencionar que a melhor sobreposição do envelope experimental com os modelos do I-TASSER não foi com o modelo de maior pontuação, mas com o segundo modelo calculado para a XAC0419. Observou-se uma similaridade impressionante entre o envelope experimental e o modelo calculado, exceto para as prováveis regiões desordenadas e de maior mobilidade. Foi observada uma excelente

¹⁵⁰ Holm, L.; Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **2010**, *38*, W545-W549.

concordância na região central que compreende as hélices- α e folhas- β calculadas. Apesar das regiões desordenadas provavelmente assumirem uma forma mais globular e compacta em solução do que o esperado pelo resultado do I-TASSER, conforme pode ser observado pelo modelo de SAXS obtido (Figura IV-37).¹⁵¹

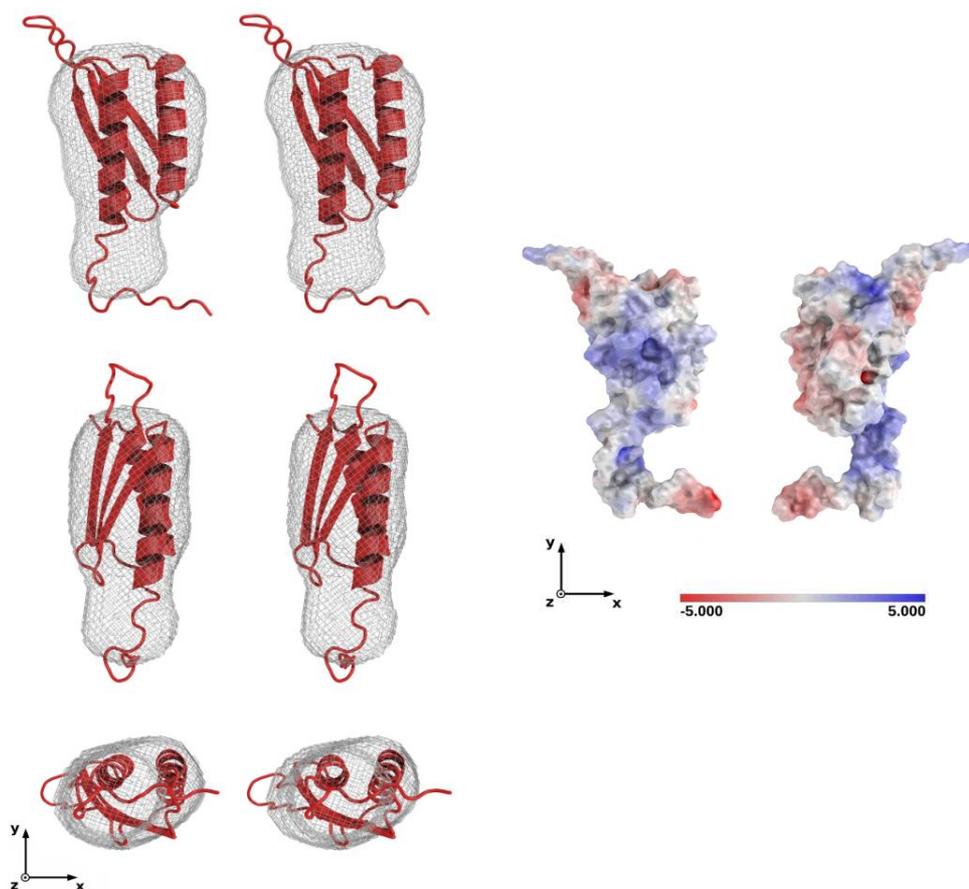


Figura IV-37. Resultados de SAXS e bioinformática. À direita três vistas diferentes do modelo gerado por SAXS (em cinza) sobreposto com a estrutura calculada pelo I-TASSER (em vermelho). À esquerda a mesma estrutura calculada referente à vista superior da direita, com o potencial de carga superficial calculado sendo, em azul potencial positivo, em vermelho negativo e em cinza os resíduos hidrofóbicos.

¹⁵¹ Fattori, J.; Prando, A.; Assis, L.H.P.; Aparicio, R.; Tasic, L. Structural insights on two hypothetical secretion chaperones from *Xanthomonas axonopodis* pv. *citri*. *Protein J.* **2011**, *30*, 324-333.

De acordo com o modelo obtido pelo I-TASSER a proteína apresenta 40% de estrutura helicoidal e 20% de folhas- β . A partir do programa k2D2, novamente foi gerado um espectro de dicroísmo circular correspondente a uma proteína com 40% de α -hélice e feita uma sobreposição com o espectro obtido experimentalmente (Figura IV-38).

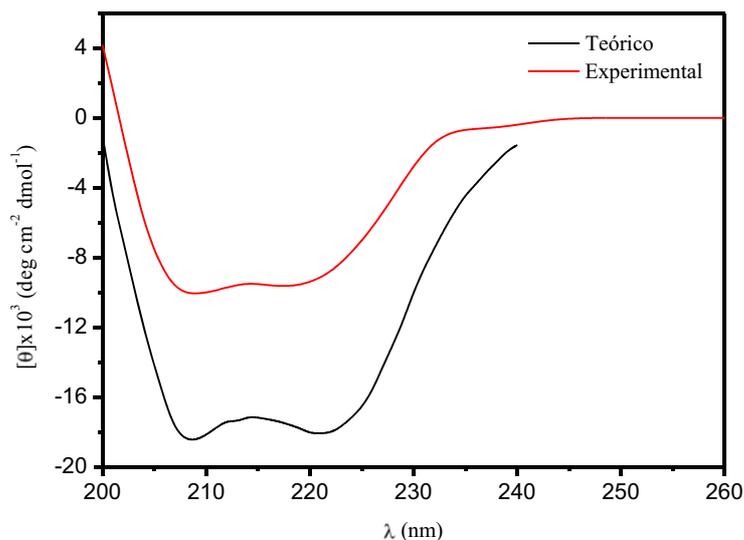


Figura IV-38. Comparação das curvas de CD, em vermelho a curva experimental e em preto a curva teórica, gerada pelo K2D2 para a estrutura obtida pelo I-TASSER.

Comparando os resultados obtidos, a distância máxima observada no modelo calculado é de cerca de 60 Å, que corresponderia a uma esfera de 30 Å. Valor este, um pouco discrepante do obtido como raio de Stokes por filtração em gel analítica (17 Å). Entretanto, este último é condizente com a distância máxima observada no envelope predito por SAXS, de 40 Å (Figura IV-34), respeitando a seguinte ordem $R_H > R_0$, indicando uma proteína de formato ligeiramente alongado.⁸⁴

A XAC0419, assim como a XAC1346, é considerada uma possível chaperona de secreção do T3SS, pois o gene que a codifica encontra-se no cluster do sistema de secreção do tipo III, próximo aos genes de

patogenicidade e de resposta de hipersensibilidade no cromossomo da *Xac* e, também, próximo ao gene que codifica a XAC1990, também alvo deste estudo.⁶³ É uma proteína pequena, com massa molar teórica de 10,7 kDa, pI de 7 e os resíduos de 67 a 91 estão envolvidos no domínio anfipático, mostrados em azul na Figura IV-35. Nos dados de CD, acima discutidos observou-se que ela apresenta 40% de estrutura helicoidal,⁶⁹ resultado bastante concordante com os dados teóricos (34%).

Pelos ensaios de filtração em gel analítica, concluiu-se que a proteína apresenta-se monomérica em soluções com cerca de 150 mmol L⁻¹ de cloreto de sódio, pela estimativa da massa molar em 12,8 kDa.

Uma vez que há vários exemplos na literatura onde os dados de SAXS têm sido usados como uma ferramenta importante para comparar e selecionar modelos de proteínas, justifica-se aplicá-los neste caso, porém, é a primeira vez que se obtêm modelos estruturais de baixa resolução para chaperonas de secreção da *Xac*.

Conforme já descrito, uma característica comum das chaperonas de secreção é a presença de domínios anfipáticos compostos de hélices- α onde aminoácidos polares e não polares estão localizados em lados opostos. Além disso, os potenciais eletrostáticos das superfícies calculados para os modelos tridimensionais preditos foram examinados e mostraram predominantemente áreas hidrofóbicas intercaladas com regiões carregadas positiva e negativamente, sugerindo trechos contínuos para interações proteína-proteína (Figura IV-37). Apesar de não existirem muitos dados estruturais de chaperonas de secreção na literatura, a maioria deles apresenta-as como diméricas. Surpreendentemente, nossos resultados indicam que a XAC0419 estava monomérica em solução, nas condições analisadas, o que pede estudos mais aprofundados sobre esta proteína.

Testes de atividade com estas proteínas seriam bem interessantes no auxílio da determinação da função das mesmas, porém, como não fazem parte do escopo do trabalho proposto não foram realizados.

Utilizando o QUARK para o cálculo da estrutura tridimensional, foi possível obter mais um modelo para a XAC0419 que é bastante semelhante ao calculado pelo I-TASSER, diferindo principalmente nas regiões randômicas (Figura IV-39). Porém com o mesmo conteúdo de estrutura helicoidal que o modelo do I-TASSER (40%).

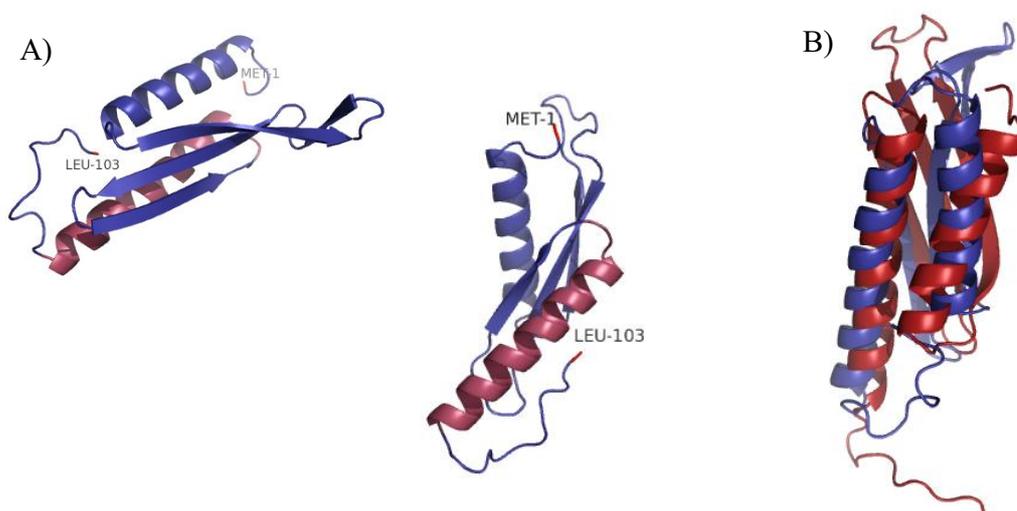


Figura IV-39. Resultados de bioinformática. A) Duas vistas diferentes da estrutura 3D da XAC0419 calculada pelo QUARK, ressaltando o N- e o C-terminal, mais a região anfipática em rosa no C-terminal. B) Sobreposição dos modelos tridimensionais da XAC0419, em vermelho a estrutura calculada pelo I-TASSER e em azul a estrutura calculada pelo QUARK.

Conforme se observa na Figura IV-39B, a estrutura calculada pelo QUARK é mais compacta que a calculada pelo I-TASSER, principalmente nas regiões desordenadas.

Fazendo um alinhamento da estrutura da XAC0419 com estruturas determinadas experimentalmente, de chaperonas de secreção das classes I e II

observou-se maior similaridade estrutural desta proteína da *Xac* com a SycE de *Yersinia pseudotuberculosis*⁵⁶ (Figura IV-40).

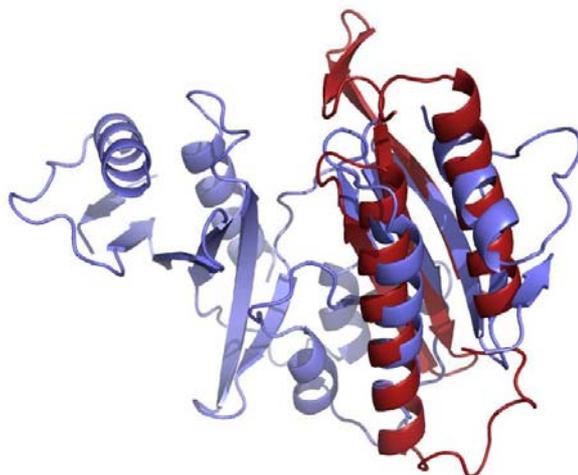


Figura IV-40. Alinhamento das estruturas: da XAC0419 calculada pelo QUARK, em vermelho, e SycE de *Yersinia pseudotuberculosis* dimérica em azul (PDB:1JYA), sendo alinhados 67 aminoácidos dos 103 da XAC0419 (*TM-Score* = 0,46).

Assim como para a XAC1346, devido à ausência de dados estruturais não se pode classificar com confiança esta proteína da *Xac*, entretanto, a partir dos resultados obtidos (Figura IV-40) pode se especular a respeito da XAC0419 como possível chaperona de secreção da classe I, sendo que as chaperonas desta classe caracterizadas estruturalmente até o momento, apresentam um enovelamento do tipo $\alpha+\beta$ com regiões de hélices- α e folhas- β segregadas ($\alpha-\beta\beta\beta-\alpha-\beta\beta-\alpha$) e se encontram diméricas em solução. Enquanto a XAC0419 apresenta um enovelamento semelhante ($\alpha-\beta\beta\beta-\alpha$), porém pelos resultados obtidos se encontra monomérica em solução, o que instiga estudos mais aprofundados, principalmente em se considerando que já foi confirmado que esta proteína é expressa pela *Xac*.

4.3. XACb0033

Esta proteína, diferentemente das anteriores, foi encontrada no precipitado da lise (Figura IV-41), portanto insolúvel.

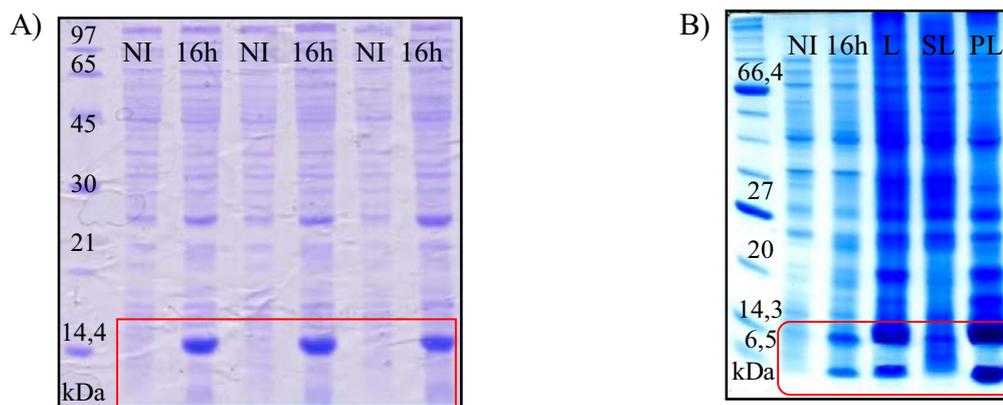


Figura IV-41. Géis de eletroforese SDS-PAGE (15%). A) Expressão da XACb0033, tendo da esquerda para a direita: marcador de massa molar de proteína (kit LMW/GE Healthcare); NI corresponde à amostra não induzida e 16h a 16 horas depois da adição de IPTG. B) Lise da XACb0033, tendo da esquerda para a direita: marcador de massa molar de proteína (P7702S/NEB); L corresponde ao lisado total, SL ao sobrenadante da lise e PL precipitado da lise.

O precipitado contendo a proteína foi então, lavado novamente com tampão de lise (10 mL) e com água destilada, sempre agitando suavemente por cerca de 5 minutos e centrifugando a suspensão em seguida (Figuras IV-41 e 42). Após a lavagem o precipitado foi tratado com uréia¹¹³ (Figura IV-42).

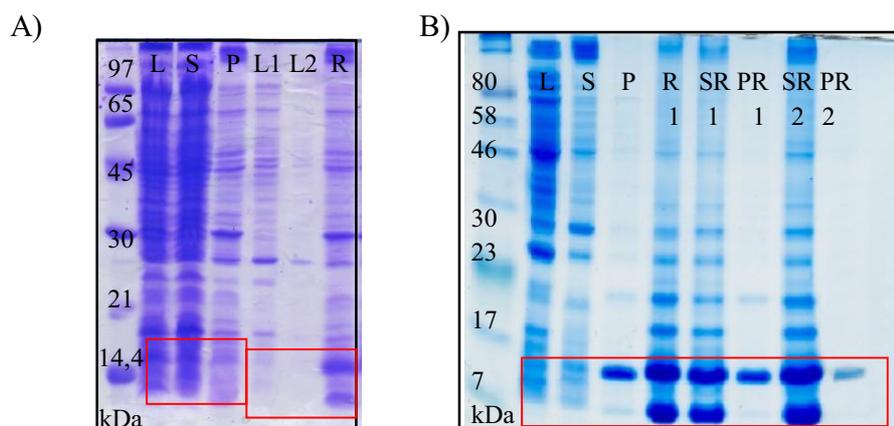


Figura IV-42. Géis de eletroforese SDS-PAGE (15%) da lavagem do precipitado (A) e do tratamento com uréia (B), tendo: marcador de massa molar de proteína (kit LMW/GE Healthcare e P7703/NEB); amostras do lisado (L); sobrenadante da lise (S); precipitado da lise (P); água das lavagens do precipitado (L1 e L2); reenovelamento (R); sobrenadante do primeiro reenovelamento (SR1); precipitado do primeiro reenovelamento (PR1); sobrenadante do segundo reenovelamento (SR2) e precipitado do segundo reenovelamento (PR2) da XACb0033.

A uréia foi utilizada para desenovelar a proteína contida no precipitado, tornando-a solúvel e em seguida foi feita lenta diminuição da concentração do desnaturante através de diálises para permitir o reenovelamento da proteína no sobrenadante (Figura IV-43).¹¹³ Este tratamento também foi acompanhado por eletroforese em gel, conforme pode ser observado na Figura IV-42, sendo que uma parte da proteína quase pura ainda ficou no precipitado.

De acordo com o que se observa na Figura IV-42, as etapas de lavagem e reenovelamento podem ser contadas como etapas de purificação, uma vez que alguns contaminantes se solubilizaram durante as lavagens e depois na adição de uréia, outros contaminantes foram precipitados.

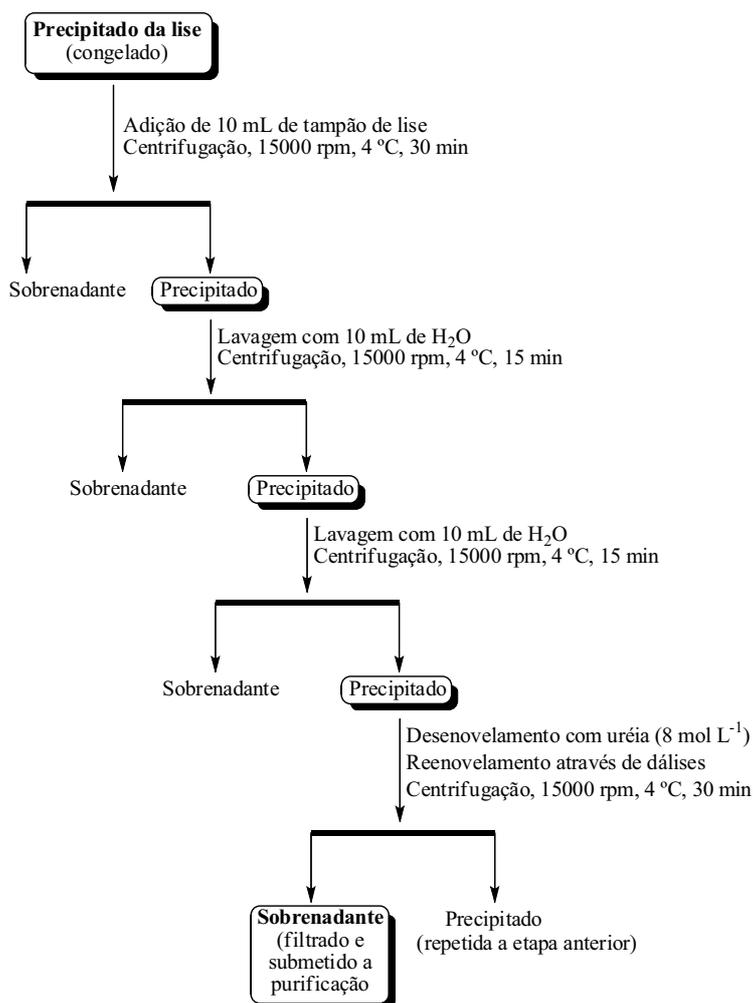


Figura IV-43. Fluxograma da obtenção da XACb0033 solúvel para posterior purificação.

Tendo em mente que as lavagens podem ser consideradas etapas de purificação, também foram feitas lavagens do precipitado do primeiro reenovelamento, antes de repetir o tratamento com uréia e obteve-se a proteína praticamente pura (Figura IV-44).

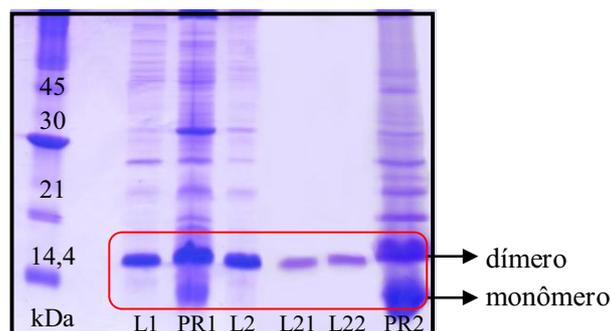


Figura IV-44. Gel de eletroforese SDS-PAGE (15%) da lavagem dos precipitados do reenovelamento da XACb0033, tendo na sequência: marcador de massa molar de proteína (kit LMW/GE Healthcare); água das lavagens do precipitado do primeiro reenovelamento (L1 e L2); precipitado do primeiro reenovelamento (PR1); água das lavagens do precipitado do segundo reenovelamento (L21 e L22) e precipitado do segundo reenovelamento (PR2).

Observa-se na Figura IV-44 que a proteína se encontra quase pura e dimérica na água de lavagem, enquanto no precipitado ela se encontra tanto na sua forma dimérica quanto monomérica. A presença da forma dimérica nos géis de eletroforese SDS-PAGE, que apresentam condições desnaturantes, pode ser explicada pelo fato de que mesmo após a redução das ligações dissulfeto pela adição de DTT, em contato com o ar as ligações dissulfeto voltam a se formar levando a dimerização da proteína. Esta questão poderia ser resolvida com a utilização de gel nativo.

As amostras de sobrenadantes do reenovelamento foram dialisadas em tampão fosfato de sódio (50 mmol L^{-1} , pH 8), contendo cloreto de sódio (150 mmol L^{-1}) e em seguida purificadas por filtração em gel (Figura IV-45).

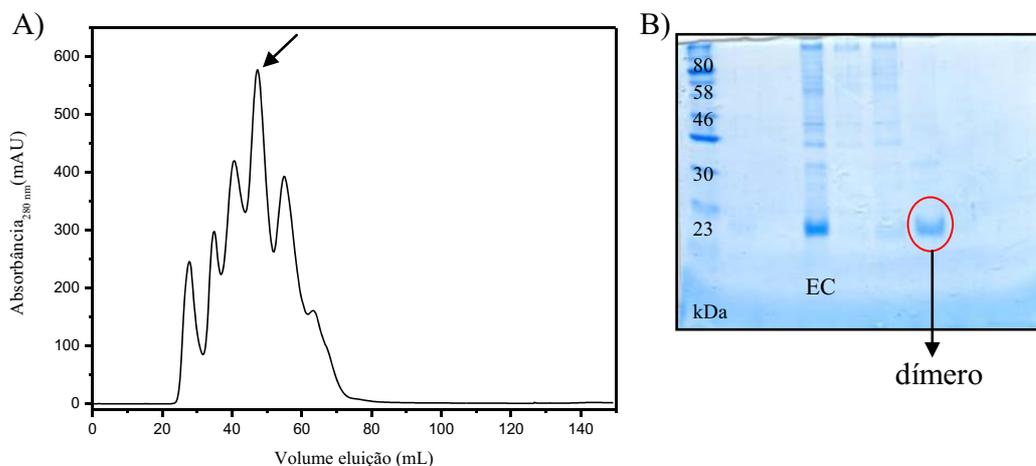


Figura IV-45. Cromatograma da purificação do sobrenadante do reenovelamento da XACb0033 (A), sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) da purificação da XACb0033, tendo na sequência: o marcador de massa molar (P7703/NEB), EC correspondendo à amostra que foi aplicada na coluna seguida de diferentes frações da purificação.

De posse da proteína pura (Figura IV-46) foram realizadas etapas de concentração e quantificação para os ensaios de caracterização estrutural.

A determinação da concentração da XACb0033 nas soluções foi feita seguindo o método de Edelhoch,¹¹⁴ medindo a absorbância em 280 nm. A quantificação através deste método foi possível, pois a proteína apresenta resíduos de triptofano em sua estrutura primária. O coeficiente de extinção molar, ou de absorvidade molar (ϵ) da mesma foi obtido por bioinformática, no ExPASy¹¹⁶ ou pelo programa sednterp®,¹¹⁵ calculados a partir do número de resíduos de triptofano, tirosina e cistinas na proteína.¹¹⁷

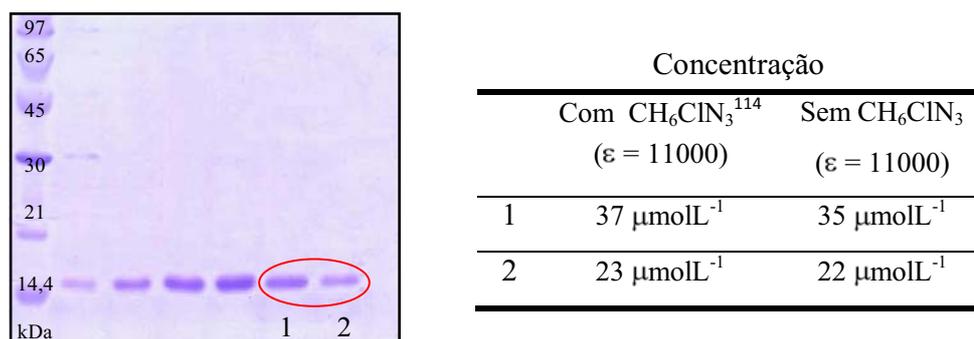


Figura IV-46. Gel de eletroforese SDS-PAGE (15%) de alíquotas da XACb0033 em solução, tendo na sequência: marcador de massa molar (kit LMW/GE Healthcare) e as diferentes frações de uma purificação, destacando as bandas cortadas para sequenciamento por Espectrometria de Massas. Ao lado tabela de medidas das concentrações das soluções marcadas.

Uma vez que a calibração da filtração em gel na coluna Superdex 200 (Figura IV-8) já havia sido realizada, as amostras do sobrenadante do reenovelamento da XACb0033, em tampão fosfato de sódio (50 mmol L^{-1} , pH 7,2), contendo cloreto de sódio (150 mmol L^{-1}), quase puras foram submetidas à purificação por filtração em gel, nas mesmas condições da calibração (Figura IV-47).

O volume de eluição da XACb0033 foi 47 mL, correspondendo a um coeficiente de partição (k_{av}) de 0,68. Plotando este valor na curva de calibração (Figura IV-8), obtemos uma massa molar de 7843 Da; sendo a massa teórica 8,4 kDa, concluiu-se portanto que a proteína se encontra em solução, nestas condições em forma monomérica.

Utilizando o fator de retenção no gel de eletroforese para estimar a massa molar da proteína, obteve-se a massa molar de 7982 Da. Assim, concluiu-se que ambas estimativas indicam presença da proteína na forma monomérica em solução

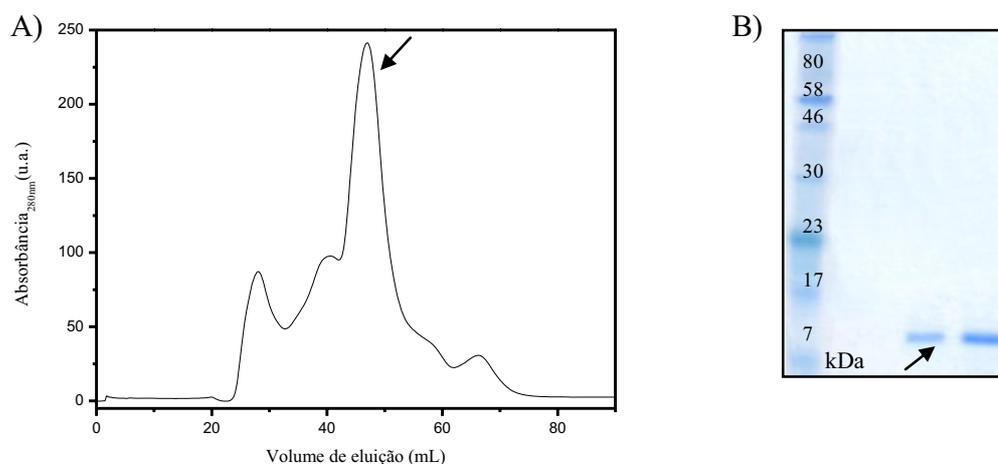


Figura IV-47. Filtração em gel da XACb0033. A) Cromatograma, sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) referente ao cromatograma tendo na sequência: marcador de massa molar (P7709S/NEB) e a amostra referente ao pico indicado no cromatograma.

De maneira similar, utilizando a curva de calibração da Figura IV-27B, realizando um novo experimento de filtração em gel analítica (Figura IV-48) obteve-se uma estimativa do raio de Stokes da proteína, em torno de 14 \AA ($13,9 \pm 0,6 \text{ \AA}$), portanto, apresentando um tamanho intermediário entre a aprotinina e a ribonuclease A. Estimando a massa molar por esta curva, obteve-se o valor de 9353 kDa, indicando que a proteína se encontra monomérica em solução.

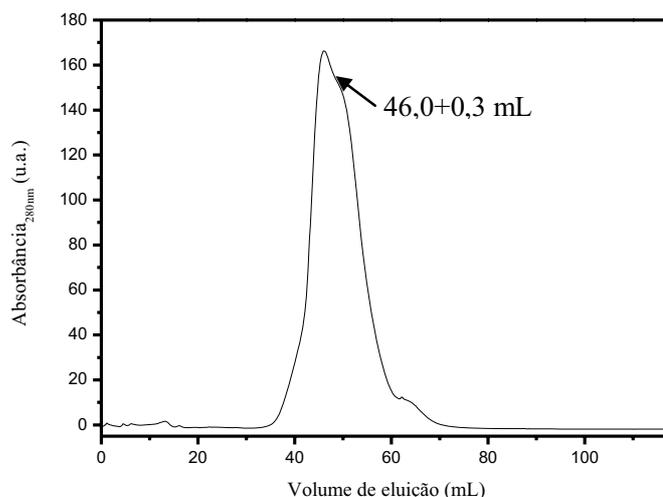


Figura IV-48. Filtração em gel analítica da XACb0033. Cromatograma, sendo o pico correspondente à proteína e seu volume de eluição indicados na figura.

Uma amostra da XACb0033 pura (Figura IV-45), em tampão fosfato de sódio 10 mmol L^{-1} , foi submetida à análise por espectrometria de massas para confirmação da massa molar exata (Figura IV-49) de onde obteve-se o valor de 8283,5 Da e observou-se também um pouco de formação do dímero com 16,6 kDa.

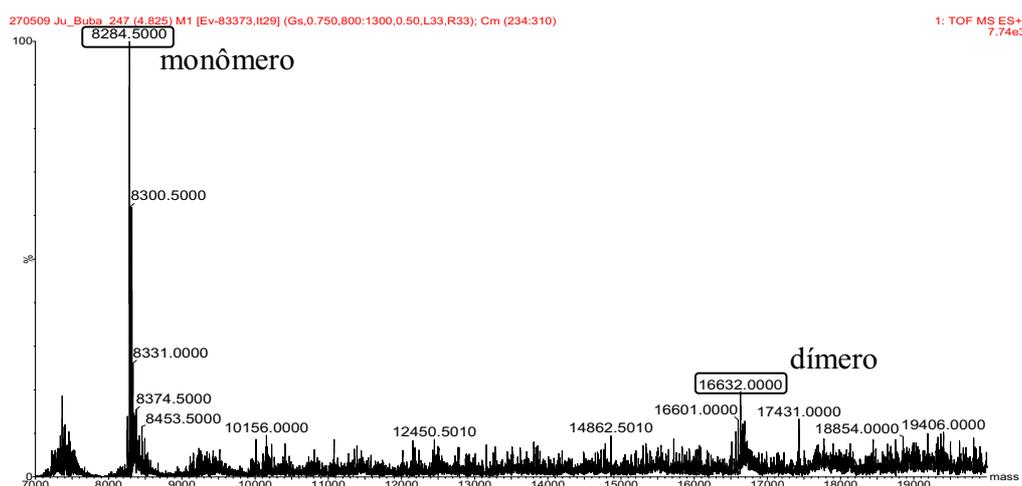


Figura IV-49. Espectro de massas da XACb0033 intacta, com massa molecular do monômero e dímero indicadas.

Comparando a massa obtida e a massa molar teórica (8418,5 kDa, Tabela IV-1) calculada para a proteína de acordo com a sequência de aminoácidos, observou-se a perda da primeira metionina (diferença de 135 Da).

Também por espectrometria de massas foram feitos experimentos para identificação da proteína, através da digestão com tripsina. Para tanto foram cortadas as bandas de proteína do gel de eletroforese (Figura IV-46), tratou-se o mesmo para retirada do corante, em seguida foi efetuado o processo de redução e alquilação para posterior digestão com tripsina. Após a digestão, a mistura de peptídeos foi passada por uma coluna C18 (ZipTip) e depois adicionada à matriz utilizando o método de evaporação da gota e foram então levadas ao espectrômetro de massas para análise por MALDI-ToF. Para a identificação das proteínas, os espectros de MS/MS dos fragmentos da digestão foram inseridos no programa de busca Mascot¹¹⁹ e a XACb0033 foi identificada (Figura IV-50).

Match to: **xac:XACb0033 hypothetical protein**
Nominal mass (M_r): **8527**; Calculated pI value: **5.61**
1 AHVNSRVQK HRDALRMAGL RPVQIWVPDT RRPDFAEECR
RQCRLAAQAD
51 MADTDMQRFM DEALADMDGW TE

Figura IV-50. Identificação da XACb0033 por MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 91% de cobertura da sequência.

Identificada a proteína, foram realizados experimentos de dicroísmo circular para avaliar se a mesma encontrava-se enovelada e obter informação sobre sua estrutura secundária. Estes experimentos indicaram que a proteína

estava enovelada (Figura IV-51) e que apresenta mais de 80% de hélice- α ^{120,152} em sua estrutura secundária (Tabela IV-7).

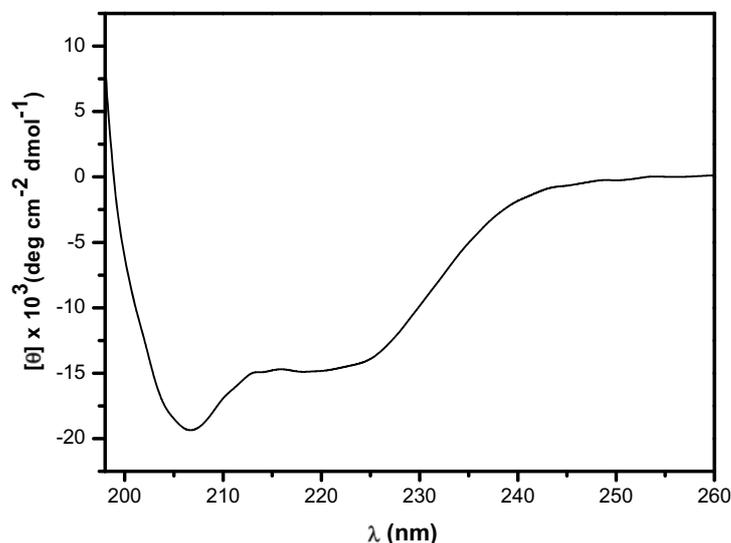


Figura IV-51. Dicroísmo Circular da XACb0033. Curva de CD experimental com a proteína 9,1 $\mu\text{mol L}^{-1}$ em tampão fosfato de sódio 10 mM pH 7,4.

Tabela IV-7. Dados de dicroísmo circular para a XACb0033

Programas utilizados	Porcentagem de hélice- α predita	Porcentagem de hélice- α calculada
GOR ¹⁴⁰	57%	-
JUFO ¹⁵³	57%	-
Jpred ¹⁴⁶	57%	-
CDNN ¹⁵²	-	23%
K2D2 ¹²¹	-	99%
	-	$[\theta]_{222\text{nm}} = 82\%$ $[\theta]_{208\text{nm}} = 99\%$

¹⁵² CDNN 2.1: Circular Dichroism Deconvolution, Neural Networks 1999.

¹⁵³ a) Meiler, J.; Mueller, M.; Zeidler, A.; Schmaeschke, F. JUFO: Secondary Structure Prediction for Proteins. *www.jens-meiler.de*, 2002. b) Meiler, J.; Baker, D. Coupled prediction of protein secondary and tertiary structure. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12105–12110.

Os dados experimentais foram utilizados em programas para calcular a quantidade de estrutura secundária, conforme indicado na Tabela IV-7. E ainda, foram utilizados programas de predição de estrutura secundária a partir da sequência de aminoácidos (Tabela IV-7) e os resultados foram comparados. Segundo os resultados obtidos por três diferentes programas, a XACb0033 apresenta 57% de estrutura helicoidal. Entretanto, os valores experimentais são discrepantes. Dois diferentes tipos de cálculo indicaram a presença de 99% de estrutura helicoidal, enquanto um terceiro tipo indicou a presença de 23% deste tipo de estrutura. Este último dado está de acordo com dados já publicados¹⁰⁸ desta proteína que indicam a presença de 25% de estrutura helicoidal.

Como a proteína apresenta dois resíduos de triptofano em sua estrutura primária, foram realizados experimentos de fluorescência intrínseca, onde se observou um máximo de emissão em 349 nm (Figura IV-52), indicando resíduo de triptofano exposto ao solvente, o que pode ser explicado pelo segundo Trp(70) encontrado no C-terminal da proteína, região caracteristicamente desordenada.

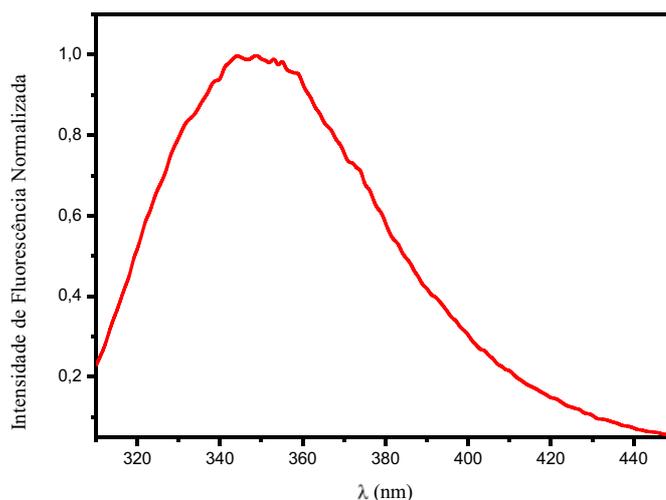


Figura IV-52. Fluorescência de emissão da XACb0033. Espectro de emissão (máximo 349 nm), da proteína $4 \mu\text{mol L}^{-1}$ em tampão fosfato de sódio 10 mmol L^{-1} , pH 7,4.

Foram realizadas medidas de DLS para obter uma estimativa do tamanho desta proteína, através do raio de Stokes. A partir das medidas de DLS, obteve-se um raio de 18 Å, com uma polidispersidade de 0,33 (Figura IV-53), indicativo de amostra monodispersa ($13 \mu\text{mol L}^{-1}$).

Comparando estes resultados com os da filtração em gel analítica, foi obtido raio de Stokes da mesma ordem (filtração em gel analítica, 14 Å; DLS, 18 Å). Provavelmente indicando uma amostra monomérica na filtração em gel analítica e dimérica na análise por DLS ($R_{0 \text{ mon.}} = 13 \text{ \AA}$ e $R_{0 \text{ dim.}} = 17 \text{ \AA}$).

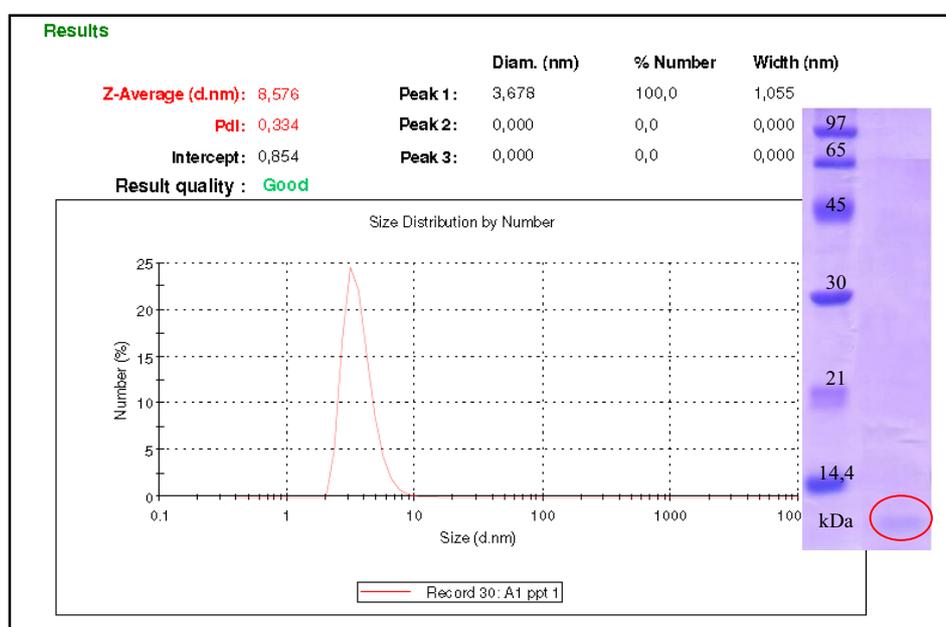


Figura IV-53. Resultados de DLS para a XACb0033. Ao lado gel de eletroforese da amostra submetida ao ensaio, tendo o marcador de massa molar (kit LMW/GE Heathcare) e a amostra da proteína.

Na sequência foram realizados experimentos de NMR para obtenção do coeficiente de difusão da molécula. Os espectros de ^1H mostraram a proteína enovelada nas condições de análise, com um coeficiente de difusão segundo os experimentos de NMR-DOSY de $0,9 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ (Figura IV-54).

No espectro da XACb0033 observa-se que o sinal da água em 4,8 ppm não foi bem suprimido, mas foi o máximo de supressão alcançado. Também se observou um sinal bem fino em torno de 1,8 ppm, provavelmente correspondente à alguma impureza do tampão fosfato de sódio. Todos os espectros de DOSY obtidos foram processados no software VNMRJ da Varian, referenciados pelo sinal da água em 4,8 ppm.

Entretanto, como o Instituto não possui o pacote de programas com a sequência de pulsos para experimentos de difusão (DOSY) com supressão do sinal do solvente, e as amostras protéicas estavam bem diluídas, os resultados obtidos por estes experimentos não foram muito satisfatórios.

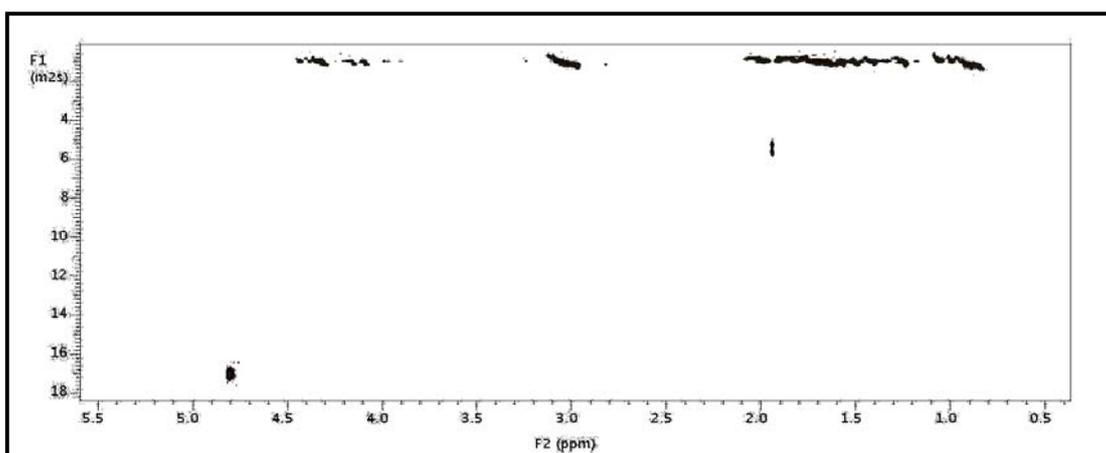


Figura IV-54. Espectro de NMR-DOSY da XACb0033 $0,12 \text{ mmol L}^{-1}$ em 10% de D_2O obtido no equipamento Varian INOVA 500 MHz, utilizando a sequência de pulso Dbppste, onde em F1 tem-se o coeficiente de difusão (D) multiplicado por 10^{-10} e a projeção do espectro de ^1H em F2.

Apesar de o experimento de NMR-DOSY ser utilizado para estudos de proteínas, existiam dúvidas a respeito da utilização do mesmo para estimar o tamanho e estado oligomérico das proteínas, dado que como a massa molar e, conseqüentemente, o tamanho das proteínas nas formas monomérica e

dimérica são relativamente da mesma ordem de grandeza, a investigação do estado oligomérico delas seria dificultado.

Para analisar este fato, foram realizados ensaios de NMR-DOSY com as mesmas proteínas utilizadas na calibração da filtração em gel analítica. Tais proteínas possuem seus raios de Stokes conhecidos e representam extremos de massa molar que englobam a massa molar das proteínas em estudo. As proteínas utilizadas foram: BSA, lisozima, ovalbumina, α -quimiotripsinogênio, ribonuclease A e aprotinina. Os espectros de ^1H destas proteínas estão representados na figuras que seguem (Figuras de IV-55).

Como os espectros obtidos para todas as proteínas padrão foram obtidos para as proteínas em D_2O 10%, para obtenção de um espectro com melhor resolução, repetiu-se o experimento para a BSA e a lisozima também em D_2O 99% (Figura IV-56 e 57).

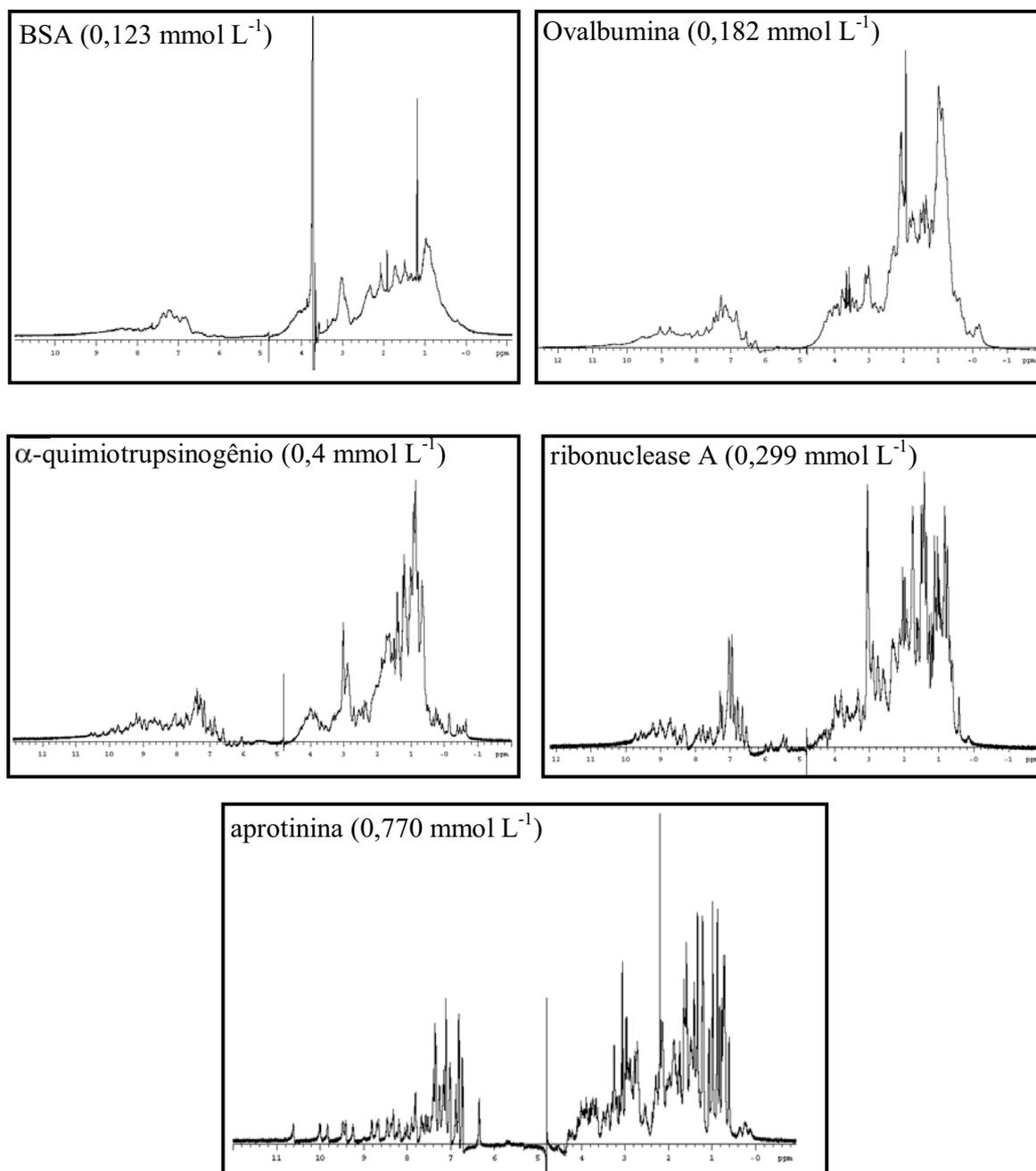


Figura IV-55. Espectros de NMR de ^1H das proteínas padrão (indicadas no espectro com as respectivas concentrações) em tampão fosfato de sódio 50 mmol L^{-1} (D_2O 10%) obtido no equipamento Varian INOVA 500 MHz (499,98 MHz), utilizando a sequência de pulso WATERGATE para supressão do sinal de água em 4,8 ppm.

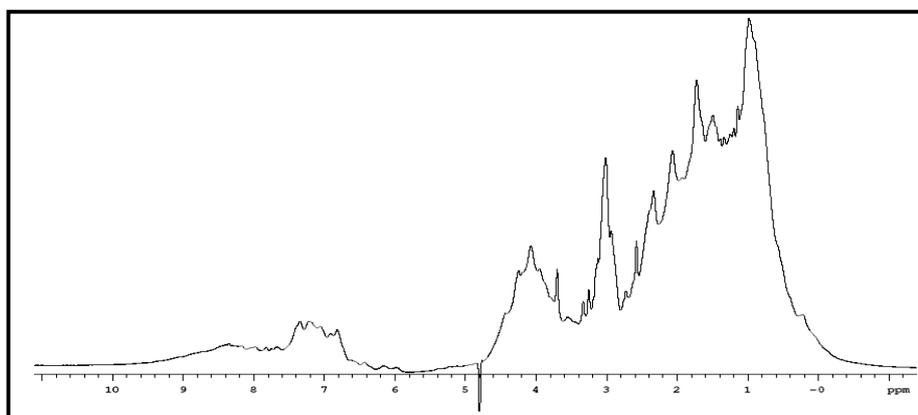


Figura IV-56. Espectro de NMR de ^1H da BSA $0,9 \text{ mmol L}^{-1}$ em D_2O obtido no equipamento de 500 MHz, utilizando a sequência WATERGATE para supressão do sinal de água em 4,8 ppm.

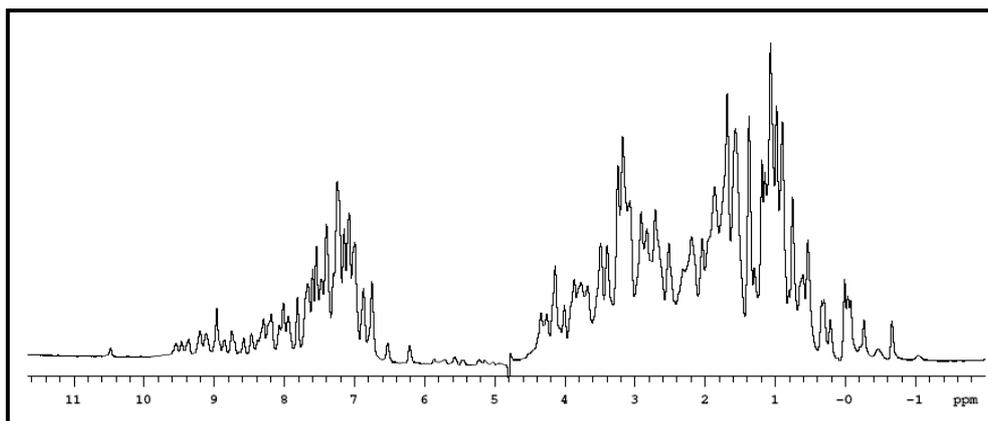


Figura IV-57. Espectro de NMR de ^1H da lisozima $4,1 \text{ mmol L}^{-1}$ em D_2O obtido no equipamento de 500 MHz, utilizando a sequência WATERGATE para supressão do sinal de água em 4,8 ppm.

Dado que o melhor resultado obtido nos experimentos de DOSY foi para a lisozima, devido a sua alta concentração, foi construído um gráfico (Figura IV-58) com os dados obtidos, aplicando a escala logarítmica para o coeficiente de difusão, uma vez que os cálculos para estrutura da molécula

envolvem uma relação logarítmica entre o coeficiente de difusão e a massa molar.¹⁵⁴

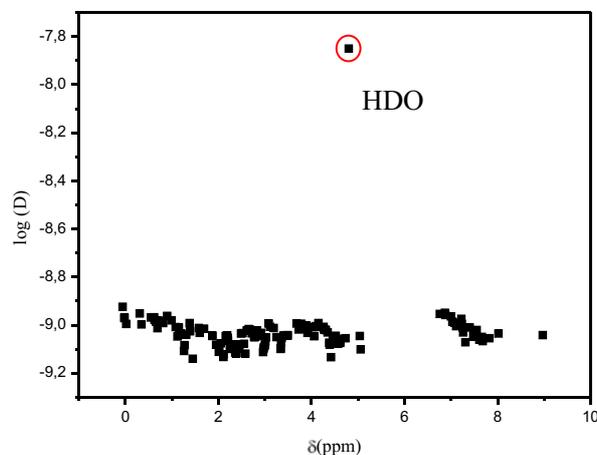


Figura IV-58. Curva com os dados de NMR tratados da lisozima 4 mmol L^{-1} em D_2O (99%) obtido no equipamento de 500 MHz, utilizando a sequência de pulso Dbppste. Sinal em -7,8 (log D) referente ao sinal da água residual (HDO).

Utilizando os dados obtidos para as proteínas padrão em D_2O 10%, foram contruídas curvas relacionando o coeficiente de difusão (log D) determinado por NMR-DOSY e os dados teóricos de massa molar e raio de Stokes das proteínas (Tabela IV-8 e Figura IV-59).

Tabela IV-8. Dados dos experimentos de NMR-DOSY

Proteína (kDa)	R_s (Å)	Concentração	$D * 10^{-10}$ (m^2/s)
Aprotinina (6,5)	9,83	$770 \mu\text{mol L}^{-1}$	$0,93 \pm 0,01$
Ribonuclease A (13,7)	16,4	$299 \mu\text{mol L}^{-1}$	$0,61 \pm 0,02$
α -quimiotripsinogênio (25)	20,9	$400 \mu\text{mol L}^{-1}$	$0,55 \pm 0,01$
Ovalbumina (44)	30,5	$182 \mu\text{mol L}^{-1}$	$0,384 \pm 0,003$
BSA (66)	35,3	$123 \mu\text{mol L}^{-1}$	$0,225 \pm 0,004$

¹⁵⁴ Groves, P.; Malgorzata, P.; Molero, M.D.; Batta, G.; Cañada, F.J.; Jiménez-Barberoa, J. Protein molecular weight standards can compensate systematic errors in diffusion-ordered spectroscopy. *Anal. Biochem.* **2004**, 331, 395-397.

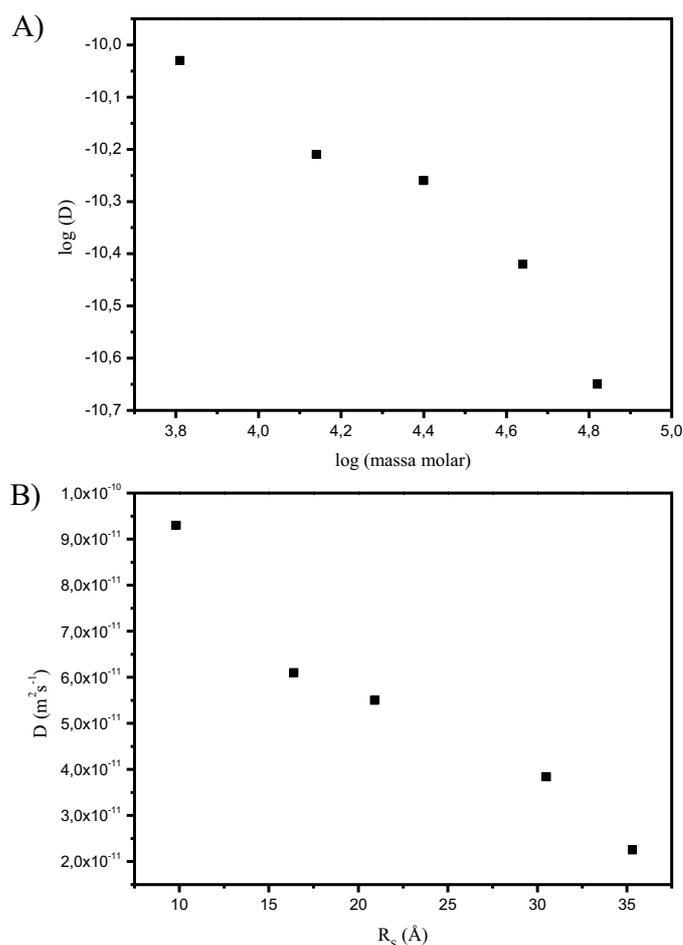
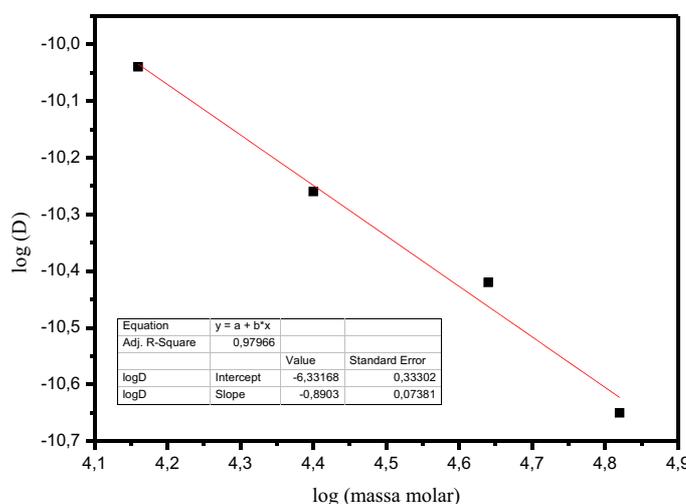


Figura IV-59. Curvas relacionando o logaritmo do coeficiente de difusão com o logaritmo da massa molar (A) o coeficiente de difusão com o raio de Stokes (B) das proteínas padrão.

Na tentativa de obter uma relação linear entre os coeficientes de difusão determinados e a massa molar das proteínas, utilizando para tanto alguns pontos da curva anterior (Figura IV-59) e adicionando os dados da lisozima (Tabela IV-9), obteve-se a curva da Figura IV-60, a partir da qual foram obtidas algumas estimativas da forma, tamanho e estado oligomérico da XACb0033.

Tabela IV-9. Dados da calibração do experimento de DOSY

Proteína (kDa)	Concentração	D * 10 ⁻¹⁰ (m ² /s)
Lisozima (14,4)	4,1 mmol L ⁻¹	0,915±0,009
α-quimiotripsinogênio (25)	400 μmol L ⁻¹	0,55±0,01
Ovalbumina (44)	182 μmol L ⁻¹	0,384±0,003
BSA (66)	900 μmol L ⁻¹	0,225±0,004

**Figura IV-60.** Curva relacionando o coeficiente de difusão com a massa molar, em escala logarítmica.

Utilizando o valor do coeficiente de difusão da XACb0033, determinado pelos ensaios de DOSY ($0,9 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$) na curva da Figura IV-60, obteve-se uma estimativa da massa molar em torno de 15 kDa, indicando portanto a presença da proteína dimérica em solução. Entretanto, comparando este mesmo valor de coeficiente de difusão, com os dados da Tabela IV-8, pode-se dizer que a XACb0033, nas condições analisadas, apresenta um raio de Stokes entre 9,83 e 16,4 Å e uma massa molar entre 6500 e 13700 Da indicando a presença da proteína monomérica em solução. Portanto, a partir destes resultados fica difícil confirmar o estado oligomérico da proteína em

solução. Confirmando a suposição de que, neste caso, a aplicação de NMR-DOSY não auxiliaria na determinação do estado oligomérico da proteína.

Pela Equação de Stokes-Einstein, utilizando os dados obtidos por DLS ($R_S = 18 \text{ \AA}$ e $D_{\text{calculado}} = 1,15 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$) e pelos experimentos de NMR-DOSY ($D = 0,9 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ e $R_{S\text{-calculado}} = 20 \text{ \AA}$) pode se considerar que os dados obtidos estão de acordo, indicando a proteína dimérica em solução, nas condições analisadas, diferentemente da filtração em gel analítica. E utilizando estes resultados de R_S , foi calculada a razão R_S/R_0 (igual a R_H/R_0) de 1,07 para o monômero e 1,1 para o dímero, em ambos casos indicando uma proteína de formato alongado.

Amostras da XACb0033 também foram submetidas à análise de SAXS na linha D02A-SAXS2 do LNLS, entretanto, pelos resultados de espalhamento observou-se que a proteína estava agregada.

Assim como para as outras duas proteínas discutidas, empregou-se a predição de estrutura terciária pelo servidor I-TASSER e obtiveram-se os modelos de coordenadas.¹³⁰ Os melhores modelos apresentaram exatidão estimada (TM-score) de $0,45 \pm 0,14$ (TM > 0,5 indicam modelos de topologia correta e TM < 0,17 indicam modelos de topologia randômica).

De acordo com o modelo (Figura IV-61), os 13 primeiros resíduos que correspondem ao domínio anfipático no N-terminal estão estruturados na forma de uma α -hélice, que conforme já discutido é uma característica comum das chaperonas de secreção.

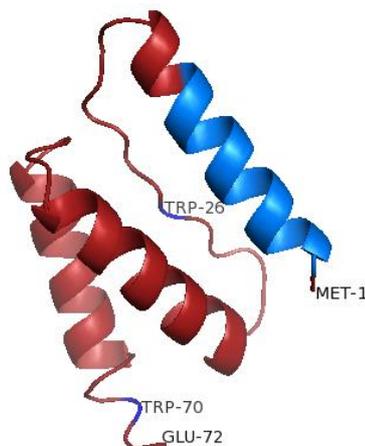


Figura IV-61. Estrutura calculada para a XACb0033. Modelo do I-TASSER (modelo de maior pontuação dentre os modelos gerados, *C-score* = -2,28) com N- e C-terminal marcados; em azul a hélice anfipática e os dois resíduos de Trp (W26 e W70) em regiões desordenadas.

De acordo com o modelo da Figura IV-60, tem-se que a proteína 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. O modelo gerado pelo I-TASSER corresponde a uma proteína com 57% de estrutura helicoidal. A partir do programa k2D2, foi gerado um espectro de dicroísmo circular correspondente a uma proteína com 57% de hélice- α e feita uma comparação com o espectro obtido experimentalmente (Figura IV-51) e dados da Tabela IV-7 (Figura IV-62).

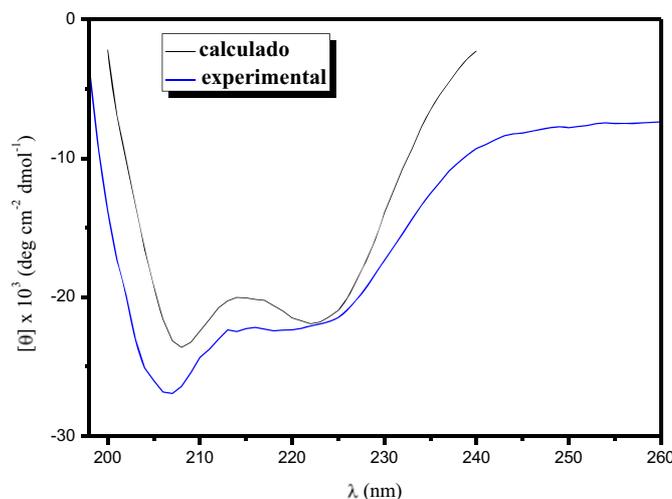


Figura IV-62. Espectros de CD da XACb0033, curva calculada em preto (200-240 nm) e curva experimental em azul.

Assim, observa-se que a XACb0033 é uma proteína pequena, expressa como uma mistura de dímero e monômero, mas após purificação se encontra monomérica em solução, nas condições analisadas. Devido a algumas características como a baixa massa molar ela é considerada uma possível chaperona de secreção. Como o gene que a codifica encontra-se no cluster do sistema de secreção do tipo IV, no plasmídeo pXAC64 da *Xac*, acredita-se que seja uma chaperona de secreção do sistema de secreção do tipo IV (T4SS). Por espectrometria de massas foi determinada a massa molar da mesma, como 8283,5 Da, indicando ausência do primeiro resíduo de metionina. Além disso, foi observada uma pequena formação do dímero na amostra (16,6 kDa). Em outros experimentos de espectrometria de massas com esta proteína, foi possível identificá-la, após a lise tríplica, obtendo 91% de cobertura da sequência.

Além da baixa massa molar, a XACb0033 apresenta outras características comuns a chaperonas de secreção, como baixo pI (5,6 - teórico) e um domínio anfipático. Segundo análise da estrutura primária por

bioinformática,⁶³ os resíduos envolvidos no domínio anfipático são os 13 primeiros resíduos (Figura IV-61).

Analisando a estrutura primária da XACb0033, observa-se que ela é uma proteína rica em metioninas (6 resíduos Met, correspondendo a 8,3%), característica de proteínas encontradas em sementes e grãos, sendo consideradas fontes nutricionais de enxofre (S).¹⁵⁵ Além disso, a presença de proteínas ricas em metioninas em plantas está relacionada, por exemplo, com a síntese de etileno. Entretanto, a presença de cinco resíduos de metionina (desconsiderando a primeira metionina, não observada no espectro de MS) na XACb0033 deve estar relacionada com a maior flexibilidade e propriedade antioxidante deste resíduo. A presença do átomo de enxofre, com maior polarizabilidade que o oxigênio, na metionina confere uma maior flexibilidade à molécula,¹⁵⁶ além de torná-la reativa frente a muitas espécies reativas de oxigênio, sendo prontamente oxidada a sulfóxido ou sulfona.¹⁵⁷ Assim, a presença de grande quantidade de metionina pode estar associada a processos de reconhecimento de proteínas parceiras, sendo que resíduos deste aminoácido levam a formação de uma superfície apolar flexível que se adapta a ligantes de vários tamanhos, além de servir como defesa antioxidativa para si mesma e para outras proteína do ambiente em que se encontra. Estas duas características são importantes para chaperonas de secreção, sendo mais um indício da função da XACb0033.

¹⁵⁵ Aragão, T.C.F.R.; Souza, P.A.S.; Uchôa, A.F.; Costa, I.R.; Bloch Jr., C.; Campos, F.A.P. Characterization of a methionine-rich protein from the seeds of *Cereus jamacaru* Mill. (Cactaceae). *Braz. J. Med. Biol. Res.* **2000**, *33*, 897-303.

¹⁵⁶ Gellman, S.H. On the Role of Methionine Residues in the Sequence-Independent Recognition of Nonpolar Protein Surfaces. *Biochemistry* **1991**, *30*, 6633-6636.

¹⁵⁷ Luo, S.; Levine, R.L. Methionine in proteins defends against oxidative stress. *FASEB J.* **2009**, *23*, 464-472.

Buscas por homologia e similaridade de sequência primária no BLAST¹²⁹ não resultaram em alinhamentos significativos. Assim como as buscas usando o PDB como banco de dados. Portanto, conforme já discutido, utilizou-se o programa I-TASSER¹³⁰ para o cálculo da estrutura tridimensional da XACb0033. A estrutura calculada pelo I-TASSER corrobora com os dados experimentais de fluorescência (Figura IV-52), nos quais se observou um máximo de emissão em 349 nm, indicando resíduos expostos ao solvente. E, de acordo com a estrutura calculada (Figura IV-61), se observa os dois resíduos de triptofano (W26 e W70) em regiões desordenadas da proteína e, portanto, expostas ao solvente.

Utilizando o programa QUARK para o cálculo de estrutura tridimensional, foi possível calcular mais um modelo tridimensional para a XACb0033 que é bem semelhante ao obtido pelo I-TASSER, diferindo principalmente nas regiões randômicas (Figura IV-63). Porém com o mesmo conteúdo de estrutura helicoidal que o modelo do I-TASSER (57%).

Como pode se observar na estrutura calculada pelo QUARK (Figura IV-63A), a principal diferença em relação à estrutura calculada pelo I-TASSER está no W70 que nesta última encontra-se numa região desordenada, enquanto que na predição do QUARK este resíduo se encontra em uma volta.

Fazendo o alinhamento da estrutura da XACb0033 com a estrutura calculada pelo QUARK da VirE1 de *Agrobacterium tumefaciens*⁶², observou-se similaridade estrutural entre as duas (Figura IV-64). Não foi possível utilizar a estrutura da VirE1 determinada experimentalmente, pois a mesma foi cristalizada no complexo com sua proteína parceira, a VirE2.

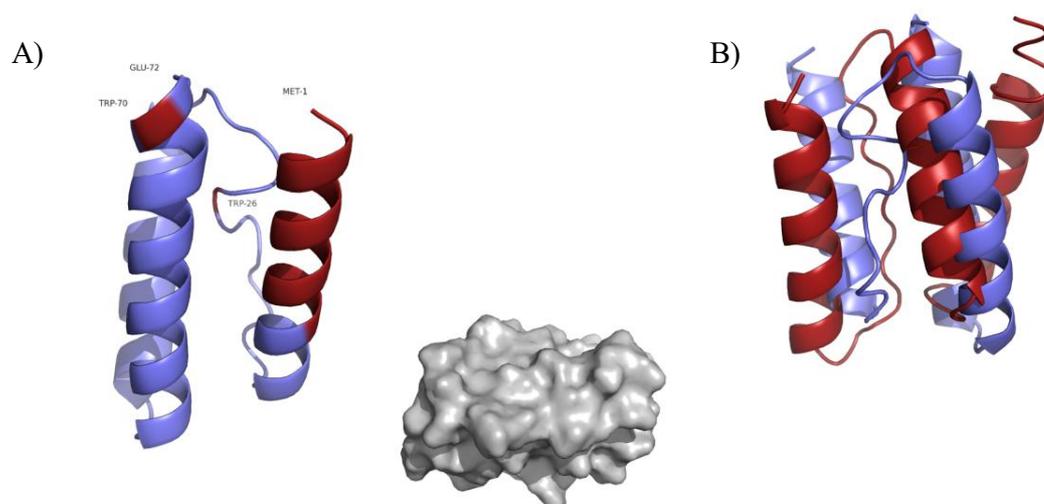


Figura IV-63. Resultados de bioinformática. A) Duas formas diferentes de representação da estrutura 3D da XACb0033 calculada pelo QUARK, ressaltando o N- e C-terminais, a região anfipática em vermelho no N-terminal e os resíduos W26 e W70 em vermelho. B) Sobreposição dos modelos tridimensionais, em vermelho a estrutura calculada pelo I-TASSER e em azul a estrutura calculada pelo QUARK.

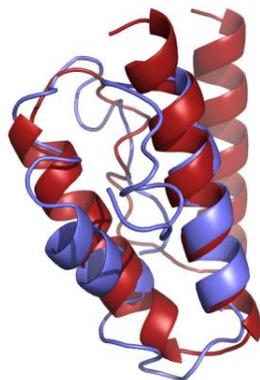


Figura IV-64. Alinhamento das estruturas: da XACb0033 calculada pelo QUARK, em vermelho, e VirE1 de *Agrobacterium tumefaciens*⁶² em azul, sendo alinhados 48 aminoácidos dos 65 da VirE1 (*TM-Score* = 0,46).

Na tentativa de resolver a estrutura da XACb0033 por cristalografia de raios-X, foram preparadas 98 condições de cristalização e mais algumas soluções não encontradas frequentemente nos kits de cristalização.¹³⁸ A

XACb0033 em tampão Tris-HCl (5 mmol L⁻¹, pH 8), na concentração de 3,6 mg mL⁻¹ foi colocada para cristalizar nestas 98 condições. De fato, ao concentrar a proteína alcançou-se a concentração de 13 mg mL⁻¹, esta amostra foi acondicionada a 4 °C por cinco dias antes de iniciar os ensaios, mas precipitou e precisou ser filtrada. A proteína foi colocada para cristalizar a 18 °C, entretanto não foi observado o crescimento de cristais protéicos, apenas inorgânicos.

Apesar de a proteína ter se apresentado agregada nos experimentos de SAXS, os resultados obtidos por DLS, filtração em gel analítica e NMR são condizentes com dados de SAXS não publicados desta proteína,^{2b} que resultaram em um raio de giro (Rg) de 23 Å e massa molar de 9,0 kDa, indicando a proteína monomérica em solução, como em alguns resultados aqui apresentados e semelhantes a VirE1 de *Agrobacterium tumefaciens*,⁶² que também foi encontrada monomérica em solução. Além de que, o alinhamento das estruturas calculadas para estas duas proteínas, indica similaridade estrutural entre as mesmas. Assim, apesar de não apresentar dados estruturais experimentais de alta resolução para confirmar a especulação da XACb0033 ser a VirE1 da *Xac*, os resultados apresentados indicam um forte indício de que ela seja uma chaperona de secreção do T4SS da *Xac*.

4.4. XAC1990 : FlgN

A XAC1990, caracterizada em trabalhos anteriores como a FlgN;¹⁵⁸ assim como a XACb0033, foi encontrada no precipitado da lise (Figuras IV-65), portanto insolúvel.

¹⁵⁸ Khater, L.; Santos, T.M.; Alegria, M.C.; Docena, C.; Tasic, L.; da Silva, A.C.R.; Farah, S.C.; Ramos, C.H.I. Identification of a flagellar chaperone FlgN in the phytopathogen *Xanthomonas axonopodis* pathovar *citri* via characterization of its interactions with FlgK. *Arch. Microbiol.* **2007**, *188*, 243-250.

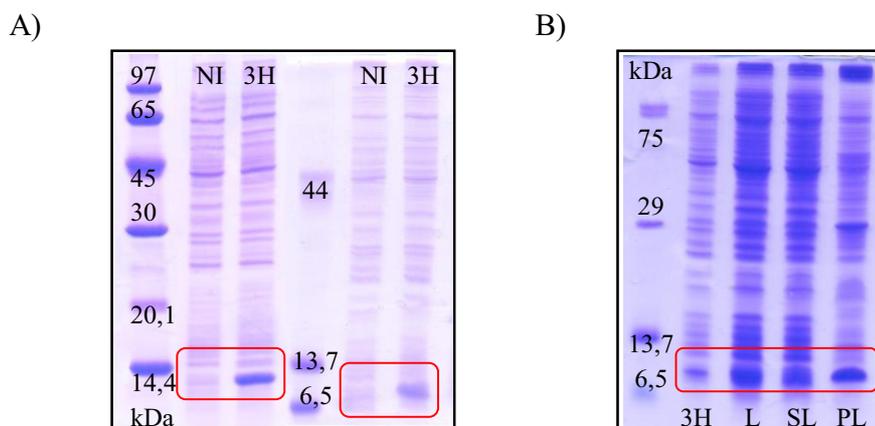


Figura IV-65. Géis de eletroforese SDS-PAGE (15%). A) Expressão da FlgN, tendo na sequência: marcador de massa molar de proteína (kit LMW/GE Healthcare); amostras não induzidas (NI) e amostras 3 horas depois da adição de IPTG (3H); mistura de proteínas com massa molar conhecida. B) Lise bacteriana, tendo na sequência: proteínas com massa molar conhecida (*Gel filtration Calibration kit LMW*); 3 horas depois da adição de IPTG (3H); lisado total (L); sobrenadante da lise (SL) e precipitado da lise (PL).

A porção da proteína FlgN presente no precipitado da lise também foi submetida ao tratamento com uréia (Figura IV-66 e 67), semelhante a XACb0033.

Vale ressaltar que a FlgN após três horas de expressão sofre degradação, portanto sua expressão é feita no tempo ótimo de três horas. O sobrenadante do reenovelamento, contendo a FlgN quase pura, foi filtrado e submetido à purificação por filtração em gel, utilizando a coluna Superdex 200 equilibrada com o tampão fosfato de sódio 50 mmol L⁻¹, pH 7,5 contendo 150 mmol L⁻¹ de cloreto de sódio (Figura IV-68).

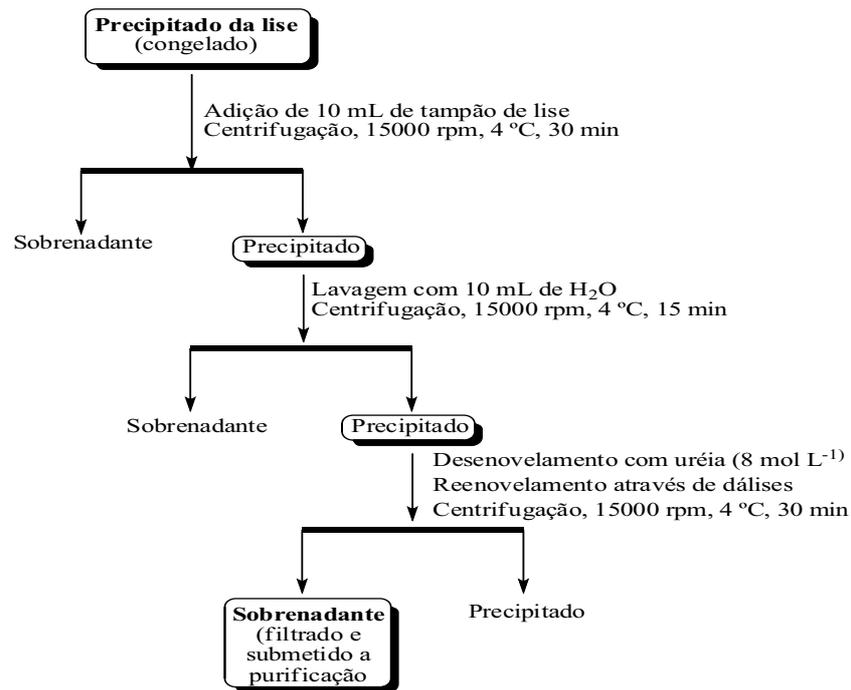


Figura IV-66. Fluxograma da obtenção da FlgN solúvel para posterior purificação.

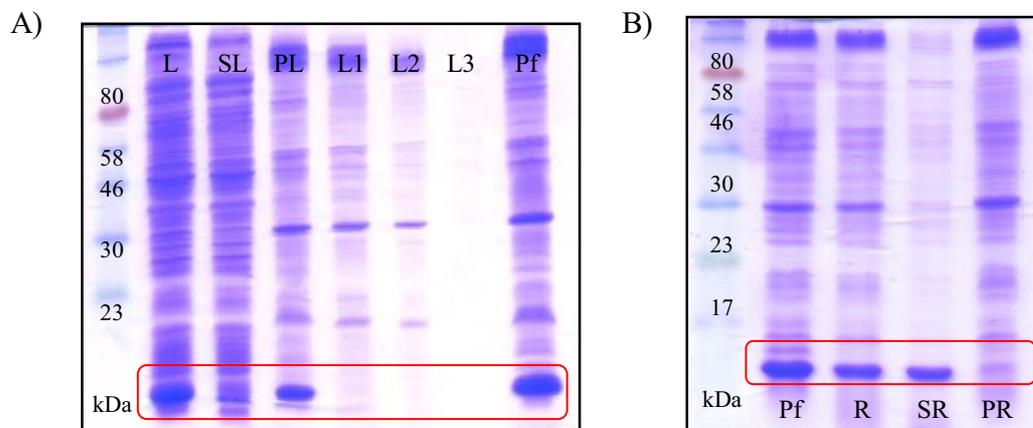


Figura IV-67. Géis de eletroforese SDS-PAGE (15%) do reenovelamento da FlgN, tendo na sequência: marcador de massa molar de proteína (P7709S/NEB); lisado total (L); sobrenadante da lise (SL) e precipitado da lise (PL), água das lavagens do precipitado da lise (L1,2 e 3), precipitado final (Pf); reenovelamento (R); sobrenadante do reenovelamento (SR) e precipitado do reenovelamento (PR).

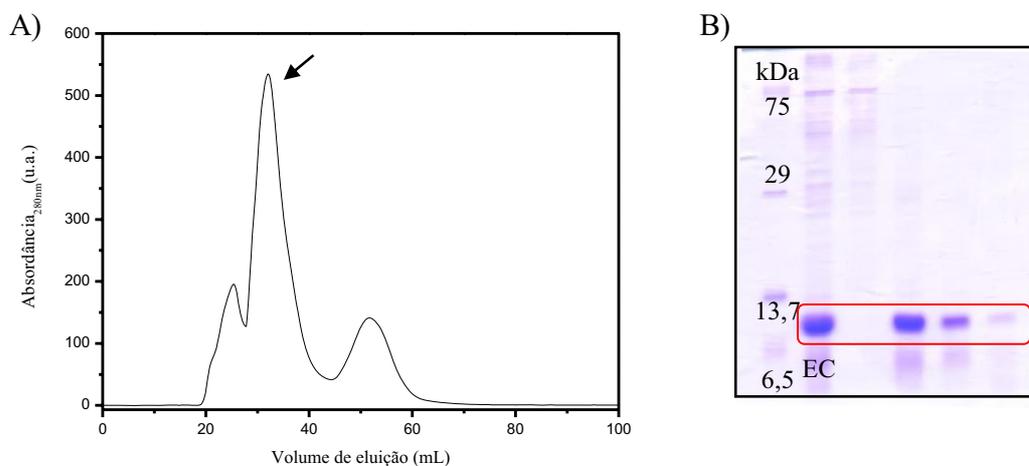


Figura IV-68. Cromatograma da purificação do sobrenadante do reenovelamento da FlgN (A), sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) da purificação, tendo na sequência: proteínas com massa molar conhecida (*Gel filtration Calibration kit LMW*), EC correspondendo à amostra que foi aplicada na coluna seguida de diferentes frações da purificação.

Para estimar a massa molar e o tamanho da FlgN, uma amostra contendo a proteína foi aplicada na coluna Superdex 200, nas mesmas condições citadas para a calibração (Figura IV-69). Utilizando o valor do coeficiente de partição da FlgN ($k_{av} = 0,62$) na curva de calibração da Figura IV-8B, obteve-se uma massa molecular de 12286 Da, sendo a massa teórica 12,12 kDa, concluiu-se portanto que a proteína se encontra em solução, nestas condições em forma monomérica.

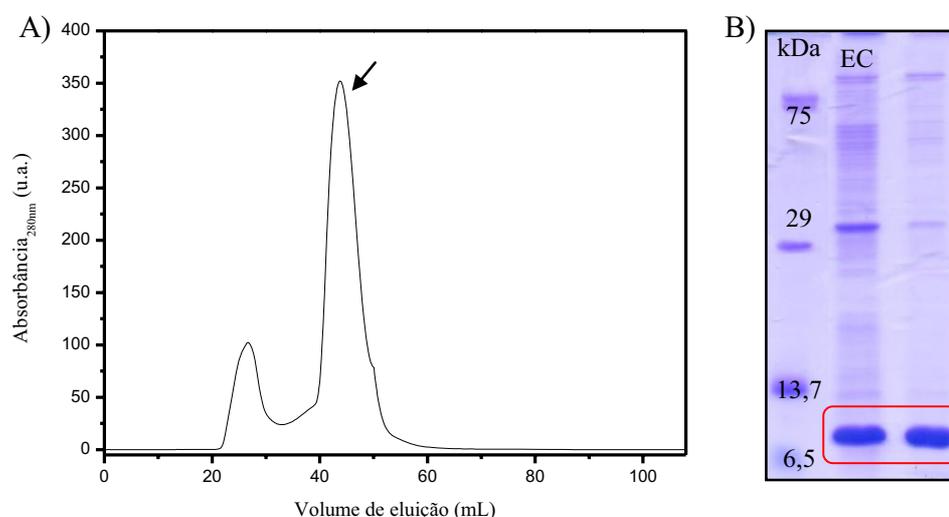


Figura IV-69. Filtração em gel analítica da FlgN. A) Cromatograma, sendo o pico correspondente à proteína indicado na figura. B) Gel de eletroforese SDS-PAGE (15%) referente ao cromatograma tendo na sequência: proteínas com massa molar conhecida (*Gel filtration Calibration kit LMW*), a amostra aplicada na coluna (EC) e a amostra referente ao pico indicado no cromatograma.

Utilizando o fator de retenção no gel de eletroforese ($R_f = 0,7$) para estimar a massa molar da proteína, obteve-se a massa molar de 10479 Da, levando à conclusão que ambas estimativas indicam presença da proteína na forma monomérica em solução.

Semelhante ao que foi feito para as outras proteínas, utilizou-se a curva de calibração da Figura IV-27B, para se obter uma estimativa do raio de Stokes da FlgN. Para tanto, foi realizado um novo experimento de filtração em gel analítica (Figura IV-70), nas mesmas condições citadas na calibração e estimou-se um raio de Stokes em torno de 19 Å. Utilizando a mesma curva para estimar a massa molar da proteína, encontrou-se o valor de 15,7 kDa, correspondendo a uma massa molar maior que a esperada, indicando portanto que a forma da proteína se desvia de uma esfera, devendo ser mais alongada, conforme já discutido para as outras proteínas.

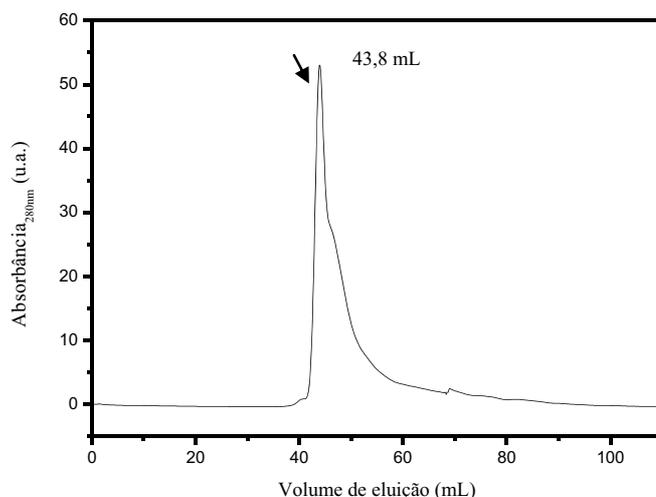


Figura IV-70. Filtração em gel analítica da FlgN. Cromatograma, sendo o pico correspondente à proteína e seu volume de eluição indicados na figura.

Como a FlgN apresenta um resíduo de triptofano em sua estrutura primária, a quantificação da mesma em solução foi realizada pelo método de Edelhoch,¹¹⁴ a partir da absorbância em 280 nm (Figura IV-71).

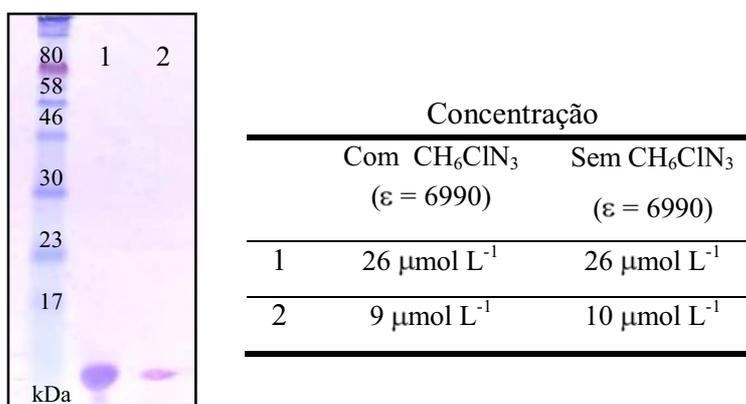


Figura IV-71. Gel de eletroforese SDS-PAGE (15%) de alíquotas da FlgN em solução, tendo na sequência: marcador de massa molar (P7709S/NEB) e diferentes frações de uma purificação. Ao lado, tabela de medidas das concentrações das soluções marcada.

Por espectrometria de massas foi feita a tentativa de confirmar a sequência da proteína através da digestão com tripsina. Para tanto, alíquotas

da proteína em solução foram submetidas à alquilação seguida de redução e depois digeridas com tripsina. Para a identificação das proteínas, os espectros de MS/MS foram inseridos no programa de busca Mascot¹¹⁹ e a FlgN foi identificada (Figura IV-72).

Match to: **xac:XAC1990** Nominal mass (M_r): **12114**; Calculated pI value: **6.74**
Matched peptides shown in **Bold Red**
MNVNEFLQRL SDALAGERQA LLENDIDGLM RHTQDKLSAL
QALEAAMPAGEEERLRELAE ANRANGALLARRRREVNWALRHL
GRTESAP SYDAKGQSSVLRGGRSLAVA

Figura IV-72. Identificação da XAC1990 por MS/MS. Em vermelho estão os peptídeos encontrados, sendo obtidos 96% de cobertura da sequência.

Alíquotas da proteína intacta foram levadas ao espectrômetro de massas para obtenção da massa exata da FlgN, entretanto a proteína degradou, pois no espectro apenas foram observadas massas em torno de 2 kDa. Esta proteína é bem instável, tanto que mesmo nas etapas de purificação observam-se sinais de degradação (Figura IV-68 e 70).

Identificada a proteína, foram realizados experimentos de dicroísmo circular para avaliar se a mesma encontrava-se enovelada e obter informação sobre sua estrutura secundária. Estes experimentos indicaram que a proteína apresentou-se enovelada (Figura IV-73) com ~80% de hélice- α ¹⁵² em sua estrutura secundária (Tabela IV-10).

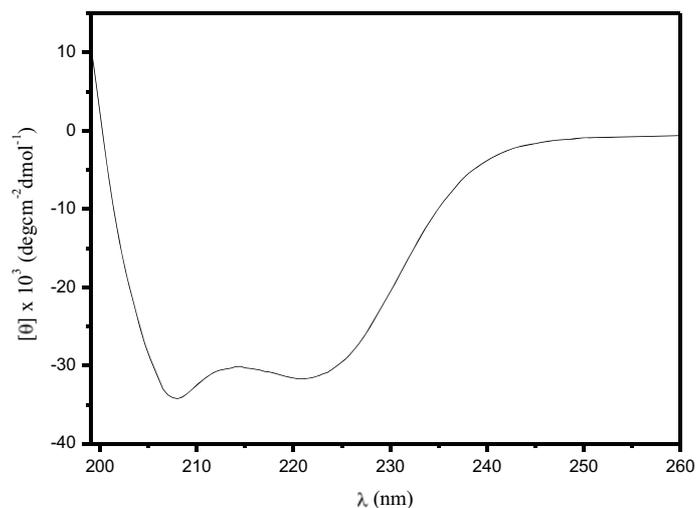


Figura IV-73. Dicroísmo Circular da FlgN. Curva de CD experimental com a proteína ($24 \mu\text{mol L}^{-1}$) em tampão fosfato de sódio (50mmol L^{-1} , pH 8).

Tabela IV-10. Dados de dicroísmo circular da FlgN

Programas utilizados	Porcentagem de hélice- α predita	Porcentagem de hélice- α calculada
GOR ¹⁴⁰	68%	-
NetSurfP ¹⁴¹	60%	-
Jpred ¹⁴⁶	52%	-
CDNN ¹⁵²	-	97%
K2D2 ¹²¹	-	99%
	-	$[\theta]_{222\text{nm}} = 89\%$ $[\theta]_{208\text{nm}} = 99\%$

Os dados experimentais foram utilizados em programas para calcular a quantidade de estrutura secundária, conforme indicado na Tabela IV-10. E ainda, foram utilizados programas de cálculo de estrutura secundária a partir da sequência de aminoácidos (Tabela IV-10) e os resultados foram comparados. Segundo os três diferentes programas de predição, a FlgN deve apresentar cerca de 60% de estrutura helicoidal. Entretanto, os valores

experimentais são discrepantes. Os três diferentes cálculos indicaram presença >97% de estrutura helicoidal.

Como a proteína apresenta um resíduo de triptofano em sua estrutura primária, foram realizados experimentos de fluorescência de emissão, onde se observou um máximo em 351 nm (Figura IV-74), indicando resíduo de triptofano exposto ao solvente.

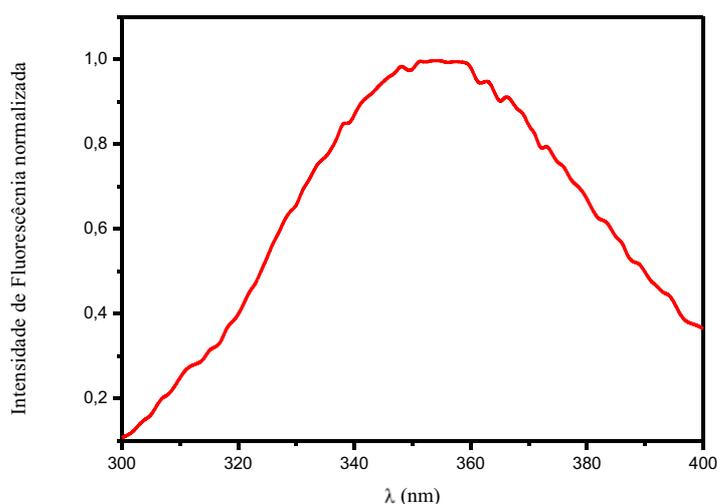


Figura IV-74. Fluorescência de emissão da FlgN. Espectros de emissão (máximo 351 nm), da proteína $2,6 \mu\text{mol L}^{-1}$ em tampão fosfato de sódio 10 mmol L^{-1} , pH 7,4.

Foram realizadas medidas de DLS para obter uma estimativa do tamanho desta proteína. A partir das medidas de DLS, obteve-se um raio de Stokes de 24 Å, com uma polidispersidade de 0,46 (Figura IV-75).

Os ensaios de DLS foram realizados com as duas amostras indicadas e em equipamentos diferentes. A amostra mais diluída apresentou um R_s de 25 Å, com uma polidispersidade de 11%, correspondendo a 99,8% das massas das partículas em solução. Portanto, uma amostra monodispersa ($59 \mu\text{mol L}^{-1}$), comparando estes resultados com os da filtração em gel analítica, obteve-se raios de Stokes da mesma ordem (filtração em gel analítica, 19 Å; DLS,

24,5 Å). Ambas as medidas, portanto, indicando a proteína monomérica em solução.

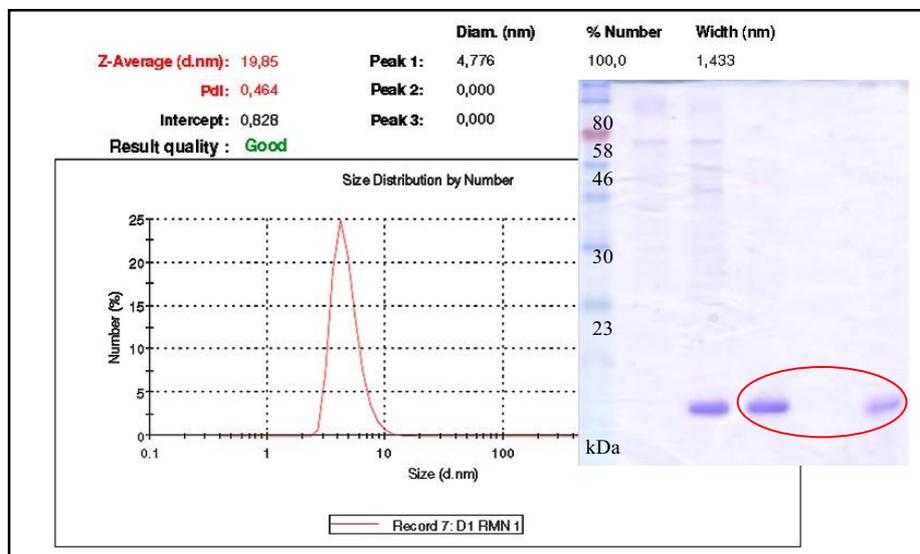


Figura IV-75. Dados de DLS da FlgN. Distribuição de tamanhos da proteína. Ao lado gel de eletroforese da amostra submetida ao ensaio, com marcador de massa molar (P7709S/NEB) e amostras da proteína.

Calculando a razão R_S/R_0 para uma proteína com esta massa molar, obteve-se um valor de 1,6, indicativo de uma proteína com formato bem alongado.

Na sequência foram realizados experimentos de NMR para obtenção do coeficiente de difusão da molécula. Entretanto, conforme observado no caso da XACb0033, os experimentos de DOSY, com as proteínas nas baixas concentrações alcançadas não permitiriam a determinação do estado de oligomerização e tamanho da proteína. Portanto, foram obtidos apenas os espectros de ^1H da FlgN, em diferentes concentrações e os melhores dados estão ilustrados na Figuras IV-76.

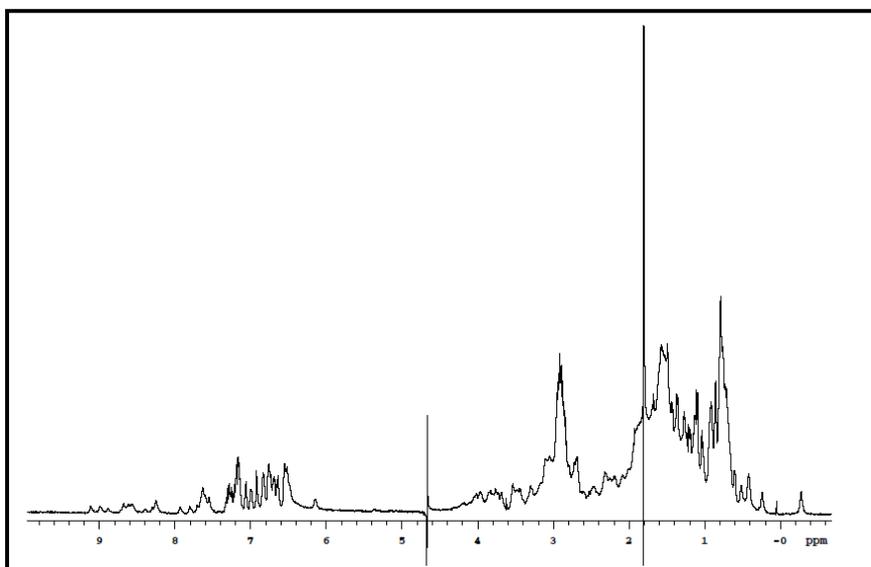
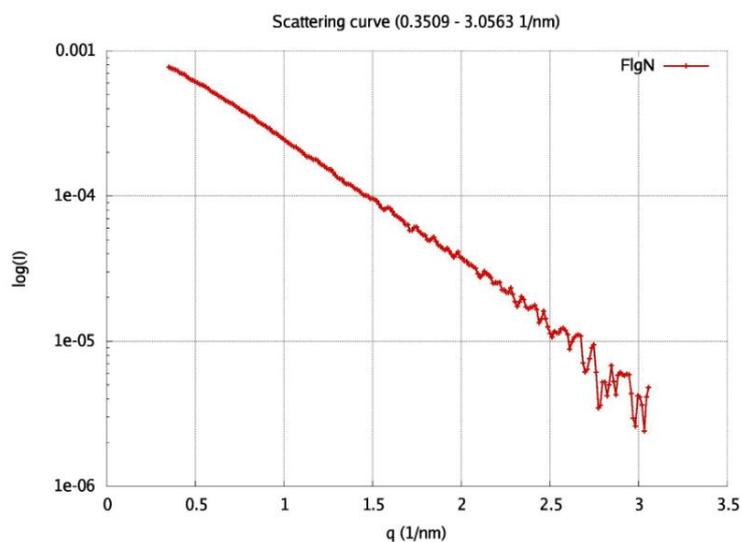


Figura IV-76. Espectro de NMR de ^1H da FlgN ($0,22 \text{ mmol L}^{-1}$) em tampão fosfato de sódio 5 mmol L^{-1} (D_2O 10%) obtido no equipamento Varian INOVA 500 MHz (499,98 MHz), utilizando a sequência de pulso WATERGATE para supressão do sinal de água em 4,8 ppm.

Amostras da FlgN foram submetidas à análise de SAXS nas linhas de SAXS do LNLS, juntamente com as duas proteínas discutidas anteriormente. As medidas foram realizadas nas condições já citadas.

Os dados obtidos foram processados utilizando o programa fit2D,¹²³ e a curva média de espalhamento foi construída usando o PRIMUS^{124b} (Figura IV-77). Em seguida construiu-se a curva de Guinier (Figura IV-76B).

A)



B)

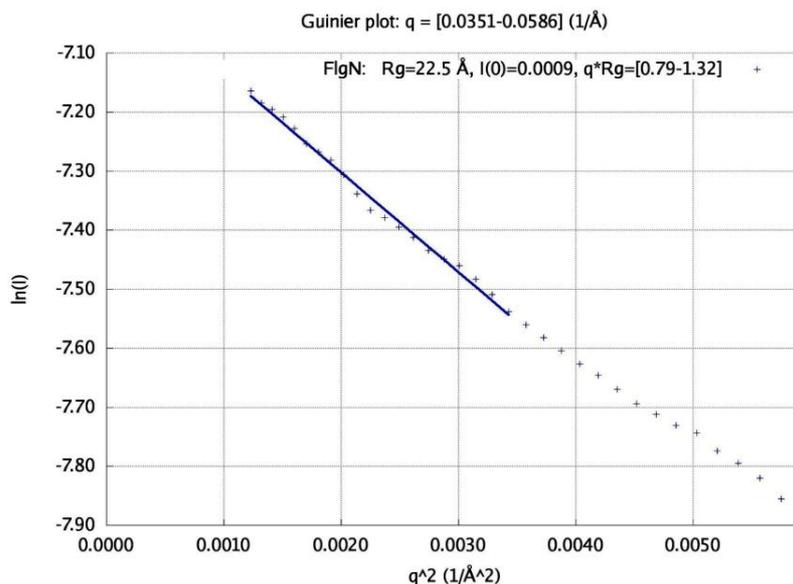


Figura IV-77. Dados de SAXS da FlgN. A) Curva de espalhamento de SAXS. B) Região de Guinier da curva de espalhamento.

O raio de giro (R_g) estimado pela regressão linear (curva de Guinier) foi de 22,5 Å. Uma nova estimativa das massas molares foi feita, usando o padrão para a calibração (BSA), sendo obtido 14,6 kDa para esta proteína, valor concordante com os obtidos nas análises anteriores. A função de distribuição

de distâncias e a curva de Kratky foram obtidas a partir das curvas experimentais e estão mostradas nas figuras a seguir (Figuras IV-78 e IV-79).

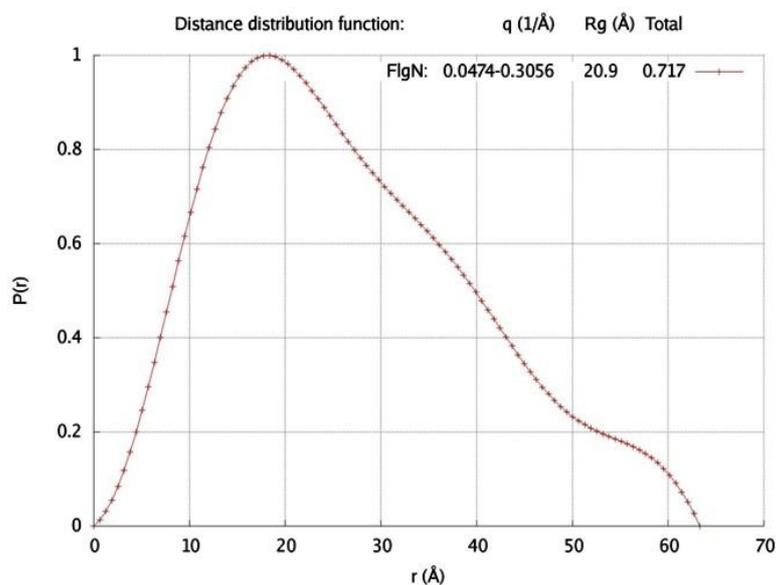


Figura IV-78. Curva de distribuição de distâncias da XAC1990, $P(r)$.

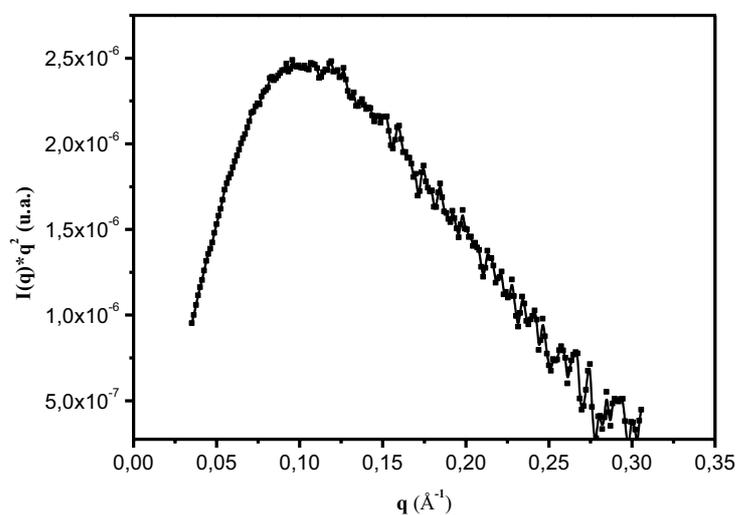


Figura IV-79. Curva de Kratky para a FlgN, obtida pelo tratamento dos dados de SAXS.

O valor do R_g derivado da função de distribuição de distâncias foi de $20,9 \text{ \AA}$ e concorda com o obtido pela análise de Guinier, pelos dados de DLS e

pela filtração em gel analítica. Uma segunda estimativa da massa molar foi de 13,0 kDa, comparável à obtida pela análise de Guinier das curvas de SAXS usando o padrão para calibração.

Estas estimativas da massa molar da FlgN concordam, dentro do erro experimental, entre si e com as estimativas obtidas nos ensaios anteriores. E todas estão de acordo com a massa molar calculada a partir da sequência de aminoácidos (12,1 kDa), indicando que a proteína em solução está na sua forma monomérica, nas condições analisadas, que incluiu baixa concentração protéica e ausência de agentes redutores. Além da análise de CD que confirmou a presença de elementos de estrutura secundária, as curvas de Kratky exibiram um máximo bem definido, característico de proteínas compactas com estrutura terciária estável. Para diferentes concentrações de proteínas e condições de tampão foi observada uma dependência não linear da região de Guinier das curvas de espalhamento para os primeiros pontos, correspondendo a valores de (q) bem baixos, o que é uma provável indicação de uma leve não-homogeneidade da amostra e pequenas quantidades de agregação. Este fato deve estar relacionado com o caráter anfipático do domínio C-terminal, que cria sítios para as interações proteína-proteína. É interessante notar que o corte de alguns pontos no começo das curvas de espalhamento não comprometem a análise total, pois as proteínas estudadas são relativamente pequenas em tamanho e massa, para as quais o conteúdo de informação se estende até ângulos maiores. A função de distribuição de distâncias não simétrica, com a forma de sino tombada para a esquerda nas distâncias menores, mostra que a FlgN possui uma forma alongada, exibindo uma distância máxima, correspondente à maior estimativa de raio de giro.

Como também não há modelos de alta resolução determinados por técnicas experimentais, a estrutura atômica da FlgN também foi calculada pelo

I-TASSER e está representada na Figura IV-80, assim como a sobreposição deste modelo no envelope experimental obtido pelos dados de SAXS.

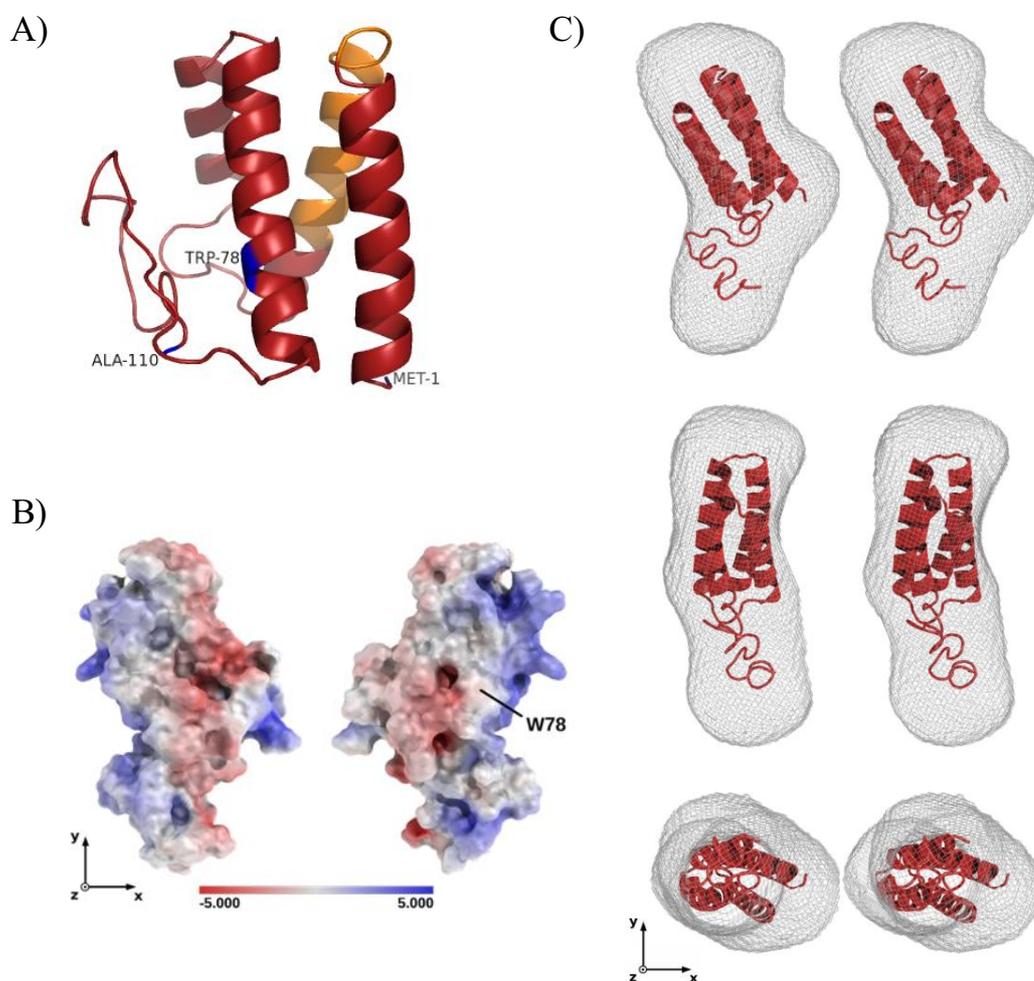


Figura IV-80. Representação em *cartoon* da estrutura da FlgN com N- e C- terminais marcados (I-TASSER, $C\text{-score} = -2,66$), a hélice anfipática destacada em laranja e o resíduo de Trp (W78) em azul (exposto ao solvente) (A). A mesma estrutura calculada, representada por superfície, com o potencial de carga superficial correspondente (potencial positivo em azul, negativo em vermelho e em cinza as regiões hidrofóbicas), destacando o W78 exposto ao solvente (B). Em (C) três vistas do modelo gerado pelos dados de SAXS da FlgN (em cinza) sobreposto com a estrutura calculada pelo I-TASSER (em vermelho).

O envelope obtido a partir das curvas experimentais é alongado, assim como o da XAC0419. O melhor modelo calculado apresentou exatidão estimada (*TM-score*) de $0,52 \pm 0,15$ sugerindo uma topologia correta para modelo. Também foram calculados outros quatro modelos, que compartilham as mesmas características estruturais gerais, diferindo principalmente nas regiões desordenadas. Os modelos obtidos para a FlgN contém um domínio compacto composto de hélices- α e a busca por similaridade estrutural no SCOP mostrou que eles equivalem à família k.9.1.1 (*de novo designed single chain three-helix bundle*).¹⁵¹

De acordo com o modelo obtido utilizando o I-TASSER, a proteína apresenta 72% de estrutura helicoidal. A partir do programa k2D2, novamente foi gerado um espectro de dicroísmo circular correspondente a uma proteína com 72% de hélice- α e feita uma sobreposição com o espectro obtido experimentalmente (Figura IV-81).

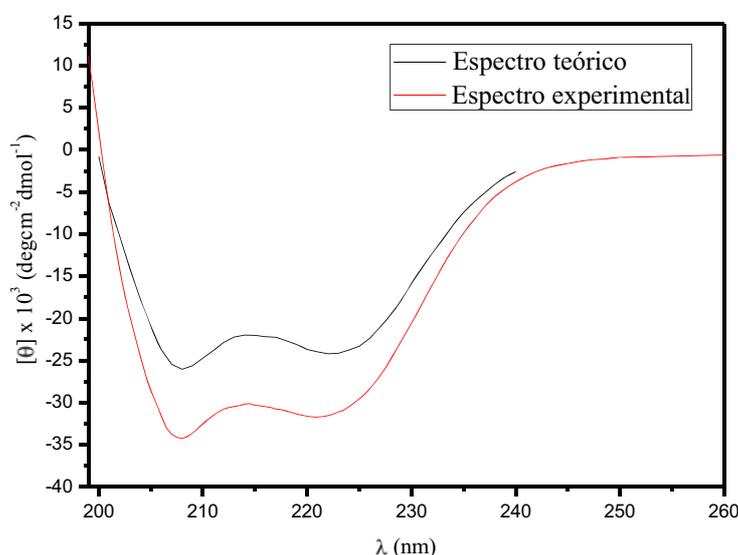


Figura IV-81. Comparação das curvas de CD da FlgN, em vermelho a curva experimental e em preto a curva teórica (200-240 nm), gerada pelo K2D2 para a estrutura calculada pelo I-TASSER.

Fazendo uma comparação, a distância máxima observada no modelo calculado é de 52 Å, que corresponderia a uma esfera de 26 Å de raio. Valor este, similar ao raio obtido pelas análises de DLS (24,5 Å) e SAXS (22,5 Å), sendo o valor obtido pela filtração em gel analítica inferior (19 Å). Utilizando estes dados de DLS e filtração em gel analítica, calculou-se a razão R_H/R_0 de 1,29, indicando uma estrutura de forma alongada.

Analisando os resultados de bioinformática e calculando o potencial de cargas superficiais desta proteína, observou-se que o resíduo de triptofano encontra-se na superfície da estrutura calculada pelo I-TASSER (Figura IV-79A). Assim, os resultados obtidos por fluorescência de emissão podem ser confirmados.

O gene que codifica a FlgN encontra-se no cluster do sistema de secreção do tipo III, próximo aos genes de patogenicidade e de resposta de hipersensibilidade no cromossomo da *Xac*. Esta proteína também é pequena, com massa molar teórica de 12121 Da, pI de 6,7 e os resíduos envolvidos no domínio anfipático são os resíduos de 22 a 38, segundo a predição. Em 2007, através de ensaios de duplo-híbrido ela foi identificada como uma chaperona flagelar, a FlgN, cuja proteína parceira é a FlgK.¹⁵⁸

Utilizando o programa QUARK, para um novo cálculo da estrutura da FlgN, obteve-se um resultado bastante similar ao do I-TASSER, apenas diferindo nas regiões randômicas (Figura IV-81). Além de um maior conteúdo helicoidal que o modelo do I-TASSER (90%).

Conforme se observa na Figura IV-82B, a estrutura calculada pelo QUARK para a FlgN apresenta algumas regiões, calculadas pelo I-TASSER como desordenadas, estruturadas em hélices. Sendo também uma estrutura mais alongada do que a calculada pelo I-TASSER.

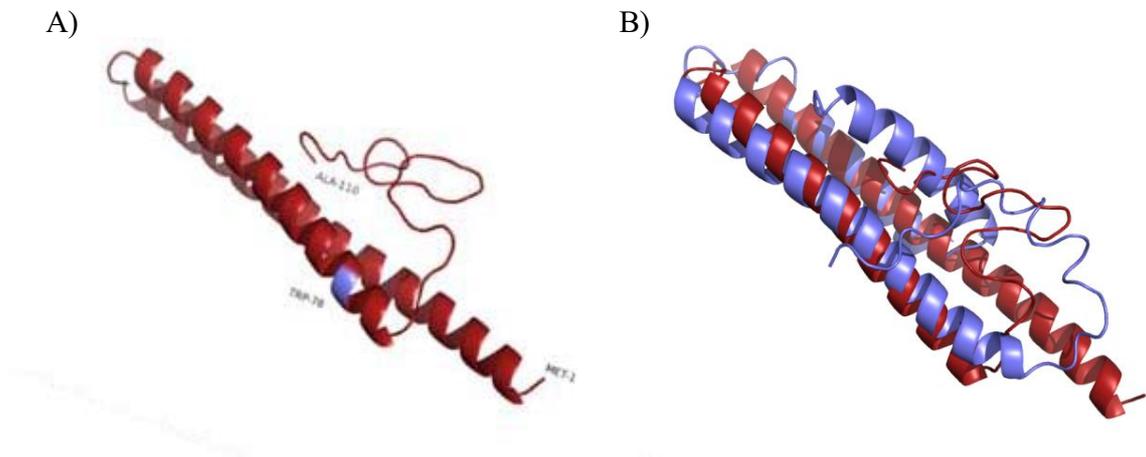


Figura IV-82. Resultados de bioinformática. A) Estrutura da FlgN calculada pelo QUARK, ressaltando o N- e o C-terminal, e o resíduo de triptofano em azul. B) Sobreposição dos modelos tridimensionais da FlgN, em azul a estrutura obtida pelo I-TASSER e em vermelho a estrutura calculada pelo QUARK.

Buscas no servidor DALI¹⁵⁰ levaram a resultados similares ao do SCOP, sendo importante ressaltar que a busca utilizando o PDB como banco de dados indicou similaridade estrutural da FlgN com a chaperona flagelar FliT de *Salmonella* sp.,⁶⁰ mais uma vez indicando a similaridade estrutural das chaperonas de secreção da classe III ou flagelar. Também se observou similaridade de 49% e 27% de identidade da FlgN da *Xac* com a FlgN de *Pseudomonas aeruginosa* e, comparando a estrutura da FlgN calculada pelo I-TASSER com a estrutura da FlgN de *Pseudomonas aeruginosa*, depositada no PDB em 2007, também observa-se semelhança estrutural entre as duas (Figura IV-83)

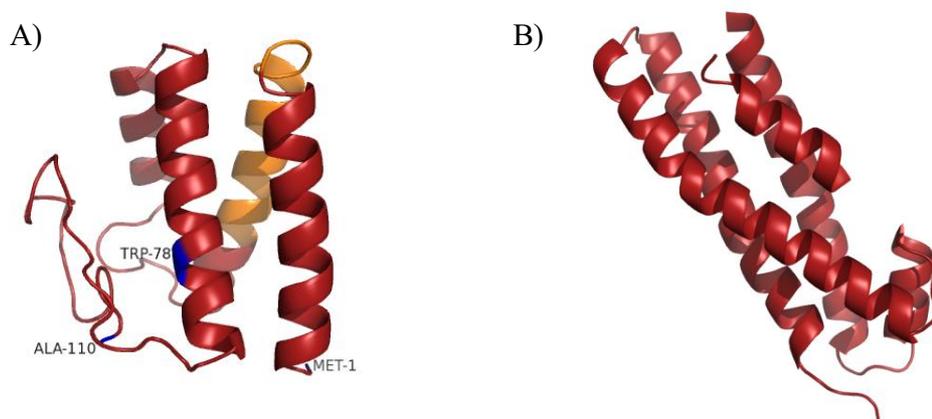


Figura IV-83. Resultados de SAXS e bioinformática. A) Estrutura da FlgN da *Xac*, calculada pelo I-TASSER. B) Estrutura cristalográfica da FlgN de *Pseudomonas aeruginosa* (PDB 2FUP, depositada em 2007, dados não publicados).

Quando se fez o alinhamento da estrutura da FlgN da *Xac* calculada pelo QUARK com a estrutura cristalográfica da FlgN de *Pseudomonas aeruginosa* (Figura IV-84), também observou-se similaridade estrutural entre as duas.

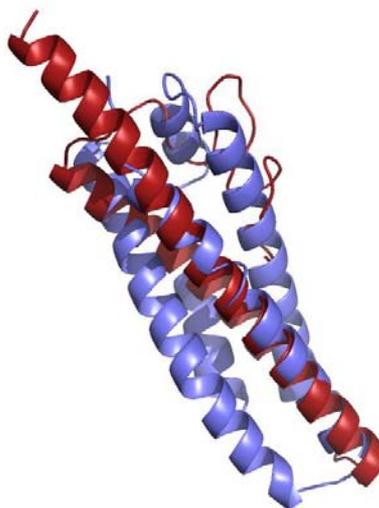


Figura IV-84. Alinhamento das estruturas: da FlgN da *Xac* calculada pelo QUARK, em vermelho, e FlgN de *Pseudomonas aeruginosa* em azul (PDB:2FUP), sendo alinhados 72 aminoácidos dos 110 da FlgN da *Xac* ($TM-Score = 0,45$).

A XAC1990 é uma chaperona flagelar e estudos anteriores acusaram que, além das interações *in vitro*, seu tamanho e ponto isoelétrico correspondem a uma FlgN.^{15,158}

Os resultados estruturais apresentados mostraram a estrutura tridimensional da proteína, porém não no nível atômico, e confirmaram suas características estruturais parecidas com a FlgN de outros microorganismos.

Capítulo 5



CONCLUSÃO E PERSPECTIVAS

*Adaptado de <http://flores.culturamix.com/blog/wp-content/gallery/flor-de-laranjeira/foto-flor-de-laranjeira-03.jpg>

5. Conclusão e Perspectivas

Considerando os resultados apresentados e discutidos, pode-se dizer que o trabalho realizado na intenção de caracterizar estruturalmente quatro possíveis chaperonas de secreção da *Xac* atingiu os objetivos propostos, sendo que para duas das proteínas obteve-se o envelope molecular através dos experimentos de SAXS, e alguns dos resultados obtidos já estão publicados.

Compilando os resultados apresentados, tem-se que:

5.1. XAC1346

- ✓ Os dados teóricos indicam uma proteína pequena, 10945,9 Da (108 aminoácidos), com baixo ponto isoelétrico (5,64) e região anfipática no C-terminal (aminoácidos de 88-100).
- ✓ A expressão em *E. coli* é baixa, sendo a proteína encontrada no sobrenadante após a lise.
- ✓ Purificada em duas etapas: troca iônica (DEAE-Sepharose) e filtração em gel (Superdex 200).
- ✓ Por filtração em gel analítica estimou-se a massa molar em 17,7 kDa, e pelo gel de eletroforese a mesma foi estimada em 17,3 kDa; ambas indicando provavelmente a proteína dimérica em solução.
- ✓ A proteína foi identificada, nas frações da filtração em gel, por espectrometria de massas, com 36% de cobertura da sequência, com massa molar calculada em 10,9 e 11,0 kDa.

- ✓ Em baixas concentrações a proteína foi obtida pura e enovelada, conforme indicado pelo espectro de CD característico de proteínas com alto conteúdo helicoidal (60% de hélice- α).
- ✓ Pelas medidas de DLS com a proteína diluída observou-se um raio hidrodinâmico de 29 Å, indicativo de uma proteína alongada.
- ✓ Pela estrutura calculada utilizando métodos computacionais, a proteína apresenta cerca de 80% de hélices- α e uma distância atômica máxima de 70 Å, o que corresponderia a um raio de Stokes em torno de 35 Å, condizente com os resultados de DLS. Além de que, a região anfipática no C-terminal (88-100) está estruturada na forma de hélice, característica de chaperonas de secreção.
- ✓ Por métodos computacionais, a estrutura calculada para a XAC1346 apresentou similaridade estrutural com a FliJ de *Salmonella enterica* serovar Typhimurium.
- ✓ Apesar dos dados obtidos reforçarem a forte indicação de que a XAC1346 é uma chaperona de secreção do T3SS da *Xac*, provavelmente uma chaperona flagelar, não foi possível confirmar esta hipótese por falta de resultados experimentais, devido à dificuldade de se obter a proteína em concentrações mais altas, necessárias para os experimentos de SAXS, NMR-DOSY e cristalografia de raios-X.

5.2. XAC0419

- ✓ Os dados teóricos também indicam uma proteína pequena, 10767,0 Da (103 aminoácidos), com pI = 7 e região anfipática no C-terminal (aminoácidos de 67-91).
- ✓ A expressão em *E. coli* também é baixa, sendo a proteína encontrada no sobrenadante após a lise.
- ✓ Purificada em duas etapas: troca iônica (CM-Sepharose) e filtração em gel (Superdex 200).
- ✓ Por filtração em gel analítica estimou-se a massa molar em 12,8 kDa, e pelo gel de eletroforese a mesma foi estimada em 10,7 kDa; ambas indicando a proteína monomérica em solução. Pelo gel de eletroforese, também se observou formação do dímero em algumas amostras, com massa molar estimada de 21,4 kDa.
- ✓ A proteína foi identificada tanto nas frações contendo a forma dimérica como nas contendo a forma monomérica por espectrometria de massas, com 11% de cobertura para a amostra dimérica e 36% de cobertura da sequência para a amostra monomérica.
- ✓ Conforme indicado pelo espectro de CD, em baixas concentrações a proteína foi obtida pura e enovelada; apresentando espectro característico de proteínas com conteúdo helicoidal (30% de hélice- α para o monômero e 54% para o dímero).

- ✓ Pelas análises de filtração em gel analítica este raio foi estimado em 17 Å, indicativo de uma proteína alongada.
- ✓ Não foram obtidas soluções da proteína em concentrações suficientemente altas para os experimentos de NMR-DOSY.
- ✓ Através dos experimentos de SAXS foram obtidos os seguintes resultados: pela aproximação de Guinier estimou-se um raio de giro de 13,9 Å e pela curva de distribuição de distâncias, o mesmo foi estimado em 13,5 Å. Tais estimativas foram concordantes entre si e com o dado de filtração em gel analítica. A massa molar foi estimada, por dois métodos diferentes, em 9,0 e 8,6 kDa. Ressaltando que todos os dados indicaram a presença da proteína monomérica em solução.
- ✓ Foi obtido um modelo da estrutura 3D por métodos computacionais, indicando que a proteína apresenta cerca de 40% de hélices- α , dados condizentes com o obtido por CD para o monômero e uma distância atômica máxima de 40 Å, o que corresponderia a um raio de Stokes de 20 Å, condizente com os resultados da filtração em gel analítica. Nesta estrutura calculada a região anfipática no C-terminal (67-91) está estruturada na forma de hélice, característica de chaperonas de secreção.
- ✓ Para avaliar o envelope molecular gerado pelo processamento dos dados de SAXS, foi feita uma sobreposição deste com a estrutura calculada. Nesta sobreposição foi observada uma excelente concordância na região central que compreende as hélices- α e folhas- β preditas, apesar das regiões desordenadas provavelmente

assumirem uma forma mais globular e compacta em solução do que o esperado pelo resultado do I-TASSER.

- ✓ Na ausência de estruturas tridimensionais determinadas por técnicas de alta resolução, não foi possível classificar a XAC0419 como chaperona de secreção. Entretanto, pode-se afirmar que existem fortes indícios de que ela seja uma chaperona de secreção do T3SS da classe I. Pois a mesma apresenta uma forma alongada com enovelamento do tipo 2α - 3β (α - $\beta\beta\beta$ - α), muito similar à estrutura de chaperonas de secreção do T3SS da classe I, como a SycE de *Yersinia pseudotuberculosis*, descritas na literatura. O fato destas chaperonas da classe I estudadas até o momento se apresentarem diméricas em solução e a XAC0419 ter se apresentado monomérica demanda estudos mais aprofundados com ela, principalmente tendo em vista que já foi comprovado que ela é expressa pela *Xac*.

5.3. XACb0033

- ✓ Os dados teóricos indicam que esta proteína também é pequena, sendo a menor das quatro estudadas, tendo 8418,5 Da (72 aminoácidos), com pI = 5,6 e os 13 primeiros resíduos formando a região anfipática no N-terminal.
- ✓ A expressão em *E. coli* é alta e a proteína é encontrada no precipitado após a lise.
- ✓ Após ser retirada do precipitado da lise através do tratamento com uréia e posterior reenovelamento – etapa que pode ser considerada

uma etapa de purificação – foi purificada por filtração em gel (Superdex 200).

- ✓ Por filtração em gel analítica estimou-se a massa molar em 7,8 kDa, e pelo gel de eletroforese a mesma foi estimada em 7,9 kDa; ambas indicando a proteína monomérica em solução.
- ✓ A proteína foi identificada nas frações contendo a forma dimérica por espectrometria de massas, com 91% de cobertura da sequência. Também por espectrometria de massas foi obtida a massa exata da proteína de 8283,5 Da (sem o primeiro resíduo de metionina) e também se observou formação do dímero em 16,6 kDa.
- ✓ A proteína foi obtida pura e enovelada, conforme indicado pelo espectro de CD característico de proteínas com alto conteúdo helicoidal (99%).
- ✓ Nas medidas de fluorescência de emissão, observou-se um máximo em 349 nm, indicando que os resíduos de triptofano (W26 e W70) estão expostos ao solvente.
- ✓ Pelas medidas de DLS, com a proteína diluída, observou-se um raio hidrodinâmico de 18 Å, porém pelas análises de filtração em gel analítica este raio foi estimado em 14 Å, correspondendo a uma massa molar de 9,3 kDa.
- ✓ Pelos ensaios de NMR-DOSY, obteve-se um coeficiente de difusão de $0,9 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$. Utilizando este resultado para o coeficiente de difusão da molécula e o dado de raio hidrodinâmico obtido pela

filtração em gel analítica calculou-se a razão R_H/R_0 da proteína (1,1), indicativo de uma estrutura alongada.

- ✓ Foi obtido um modelo da estrutura 3D por métodos computacionais, indicando que a proteína apresenta 57% de hélices- α , dados condizentes com o obtido por CD para o monômero e uma distância atômica máxima de 33 Å, o que corresponderia a um raio de Stokes de 17 Å, condizente com os resultados de DLS e da filtração em gel analítica. Nesta estrutura calculada a região anfipática no N-terminal (1-13) está estruturada na forma de hélice, característica de chaperonas de secreção. E os dois resíduos de triptofano (W26 e W70) encontram-se em regiões desordenadas no modelo, o que corrobora com os dados de fluorescência de emissão obtidos. Além dos resultados obtidos também corroborarem com dados de SAXS desta proteína não publicados.^{2b}

- ✓ Na ausência de estruturas tridimensionais determinadas por técnicas de alta resolução, não foi possível classificar a XACb0033 como chaperona de secreção. Entretanto, especula-se que ela seja uma chaperona de secreção do T4SS, mais especificamente a VirE1 da *Xac*. Pois a mesma apresenta baixa massa molar, baixo pI, uma hélice anfipática no N-terminal e possui uma forma alongada com estrutura principalmente helicoidal, características comuns a chaperonas de secreção, e como discutido, muito similar em todos os elementos à VirE1 da *Agrobacterium tumefaciens*.

5.4. XAC1990: FlgN

- ✓ Os dados teóricos correspondem a uma proteína pequena, 12121,4 Da (110 aminoácidos), pI baixo (6,7) e região anfipática no N-terminal (aminoácidos de 22-38).
- ✓ Os níveis de expressão em *E. coli* são altos, sendo a proteína encontrada no precipitado após a lise. Ressaltando que a expressão da mesma é feita em 3 horas, devido a sua instabilidade.
- ✓ Retirada do precipitado da lise também pelo tratamento com uréia. Sendo que este processo de reenovelamento pode ser considerado uma etapa de purificação. Purificada por filtração em gel (Superdex 200).
- ✓ Por filtração em gel analítica estimou-se a massa molar em 12,3 kDa, e pelo gel de eletroforese a mesma foi estimada em 10,5 kDa; ambas indicando a proteína monomérica em solução.
- ✓ A proteína foi identificada por espectrometria de massas, com 96% de cobertura da sequência para a amostra monomérica.
- ✓ Em baixas concentrações a proteína foi obtida pura e enovelada, conforme indicado pelo espectro de CD característico de proteínas com alto conteúdo helicoidal (99% de hélice- α).
- ✓ Nas medidas de fluorescência de emissão, observou-se um máximo em 351 nm, indicando que o triptofano (W78) desta proteína também está exposto ao solvente.

- ✓ Pelas medidas de DLS com a proteína diluída observou-se um raio hidrodinâmico de 24,5 Å, porém pelas análises de filtração em gel analítica este raio foi estimado em 19 Å, correspondendo a uma massa molar de 15,7 kDa. E pela a razão R_H/R_0 a proteína deve possuir uma estrutura alongada.
- ✓ Não foram obtidas soluções da proteína em concentrações suficientemente altas para os experimentos de NMR-DOSY.
- ✓ Através dos experimentos de SAXS foram obtidos os seguintes resultados: pela aproximação de Guinier estimou-se um raio de giro de 22,5 Å e pela curva de distribuição de distâncias, o mesmo foi estimado em 20,9 Å. Tais estimativas foram concordantes entre si e com o dado de filtração em gel analítica e de DLS. A massa molar foi estimada, por dois métodos diferentes, em 14,6 e 13 kDa. Ressaltando que todos os dados indicaram a presença da proteína monomérica em solução.
- ✓ Foi obtido um modelo da estrutura 3D por métodos computacionais, indicando que a proteína apresenta cerca de 70% de hélices- α , sendo que os dados de CD indicaram um conteúdo helicoidal maior; o modelo apresenta uma distância atômica máxima de 52 Å, o que corresponderia a um raio de Stokes de 26 Å, condizente com os resultados de diferentes técnicas. Nesta estrutura calculada a região anfipática, no N-terminal (22-38), também está estruturada na forma de hélice, como nas outras proteínas estudadas.
- ✓ Para avaliar o envelope molecular gerado pelo processamento dos dados de SAXS, foi feita uma sobreposição deste com a estrutura

calculada. Onde se observou que o domínio compacto composto das hélices- α se sobrepõe muito bem à porção central do envelope experimental, apesar das regiões desordenadas provavelmente assumirem uma forma mais compacta em solução do que o esperado pelo resultado do I-TASSER.

- ✓ O modelo gerado pelo programa QUARK foi mais condizente com os dados experimentais, sendo que nele algumas regiões calculadas pelo I-TASSER como desordenadas estão compactadas em hélices. Este modelo é mais alongado, condizente com os dados experimentais e possui 90% de estrutura helicoidal, concordante com os dados obtidos por CD. Assim como na estrutura obtida pelo I-TASSER, neste modelo o resíduo de triptofano (W78) se encontra em uma hélice, mas exposto ao solvente, corroborando os resultados de fluorescência de emissão.

- ✓ Todos os dados estruturais obtidos confirmam a identificação da XAC1990 como a FlgN da *Xac* feita em 2007. Buscas por similaridade de sequência indicaram 27% de identidade com a FlgN de *Pseudomonas aeruginosa* e, comparando a estrutura da FlgN calculada pelo I-TASSER com a estrutura da FlgN de *Pseudomonas aeruginosa*, depositada no PDB em 2007, também observa-se semelhança estrutural entre as duas. Além disso, a estrutura calculada pelo I-TASSER apresentou similaridade estrutural com a chaperona flagelar FliT de *Salmonella enterica* serovar Typhimurium. Sendo estes os primeiros dados estruturais da FlgN da *Xac*, classificada como uma chaperona de secreção da classe III.

Como perspectivas deste trabalho, dadas as dificuldades em se obter as quatro proteínas estudadas em concentrações razoavelmente altas para resolução da estrutura atômica por cristalografia de raios-X ou NMR, tem-se o uso da Espectrometria de Massas, em colaboração com o grupo do Prof. Dr. Fábio Gozzo, através de experimentos de ligação cruzada para a obtenção de maiores informações estruturais.

Graças à concentração relativamente baixa de proteína, necessária para a realização dos experimentos de ligação cruzada, este estudo parece bastante promissor. Tais experimentos podem ser aplicados para as quatro proteínas, sendo que, no caso das XAC1346 e XAC0419, pode ser utilizado o agente de ligação cruzada mais comum destes experimentos (Disuccinimidil Suberato – DSS), pois tais proteínas apresentam alto conteúdo de lisinas, resíduos que reagem com o DSS. Já as proteínas XACb0033 e FlgN devem ser estudadas utilizando como agente de ligação cruzada o EDC (1-etil-3-(3-dimetilaminopropil) carboimida), que se liga a resíduos de ácido glutâmico, presentes nas duas proteínas.

Estes experimentos permitem a determinação de restrições espaciais entre os pares ligados pelo agente de ligação cruzada e em seguida podem ser obtidas, por meio de modelagem molecular, estruturas tridimensionais que satisfaçam tais distâncias definidas experimentalmente.¹⁵⁹ Os resultados destes experimentos podem ser úteis na classificação da XACb0033, XAC0419 e XAC1346 como chaperonas de secreção da *Xac*.

¹⁵⁹ Figueredo, A.R. 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.

ANEXOS

- ✓ Tabelas das condições de cristalização, onde a concentração molar (mol L^{-1}) foi referida como M por questão de espaço.
- ✓ Revisão sobre Chaperonas de Secreção de bactérias publicada em 2011 no periódico *Protein and Peptide Letters*.
- ✓ Artigo publicado em 2011 com os dados estruturais obtidos para as proteínas XAC0419 e FlgN (*The Protein Journal*).

Tabela A1. Condições de cristalização utilizadas, referentes ao kit 1 da Hampton

Condição	Sal	Conc.	Tampão	Conc.	pH	Precipitante 1	Conc.	Precipitante 2	Conc.
1	Cloreto de cálcio dihidratado	0,02 M	Acetato de sódio trihidratado	0,1 M	4,6	(+/-)-2-metil-2,4-pentanolol	30%		
2						Tartarato de sódio e potássio tetrahidratado	0,4 M		
3						Fosfato de amônio monobásico	0,4 M		
4			Tris-HCl	0,1 M	8,5	Sulfato de amônio	2,0 M		
5	Citrato de sódio tribásico dihidratado	0,2 M	HEPES sódico	0,1 M	7,5	(+/-)-2-metil-2,4-pentanolol	30%		
6	Cloreto de magnésio Hexahidratado	0,2 M	Tris-HCl	0,1 M	8,5	Poli(etilenoglicol) 4000	30%		
7			Cacodilato de sódio trihidratado	0,1 M	6,5	Acetato de sódio trihidratado	1,4 M		
8	Citrato de sódio tribásico dihidratado	0,2 M	Cacodilato de sódio trihidratado	0,1 M	6,5	2-Propanol	30%		
9	Acetato de amônio	0,2 M	Citrato de sódio tribásico dihidratado	0,1 M	5,6	Poli(etilenoglicol) 4000	30%		
10	Acetato de amônio	0,2 M	Acetato de sódio trihidratado	0,1 M	4,6	Poli(etilenoglicol) 4000	30%		
11			Citrato de sódio tribásico dihidratado	0,1 M	5,6	Fosfato de amônio monobásico	1,0 M		
12	Cloreto de magnésio Hexahidratado	0,2 M	HEPES sódico	0,1 M	7,5	2-Propanol	30%		
13	Citrato de sódio tribásico dihidratado	0,2 M	Tris-HCl	0,1 M	8,5	Poli(etilenoglicol) 400	30%		
14	Cloreto de cálcio dihidratado	0,2 M	HEPES sódico	0,1 M	7,5	Poli(etilenoglicol) 400	28%		
15	Sulfato de amônio	0,2 M	Cacodilato de sódio trihidratado	0,1 M	6,5	Poli(etilenoglicol) 8000	30%		
16			HEPES sódico	0,1 M	7,5	Sulfato de lítio monohidratado	1,5 M		
17	Sulfato de lítio monohidratado	0,2 M	Tris-HCl	0,1 M	8,5	Poli(etilenoglicol) 4000	30%		
18	Acetato de amônio trihidratado	0,2 M	Cacodilato de sódio trihidratado	0,1 M	6,5	Poli(etilenoglicol) 8000	20%		

Continuação Tabela A1. Condições de cristalização utilizadas, referentes ao kit 1 da Hampton

19	Acetato de amônio	0,2 M	Tris-HCl	0,1 M	8,5	2-Propanol	30%	
20	Sulfato de amônio	0,2 M	Acetato de sódio trihidratado	0,1 M	4,6	Poli(etilenoglicol) 4000	25%	
21	Acetato de magnésio tetrahidratado	0,2 M	Caocilato de sódio trihidratado	0,1 M	6,5	(+/-)-2-metil-2,4-pentanodiol	30%	
22	Acetato de sódio trihidratado	0,2 M	Tris-HCl	0,1 M	8,5	Poli(etilenoglicol) 4000	30%	
23	Cloreto de magnésio Hexahidratado	0,2 M	HEPES sódico	0,1 M	7,5	Poli(etilenoglicol) 400	30%	
24	Cloreto de cálcio dihidratado	0,2 M	Acetato de sódio trihidratado	0,1 M	4,6	2-Propanol	20%	
25			Imidazol	0,1 M	6,5	Acetato de sódio trihidratado	1,0 M	
26	Acetato de amônio	0,2 M	Citrato de sódio tribásico dihidratado	0,1 M	5,6	(+/-)-2-metil-2,4-pentanodiol	30%	
27	Citrato de sódio tribásico dihidratado	0,2 M	HEPES sódico	0,1 M	7,5	2-Propanol	20%	
28	Acetato de sódio trihidratado	0,2 M	Caocilato de sódio trihidratado	0,1 M	6,5	Poli(etilenoglicol) 8000	30%	
29			HEPES sódico	0,1 M	7,5	Tartarato de sódio e potássio tetrahidratado	0,8 M	
30	Sulfato de amônio	0,2 M				Poli(etilenoglicol) 8000	30%	
31	Sulfato de amônio	0,2 M				Poli(etilenoglicol) 4000	30%	
32						Sulfato de amônio	2,0 M	
33						Formato de sódio	4,0 M	
34			Acetato de sódio trihidratado	0,1 M	4,6	Formato de sódio	2,0 M	
35			HEPES sódico	0,1 M	7,5	Fosfato de sódio monobásico monohidratado	0,8 M	Fosfato de potássio monobásico 0,8 M
36			Tris-HCl	0,1 M	8,5	Poli(etilenoglicol) 8000	8%	
37			Acetato de sódio trihidratado	0,1 M	4,6	Poli(etilenoglicol) 4000	8%	
38			HEPES sódico	0,1 M	7,5	Citrato de sódio tribásico dihidratado	1,4 M	
39			HEPES sódico	0,1 M	7,5	Poli(etilenoglicol) 400	2%	Sulfato de amônio 2,0 M
40			Citrato de sódio tribásico dihidratado	0,1 M	5,6	2-Propanol	20%	Poli(etilenoglicol) 4000 20%

Continuação Tabela A1. Condições de cristalização utilizadas, referentes ao kit 1 da Hampton

		HEPES sódico	0,1 M	7,5	2-Propanol	10%	Poli(etileno)glicol 4000	20%
41								
42	Fosfato de potássio monobásico				Poli(etileno)glicol 8000	20%		
43					Poli(etileno)glicol 1500	30%		
44					Formato de magnésio monohidratado	0,2 M		
45	Acetato de zinco dihidratado	Cacodilato de sódio trihidratado	0,1 M	6,5	Poli(etileno)glicol 8000	18%		
46	Acetato de cálcio hidratado	Cacodilato de sódio trihidratado	0,1 M	6,5	Poli(etileno)glicol 8000	18%		
47		Acetato de sódio trihidratado	0,1 M	4,6	Sulfato de amônio	2,0M		
48		Tris-HCl	0,1 M	8,5	Fosfato de amônio monobásico	2,0M		
49	Sulfato de lítio monohidratado				Poli(etileno)glicol 8000	2%		
50	Sulfato de lítio monohidratado				Poli(etileno)glicol 8000	15%		

Tabela A2. Condições de cristalização utilizadas, referentes ao kit 2 da Hampton

Condição	Sal	Conc.	Tampão	Conc.	pH	Precipitante 1	Conc.	Precipitante 2/Sal 2	Conc.
51	Cloreto de Sódio	2,0 M				Poli(etileno)glicol 6000	10%		
52	Cloreto de Sódio	0,5 M				Brometo de hexadeciltrimetilamônio	0,01 M	Cloreto de magnésio hexahidratado	0,01 M
53						Etileno)glicol	25%		
54						1,4-Dioxano	35%		

Continuação da Tabela A2. Condições de cristalização utilizadas, referentes ao kit 2 da Hampton

55	Sulfato de amônio	2,0 M					2-Propanol	5%	
56							Imidazol pH 7.0	1,0 M	
57							Poli(etilenoglicol) 1000	10%	Poli(etilenoglicol) 8000 10%
58	Cloreto de Sódio	1,5 M					Etanol	10%	
59						Acetato de sódio trihidratado	Cloreto de sódio	2,0 M	
60	Cloreto de Sódio	0,2 M			0,1 M	Acetato de sódio trihidratado	(+/-)-2-Metil-2,4-pentanodiol	30%	
61	Cloreto de cobalto(II) hexahidratado	0,01 M			0,1 M	Acetato de sódio trihidratado	1,6-Hexanodiol	1,0M	
62	Cloreto de cádmio hidratado	0,1 M			0,1 M	Acetato de sódio trihidratado	Poli(etilenoglicol) 400	30%	
63	Sulfato de amônio	0,2 M			0,1 M	Acetato de sódio trihidratado	Poli(etilenoglicol) monometiléter 200	30%	
64	Tartarato de sódio e potássio tetrahidratado	0,2 M			0,1 M	Citrato de sódio tribásico dihidratado	Sulfato de amônio	2,0 M	
65	Sulfato de amônio	0,5 M			0,1 M	Citrato de sódio tribásico dihidratado	Sulfato de lítio monohidratado	1,0 M	
66	Cloreto de Sódio	0,5 M			0,1 M	Citrato de sódio tribásico dihidratado	Polímero etileno-imina	2%	
67					0,1 M	Citrato de sódio tribásico dihidratado	<i>tert</i> -Butanol	35%	

Continuação da Tabela A2. Condições de cristalização utilizadas, referentes ao kit 2 da Hampton

69				Citrato de sódio tribásico dihidratado	0,1 M	5,6	1,6-Hexanodiol	2,5 M		
70				MES monohidratado	0,1 M	6,5	Sulfato de magnésio heptahidratado	1,6 M		
71	Fosfato de sódio monobásico monohidratado	0,1 M		MES monohidratado	0,1 M	6,5	Cloreto de sódio	2,0 M	Fosfato de potássio monobásico	0,1 M
72				MES monohidratado	0,1 M	6,5	Poli(etilenoglicol) 20000	12%		
73	Sulfato de amônio	1,6 M		MES monohidratado	0,1 M	6,5	1,4-Dioxano	10%		
74	Cloreto de céσιο	0,05 M		MES monohidratado	0,1 M	6,5	Jeffamina M-600	30%		
75	Cloreto de cobalto(II) hexahidratado	0,01M		MES monohidratado	0,1 M	6,5	Sulfato de amônio	1,8 M		
76	Sulfato de amônio	0,2 M		MES monohidratado	0,1 M	6,5	Poli(etilenoglicol) monometiléter 5000	30%		
77	Sulfato de zinco heptahidratado	0,01 M		MES monohidratado	0,1 M	6,5	Poli(etilenoglicol) monometiléter 550	25%		
78							Citrato de sódio tribásico dihidratado pH 6,5	1,6 M		
79	Sulfato de amônio	0,5 M		HEPES	0,1 M	7,5	(+/-)-2-Metil-2,4-pentanodiol	30%		
80				HEPES	0,1 M	7,5	Poli(etilenoglicol) 6000	10%	(+/-)-2-Metil-2,4-pentanodiol	5%

Continuação Tabela A2. Condições de cristalização utilizadas, referentes ao kit 2 da Hampton

81			HEPES	0,1 M	7,5	Jeffamina M-600	20%		
82	Cloreto de Sódio	0,1 M	HEPES	0,1 M	7,5	Sulfato de amônio	1,6 M		
83			HEPES	0,1 M	7,5	Sulfato de amônio	2,0 M		
84	Sulfato de cádmio hidratado	0,05 M	HEPES	0,1 M	7,5	Acetato de sódio trihidratado	1,0 M		
85			HEPES	0,1 M	7,5	(+/-)-2-Metil-2,4-pentanodiol	70%		0,8 M
86			HEPES	0,1 M	7,5	Cloreto de sódio	4,3 M		
87			HEPES	0,1 M	7,5	Poli(etilenoglicol) 8000	10%		Etilenoglicol 8%
88			HEPES	0,1 M	7,5	Poli(etilenoglicol) 10000	20%		
89	Cloreto de magnésio hexahidratado	0,2 M	Tris	0,1 M	8,5	1,6-Hexanodiol	3,4 M		
90			Tris	0,1 M	8,5	<i>tert</i> -Butanol	25%		
91	Cloreto de níquel(II) hexahidratado	0,01 M	Tris	0,1 M	8,5	Sulfato de lítio monohidratado	1,0 M		
92	Sulfato de amônio	1,5 M	Tris	0,1 M	8,5	Glicerol	12%		
93	Fosfato de amônio monobásico	0,2 M	Tris	0,1 M	8,5	(+/-)-2-Metil-2,4-pentanodiol	50%		
94			Tris	0,1 M	8,5	Etanol	20%		
95	Cloreto de níquel(II) hexahidratado	0,01 M	Tris	0,1 M	8,5	Poli(etilenoglicol) monometiléter 2000	20%		

Continuação Tabela A2. Condições de cristalização utilizadas, referentes ao kit 2 da Hampton

96	Cloreto de Sódio	0,1 M	Bicina	0,1 M	9,0	Poli(etileno)glicol monometiléter 550	20%		
97			Bicina	0,1 M	9,0	Cloreto de magnésio hexahidratado	2,0 M		
98			Bicina	0,1 M	9,0	1,4-Dioxano	2%	Poli(etileno)glicol 20000	10%

Tabela A3. Condições de cristalização utilizadas, diferentes dos kits da Hampton, denominadas exóticas

Condição	Sal	Conc.	Tampão	Conc.	pH	Precipitante 1	Conc.	Precipitante 2/Sal 2	Conc.
1	Sulfato de amônio	0,1 M	Glicina	0,1 M	3,8			Tri(etileno)glicol	25%
2	Cloreto de magnésio	0,2 M	HEPES	0,1 M	7,5	Ácido Poli(acrílico) 5100 Sal de sódio	22%		
3	N-óxido de trimetilamina	0,2 M	Tris	0,1 M	8,5	Ácido Poli(acrílico) 5100 Sal de sódio	22%		
4	Nitrato de ítrio	0,1 M				polivinilpirrolidona K15	5%	Cloreto de magnésio	0,005 M
6	cloreto de cobalto	0,005 M	HEPES	0,1 M	7,5	polivinilpirrolidona K15	20%	Cloreto de cádmio	0,005 M
7			D,L-Ácido Máltico	2,1 M	7			Cloreto de níquel	0,005 M
10	Cloreto de cálcio	0,03 M	L-Ácido Máltico	0,05 M	5,0	Poli(etileno)glicol 6000	20%		
11	Malonato de sódio	1,1 M	HEPES	0,1 M	7	D,L-Ácido Máltico	2,1 M		
12	Cloreto de sódio	0,35 M	Tricina	0,1 M	8	Poli(etileno)glicol 1000	28%	Glicerol	10%
13			Tricina	0,05 M	8	Poli(etileno)glicol 4000	22%		

Continuação da Tabela A3. Condições de cristalização utilizadas, diferentes dos kits da Hampton, denominadas

14	Sulfato de lítio	0,05 M	Tricina	0,1 M	7,4	Polietilenoglicol 3000	7%		
15	Cloreto de sódio	0,35 M	Tricina	0,1 M	8	Polietilenoglicol 1000	31%		
16	Cloreto de amônio	0,2 M	Tricina	0,15 M	8	Polietilenoglicol 6000	15%		
17			Bicina	0,1 M	9	Polietilenoglicol 20000	10%	1,4-Dioxano	2%
18			Bicina	0,1 M	9,5	Polietilenoglicol 10000	10%	1,4-Dioxano	2%
20			Bicina	0,1 M	9	Jeffamina M-600	20%		
21	Cloreto de magnésio	0,3 M	Bicina	0,1 M	9	Polietilenoglicol 2000	28%		
22	Sulfato de zinco	0,8 M	Bicina	0,1 M	8,5				
23	Cloreto de Lítio	1,0 M	Bicina	0,1 M	9	Polietilenoglicol 6000	20%		
26			Ácido Succínico	0,8 M	7				
27			HEPES	0,1 M	7	Polietilenoglicol monometiléter 2000	1%	Ácido succínico	1,0 M

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 hexameric 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

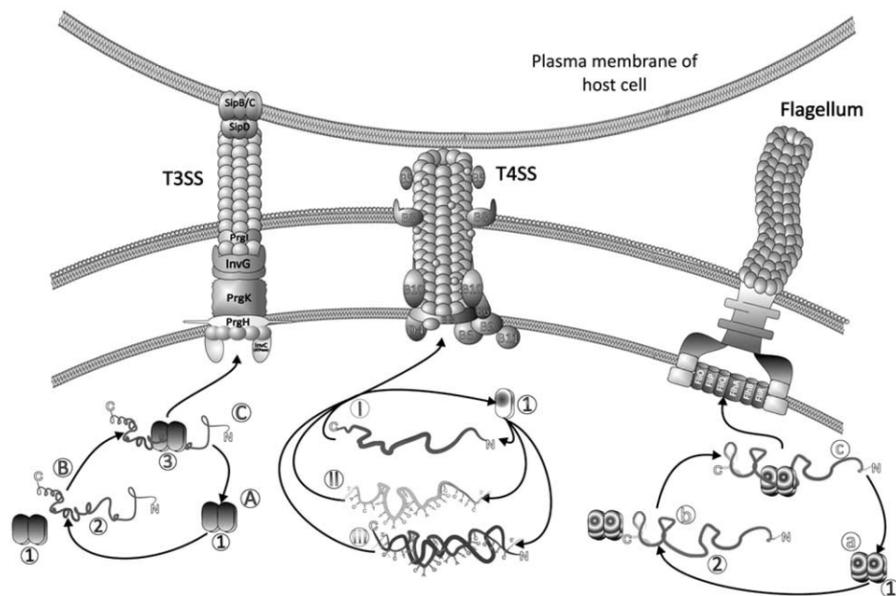


Figure 1. Illustration of bacterial secretion systems: T3SS (left), T4SS (middle), and flagellar (right). In T3 system secretion, three steps are involved: (A) the chaperone dimer (1) is free in the cytosol, (B) the dimer encounters its effector protein (2), and (C) the chaperone/effector complex (3) 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 infection 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 FlhS) [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-

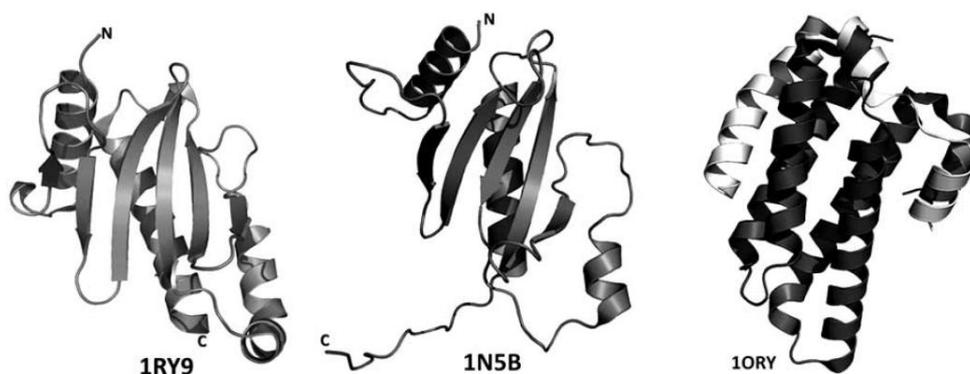


Figure 2. Tertiary structure representation of (A) two T3SS (Protein Data Bank entries: 1RY9 and 1N5B, PyMol) chaperones, from *Shigella flexneri* (Spa15: 1RY9) and *Yersinia enterocolitica* (SycE: 1N5B), respectively. Secondary structure similarities can be clearly observed, since both show similar 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 2.5 Å 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.

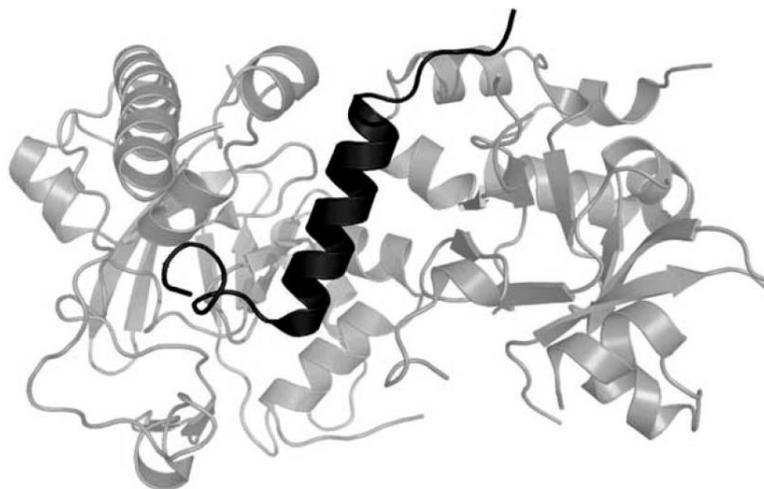


Figure 3. Tertiary structure of VirE1/VirE2 complex from *Agrobacterium tumefaciens* T4SS, with small VirE1 chaperone (dark grey) preventing VirE2 (shade grey) aggregation and premature precipitation (Protein Data Bank entry: 3BTP).

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 injectosome. *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-

Bacterial Secretion Chaperones

Protein & Peptide Letters, 2011, Vol. 18, No. 2 165

- 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**, *72*, 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.; Brummell, 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 FljI 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 FlgN 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. Pur.*, **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.

166 *Protein & Peptide Letters*, 2011, Vol. 18, No. 2

Fattori et al.

- [61] Chou, J.; Delaglio, F.; Bax, A. Measurement of one-bond ^{15}N - ^{13}C 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.; Mukerjee, 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.

Protein J (2011) 30:324–333
DOI 10.1007/s10930-011-9335-z

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 CesA 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 $\alpha\beta$ 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_{600nm} 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 centrifugated 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: $\epsilon_{280nm} = 6,990 \text{ L mol}^{-1} \text{ cm}^{-1}$ [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 PDBFold [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.

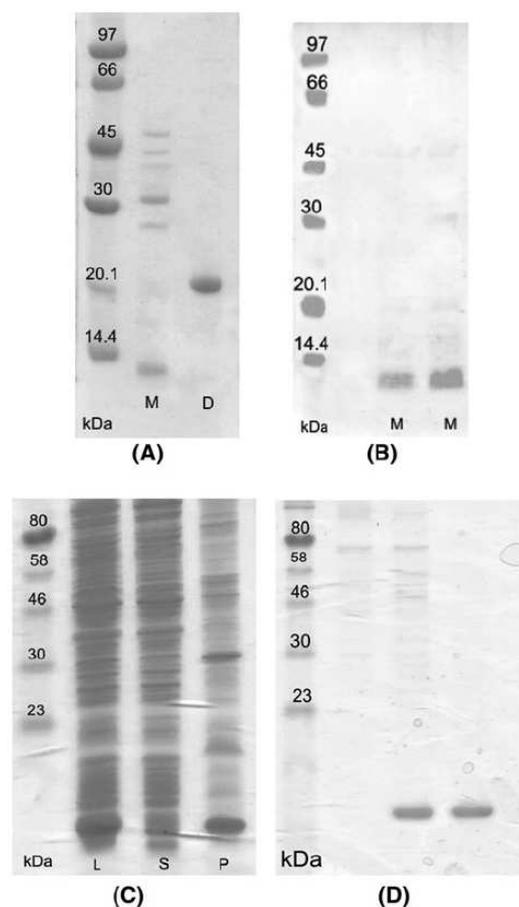


Fig. 1 SDS-PAGE electrophoresis gels. **a** XAC0419 after first purification procedure showing the pure dimer (D) and monomer (M) fractions; **b** XAC0419 after second purification procedure showing the pure monomer (M); **c** FlgN after lysis with samples showing L—total lysate, S—supernatant of lysis 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).

328

J. Fattori et al.

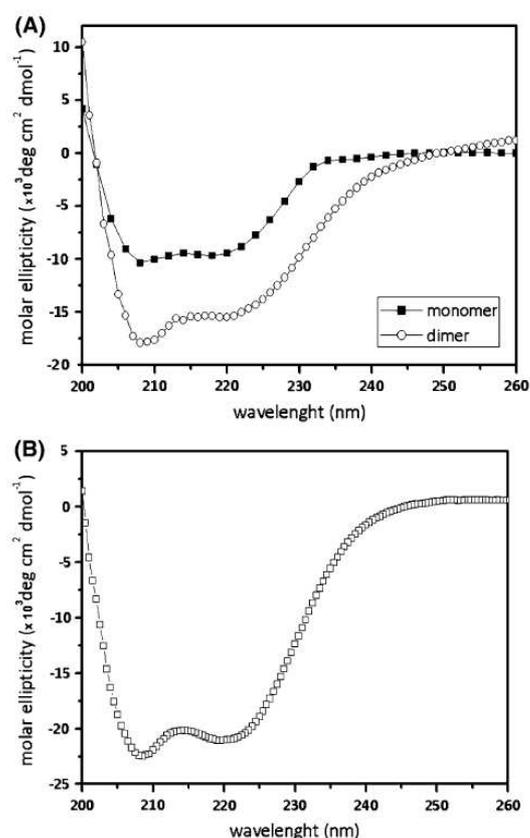


Fig. 2 Spectroscopic data. **a** CD spectra of the XAC0419 at $4 \mu\text{mol L}^{-1}$ in both forms: (open circle) dimeric and (filled square) monomeric; **b** CD spectra of the FlgN at $4 \mu\text{mol L}^{-1}$ (open square). All of them recorded as described in Materials and methods

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.

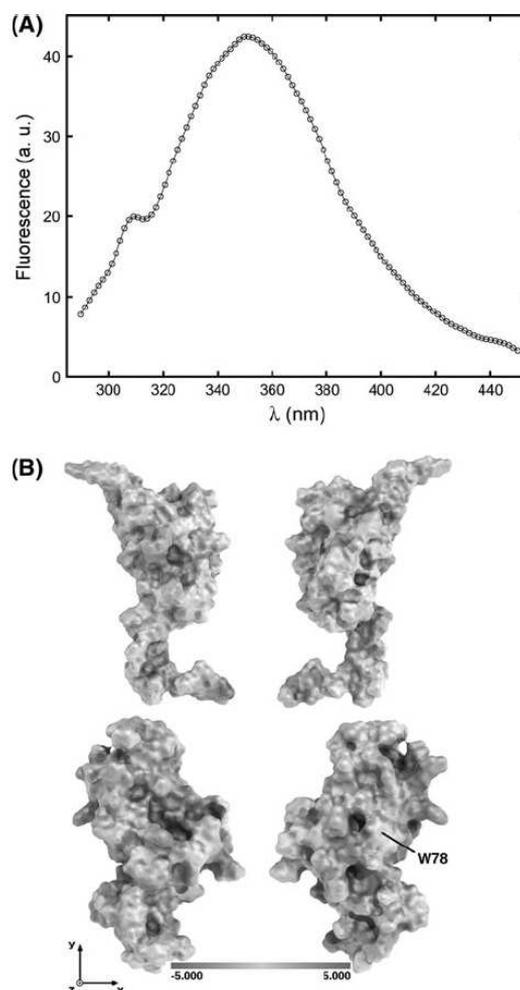


Fig. 3 Spectroscopic and bioinformatics data for target proteins. **a** Emitted fluorescence spectrum of FlgN measured at 25 °C. **b** The electrostatic surface potentials (310 K) for the predicted models of XAC0419 (top row) and FlgN (bottom row). The right column is rotated clockwise by 180° around 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)

R_g 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

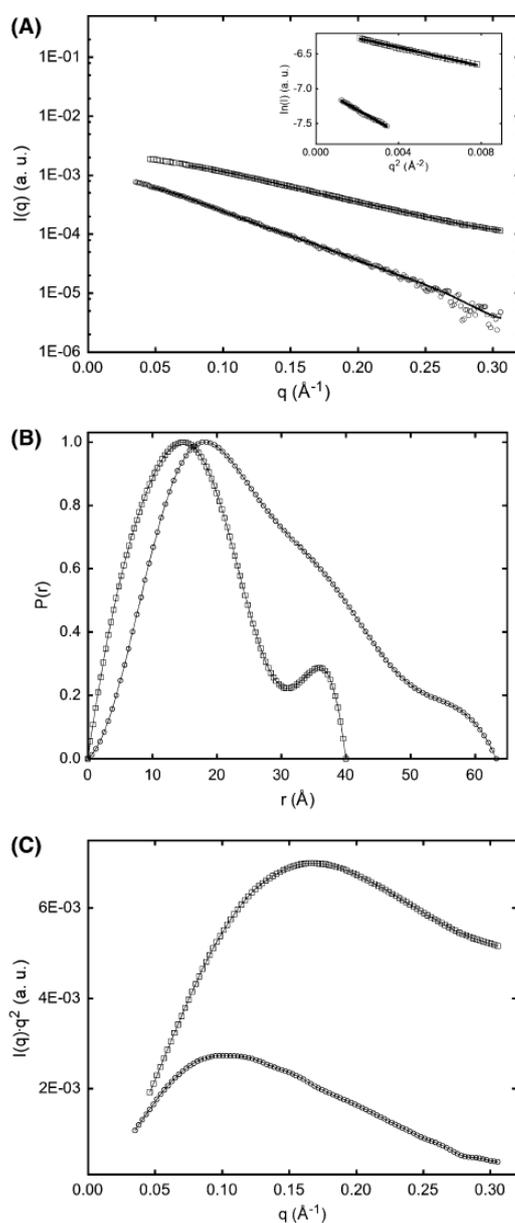


Fig. 4 SAXS results. **a** SAXS scattering curves corresponding to the average of 3 consecutive frames of XAC0419 (*open square*) and FlgN (*open circle*). In each case, the solid line represents the curve fitting corresponding to the distance distribution function obtained with AUTOGNOM [36]. The *inset* 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

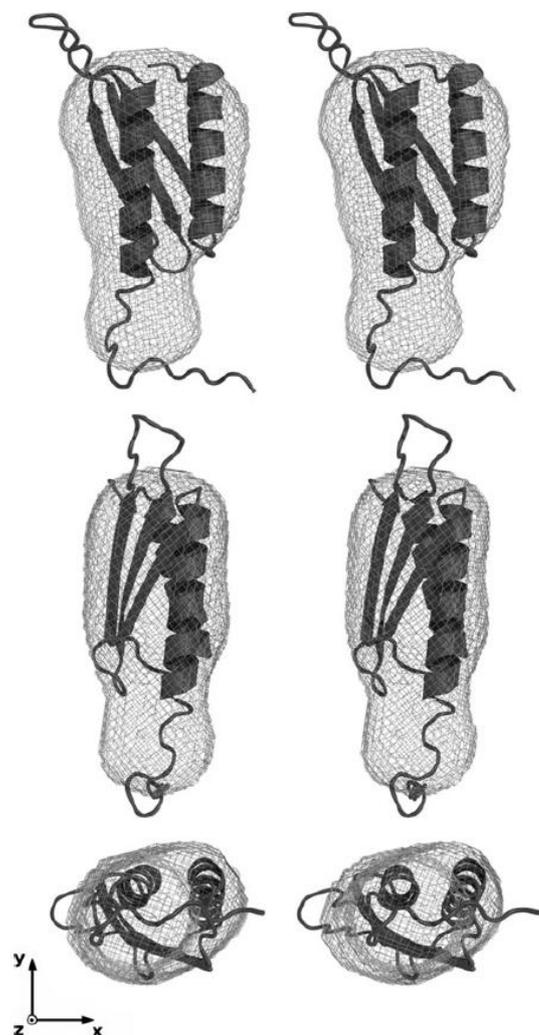


Fig. 5 Stereo views showing the superposition of the molecular envelope of XAC0419 derived from the experimental curves (*in light gray*) superposed onto a cartoon representation of the theoretical atomic model (*in dark gray*) obtained by computational modeling. The middle and bottom rows are rotated 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

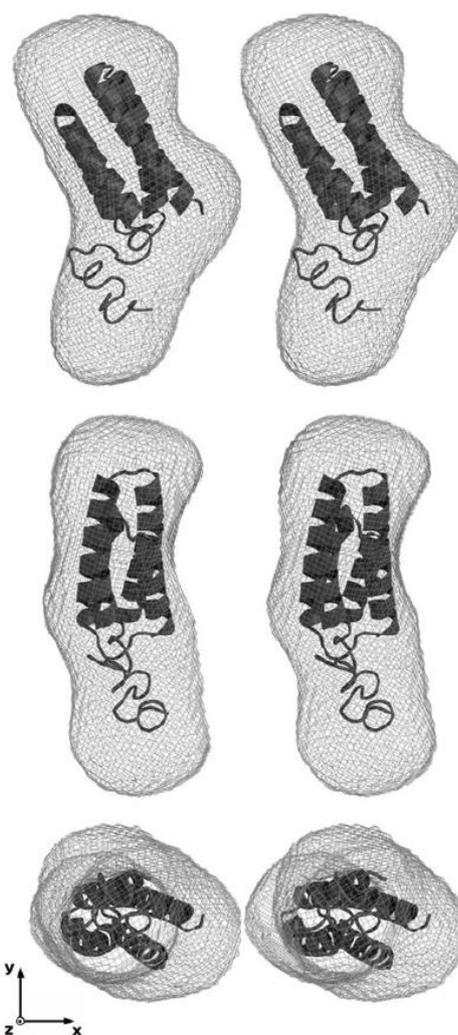


Fig. 6 Stereo view of the molecular envelope of FlgN recovered from the scattering curve (*in light gray*) superposed onto a cartoon representation of the atomic model predicted by computational modeling techniques (*in dark gray*). The middle and bottom rows are rotated clockwise by 90° around the y-axes 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

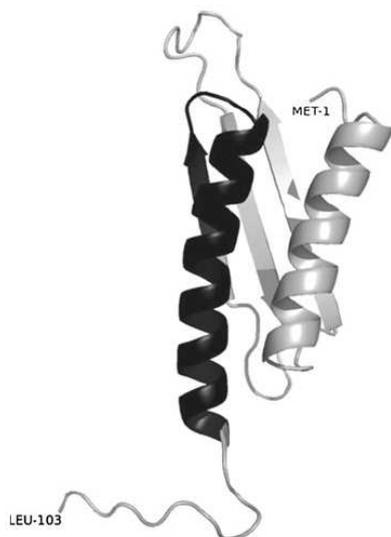


Fig. 7 Cartoon representation of the atomic model for XAC0419 as predicted by I-TASSER. This is the second scored model (C-score -1.36) drawn using PyMOL with N- and C-terminal residues marked (Met-1 and Leu-103), and the predict amphipathic helix (67–91) showed in dark gray

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

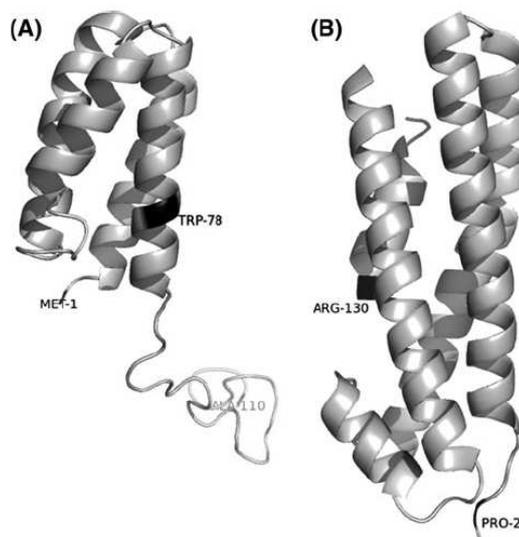


Fig. 8 Cartoon representation of the atomic model predicted for FlgN (I-TASSER) and PDB structure of the FlgN from *Pseudomonas aeruginosa*. **a** This is the third scored model with I-TASSER C-score of -2.66 and drawn using PyMOL with N- and C-terminal marked (Met-1 and Ala-110), and showing once again 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 FliT 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 investigated.

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

1. Alegria MC, Souza DP, Andrade MO, Docena C, Khater L, Ramos CHI, da Silva AC, Farah CS (2005) *J Bact* 187: 2315–2325
2. Khater L, Santos T, Alegria MC, Docena C, da Silva AC, Ramos CHI (2005) *Genet Mol Biol* 28:321–327
3. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402
4. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG (2008) *Nucleic Acids Res* 36:D419–D425
5. Aparicio R, Fischer H, Scott DJ, Verschueren KH, Kulminskaya AA, Eneiskaya EV, Neustroev KN, Golubev AM, Polikarpov I (2002) *Biochemistry* 41:9370–9375
6. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) *Proc Natl Acad Sci USA* 98:10037–10041
7. Böhm G, Muhr R, Jaenicke R (1992) *Protein Eng* 5:191–195
8. Bradford MM (1976) *Anal Biochem* 72:248–254
9. Christie PJ (1997) *J Bacteriol* 179:3085–3094
10. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr Sect D: Biol Crystallogr* 50:760–763
11. Cooper CA, Zhang K, Andres SN, Fang Y, Kaniuk NA, Hannemann M, Brumell JH, Foster LJ, Junop MS, Coombes BK (2010) *PLoS Pathog* 6:e1000751
12. Cornelis GR (2006) *Nat Rev Microbiol* 4:811–825
13. Cornelis GR, Van Gijsegem F (2000) *Annu Rev Microbiol* 54:735–774
14. 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, Chambergio 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, Comelis 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