



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**

DIEGO JOSÉ PIMENTA DE CARVALHO

**UMA PERSPECTIVA DINÂMICA SOBRE OS DADOS DE LIGAÇÃO CRUZADA
XL-MS**

CAMPINAS

2022

DIEGO JOSÉ PIMENTA DE CARVALHO

**UMA PERSPECTIVA DINÂMICA SOBRE OS DADOS DE LIGAÇÃO CRUZADA
XL-MS**

**Dissertação de Mestrado apresentada ao Instituto de
Química da Universidade Estadual de Campinas como
parte dos requisitos exigidos para a obtenção do título
de Mestre em Química na área de Físico-Química**

Orientador: Prof. Dr. Leandro Martínez

**O arquivo digital corresponde à versão final da Dissertação defendida pelo aluno Diego
José Pimenta de Carvalho e orientada pelo Prof. Dr. Leandro Martínez.**

CAMPINAS

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Química
Simone Luiz Alves - CRB 8/9094

Carvalho, Diego José Pimenta de, 1995-
C253p Uma perspectiva dinâmica sobre os dados de ligação cruzada XL-MS /
Diego José Pimenta de Carvalho. – Campinas, SP : [s.n.], 2022.

Orientador: Leandro Martínez.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de
Química.

1. Cross-linking. 2. Dinâmica molecular. 3. Proteínas. I. Martínez, Leandro,
1979-. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: A dynamic perspective on cross-linking data XL-MS

Palavras-chave em inglês:

Cross-linking

Molecular dynamics

Proteins

Área de concentração: Físico-Química

Titulação: Mestre em Química na área de Físico-Química

Banca examinadora:

Leandro Martínez [Orientador]

Munir Salomão Skaf

Paulo Costa Carvalho

Data de defesa: 14-03-2022

Programa de Pós-Graduação: Química

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-3852-3806>

- Currículo Lattes do autor: <http://lattes.cnpq.br/5360775451172074>

BANCA EXAMINADORA

Prof. Dr. Leandro Martínez (Orientador)

Prof. Dr. Munir Salomão Skaf (Instituto de Química - UNICAMP)

Dr. Paulo Costa Carvalho (Fundação Oswaldo Cruz do Paraná)

A Ata da defesa assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Este exemplar corresponde à redação final da Dissertação de Mestrado defendida pelo aluno Diego José Pimenta de Carvalho, aprovada pela Comissão Julgadora em 14/03/2022.

*“ E disse Deus: Haja luz; e houve luz.
E viu Deus que era boa a luz;
e fez Deus separação entre a luz e as trevas.”*
(Gênesis 1:3,4)

*“ Somente um principiante que não sabe nada sobre ciência
diria que a ciência descarta a fé.
Se você realmente estudar a ciência,
ela certamente o levará para mais perto de Deus.”*

(James Clerk Maxwell)

*“ E agora, José?
A festa acabou, a luz apagou,
o povo sumiu, a noite esfriou,
e agora José?”*

(Carlos Drummond De Andrade)

*“ ...Deus me proteja de mim e da maldade de gente boa
Da bondade da pessoa ruim
Deus me governe e guarde ilumine e zele assim
Caminho se conhece andando
Então vez em quando é bom se perder
Perdido fica perguntando
Vai só procurando
E acha sem saber
Perigo é se encontrar perdido
Deixar sem ter sido
Não olhar, não ver
Bom mesmo é ter sexto sentido
Sair distraído espalhar bem-querer...”*
(Chico César)

Dedico este trabalho aos meus sobrinhos Catherine, James e Zoé que apesar de pequenos possam se entusiasmar com a ciência e possivelmente contribuir com a mesma quando maiores. A dedicação também se estende a todos aqueles que um dia eu possa influenciar para a ciência. Não teremos futuro sem ciência, ela é a melhor ferramenta que temos para assegurar um futuro melhor.

Agradecimentos

Agradeço primeiramente a Deus. Crendo que é provedor de todas as coisas, a primeira menção de gratidão será sempre destinada a Ele.

Em segundo lugar, todo meu reconhecimento e gratidão para o Prof. Dr. Leandro Martínez que me apoiou, orientou e me incentivou ao longo do caminho deste projeto. O professor Leandro é um acadêmico de excelência e não mede esforços para incentivar e ensinar aqueles que o permitem compartilhar seus conhecimentos. Por ele cultivo grande admiração e o tenho como exemplo a ser seguido.

Estendo minha gratidão também aos meus colegas de grupo Alexandre, Maurício, Eduard, Felipe, Pamela, Ander, Brenda, Állan, Guilherme e Vinícius. A eles sou grato pelo regime de cooperação e críticas construtivas que desenvolvemos em todas as reuniões e nos períodos presenciais também. Alguns destes, os últimos listados, são aqueles com quem mais tempo convivi e compartilhei perrengues em congressos e no desenvolvimento de experimentos computacionais, e por eles uma maior porção de gratidão. Totalmente relevante mencionar o nome do Manuel, com quem dividi espaço de trabalho antes da pandemia nos atingir, que apesar de não ser do nosso grupo M3G, ele me auxiliou em aspectos científicos muito pertinentes em que o conhecimento dele era mais vasto do que o meu.

Indispensável mencionar as organizações e instituições que fomentaram ou forneceram infraestrutura para que eu pudesse desenvolver os resultados presentes neste documento. A estas sou grato pelo apoio financeiro que permitiu que eu tivesse acesso a congressos científicos e tempo dedicação exclusiva à pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 (Proc. 88882.329147/2019-01).

Obrigado à FAPESP que promoveu apoio financeiro ao CCES (Center for

Computing in Engineering and Sciences) (Proc. 2010/16947-9, 2013/05475-7, 2013/08293-7). Foi através da estrutura do CCES que foram realizados os procedimentos computacionais deste trabalho. A FAPESP também auxilia financeiramente o nosso grupo de pesquisa M3G (2018/24293-0, 2018/14274-9), sendo assim estendo minha gratidão a este apoio no qual eu usufruí através da infraestrutura do grupo e de recursos para a participação de congressos.

Obrigado ao CNPq pelo apoio financeiro ao grupo M3G (Proc. 302332/2016-2), sem tal apoio a produtividade do grupo não seria a mesma.

Agradeço também a comunidade que desenvolve programas e os compartilham de forma gratuita. Este trabalho não seria possível sem as contribuições desta comunidade.

Não poderia deixar de mencionar os elementos presentes na mistura heterogênea que é a minha vida que fizeram com que eu chegasse até aqui e concluisse esta etapa, mesmo não tendo atuação direta sobre este trabalho.

Muito obrigado ₅₂Te (Túlia Estanislau) por acreditar em mim, me incentivar e prover tudo que eu necessitei ao longo de todos esses anos.

Muito obrigado ₆₈Er (Elise Renata) por ser meu exemplo, me ajudar em momentos difíceis e pela família maravilhosa que você construiu.

Muito obrigado ₁H (Heyk) por ser a influencia cultural alternativa que sempre fez mais humano o menino de exatas que sempre fui.

Muito obrigado a família ₃₈Sr por sempre me acolherem e me tratar como um de vocês.

Muito obrigado a família ₄₂Mo por me proporcionarem momentos familiares como se eu fizesse parte de vocês.

Muito obrigado ₇₁Lu por me impactar tanto, sem você eu não seria a pessoa que sou hoje e, apesar dos poucos, vou me lembrar de cada momento que tivemos.

Muito obrigado ₉₀Th por me fazer acreditar em mim, por ser uma pessoa tão amável e por me fazer uma pessoa melhor.

Obrigado aos meus amigos de colégio, principalmente a ^{105}Db e sua família que me acolheram por tanto tempo em sua casa, vocês são as pessoas que trazem a nostalgia que me move nos momentos mais difíceis.

Obrigado ^{74}W e ^{23}V que mesmo depois de tanto tempo, somos próximos como se ainda estivéssemos no ensino médio.

Obrigados aos meus amigos de jogatina e aos de passeios improvisados todos vocês fazem com que a vida se torne mais leve e menos cinza.

Obrigado ^{20}Ca , ^{3}Li e ^{8}O que sempre me apoiaram profissionalmente.

UNICAMP

Resumo

Instituto de Química
Departamento de Físico-Química

Mestre em Ciências

Uma perspectiva dinâmica sobre os dados de ligação cruzada XL-MS

por Diego PIMENTA

O experimento de ligação cruzada XL-MS consiste na promoção de ligações interresiduais e suas identificações em uma proteína. Mediante o conhecimento do agente de ligação cruzada utilizado, informações relevantes sobre a estrutura da proteína podem ser obtidas. Neste estudo analisaram-se os dados experimentais de ligação cruzada XL-MS e as estruturas obtidas por simulação para a SalBIII e para a Calmodulina. A compatibilidade dos cross-links com as estruturas da simulação em função do parâmetro de qualidade de espectros nos mostrou uma correlação positiva, indicando o parâmetro de score como mais significativo para a seleção de cross-links compatíveis tanto com a estrutura cristalográfica quanto com as estruturas da simulação. Na análise de compatibilidade das ligações cruzadas com estruturas da simulação mostrou-se importante o estudo das distâncias topológicas e euclidianas em conjunto, sendo ambas relevantes e pertinentes em muitos dos casos onde os resultados fugiam do esperado. No caso da Calmodulina, as simulações mostraram que a mudança de conformação da proteína, mediante a complexação com cálcio, restringe a compatibilidade da estrutura com cross-links relacionados a conexão dos domínios afastados pela formação da α -hélice. Este estudo permitiu uma maior compreensão da abrangência dos resultados e das consequências das condições que expomos a proteína no experimento de XL-MS.

UNICAMP

Abstract

Chemistry Institute
Physical-Chemistry Department

Master in Science

A dynamic perspective on cross-linking data XL-MS

by Diego PIMENTA

The XL-MS cross-linking experiment consists of the promotion of interresidual bonds and their identification in a protein. By knowing the cross-linking agent used, relevant information about the structure of the protein can be obtained. In this study, experimental crosslinking data and structures obtained by simulation for SalBIII and Calmodulin were analyzed. The compatibility of the cross-links with the simulation structures as a function of the spectra quality parameters showed us a positive correlation, indicating the score parameter as most significant for the selection of cross-links compatible with both the crystallographic structure and the simulation structures. In the crosslink compatibility analysis with simulation structures it was shown to be important to study the topological and Euclidean distances together, both being relevant and pertinent in many of the cases where the results deviated from what was expected. In the case of Calmodulin, the simulations showed that the change in conformation of the protein, upon complexation with calcium, restricts the compatibility of the structure with cross-links related to the connection of domains separated by the formation of the α -helix. This study provides a greater understanding of the scope of the results and the consequences of the conditions we expose the protein to in the XL-MS experiment.

Figuras

1.1	Representação de ligação cruzada produzida pela ação do linker DSS no experimento de Ligação Cruzada em Espectrometria de Massa (XL-MS) para a SalBIII, reproduzida no software PyMol.	28
1.2	Estrutura geral de um derivado de NHS exemplificando Suberato de Disuccinimida (DSS) e Suberato de bis(sulfosuccinimida) (BS3). Ambos com o comprimento de 11.4 Å.	29
1.3	Representação estrutural de diferentes grupos de Agente de Ligação Cruzadas (ALCs). (a) Ácido ligantes. (b) Fotoativável.	29
1.4	Representação estrutural de diferentes grupos de ALCs cliváveis.	30
2.1	Esquema visual do integrador leap-frog.	36
2.2	Representação da descrição coarse-grained em uma molécula proteica. Na metade inferior esquerda da figura são ilustrados todos os átomos dessa metade da estrutura, utilizando a representação Licorice. Já na metade superior, as partículas que são ilustradas são aquelas que substituíram os grupos de átomos.	43
3.1	Representação das estruturas para a Calmodulina livre - PDB 1DMO. . .	51
4.1	Representação dos cross-links compatíveis com a estrutura cristalográfica da SalBIII.	54
4.2	Quantidade de cross-links compatíveis com estruturas presentes na simulação em função do tempo de simulação - <i>All-atom</i>	54

4.3	Probabilidade de Validação em uma Estrutura da Simulação dos cross-links compatíveis com estruturas presentes na simulação - <i>All-atom</i> . – Em verde são destacados os cross-links compatíveis com a estrutura cristalográfica.	55
4.4	Representação de cross-links na estrutura cristalográfica - SalBIII. (a) M1–K8, M1–K17, M1–K113 e E13–E111 em ciano, magenta, limão e amarelo, respectivamente. (b) E32–K113, E37–E46, E53–57 e E62–E86 em oliva, musgo, azul petróleo e bege, respectivamente.	56
4.5	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links da Figura 4.4. - A referência pontilhada é o limite estatístico.	57
4.6	Distância Excedida Euclidiana, Distância Excedida Topológica e Probabilidade de Validação em uma Estrutura da Simulação (PVES) dos cross-links compatíveis com estruturas presentes na simulação mas não compatíveis com a estrutura cristalográfica - <i>All-atom</i> . – Em vinho, amarelo e azul, respectivamente. Somente a PVES tem correspondência com o eixo y direito, as demais grandezas se relacionam com o da esquerda.	60
4.7	Representação das estruturas secundárias da SalBIII ao longo da sequencia de aminoácidos - Sendo α -hélice e folha- β representadas em vermelho e amarelo, respectivamente. (Fonte: Protein Data Bank, 5CXO)	60
4.8	Representação de cross-links não validados na estrutura cristalgráfica mas compatíveis com estruturas da simulação - SalBIII. (a) K6–E13, E9–E28 e E9–E66 em gelo, magenta e verde limão, respectivamente. (b) K42–S133, S57–D125, e K99–S133 em verde floresta, azul piscina e verde, respectivamente.	61
4.9	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links da Figura 4.8. - A referência pontilhada é o limite estatístico.	62

4.10	Probabilidade de Validação em uma Estrutura da Simulação convertida para unidades de energia livre para cada cross-link - <i>All-atom</i> . – Em verde são destacados os cross-links validados na estrutura cristalográfica.	63
4.11	Probabilidade de Validação em uma Estrutura da Simulação dos cross-links compatíveis com estruturas presentes na simulação - <i>Coarse-grained</i> . – Em verde são destacados os cross-links compatíveis com a estrutura cristalográfica.	64
4.12	Distância Excedida Topológica dos cross-links compatíveis com estruturas presentes na simulação <i>all-atom</i> e <i>coarse-grained</i> , em amarelo e vermelho, respectivamente.	65
4.13	Diagrama de coordenadas dos cross-links validados em simulação. – Em azul são representados os cross-links comuns entre as simulações dos dois campos de força, em vermelho são os cross-links validados apenas nas simulações <i>coarse-grained</i> , em magenta aqueles validados apenas na simulação <i>all-atom</i> e em cinza o par de resíduos que possui distância igual ou menor que 8.0 Å na estrutura cristalográfica (mapa de contatos).	66
4.14	Correlações entre os parâmetros de qualidade de espectros e a compatibilidade dos conjuntos gerados (A-C) com a estrutura cristalográfica (I-III) com as estruturas da simulação <i>all-atom</i> . – O quadrado dos parâmetros de correlação (R^2) para as distribuições de pontos são 0.98, 0.95, 0.95, 0.88, 0.91 e 0.97 para os gráficos A, B, C, I, II e III, respectivamente.	68
4.15	Fração de Estruturas Compatíveis para cada cross-link validado nas estruturas de referência, Calmodulina livre (30 conformações).	72
4.16	Representação de cross-links nas estruturas de referência - Calmodulina livre. (a) Conformação 7 - Destacado em azul o cross-link S82-K116, amarelo o K78-K116 e verde o K95-S102. (b) Conformação 25 - Destacado em azul o cross-link M1-K95, amarelo o K22-K31, verde o S18-K22, vermelho o K76-K78 e magenta o K78-S82. (c) Conformação 29 - Destacado em azul o cross-link M1-K76, amarelo o M1-K78 e verde o K76-K116.	73

4.17	PVES para cada cross-link compatível com as estruturas das simulações <i>all-atom</i> da Calmodulina livre.	74
4.18	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links compatíveis com as estruturas de referência. - Calmodulina livre, <i>all-atom</i> . A referência pontilhada é o limite estatístico.	75
4.19	Representação de cross-links na estrutura de referência - Calmodulina complexada com cálcio. (a) Destacado em azul o cross-link M1–K78, em amarelo o K22–K31, em verde o K95–S102 e em vermelho o S18–K22. (b) Destacado em azul o cross-link K76–K78 e em amarelo o K78–S82.	76
4.20	PVES para cada cross-link compatível com as estruturas das simulações <i>all-atom</i> da Calmodulina complexada com cálcio.	76
4.21	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links compatíveis com a estrutura cristalográfica. - Calmodulina complexada com cálcio, <i>all-atom</i> . A referência pontilhada é o limite estatístico.	77
4.22	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links S18–K116 e K22–S102. - Calmodulina livre, <i>all-atom</i> . A referência pontilhada é o limite estatístico.	79
4.23	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links K14–S39, K14–K76, K14–K78 e S39–K78. - Calmodulina complexada com cálcio, <i>all-atom</i> . A referência pontilhada é o limite estatístico.	80
4.24	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links M1–K76 e M1–K78. - Calmodulina complexada com cálcio, <i>all-atom</i> . A referência pontilhada é o limite estatístico.	81
4.25	PVES para cada cross-link compatível com as estruturas das simulações <i>coarse-grained</i> da Calmodulina complexada com cálcio.	82

4.26	Representação do comportamento dos clusters identificados pelo software QuinXL. Onde a condição biológica 1 é a ausência de íons cálcio e a 2 a presença de íons cálcio. Os Clusters 1 e 2 são representados pelas cores vermelho e azul, respectivamente.	84
4.27	Representação dos cross-links na sequência da proteína coloridos em função do cluster identificado. Os Clusters 1 e 2 são representados pelas cores vermelho e azul, respectivamente. Destaque em azul para a parte da sequência na qual é formada a α -hélice inter-domínio.	84
A.1	Gráfico da energia livre da conformação pelo tempo de simulação com o campo de força SIRAH - 1 μ s	89
A.2	Gráfico da energia livre da conformação pelo tempo de simulação com o campo de força Martini	90
A.3	Desvio quadrático médio (RMSD) - 1 μ s, Amber	91
A.4	Desvio quadrático médio (RMSD) - 1 μ s, Martini (A) SIRAH (B)	92
A.5	Flutuação quadrática média (RMSF) dos carbonos- α dos três campos de força	93
A.6	Flutuação quadrática média (RMSF) dos carbonos- α em comparação a troca de hidrogênio-deutério (HDX) e ao B-fator - Amber.	93
A.7	RMSD em função da fração selecionada da proteína	95
A.8	Desvio quadrático médio com $\phi = 1$ e $\phi = 0.7$	95
A.9	Sobreposição de conformações da simulação <i>all-atom</i> de 1 μ s	96
II.1	Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links dà Figura 4.4, a referênciа pontilhada é o limite estatístico.	101
II.2	Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links dà Figura 4.8, a referênciа pontilhada é o limite estatístico.	102

II.3	Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links dà Figura 4.6, a referênciapontilhada é o limite estatístico.	103
II.4	Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links compatíveis com a estrura cristalográfica.- , <i>all-atom</i> . A referênciapontilhada é o limite estatístico.	104
II.5	Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links M1–K76 e M1–K78.- , <i>all-atom</i> . A referênciapontilhada é o limite estatístico.	104

Tabelas

4.1	Ligações cruzadas validadas na estrutura cristalográfica - SalBIII	54
4.2	Ligações cruzadas não validadas na estrutura cristalográfica mas compatíveis com estruturas encontradas na simulação <i>all-atom</i> - SalBIII.	59
4.3	Cross-links experimentais não validados em simulações	67
4.4	Cross-links experimentais para Calmodulina livre (exclusivamente), Calmodulina complexada com cálcio (exclusivamente) e comuns entre os experimentos.	70
4.5	Cross-links validados nas respectivas estruturas de referência para a Calmodulina livre (30 conformações de RMN) e Calmodulina complexada com cálcio (1 conformação cristalográfica)	71
4.6	Cross-links compatíveis com estruturas das simulações <i>all-atom</i> mas não validados nas estruturas de referência para a Calmodulina livre (30 estruturas de RMN) e Calmodulina complexada com cálcio (1 estrutura cristalográfica)	78
4.7	Cross-links do conjunto experimental não compatíveis com as estruturas de referência nem com as das simulações para a Calmodulina complexada com cálcio.	83
4.8	Cross-links compatíveis com as estruturas das simulações <i>all-atom</i> comuns entre Calmodulina livre e Calmodulina complexada com cálcio agrupados pelos clusters em que eles se encontram na identificação do software QuinXL.	85

Abreviações

ALC	Agente de Ligação Cruzada.
BS3	Suberato de bis(sulfosuccinimida).
CDRX	Cristalografia por Difração de Raios X.
CG	Cromatografia Gasosa.
Crio-ME	Crio-Microscopia Eletrônica.
DEE	Distância Excedida Euclidiana.
DET	Distância Excedida Topológica.
DM	Dinâmica Molecular.
DSS	Suberato de Disuccinimida.
EDC	1-etil-3(3-(dimetilamino)propil)carbodiimida.
ESI	Ionização por Electrospray.
FEC	Fração de Estruturas Compatíveis.
HDX	Troca de Hidrogênio-Deutério.
HOBt	Hidroxibenzotriazol.
HPLC	Cromatografia Líquida de Alta Eficiência.
IM	Mobilidade Iônica.
MALDI	Ionização e Dessorção a Laser Assistida por Matriz.
NHS	N-hidroxisuccinimida.
PME	Particle Mesh Ewald.
PVES	Probabilidade de Validação em uma Estrutura da Simulação.
QqTOF	Duplo Quadrupolo com Tempo de Voo.

RMN Ressonância Magnética Nuclear.

XL-MS Ligação Cruzada em Espectrometria de Massa.

Constantes Físicas

Constante Universal dos Gases $R = 8.315 \text{ J mol}^{-1} \text{ K}^{-1}$

Permissividade do vácuo $\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$

Constante de Boltzmann $k = 1.381 \times 10^{-23} \text{ J K}^{-1}$

Símbolos

ϵ_i	Energia de estado	J
δ	Função Dirac	
G_σ	Função Gaussiana	

Sumário

Agradecimentos

Resumo

Abstract

1 Cross-linking	26
1.1 O Equipamento	26
1.2 O Experimento	27
1.2.1 Agente de Ligação Cruzada	28
XPlex	30
Limite estatístico de validação	30
1.3 O Resultado	31
2 Dinâmica Molecular	33
2.1 Os Princípios	33
2.1.1 Métodos de cálculo na Dinâmica Molecular (DM)	35
2.2 Campo de força	41
2.2.1 Coarse-grained	43
3 Objetivos e Métodos	45
3.1 Objetivos	45
3.2 Metodologia	45
3.2.1 Proteínas	45
SalBIII	45

Calmodulina	46
3.2.2 Cross-linking	46
Procedimentos Experimentais	46
Procedimentos Computacionais	49
3.2.3 Dinâmica Molecular	50
SalBIII	50
Calmodulina	51
4 Resultados e Discussão	53
4.1 SalBIII	53
4.1.1 Cross-links compatíveis com a estrutura cristalográfica	53
4.1.2 Cross-links não compatíveis com a estrutura cristalográfica . . .	58
Cross-links compatíveis com estruturas da simulação all-atom . .	58
Cross-links compatíveis com estruturas da simulação coarse-grained	64
4.1.3 O conjunto de cross-links compatíveis com as estruturas das si- mulações	65
4.1.4 Qualidade dos filtros e o conjunto de cross-links	66
4.2 Calmodulina	69
4.2.1 Cross-links compatíveis com as estruturas de referência	71
Calmodulina livre	71
Calmodulina complexada com cálcio	75
4.2.2 Cross-links não compatíveis com a estrutura de referência . . .	78
Calmodulina livre	78
Calmodulina complexada com cálcio	79
4.2.3 Cross-links não compatíveis com as estruturas das simulações .	82
Calmodulina livre	82
Calmodulina complexada com cálcio	82
4.2.4 Agrupamento por comportamento das ligações cruzadas em dife- rentes condições biológicas/experimentais	83

5 Considerações Finais	87
A Análises preliminares dos campos de força	89
A.1 Campo de força	89
A.2 Coarse-grained vs All-Atom	91
A.3 RMSD	94
I Anexo A	97
II Anexo B	101

Capítulo 1

Cross-linking

O experimento de Ligação Cruzada em Espectrometria de Massa (XL-MS) é uma técnica de proteômica estrutural por espectrometria de massas. Esta abordagem, além de poder complementar os resultados de outros métodos, permite a obtenção de informações estruturais da proteína contornando algumas dificuldades. A pureza e disponibilidade do analito são algumas destas dificuldades que são contornadas mas são encontradas nos métodos determinação estrutural com maior resolução como a Ressonância Magnética Nuclear (RMN), a Cristalografia por Difração de Raios X (CDRX) e Crio-Microscopia Eletrônica (Crio-ME) [1]. Fundamentado no poder de análise do espectrômetro de massas, a técnica não tem muitas fraquezas associadas às características da proteína, já que com a grande gama de ionizadores e analisadores hoje comercializados, o tamanho, a baixa solubilidade, não cristalinidade ou pouca disponibilidade de proteína podem ser contornados por módulos mais especializados do equipamento.

1.1 O Equipamento

O espectrômetro de massas utiliza a razão massa-carga das partículas analisadas em fase gasosa para separá-las por terem diferentes tempos de translação quando aceleradas em um campo elétrico. O equipamento é dividido em três partes gerais: o ionizador, o analisador e o detector.

O ionizador é onde as partículas que serão analisadas ganham carga, tornando possível a condução, análise e detecção através do equipamento, o analisador é onde há a separação das espécies e o detector é onde são detectados os íons. Particularmente, a análise de proteínas por espectrometria de massas tomou força ao final dos anos 80, quando surgiram técnicas de ionização mais brandas como a Ionização por Electrospray (ESI) [2] e a Ionização e Dessorção a Laser Assistida por Matriz (MALDI) [3], permitindo a ionização de peptídeos e ácidos nucleicos sem a fragmentação dessas moléculas [4].

Os equipamentos podem ser acoplados a métodos de separação como Cromatografia Líquida de Alta Eficiência (HPLC) ou Cromatografia Gasosa (CG), melhorando a separação das espécies que entram no sistema, e os analisadores modernos podem ser sequenciados, agregando vários módulos de fragmentação, seleção de massa-carga e/ou alta resolução, como por exemplo um Duplo Quadrupolo com Tempo de Voo (QqTOF). Com isso, através de configurações como esta, é possível fazer a separação dos precursores, ionizá-los, identificá-los, posteriormente fragmentar cada um desses precursores e identificar seus fragmentos para melhor identificar cada espécie, impulsionando métodos relacionados a proteínas como a Troca de Hidrogênio-Deutério (HDX) [5], Mobilidade Iônica (IM) [6], e XL-MS [7].

1.2 O Experimento

O experimento de XL-MS consiste em reagir o Agente de Ligação Cruzada (ALC) (*linker*) com a(s) proteína(s) de estudo em solução, nas condições adequadas de temperatura, pH e força iônica para que este forme ligações covalentes entre os resíduos reativos [8]. A Figura 1.1 representa a conexão produzida entre os resíduos de uma proteína.

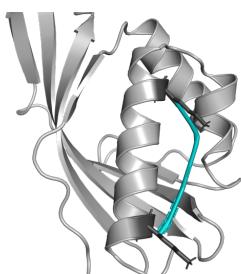


Figura 1.1: Representação de ligação cruzada produzida pela ação do linker DSS no experimento de XL-MS para a SalBIII, reproduzida no software PyMol.

Todavia, para que essa etapa seja realizada é necessária a solubilização da proteína, mesmo que em concentrações baixas, pois a ação do linker é dependente do acesso, através do solvente, à superfície da proteína. Sendo assim, mesmo não sendo limitação do equipamento [9], o experimento necessita de proteínas em solução para a produção das ligações interresiduais através do ALC.

Passado o processo de ligação entre os resíduos, a proteína é digerida, normalmente por tripsina, e os peptídeos resultantes são injetados no espectrômetro de massas. Estes então, produzem um conjunto de espectros que, com o tratamento adequado, permite a identificação dos resíduos que foram interligados (*cross-link*) [10]. Consequentemente, devido ao linker utilizado, pode-se determinar também a distância máxima na qual os resíduos se encontravam no momento da ligação com o ALC.

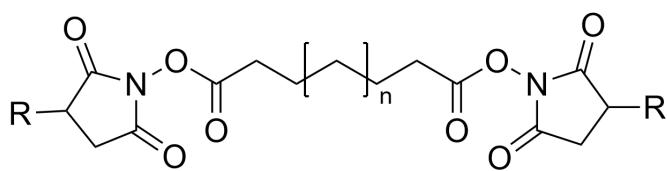
1.2.1 Agente de Ligação Cruzada

Os ALCs são os compostos orgânicos multifuncionais que promovem a ligação entre resíduos da proteína. Cada ALC tem um comprimento específico, consequência do número de átomos em sua cadeia espaçadora. Este comprimento limita a ação do agente a partir da formação da primeira conexão, fazendo com que as ligações cruzadas identificadas posteriormente, tenham um sentido estrutural muito relevante.

Os compostos desta classe mais utilizados pela comunidade do experimento de ligação cruzada certamente são os disponíveis comercialmente e derivados de esteres de N-hidroxisuccinimida (NHS). Estes são reativos frente a grupos nucleofílicos, principalmente aminas primárias presentes no N-terminal e na cadeia lateral de lisinas.

Em pH 7-9, os esteres de N-hidroxisuccinimida (NHS) reagem com estas aminas formando amidas estáveis e eliminando o grupo NHS.

A Figura 1.2 mostra a estrutura dos ALCs DSS e de sua versão sulfonada BS3 que apresenta melhor solubilidade em solventes polares. Estes são os ALCs mais usados em experimentos de ligação cruzada.



DSS: R = H, n = 2

BS3: R = SO₃Na, n = 2

Figura 1.2: Estrutura geral de um derivado de NHS exemplificando DSS e BS3. Ambos com o comprimento de 11.4 Å.

Visando ampliar a aplicação da técnica de ligação cruzada, novos reagentes disponíveis como ALCs têm surgido. Compostos que ligam grupos de resíduos ácidos [11], reagentes que promovem a ligação entre resíduos ácidos e resíduos básicos (zero-length) [12], compostos fotoativáveis que formam espécies de carbeno quando irradiados com luz ultravioleta [13], e ALC cliváveis que permitem a clivagem da cadeia do agente ligado no experimento de MS2, facilitando a identificação pela redução do espaço de busca [14].

As Figuras 1.3 e 1.4 mostram exemplos de ALCs para aqueles citados acima.

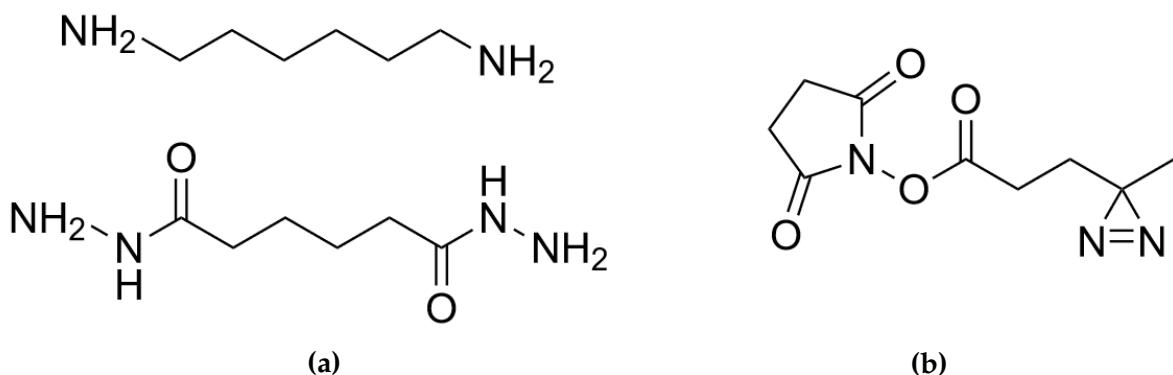


Figura 1.3: Representação estrutural de diferentes grupos de ALCs. (a) Ácido ligantes. (b) Fotoativável.

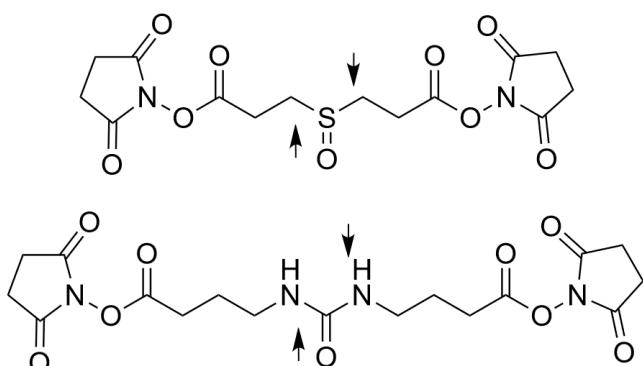


Figura 1.4: Representação estrutural de diferentes grupos de ALCs cliváveis.

XPlex

O XPlex é um ALC com múltipla química de ligação cruzada desenvolvido pelo grupo Dalton e seus colaboradores [12].

Ativando os grupos de ácido carboxílico por meio da adição de 1-etil-3(3-(dimetilamino)propil)carbodiimida (EDC) seguido de Hidroxibenzotriazol (HOBT), estes grupos se tornam reativos frente a nucleófilos. Se os nucleófilos que os grupos de ácido carboxílico encontram são grupos amino ou hidroxi provenientes de lisinas ou serinas, o tipo de ligação cruzada que se forma é o zero-length. Porém, caso nucleófilos do tipo diaminas (Figura 1.3a), 1,6-hexanodiamina no caso do XPlex, são os primeiros a reagir com os grupos ativados, a ligação cruzada entre resíduos ácidos é formada. Sendo assim, a abordagem com XPlex se torna versátil abrangendo, em média, 24% de todos os resíduos em uma proteína.

Limite estatístico de validação

Para diminuir a influência da flexibilidade das cadeias laterais dos resíduos, neste trabalho utilizou-se a distância entre os $\text{C}\beta$ para verificação da compatibilidade do ALC com a estrutura. Com esta consideração, os comprimentos das cadeias laterais dos resíduos também devem ser contabilizados juntamente ao comprimento do ALC ($L_{max} = L_{R_a} + L_{R_b} + L_{XL}$) para melhor representação da realidade de ligação.

A partir deste cenário, foi computada a distribuição de frequências das distâncias topológicas entre $\text{C}\beta$'s dos resíduos de proteínas conhecidas sob a condição

$d_{top}(R_a, R_b) \leq L_{XL}$, onde $d_{top}(R_a, R_b)$ é a distância topológica entre os átomos reativos dos resíduos em questão. Com estas distribuições, ao excluir os casos pouco prováveis (1% das maiores distâncias compatíveis com a condição), define-se como limite estatístico de validação o L_{max} efetivo. Sendo este a distância topológica entre $C\beta$'s que contempla 99% das estruturas deste conjunto [15, 16].

1.3 O Resultado

Os dados de ligação cruzada são reflexo da estrutura tridimensional na proteína. Entretanto, considerando que estas ligações também podem ser reflexo da variabilidade conformacional em solução ou até mesmo de erros experimentais, nem todas provém da estrutura cristalográfica, podendo fazer com que as suas aplicações levem a resultados que não condizem com a estrutura encontrada pelos métodos de referência. Estas ligações cruzadas que decorrem da variabilidade conformacional da proteína em solução, são muito interessantes e podem ser de grande utilidade e agregam valor aos resultados que a técnica de XL-MS traz. Sendo assim, um estudo sobre a compatibilidade dos resultados de cross-link com dados de dinâmica molecular de proteínas, pode contribuir para o entendimento mais detalhado dos resultados que esse experimento oferece [17, 18]. Este estudo sobre a dinâmica da proteína pode gerar melhorias nas interpretações dos dados de XL-MS e, consequentemente, em suas aplicações como o estudo da dinâmica de proteínas por experimentos de cross-link quantitativos [19, 20], a determinação de redes de interação entre proteínas [8, 21, 22] e outras vertentes de pesquisa com a abordagem cross-link ou suas integrações [23].

Outra aplicação relevante para as ligações cruzadas obtidas por XL-MS é a assistência dos métodos computacionais de determinação de estrutura de proteínas. Os dados experimentais trazem mais contribuições para o cálculo do potencial utilizado para o enovelamento *in silico* de proteínas [24], enviesando o processo. Entretanto, com a iniciativa AlphaFold [25], um método computacional de predição de estruturas

de proteínas com resultados impressionantes e que não utiliza de dados experimentais como as restrições de cross-link, qualquer método dessa classe de determinação estrutural que não ofereça vantagens significantes sobre esta abordagem tende a perder força, principalmente depois da disponibilização dos códigos pelos desenvolvedores.

Capítulo 2

Dinâmica Molecular

Imagine um método o qual te permita obter resolução atômica de todas partículas de um sistema ao longo do tempo. Um método como esse possibilitaria a elucidação da dinâmica de macromoléculas tornando possível a resolução de questionamentos estruturais como a compatibilidade de ALCs com proteínas solvatadas, por exemplo.

A simulação computacional dos movimentos moleculares permite a obtenção de resolução atômica e temporal das partículas simuladas. Mesmo que não experimental, a Dinâmica Molecular (DM) devidamente descrita proporciona resultados relevantes e compatíveis com dados experimentais [26]. Partindo da descrição do estado inicial do sistema (posição e velocidade de todas as partículas que o compõe) e das interações que as partículas estão sujeitas neste sistema, é computacionalmente possível, até mesmo com pouco poder computacional, propagar estas coordenadas iniciais para instantes de tempo posteriores.

2.1 Os Princípios

A Dinâmica Molecular (DM) é construída através da determinação da configuração do sistema para cada intervalo de tempo t , produzindo o que chamamos de trajetória [27].

Num procedimento geral, as posições das partículas são determinadas aleatoriamente, sem sobreposição, ao colocá-las no espaço delimitado para o sistema que chamamos de caixa. Para moléculas complexas com estruturas definidas como proteínas, essas posições são obtidas a partir de suas estruturas cristalográficas ou de RMN. As velocidades iniciais são designadas seguindo a distribuição energética de Maxwell-Boltzmann para a temperatura escolhida, tornando a energia cinética do sistema correspondente à temperatura termodinâmica. Em seguida, sendo a descrição das interações do sistema uma função de potencial em relação as posições das partículas, agora é possível calcular a força resultante sobre estas a partir da derivação deste potencial com o tempo, $\vec{F} = -\nabla \vec{V}(t)$.

Podemos determinar a aceleração correspondente a força calculada através da segunda lei de Newton sobre cada átomo

$$-\nabla \vec{V}(t) = m\vec{a}(t) \quad (2.1)$$

onde m é a massa do átomo. Desta forma, por meio das equações do movimento uniformemente variado correspondentes é possível propagar as coordenadas de velocidade e posição do instante t , neste caso t_0 , para um instante posterior $t + \Delta t$, usando

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \vec{a}(t)\Delta t \quad (2.2)$$

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{\vec{a}(t)\Delta t^2}{2} \quad (2.3)$$

Calculadas as novas posições e velocidades, o ponto previsto se torna o novo instante t e o procedimento do cálculo do potencial das interações e propagação das coordenadas recomeça. Este ciclo se repete n vezes até que o tempo de simulação t_f seja atingido, $t_f - t_0 = n\Delta t$.

Ao concluir a simulação, o produto é a trajetória de todas as partículas do sistema. Este resultado contém a posição de cada partícula em cada intervalo de tempo simulado, que permite visualizar o comportamento dos componentes e conectar

propriedades microscópicas e macroscópicas do sistema [28] ou calcular propriedades estatísticas como distribuições de mínima distância, por exemplo [29].

Todavia, como o resultado de uma DM não é obtido através de um experimento com o sistema real, comparações com dados experimentais devem ser feitos para atestar a representatividade do método. Para isso, conceitos da mecânica estatística são utilizados para transformar as propriedades microscópicas observadas na trajetória em gradezas macroscópicas observáveis [30]. Em princípio, quanto melhor for a descrição das interações entre as partículas do sistema, melhor serão os resultados provenientes da DM. Esta descrição, em prol de uma maior eficiência e/ou fidelidade com os dados experimentais para um certo tipo de sistema ou ambiente, forma o que chamamos de campo de força.

2.1.1 Métodos de cálculo na DM

Integrador leap-frog. O integrador leap-frog é o método padrão de integração do GROMACS [27] e é utilizado para resolução de equações diferenciais do tipo

$$x'' = f(x) ,$$

onde x é função de t . Exatamente como no caso em questão de posição (x) e tempo (t)

As equações que descrevem esse método são

$$v_{t+\frac{1}{2}\Delta t} = v_{t-\frac{1}{2}\Delta t} + \frac{\Delta t}{m} F(t) , \quad (2.4)$$

onde v é a velocidade do corpo, m a massa do corpo e $F(t)$ a função força em relação as posições das partículas do sistema no tempo t , e

$$x_{t+\Delta t} = x_t + \Delta t \cdot v_{t+\frac{1}{2}\Delta t} , \quad (2.5)$$

onde x é a posição da partícula.

Seguindo as equações, o método calcula as posições e velocidade na forma alternada como representado na Figura 2.1, como se os cálculos de x e v remetessem a sapos pulando um sobre o outro, justificando o nome do método.

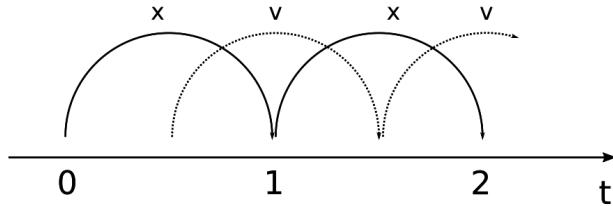


Figura 2.1: Esquema visual do integrador leap-frog.

Com esse método há a vantagem da reversibilidade temporal, ou seja, partindo de um enésimo ponto de posição x_n pode-se, apenas trocando o sentido temporal, voltar a exatamente o mesmo ponto de posição inicial x_0 . Essa propriedade permite a conservação da energia no sistema, diferentemente de métodos como o Mid-point ou Runge-Kutta-4 [31]. Estes métodos mencionados como exemplo geram, apesar de pequenos, erros sistemáticos que se somam significativos em longas integrações e não se anulam no sentido temporal contrário como o leap-frog.

Um método como o leap-frog se torna valioso uma vez que não há o acúmulo desses erros, principalmente em casos onde o interesse do estudo repousa nas pequenas alterações de propriedades de órbita quase periódica, mesmo fornecendo somente uma precisão de segunda ordem.

Somatório de Ewald. A energia eletrostática total de um sistema periódico de partículas carregadas é

$$E_q = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^N \sum_{j=1}^N' \frac{q_i q_j}{|r_{ij} + \mathbf{n}L|}, \quad (2.6)$$

onde ϵ_0 é a constante de permissividade do vácuo, \mathbf{n} é o vetor contendo os índices das celas de repetição, L o vetor contendo as distâncias da partícula até a sua repetição nas 3 coordenadas e o símbolo ' $'$ no somatório mais interno é para condicionar a exclusão do termo $j = i$, se e somente se $\mathbf{n} = 0$. Sendo assim, com essa descrição uma partícula com

carga q_i localizada em r_{ij} apresentaria repetições em $r_i + \mathbf{n}L$, isto é, $r_i + n_1c_1 + n_2c_2 + n_3c_3$.

O somatório de Ewald, no objetivo de obter uma soma mais rapidamente convergente, divide a Equação 2.6 em três termos: a contribuição eletrostática de curta distância, de longa distância e um termo constante.

$$E_q = E_C + E_L + E_0 , \quad (2.7)$$

A primeira parte da estratégia para alcançar esta soma computacionalmente mais implementável é a atribuição das cargas como uma combinação de funções gaussianas.

Considerando a distribuição de cargas do sistema como uma coleção de funções de Dirac, a densidade de carga para um ponto carregado q_i é

$$\rho_i(r) = q_i\delta(r - r_i)$$

em que podemos subtrair e somar uma função gaussiana, obtendo

$$\rho_i(r) = q_i\delta(r - r_i) - q_iG_\sigma(r - r_i) + q_iG_\sigma(r - r_i) , \quad (2.8)$$

onde G_σ é uma função gaussiana do tipo

$$G_\sigma(r) = \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} \exp\left(-\frac{|r|^2}{2\sigma^2}\right) ,$$

em que σ é o desvio padrão da distribuição.

Denominam-se os dois primeiros termos da Equação 2.8 como sendo a contribuição de curta distância para densidade de carga q_i e o terceiro a contribuição de

longa distância.

$$\begin{aligned}\rho_i(r) &= \rho_i^C(r) + \rho_i^L(r) \\ \rho_i^C(r) &= q_i\delta(r - r_i) - q_iG_\sigma(r - r_i) \\ \rho_i^L(r) &= q_iG_\sigma(r - r_i)\end{aligned}$$

Por meio da equação de Poisson, o potencial de curta distância toma a forma de

$$\phi_i^C(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i}{|r - r_i|} \operatorname{erfc}\left(\frac{|r - r_i|}{\sqrt{2}\sigma}\right), \quad (2.9)$$

onde $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$.

Este potencial então pode ser transformado na energia eletrostática, como

$$\begin{aligned}E_C &= \frac{1}{2} \sum_i^N q_i \phi_{[i]}^S(r_i) \\ &= \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|r_i - r_j + \mathbf{n}L|} \operatorname{erfc}\left(\frac{|r_i - r_j + \mathbf{n}L|}{\sqrt{2}\sigma}\right).\end{aligned} \quad (2.10)$$

Para a energia de longa distância não é tão prático. Através dos mesmos passos que originaram a Equação 2.9, pode-se fazer o mesmo para o potencial de longa distância, obtendo

$$\phi_i^L(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i}{|r - r_i|} \operatorname{erf}\left(\frac{|r - r_i|}{\sqrt{2}\sigma}\right). \quad (2.11)$$

Pela natureza da expressão, a interação de longa distância não pode ser calculada diretamente no espaço real, então usam-se de alguns artifícios matemáticos para contornar essa situação. O primeiro deles é a adição e subtração do termo de interação própria dentro das condições periódicas de contorno, tornando o termo de longa distância periódico e, consequentemente, passível de ser Fourier transformado ao mesmo passo que dá origem ao termo E_0 da Equação 2.7. Este último pode ser

determinado tomando o limite de r_i a zero na Equação 2.11.

$$\begin{aligned}\lim_{z \rightarrow 0} \operatorname{erf}(z) &= \frac{2}{\sqrt{\pi}} z \\ \phi_i^L(r) &= \frac{q_i}{4\pi\epsilon_0\sigma} \sqrt{\frac{2}{\pi}} \\ E_0 &= \frac{q_i}{4\pi\epsilon_0\sigma} \frac{1}{\sqrt{2\pi}} \sum_i^N q_i^2\end{aligned}\quad (2.12)$$

O segundo dos artifícios é a resolução do problema no espaço da transformada de Fourier. Conhecida a equação de Poisson no espaço da transformada como

$$k^2 \hat{\phi}_i^L(k) = \frac{\hat{\rho}_i^L(r)}{\epsilon_0} .$$

O conceito é calcular a transformada de Fourier da densidade de carga, dividir este resultado por k^2 (parâmetro oriundo da transformada) e então fazer a transformada inversa para obter o potencial de longa distância no espaço real.¹

Este potencial no espaço real é

$$\phi^L(r) = \frac{1}{V\epsilon_0} \sum_{k \neq 0} \sum_{j=1}^N \frac{q_j}{k^2} e^{ik \cdot (r - r_j)} e^{-\sigma^2 k^2 / 2} . \quad (2.13)$$

Consequentemente, a energia respectiva é dada por:

$$\begin{aligned}E_L &= \frac{1}{2} \sum_{i=1}^N q_i \phi^L(r_i) \\ &= \frac{1}{2V\epsilon_0} \sum_{k \neq 0} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{k^2} e^{ik \cdot (r_i - r_j)} e^{-\sigma^2 k^2 / 2} .\end{aligned}\quad (2.14)$$

O método descrito pela expressões acima que detalham a soma da Equação 2.7 é chamado Ewald clássico. Neste método, com a otimização de σ em função de N , o custo computacional escala na ordem de $O(N^{3/2})$, menor que o escalonamento

¹Uma dedução mais detalhada deste método pode ser encontrada na referência [32].

quadrático mas ainda custoso para sistemas com maior número de partículas.

Particle Mesh Ewald. O método Particle Mesh Ewald (PME) é um aprimoramento do Ewald clássico sobre o termo da soma no espaço da transformada (ϕ^L e E_L). A proposta de Tom Darden [33] é de atribuir as cargas a uma malha usando interpolação ao invés de somar ondas diretamente antes da transformação. Com essa nova abordagem o custo computacional passa a escalar na ordem de $O(N \cdot \log N)$, mais eficiente que o Ewald clássico em sistemas com maior número de partículas.

Particularmente, quando a interpolação utilizada é do tipo B-spline, o método passa a ser referido como smooth PME [34] e é essa variante que é implementada no GROMACS.

Controle de Temperatura. O controle da temperatura em um sistema de simulação é dependente da precisão na qual se calcula essa grandeza. O algorítimo de Berendsen suprime as flutuações de energia cinética devido a conexão que promove entre T_t e T_{t+1} dada da seguinte forma:

$$\frac{dt}{d\bar{t}} = \frac{T_t + T_{t+1}}{\tau} ,$$

onde τ é uma constante de tempo que pode ser ajustada de acordo com a situação, uma vez que a incerteza decai exponencialmente com sua magnitude.

Todavia, a supressão dessas flutuações tornam o sistema não canônico, fazendo com que a amostragem do sistema não seja termodinamicamente apropriada. A modificação deste algorítimo que foi utilizada neste estudo é o reescalonamento das velocidades (V-rescale). Esta modificação garante uma distribuição correta da energia cinética por meio do termo estocástico

$$dK = (K_t - K_{t+1}) \frac{dt}{\tau_T} + 2 \sqrt{\frac{K_t K_{t+1}}{N_f}} \frac{dW}{\sqrt{\tau_T}} ,$$

onde K é a energia cinética, N_f o número de graus de liberdade e dW um processo Wiener [35].

Desta forma, o sistema se torna canônico e ainda preserva as vantagens de não apresentar oscilações de energia e o decaimento de primeira ordem da incerteza [27].

Parrinello-Rahman. A pressão é relacionada com o volume do sistema, consequentemente, o controle dessa grandeza também. A relação entre tais termos, para um gás ideal, é dada pela equação virial

$$pV = nRT ,$$

onde p é a pressão, V o volume, n o número de mols, R a constante universal dos gases e T a temperatura.

Ao modificar o volume da caixa de simulação, a pressão pode ser controlada e, com o trabalho realizado pelo barostato devidamente subtraído da energia total, a energia conservada.

O algorítmico Parrinello-Rahman se caracteriza por conseguir representar um sistema NPT verdadeiro. Entretanto, se a pressão do sistema se encontra longe do equilíbrio, este método de acoplamento pode gerar grandes oscilações nas dimensões da caixa, podendo resultar na interrupção da simulação [27, 36, 37].

2.2 Campo de força

Moléculas são grupos de átomos unidos por potenciais com graus de liberdade de torção, ângulos e estiramento de ligação, interações intramoleculares. Além disso, há interações entre as partículas de moléculas diferentes, com potenciais de longa distância e forças dispersivas, interações intermoleculares. A descrição destas interações potenciais são o campo de força. Especificamente para a DM clássica, esses potenciais são representados classicamente em potenciais elásticos e/ou harmônicos, e os parâmetros são as constantes de força dessas equações, podendo ser ajustados empiricamente ou com cálculos quânticos.

Uma representação da equação típica de um campo de força é dado pela Equação 2.15:

$$\vec{V}(r) = \sum \vec{V}_{estir} + \sum \vec{V}_{ang} + \sum \vec{V}_{tor} + \sum \vec{V}_{vdW} + \sum \vec{V}_{elet} \quad (2.15)$$

onde o potencial final é resultante da soma dos potenciais de estiramento, ângulo de ligação, torção de ligação, van der Waals, e atração e repulsão eletrostática, respectivamente [38]. É relevante apontar que fenômenos como a quebra e/ou formação de ligação química não são abordados por esse tipo de campo de força.

Em um cenário ideal, um mesmo campo de força, bem parametrizado, seria robusto e consistente com dados experimentais para todos os tipos de sistemas simulados. Entretanto, essa grande abrangência com ótima precisão ainda não foi alcançada, fazendo com que campos de força parametrizados especificamente para reproduzir os dados de sistemas específicos sejam o atual regime dos protocolos de DM. Campos de força como o OPLS (Optimized Potentials for Liquid Simulations) foram desenvolvidos e são utilizados para melhor reproduzir o comportamento de líquidos [39], assim como o AMBER03 para o comportamento de proteínas em fase condensada [40].

Em uma simulação, o número de partículas chega facilmente à marca de dezenas de milhares, consequentemente, o trabalho computacional necessário para propagar tantas coordenadas é grande. Ao aumentar o número de átomos, o tempo de simulação, ou até mesmo diminuir o intervalo de tempo no qual as coordenadas são propagadas, o tempo necessário para realizar os cálculos, para um mesmo poder computacional, aumenta proporcionalmente, mesmo com implementações sofisticadas dos algorítimos que calculam as interações. Pensando nisso, para a abordagem de sistemas maiores e/ou de maiores tempos de simulação, campos de força *coarse-grained* surgiram para compensar falta de poder computacional e/ou tempo hábil de produção.

2.2.1 Coarse-grained

Diferentemente dos campos de força mais tradicionais, nos campos de força coarse-grained as partículas descritas ao se construir o sistema não são os átomos. Nesse método, grupos de átomos são descritos por partículas maiores e com características específicas (carga, raio, polarizabilidade...) que buscam representar da melhor forma os átomos substituídos [41, 42]. Desta forma, o número de graus de liberdade e o número de partículas diminuem, exigindo menos recursos computacionais para a simulação de um mesmo sistema.

A descrição das interações também são clássicas, utilizando funções como a do oscilador harmônico para as ligações covalentes e, no caso do estudo de proteínas, funções de Lennard-Jones para interações dispersivas. As constantes destas funções são parametrizados semelhantemente aos campos de força tradicionais, entretanto não são os átomos que terão as coordenadas propagadas e sim essas partículas que os substituem.

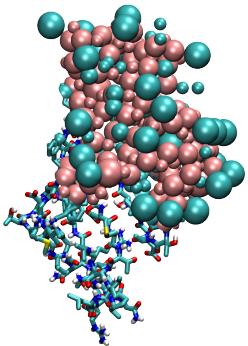


Figura 2.2: Representação da descrição coarse-grained em uma molécula proteica. Na metade inferior esquerda da figura são ilustrados todos os átomos dessa metade da estrutura, utilizando a representação Licorice. Já na metade superior, as partículas que são ilustradas são aquelas que substituíram os grupos de átomos.

Com estas substituições, o número de cálculos a serem realizados para a produção da simulação diminui e, consequentemente, o tempo necessário para a realização de cada passo da simulação também. Possibilitando a exploração de sistemas e tempos ainda maiores para simulação [43], campos de força com essa abordagem ganham mais usuários e podem ser utilizados na maioria dos softwares de DM. Exemplos desses campos de força com aspectos voltados para as biomoléculas são:

MARTINI [44], desenvolvido na Universidade de Groningen - Países Baixos, e SIRAH [45], do Instituto Pasteur - Uruguai.

Capítulo 3

Objetivos e Métodos

3.1 Objetivos

- Comparar os dados de cross-link com as estruturas obtidas por dinâmica molecular por método all-atom e coarse-grained.
- Verificar a possível existência de uma correlação entre os parâmetros de qualidade de espectro das análises dos dados de cross-link e o conjunto de restrições satisfeitas na simulação para a SalBIII.
- Comparar os conjuntos de restrições encontradas para os diferentes estados da Calmodulina com as estruturas obtidas por suas respectivas simulações.
- Demonstrar a relevância de uma interpretação de distância acessível ao solvente em conjunto com a distância euclidiana para a análise da satisfação de restrições.

3.2 Metodologia

3.2.1 Proteínas

SalBIII

A primeira proteína abordada no estudo foi a SalBIII (código Protein Data Bank 5CXO), proteína bacteriana (*Streptomyces albus*), dimérica em solução e participante

na síntese da salinomicina, mais especificamente na transformação de um anel pirano em tetraidropirano. A SalBIII faz parte de uma classe de hidrolases/ciclases e apresenta um mecanismo de ação diferente na produção desses compostos policetídeos comumente usados como antibióticos [46].

Esse polipeptídeo participa deste estudo pois os dados experimentais foram gentilmente concedidos pelo grupo experimental colaborador (Grupo Dalton). Contudo, olhando por uma perspectiva alternativa, este composto além de possuir dois tipos de estruturas secundárias, sua sequência de aminoácidos possui baixa similaridade com outras proteínas do mesmo grupo, tornando-o um candidato interessante para aplicações dos dados de cross-link em técnicas de determinação estrutural.

Calmodulina

A Calmodulina tem papel significativo nesse estudo pois é um exemplo de conformação dupla, podendo apresentar as conformações livre ou complexada com cálcio (código Protein Data Bank 1DMO [47] e 3CLN [48]) em função do ambiente químico. Presente em todas as células eucarióticas, essa proteína cálcio-moduladora atua como receptor de íons cálcio e converte esse tipo de sinal em uma mudança bioquímica [49, 50].

3.2.2 Cross-linking

Procedimentos Experimentais

Um panorama geral do procedimento instrumental será relatado, necessário para a interpretação dos dados desta dissertação. Os experimentos foram realizados pelo grupo experimental de espectrometria de massas do laboratório Dalton. Aqueles associados à SalBIII foram experimentalmente conduzidos por Hugo César Ramos de Jesus e trabalhados por Állan Jhonathan Ramos Ferrari para gerar o conjunto de ligações cruzadas [12, 15, 16]. Já para a Calmodulina, os experimentos foram conduzidos por Luana Oliveira dos Santos e os dados trabalhados pela mesma [51].

SalBIII. No experimento de cross-link para essa proteína, o grupo colaborador utilizou duas abordagens para a produção de ligações cruzadas, com DSS e com XPlex [15].

Reação de ligação cruzada com DSS. Em 500 μL da solução de proteína (10 μM) em tampão fosfato 200 mM, adicionaram-se 35 μL da solução de DSS em DMF (N,N-dimetilformamida anidro) e ajustou-se para pH 7. A reação foi conduzida por 2 horas a 25 °C e 300 rpm.

Reação de ligação cruzada XPlex. Em 500 μL da solução de proteína (10 μM) em tampão MES (ácido 2-(N-morfolino) etanossulfônico) 200 mM e pH 6.5, adicionaram-se 10 μL de uma solução 500 mM de 1,6-hexanodiamina, 5 μL de uma solução 500 mM de HOBt (1-hidroxi-benzotriazol) e 20 μL de uma solução 500 mM de EDC (1-etyl-3-(3-dimetilaminopropil)carbodiimida). A reação foi conduzida por 2 horas a 25 °C e 300 rpm.

Procedimento pós reação. Após o período de reação, medidas a redução de pontes de dissulfeto foram executadas e as amostras foram submetidas à digestão enzimática com tripsina por 18 horas.

A análise dos produtos de ligação cruzada foi realizada com uma abordagem de aquisição dependente de dados (ADD) em um sistema de cromatografia líquida ligado a um espectrômetro de massas com ionização por electrospray (nanoLC-nESI-MS/MS). Sistema de cromatografia Dionex UltiMateTM 3000 RSLC nano com acoplamento online e um espectrômetro de massas Q-Exactive Plus (Thermo Scientific Quadrupole-Orbitrap).

A identificação das espécies de ligação cruzada foi efetuada através do programa SIM-XL [10], com os dados brutos de spectrometria e com os parâmetros adequados para os agentes de ligação cruzada XPlex C6N2 e DSS/BS3. Com um conjunto mais abrangente, características dos espectros foram utilizadas posteriormente para filtrar as restrições, são essas características: contagem de espectros, número de espécies (contagem de fragmentos que identificam a mesma restrição) e escore.

Calmodulina. A Calmodulina fornecida para o grupo experimental estava complexada com íons cálcio e um procedimento de remoção foi realizado. Na solução de proteína complexada, adicionou-se uma solução de tampão MOPS (ácido 3-(N-morfolino) propanossulfônico) 100 mM contendo 5 mM de EGTA (ácido etilenoglicol-bis(2-aminoetil-éter)-N,N,N',N'-tetracético) e, depois de 30 minutos de reação, esta solução foi lavada utilizando filtro de centrífuga Amicon 500 com exclusão por peso molecular de 10 kDa (Milipore) e solução tampão MOPS com EGTA 5 μ M. A partir do produto dessa remoção, parte da solução foi induzida a mudança de conformação pela adição de íons cálcio (CaCl_2) obtendo uma concentração de 12 mM e a outra parte recebeu íons sódio (NaCl) suficiente para atingir a mesma força iônica.

No experimento de cross-link para essa proteína, para ambas as espécies, o grupo colaborador utilizou o agente de ligação cruzada DSS para a formação das ligações [51].

Reação de ligação cruzada com DSS. Na solução de proteína, adicionou-se uma alíquota da solução de DSS em DMF (N,N -dimetilformamida anidro). A reação foi conduzida por 2 horas a 27 °C.

Procedimento pós reação. Após o período de reação, as amostras foram submetidas à digestão enzimática com tripsina em duas etapas totalizando um período de 24 horas.

A análise dos produtos de ligação cruzada foi realizada com uma abordagem de aquisição dependente de dados (ADD) em um sistema de cromatografia líquida ligado a um espectrômetro de massas com ionização por electrospray (nanoLC-nESI-MS/MS). Sistema de cromatografia de nanofluxo Easy-nLC 1200 (Thermo Scientific) acoplado a um espectrômetro de massas Orbitrap Fusion Lumos Tribrid (Thermo Scientific).

A identificação das espécies de ligação cruzada foi efetuada através do programa SIM-XL [10], com os dados brutos de espectrometria e com os parâmetros adequados para o agente de ligação cruzada DSS/BS3. Com um conjunto mais abrangente, características dos espectros foram utilizadas posteriormente para filtrar as restrições, são essas características: contagem de espectros, número de espécies (contagem de fragmentos que identificam a mesma restrição) e escore.

Procedimentos Computacionais

Topolink. Para a verificação de compatibilidade das restrições, as estruturas vindas de simulação foram analisadas com o pacote Topolink [16]. Este pacote calcula a distância acessível ao solvente entre os resíduos reativos, neste estudo com o C β de cada resíduo, e verifica se essa distância é compatível com o limite estatístico respectivo. No caso particular de cross-links do tipo zero-length, apesar de não haver nenhuma cadeia espaçadora, uma distância de aproximadamente 4 Å, justificada pela interação polar iônica entre os resíduos, foi considerada para a determinação dos limites estatísticos respectivos [12].

Baseado nas informações do Topolink, pôde-se determinar a probabilidade (ρ_i) de que uma estrutura da simulação seja compatível com o cross-link:

$$\rho_i = \frac{n^o \text{ de estruturas compatíveis com o cross-link}}{n^o \text{ total de estruturas}} \quad (3.1)$$

A probabilidade pode ser convertida em unidades de energia livre por meio da relação de Boltzmann:

$$\rho_i \propto e^{-\frac{\varepsilon_i}{kT}} \implies \varepsilon_i \propto kT \ln \rho_i \quad (3.2)$$

A energia livre calculada desta forma estima a estabilidade do conjunto de conformações da proteína que são consistentes com o cross-link estudado.

QuinXL. Para analisar a diferença entre os resultados do experimento de ligação cruzada da Calmodulina nas duas condições (Calmodulina livre e Calmodulina complexada com cálcio), fez-se o uso do software QuinXL [52]. Com o objetivo de analisar dados de ligações cruzadas em espectrometria de massas, o programa utiliza de agrupamentos não supervisionados de diferentes experimentos para identificar mudanças na conformação na proteína em diferentes condições experimentais/biológicas.

3.2.3 Dinâmica Molecular

As produções de dinâmica molecular *all-atom* foram antecedidas por uma minimização e duas equilibrações, uma a volume constante e outra a pressão constante. Em todas as etapas, utilizou-se 1.0 nm como cutoff para interações eletrostáticas e de van der Waals de curta distância.

Já para as produções *coarse-grained*, foram executadas duas minimizações e duas equilibrações, sendo ambas as equilibrações a volume constante, e como medida de cutoff para interações eletrostáticas e de van der Waals de curta distância, utilizou-se 1.2 nm

Em ambos os casos o software de DM utilizado foi o GROMACS (5.1.4 - CUDA) [27], como método integrador o leap-frog, para o cálculo de forças eletrostáticas o PME [33], para o termostato o algorítimo modificado de Berendsen (Velocity-rescale) [53] e para o barostato o Parrinello-Rahman [37].

SalBIII

A estrutura utilizada para as simulações de dinâmica molecular da SalBIII foi o monômero B do arquivo pdb 5CXO [46]. Foram produzidas quatro simulações com essa proteína, todas com a duração de 5 μ s. e através do software GROMACS (5.1.4 - CUDA) [27]. Sendo essas simulações 1 simulação *all-atom*, com o campo de força Amber03 [40], e 3 simulações *coarse-grained* com o SIRAH [41].

Calmodulina

A estrutura da Calmodulina livre utilizada nesse estudo foi obtida por meio de experimentos de RMN em solução. Seu arquivo no Protein Data Bank (*1DMO*) contém um conjunto de 30 conformações [47] e cada uma dessas estruturas foi utilizada como conformação possível e, consequentemente, estado inicial para as simulações. A Figura 3.1 apresenta este conjunto de estruturas.



Figura 3.1: Representação das estruturas para a Calmodulina livre - PDB *1DMO*.

Para a Calmodulina complexada com cálcio a estrutura encontrada no arquivo do Protein Data Bank é única e, além de começar no resíduo de treonina com o número 5, apresenta somente 148 resíduos, diferentemente dos 149 apresentados pela estrutura livre de íons cálcio e pela sequência relatada pelos dados experimentais de cross-linking. Sendo assim, adicionaram-se na estrutura Calmodulina complexada com cálcio, via *pymol*, os 6 resíduos faltantes (*MADQL-*) para torná-la igual a Calmodulina livre.

Calmodulina livre. As estruturas utilizadas para as simulações de dinâmica molecular da Calmodulina livre foram as 30 conformações no PDB *1DMO*. Foram produzidas duas simulações para cada uma das 30 estruturas, uma de 200 ns com o campo de força *all-atom Amber03* [40] e a outra de 1 μ s com o campo de força *coarse-grained SIRAH* [41].

Calmodulina complexada com cálcio. A estrutura utilizada para as simulações de dinâmica molecular da Calmodulina complexada com cálcio foi a do arquivo pdb 3CLN. Foram produzidas duas simulações de $4.6 \mu\text{s}$ para essa proteína, uma para cada campo de força, *all-atom* Amber03 [40] e *coarse-grained* SIRAH [41].

Capítulo 4

Resultados e Discussão

4.1 SalBIII

Os dados abordados nesta parte do estudo foram fruto de experimentos conduzidos com dois tipos de abordagem de ligação cruzada. A com DSS, ligando resíduos básicos (K–K, K–S, (N terminal)–K e (N terminal)–S). E a com Xplex, ligando os resíduos ácidos e básicos (D–D, E–E, D–E, D–K, D–S, E–K e E–S).

Outro ponto a ser mencionado sobre os dados logo mais analisados é a seleção por qualidade de espectro neles realizada. Esse parâmetro foi computado pelo software SIM-XL [10] em função de sinais de massa-carga compatíveis com os picos presentes no espectro e é denotado como *score*. Os resultados a serem relatados foram obtidos através de um conjunto de ligações cruzadas experimentais com valor de *score* maior que 2. Sendo assim, um conjunto de 182 cross-links foi filtrado do total de 408 reportados pelo software.

4.1.1 Cross-links compatíveis com a estrutura cristalográfica

Na Figura 4.1 são representados os 24 cross-links do conjunto de 182 trabalhado que são compatíveis com a estrutura cristalográfica da SalBIII de acordo com o Topolink [16]. Na Tabela 4.1 são expostos estes mesmos cross-links.

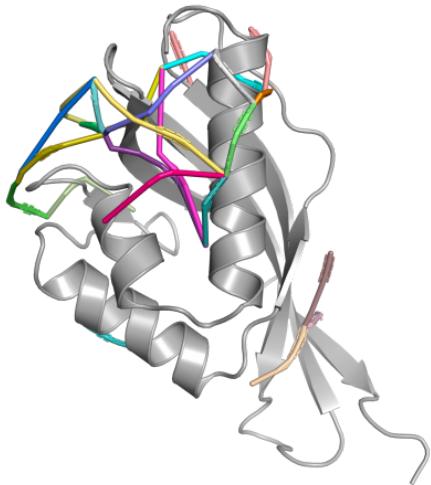


Figura 4.1: Representação dos cross-links compatíveis com a estrutura cristalográfica da SalBIII.

Table 4.1: Ligações cruzadas validadas na estrutura cristalográfica - SalBIII

M1-K8	K6-E9	E13-K17	E32-E46	E53-S57	E86-K99
M1-K17	K6-K113	E13-E28	E32-E111	E62-E86	E111-K113
M1-K113	K8-E9	E13-E111	E32-K113	E66-E86	E111-D115
E4-K8	E9-E13	K17-K113	E37-E46	D75-S78	K113-D115

Com apenas 13% dos cross-links do conjunto filtrado apresentando compatibilidade com a estrutura cristalográfica da proteína, se ergue o questionamento da compatibilidade estrutural requerida pelos outros 87%.

Na simulação, novas estruturas são amostradas e novos cross-links são validados. A Figura 4.2 mostra a evolução da quantidade de cross-links únicos (aqueles que não eram compatíveis com nenhuma das estruturas anteriores) que foram validados ao longo do tempo de simulação de 5 μ s.

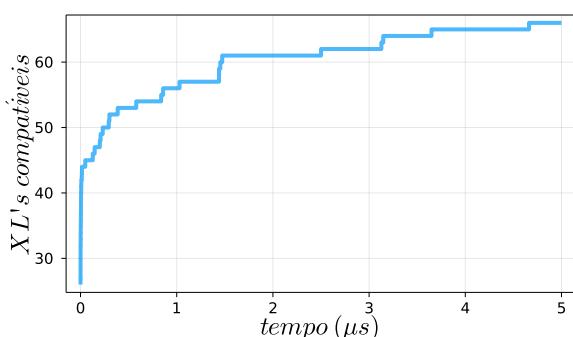


Figura 4.2: Quantidade de cross-links compatíveis com estruturas presentes na simulação em função do tempo de simulação - All-atom.

A Figura 4.3 mostra os cross-links compatíveis com estruturas encontradas na simulação de 5 μ s com campo de força *all-atom* e sua respectiva Probabilidade de Validação em uma Estrutura da Simulação (PVES).

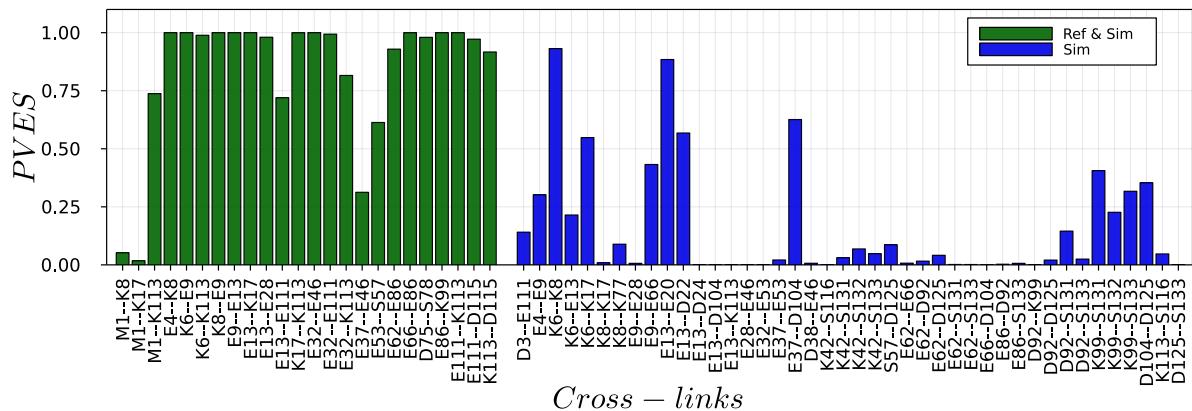


Figura 4.3: Probabilidade de Validação em uma Estrutura da Simulação dos cross-links compatíveis com estruturas presentes na simulação - *All-atom*. – Em verde são destacados os cross-links compatíveis com a estrutura cristalográfica.

Apesar de passar por etapas de minimização de energia que poderiam afetar a validação das ligações cruzadas na estrutura cristalográfica, a Figura 4.3 apresenta todos os 24 cross-links da Tabela 4.1 e, para a maioria deles, a proporção das conformações em que elas se encontram é quase sempre muito próxima de 1, mostrando que a maioria das conformações ao longo da simulação possui distância compatível com essas ligações. Apenas 7 cross-links não apresentam proporções tão elevadas: M1–K8, M1–K17, M1–K113, E13–E111, E32–K113, E37–E46 e E53–S57.

Na Figura 4.4 são representadas as ligações cruzadas mencionadas sobre a estrutura cristalográfica. Nesta figura também foi representada a ligação E62–E86, que faz parte do conjunto com PVES mais elevada, a fins comparativos.

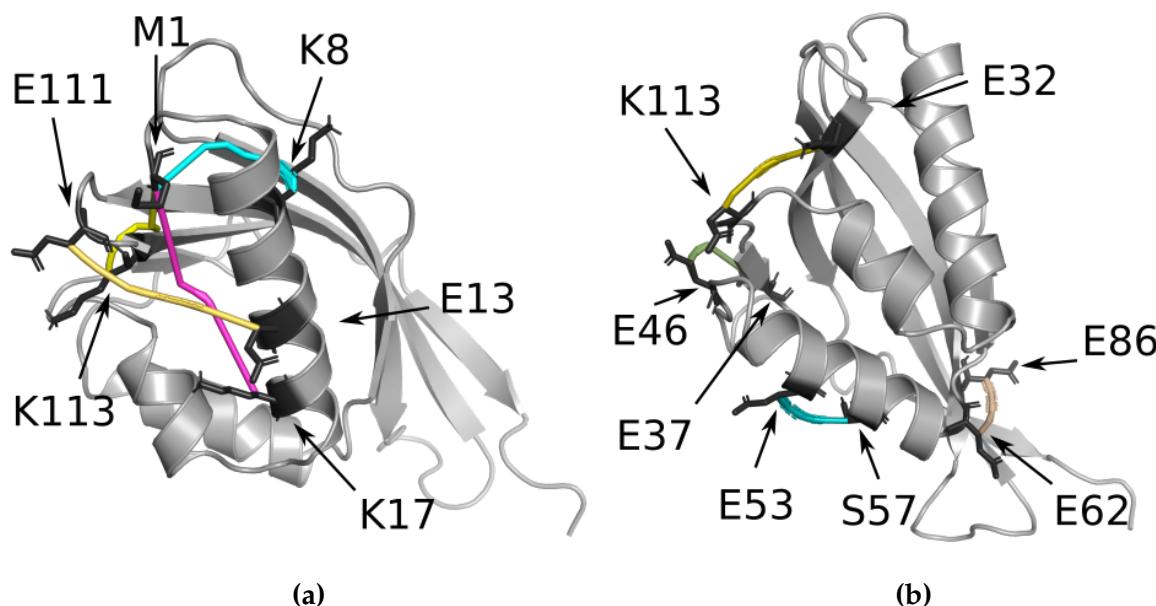


Figura 4.4: Representação de cross-links na estrutura cristalográfica - SalBIII. **(a)** M1–K8, M1–K17, M1–K113 e E13–E111 em ciano, magenta, limão e amarelo, respectivamente. **(b)** E32–K113, E37–E46, E53–57 e E62–E86 em oliva, musgo, azul petróleo e bege, respectivamente.

Para o cross-link E13–E111, a distância entre resíduos é 14.355 Å e seu limite estatístico 15.100 Å, com isso, calculando a diferença desses valores tem-se uma margem de 0.765 Å de tolerância para variações na conformação e ainda validação do cross-link. Entretanto, esse cross-link liga o resíduo de uma α -hélice ao de um loop de folhas- β , considerando a flexibilidade dessa segunda estrutura, é rationalizada a não validação do mesmo em parte das conformações da simulação, já que essa margem não é muito ampla. Tendo uma margem também pequena, 0.600 Å, a validação da ligação cruzada E32–K113 nas estruturas da simulação sofre de maneira semelhante ao que foi argumentado para a E13–E111.

Para as ligações cruzadas envolvendo a metionina 1 (M1), mesmo apresentando margens de tolerância como 5.152 Å (M1–K8) e 4.534 Å (M1–K113), como este resíduo é o N-terminal da estrutura já é esperado uma flexibilidade mais acentuada nesta região, consequentemente fazendo com que as distâncias entre estes aminoácidos varie mais facilmente.

No caso dos resíduos de ácido glutâmico 37 e 46, o valor da distância entre eles é 12.676 Å e a margem de tolerância calculada é 2.424 Å. Apesar de ser uma margem relativamente grande, este cross-link liga dois resíduos do mesmo loop com 8 resíduos entre eles. Sendo assim, fica justificada a movimentação dessa região afetando significativamente a validação desse cross-link ao longo da simulação.

Finalmente, apesar do cross-link E53–S57 se estabelecer em resíduos da mesma α -hélice, sugerindo maior estabilidade estrutural ao longo da simulação, a margem de tolerância para este apresenta o valor de apenas 0.441 Å tornando o cross-link mais sensível mesmo em uma estrutura mais estável.

Ao agrupar as estruturas de acordo com a distância entre os resíduos de um cross-link específico e dividirmos este valor pelo número total de estruturas, podemos visualizar como a PVES de um determinado cross-link varia em função da distância topológica entre os resíduos da ligação cruzada. A Figura 4.5 mostra esta distribuição para os cross-links representados na Figura 4.4.

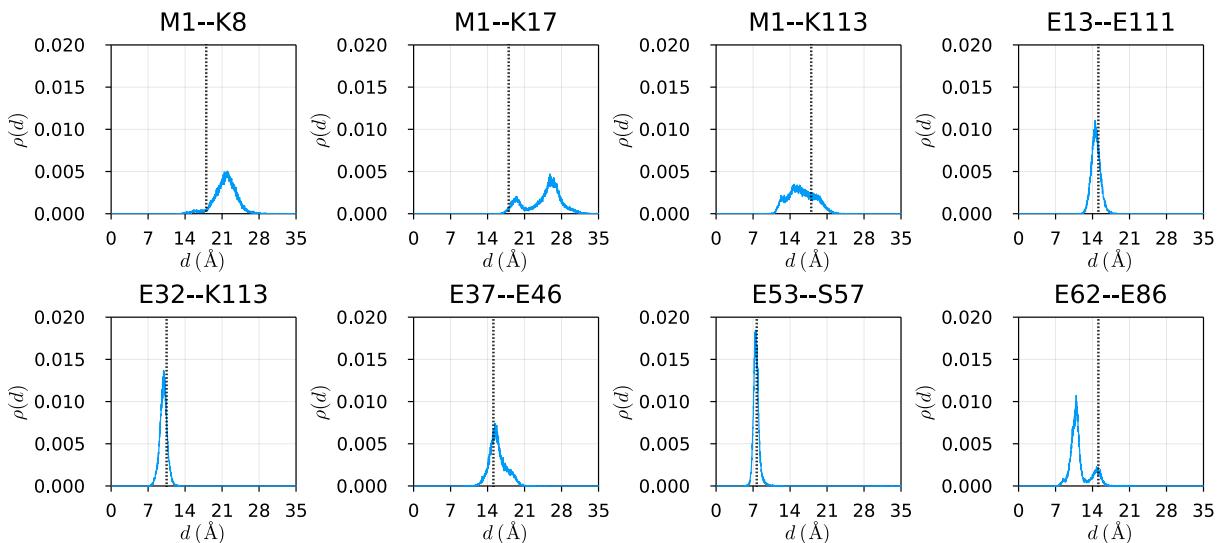


Figura 4.5: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links da Figura 4.4. - A referência pontilhada é o limite estatístico.

A primeira característica destas distribuições que se mostra relevante para a discussão são as larguras das curvas. Se a curva se estabelece sobre uma ampla

faixa de distâncias, isso remete uma flexibilidade na distância entre os resíduos. Esta flexibilidade pode ser referente à porção da estrutura que um dos resíduos se encontra ou referente a ambas as porções.¹ Assim, dois dados destas distribuições são bem pertinentes para a discussão, a largura da curva de distribuição e a posição dessa curva em relação ao limite estatístico de validação (linha pontilhada). A integral à esquerda da linha pontilhada é a porção das conformações que são compatíveis com o cross-link correspondente, PVES do gráfico anterior. A distribuição do cross-link E62–E86 nos mostra porque a PVES dele é tão mais próxima de 1 em comparação com as demais.

Outro ponto interessante é que apesar da ligação cruzada E53–E57 apresentar uma distribuição estreita, sua PVES é 0.6 por conta da sua posição em relação ao limite estatístico.

4.1.2 Cross-links não compatíveis com a estrutura cristalográfica

Cross-links compatíveis com estruturas da simulação all-atom

Na Tabela 4.2 estão apresentados os cross-links não validados na estrutura cristalográfica mas encontrados na simulação *all-atom*, cross-links com barras azuis na Figura 4.3.

¹Para ser mais rigoroso, a largura da curva é mais diretamente relacionada com a aproximação ou afastamento dos resíduos se a distribuição correspondente fosse em termos de distância euclidiana. Como a argumentação sobre flexibilidade nesta secção é uma questão mais ampla, considerando também o arranjo de cadeias laterais próximas, e a integração sobre a curva é mais relevante, é a distribuição com distâncias topológicas que é apresentada. Todavia, a distribuição euclidiana se encontra na Figura II.1.

Table 4.2: Ligações cruzadas não validadas na estrutura cristalográfica mas compatíveis com estruturas encontradas na simulação *all-atom* - SalBIII.

D3–E111	K8–K77	E13–D104	D38–E46	E62–E66	E86–D92	K99–S131
E4–E9	E9–E28	E13–K113	K42–S116	E62–D92	E86–S133	K99–S132
K6–K8	E9–E66	E28–E46	K42–S131	E62–D125	D92–K99	K99–S133
K6–E13	E13–E20	E32–E53	K42–S132	E62–S131	D92–D125	D104–D125
K6–K17	E13–D22	E37–E53	K42–S133	E62–S133	D92–S131	K113–S116
K8–K17	E13–D24	E37–D104	S57–D125	E66–D104	D92–S133	D125–S133

Somando as 42 ligações cruzadas na tabela acima aos 24 cross-links compatíveis com estrutura cristalográfica, totalizam-se 66 cross-links encontrados na simulação, 36% do conjunto filtrado.

Para os cross-links compatíveis com estruturas da simulação mas não validados na estrutura cristalgráfica, fizemos uma análise semelhante porém na forma de um gráfico de barras agrupadas, considerando a PVES, a diferença entre a distância topológica encontrada pelo Topolink na estrutura cristalgráfica e o limite estatístico de validação, que chamaremos de Distância Excedida Topológica (DET) (o negativo da margem de tolerância), e a diferença entre a distância euclidiana encontrada, também na estrutura cristalgráfica, e o limite estatístico de validação, que chamaremos de Distância Excedida Euclidiana (DEE). A Figura 4.6 coloca estes dados em forma de um gráfico de barras.

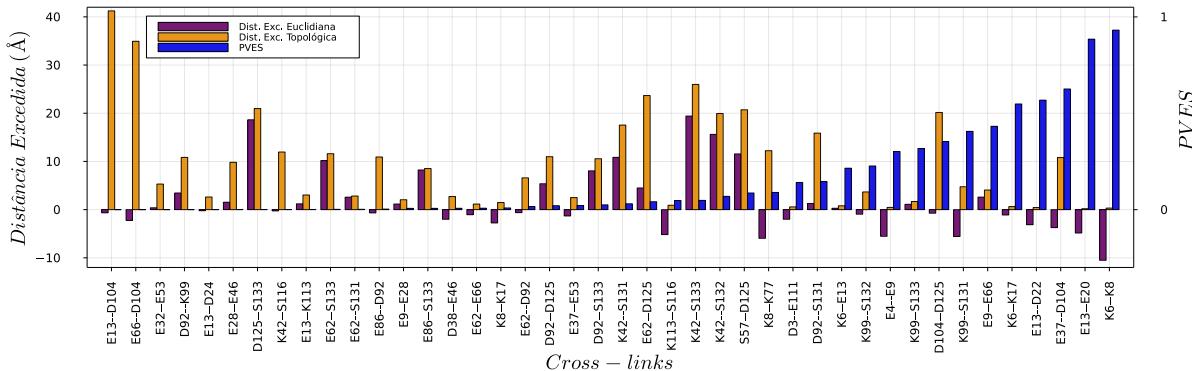


Figura 4.6: Distância Excedida Euclidiana, Distância Excedida Topológica e PVES dos cross-links compatíveis com estruturas presentes na simulação mas não compatíveis com a estrutura cristalográfica - *All-atom*. – Em vinho, amarelo e azul, respectivamente. Somente a PVES tem correspondência com o eixo y direito, as demais grandezas se relacionam com o da esquerda.

Pela razão dos resíduos estarem euclidianamente próximos na estrutura cristalográfica, ao decorrer da simulação um caminho acessível ao solvente foi possibilitado, explicando assim valores significativos de PVES. Na Figura 4.6 podemos ver que as ligações cruzadas que apresentam maiores de PVES são aquelas que têm DEE negativas.

Há também na Figura 4.6, ligações cruzadas as quais apresentam DEE positivas. Ou seja, na estrutura cristalográfica os resíduos são euclidianamente mais distantes que o limite estatístico de validação. Contudo, estes cross-links foram compatíveis com estruturas da simulação em números significativos. Os cross-links E9–E28, E9–66, K6–E13, K42–S133, K99–S133 e S57–D125 são exemplos destes casos.²



Figura 4.7: Representação das estruturas secundárias da SalBIII ao longo da sequencia de aminoácidos - Sendo α -hélice e folha- β representadas em vermelho e amarelo, respectivamente. (Fonte: Protein Data Bank, 5CXO)

A compatibilidade com estruturas da simulação para os cross-links K42–S133 e K99–S133 é influenciada pela movimentação do C-terminal da estrutura (serina 133). A

²A distribuição das estruturas em função da distância topológica entre esses resíduos para todos os cross-links da Figura 4.6 se encontram na Figura II.3.

aproximação do C-terminal para a conexão destes resíduos é o que se observa ao longo da simulação, promovendo distâncias compatíveis com o limite estatístico de validação. Mesmo com valores significativos de DET, a mobilidade do C-terminal é suficiente para superá-los. A Figura 4.8b apresenta os cross-links em questão sobre a estrutura cristalográfica e a Figura 4.9 mostra a distribuição da PVES função da distância entre estes resíduos.

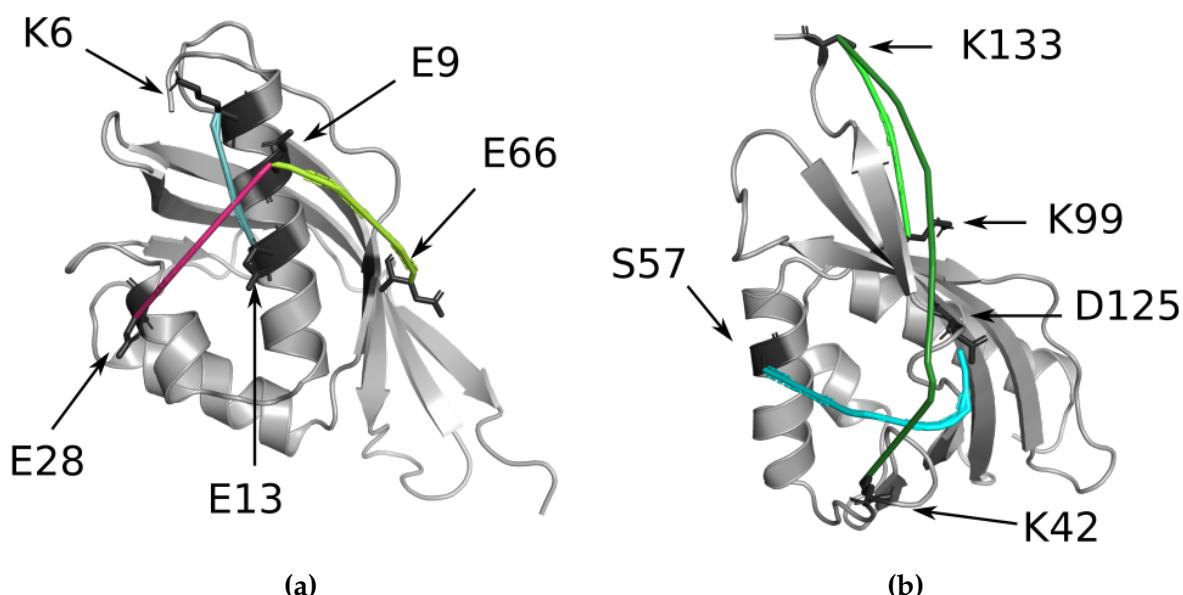


Figura 4.8: Representação de cross-links não validados na estrutura cristalográfica mas compatíveis com estruturas da simulação - SalBIII. **(a)** K6–E13, E9–E28 e E9–E66 em gelo, magenta e verde limão, respectivamente. **(b)** K42–S133, S57–D125, e K99–S133 em verde floresta, azul piscina e verde, respectivamente.

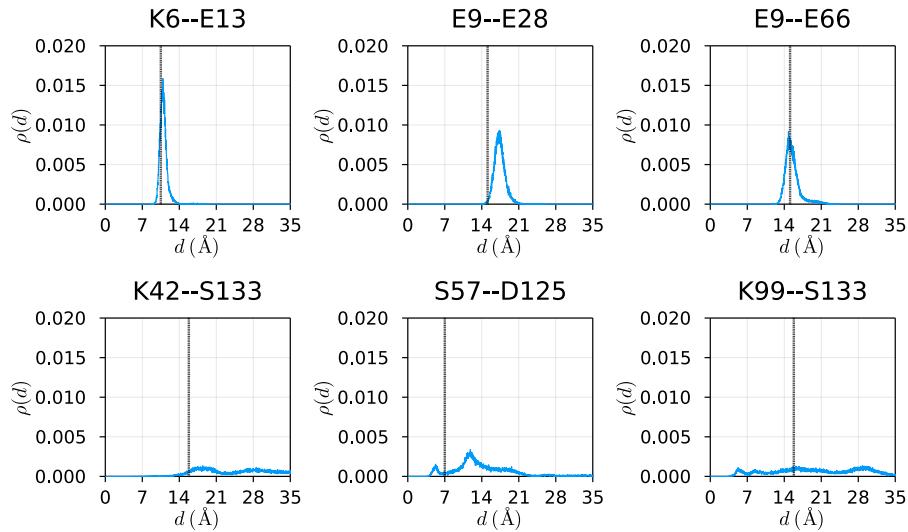


Figura 4.9: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links da Figura 4.8. - A referência pontilhada é o limite estatístico.

Na Figura 4.9 vemos que a distância entre os resíduos da ligação cruzada S57–D125 também apresenta uma distribuição larga como os casos com a serina 133, podendo ser validada em todas as estruturas agrupadas a esquerda do limite estatístico. Todavia, diferentemente da situação do C-terminal, vemos na simulação que é a mobilidade das duas partes, região do resíduo 57 e do resíduo 125, que influenciam na largura da curva de distribuição.

Localizado em uma α -hélice e no resíduo 9, o ácido glutâmico demonstra nas simulações uma menor mobilidade que o caso anterior, mesmo estando próximo ao N-terminal da proteína. Na Figura 4.9 podemos observar que, mesmo não sendo uma distribuição estreita como encontramos para K6–E13, as curvas referentes as ligações E9–E28 e E9–E66 não são tão amplas em sua faixa de ocorrência majoritária como as distribuições envolvendo a serina 133 (S133). Sendo assim, a razão pela qual estes cross-links apresentam compatibilidade com estruturas da simulação é a proximidade entre os resíduos.

A simulação nos mostra que para o cross-link E9–E66, que apresentou uma DET de 4.07 Å, a distância entre os resíduos mais comum é bem próxima ao limite

estatístico de validação. Já para a ligação cruzada E9–E28, com DET de 2.01 Å, podemos observar que esta é, aproximadamente, a diferença entre as distâncias topológicas mais comuns dos resíduos de ácido glutâmico (9 e 28) na simulação e a sua validação em estruturas foi acarretada por raras perturbações.

O cross-link K6–K13 é localizado na α -hélice do N-terminal. Como esta é uma estrutura significativamente rígida na proteína, essa ligação cruzada se torna compatível com as estruturas da simulação pois, além de apresentar uma DET pequena, 0.8 Å, as distâncias topológicas mais comuns entre estes resíduos na simulação são próximas o suficiente do limite estatístico para promover a validação do cross-link com as pequenas deformações da estrutura.

Energia livre. Apesar de certos cross-links serem compatíveis com poucas estruturas da simulação, convertendo a PVES para unidades de energia (Equação 3.2), pode-se analisar com uma nova perspectiva as variáveis energéticas envolvidas no experimento de espectrometria.

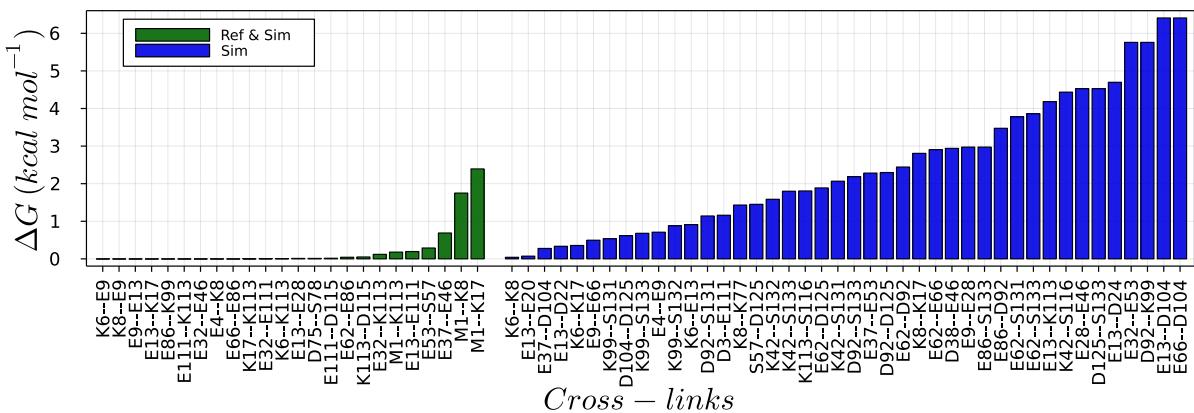


Figura 4.10: Probabilidade de Validação em uma Estrutura da Simulação convertida para unidades de energia livre para cada cross-link - *All-atom*. – Em verde são destacados os cross-links validados na estrutura cristalográfica.

O cross-link E13–D104 é compatível com estruturas com ΔG de 6.4 kcal mol⁻¹. Esse valor de energia significa que cerca de 1 a cada 50,000 estruturas são compatíveis com esta ligação. Considerando que o limite de detecção comum de um

espectrômetro de massas chega facilmente na ordem de $1/10^6$, é justificável afirmar que, nestas condições, a presença de uma estrutura com este nível de particularidade é numerosa o suficiente para acarretar na detecção de seu cross-link correspondente experimentalmente, se a mesma for ligada pelo ALC.

Cross-links compatíveis com estruturas da simulação coarse-grained

Combinados, os resultados das 3 simulações *coarse-grained* contemplam estruturas compatíveis com 105 cross-links. Sendo 58 destes cross-links comuns com a simulação *all-atom* e 47 novos. Totalizando 113 ligações cruzadas compatíveis com estruturas de simulação, 61% do conjunto filtrado.

A Figura 4.11 mostra os valores de PVES para cada cross-link validado nas simulações *coarse-grained*.

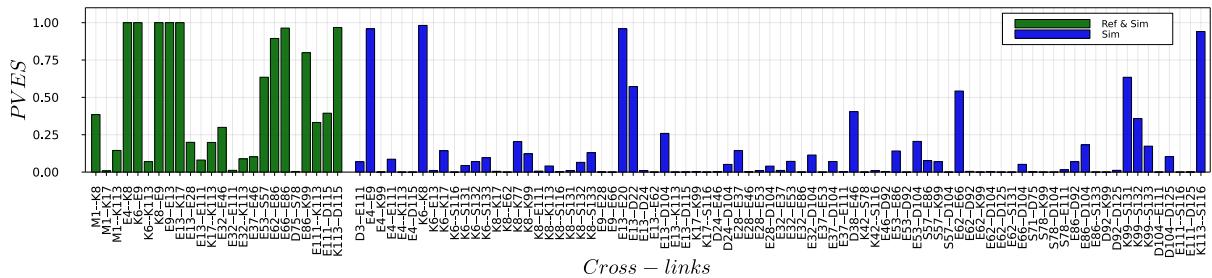


Figura 4.11: Probabilidade de Validação em uma Estrutura da Simulação dos cross-links compatíveis com estruturas presentes na simulação - *Coarse-grained*. – Em verde são destacados os cross-links compatíveis com a estrutura cristalográfica.

Na Figura 4.12, observam-se os valores de DET para os cross-links validados nas simulações *all-atom* e *coarse-grained*, sendo que para este segundo conjunto foram excluídos aqueles já reportados pelo primeiro.

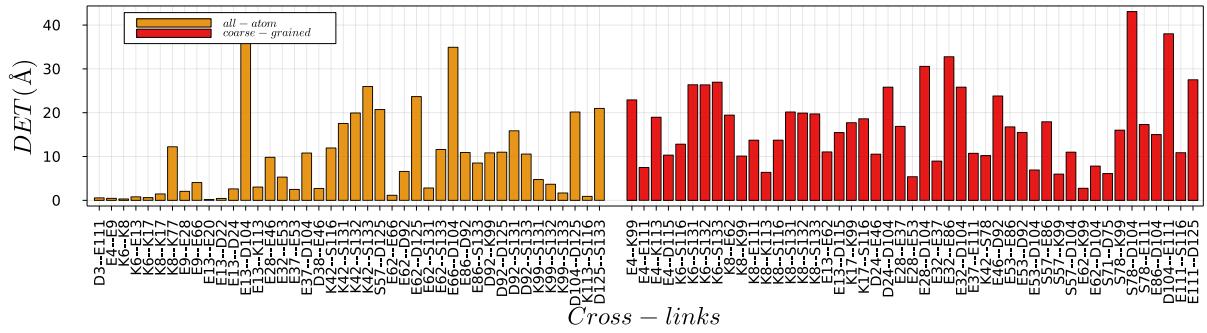


Figura 4.12: Distância Excedida Topológica dos cross-links compatíveis com estruturas presentes na simulação *all-atom* e *coarse-grained*, em amarelo e vermelho, respectivamente.

Comparando os dois grupos da Figura 4.12 percebe-se que os valores de DET apresentados pelo primeiro são, em média, menores que os reportados pelo segundo, 9 ± 9.85 e 17 ± 8.88 Å, respectivamente. Implicando uma maior variabilidade conformacional nas simulações *coarse-grained*, satisfazendo assim cross-links com maior DET.

4.1.3 O conjunto de cross-links compatíveis com as estruturas das simulações

Apesar das simulações *coarse-grained* apresentarem um número maior de cross-links validados, 8 conexões reportadas pela simulação *all-atom* não fazem parte desse conjunto. São essas: K42–S131, K42–S132, K42–S133, S57–D125, E62–S133, D92–S131, D92–S133 e D125–S133.

A Figura 4.13 é um diagrama que apresenta as coordenadas (x,y) de cada cross-link validado nas simulações, sendo x e y os respectivos resíduos ligados. Aproveitando a simetria do mapa, na metade diagonal superior são representados os cross-links validados nas simulações *coarse-grained* e na metade inferior os validados na simulação *all-atom*.

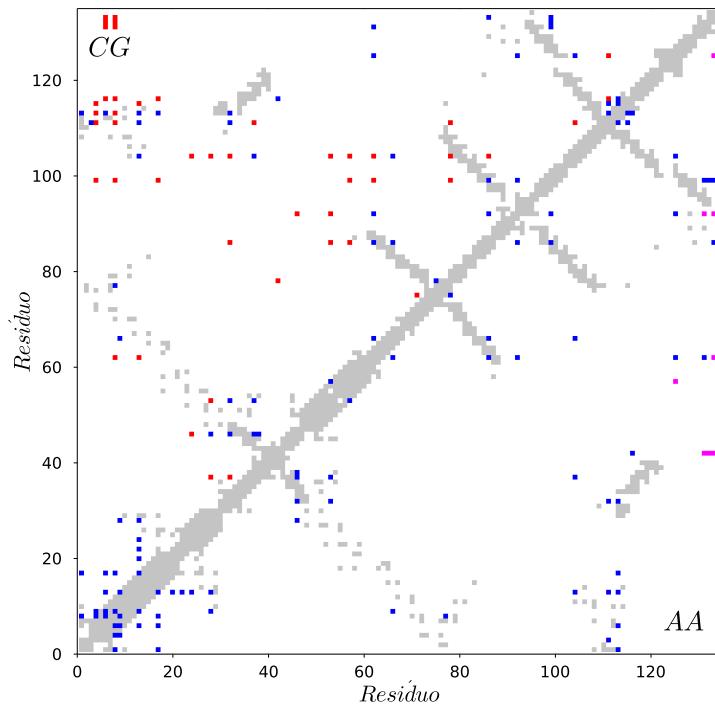


Figura 4.13: Diagrama de coordenadas dos cross-links validados em simulação. – Em azul são representados os cross-links comuns entre as simulações dos dois campos de força, em vermelho são os cross-links validados apenas nas simulações *coarse-grained*, em magenta aqueles validados apenas na simulação *all-atom* e em cinza o par de resíduos que possui distância igual ou menor que 8.0 Å na estrutura cristalográfica (mapa de contatos).

4.1.4 Qualidade dos filtros e o conjunto de cross-links

A Tabela 4.3 lista os cross-links encontrados experimentalmente, porém não validados em nenhuma das simulações.

Table 4.3: Cross-links experimentais não validados em simulações

D3–E53	E9–D125	E20–E37	E46–S131	S57–K77	D75–E86	D104–K113
D3–E62	E13–E32	D22–E37	E53–E66	S57–D92	D75–D92	E111–S131
E4–D24	E13–E37	D24–E86	E53–S71	S57–E111	S78–E86	K113–D125
E4–E53	E13–E53	E32–S57	E53–K99	S57–K113	S78–D92	K113–S131
E4–E62	E13–S57	E32–S78	E53–E111	E62–E111	E86–E111	K113–S132
E4–S78	E13–S116	E32–K99	E53–D115	E62–D115	D92–D104	K113–S133
K6–S57	E13–D125	E37–K113	E53–S116	E62–S116	D92–K113	D115–D125
E9–K17	K17–E32	E46–S57	S57–E62	E66–D92	K99–D104	S116–D125
E9–D22	K17–E37	E46–E62	S57–E66	E66–E111	K99–E111	D125–S132
E9–D115	K17–S133	E46–S116	S57–D75	S71–E86	K99–S116	

Com as ligações cruzadas da tabela acima, completa-se em totalidade o conjunto filtrado ($\text{score} \geq 2$). Não sendo compatível com nenhuma estrutura das simulações, este conjunto de cross-links não compatíveis apresenta uma DET média de $23 \pm 10.40 \text{ \AA}$, indicando uma variabilidade conformacional ainda maior que a observada nas simulações *coarse-grained*.

Seria a quantidade de cross-links não compatíveis relacionada a qualidade do parâmetro utilizado para filtrar o conjunto de ligações cruzadas? Filtros com parâmetros de qualidade espectral piores gerariam conjuntos com mais cross-links não compatíveis com a estrutura cristalográfica? E com as estruturas da simulação?

A Figura 4.14 (A-C) mostra a correlação entre a qualidade do filtro e a fração do conjunto filtrado compatível com a estrutura cristalográfica. De forma equivalente, na mesma figura (I-III) faz a mesma representação para a compatibilidade com as estruturas da simulação *all-atom*.

Os parâmetros de qualidade disponíveis para filtragem de espectros no programa SIM-XL [10] que serão analisados são: a contagem de espectros, a contagem de espécies e o score.

A contagem de espectros se relaciona com a quantidade de espectros que corroboraram para a identificação do cross-link, a contagem de espécies com a quantidade de diferentes peptídeos que identificaram o mesmo cross-link e o score é a "nota" que o software dá aos espectros relacionados a identificação do cross-link baseado na correlação entre a detecção experimental e a previsão teórica de fragmentação da proteína.

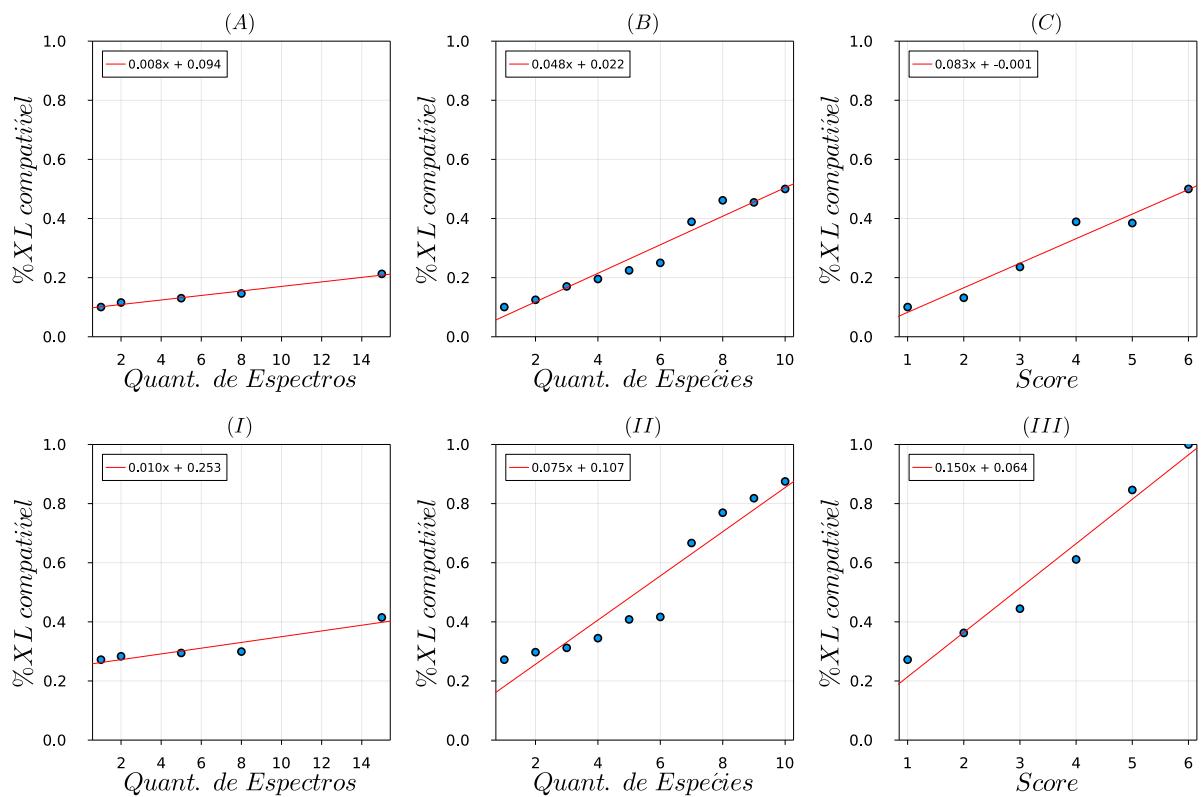


Figura 4.14: Correlações entre os parâmetros de qualidade de espectros e a compatibilidade dos conjuntos gerados (A-C) com a estrutura cristalográfica (I-III) com as estruturas da simulação *all-atom*. – O quadrado dos parâmetros de correlação (R^2) para as distribuições de pontos são 0.98, 0.95, 0.95, 0.88, 0.91 e 0.97 para os gráficos A, B, C, I, II e III, respectivamente.

A quantidade de espectros mostra ser a distribuição mais fortemente correlacionada entre os três parâmetros no caso da estrutura cristalográfica, porém a com correlação mais fraca no caso das estruturas da simulação *all-atom*. Esta distribuição

também apresenta o menor coeficiente angular da reta, para ambos os casos de comparação, significando que o aumento da quantidade de espectros acarreta na menor variação da porcentagem de cross-links compatíveis dentre os três parâmetros.

O score é o parâmetro que apresenta maior coeficiente de angular, para ambos os casos. O coeficiente de correlação de 0.95 e 0.97 para os casos estudados torna o parâmetro de score o mais robusto para escolher cross-links compatíveis com a estrutura cristalográfica e a variabilidade conformacional apresentada pela simulação *all-atom*.

4.2 Calmodulina

A Tabela 4.4 mostra os cross-links reportados para o experimento de ligação cruzada por espectrometria de massas para as espécies da Calmodulina livre (exclusivamente), Calmodulina complexada com cálcio (exclusivamente) e os comuns entre os experimentos.

Um ponto importante a ser mencionado é que os dados apresentados na tabela abaixo são considerados pelo grupo experimental como dados brutos com pouco tratamento, ou seja, mesmo cross-links oriundos de espectros com baixa qualidade foram reportados neste conjunto [51].

Table 4.4: Cross-links experimentais para Calmodulina livre (exclusivamente), Calmodulina complexada com cálcio (exclusivamente) e comuns entre os experimentos.

Calm. livre		Calm. Ca ²⁺	Calm. livre ∩ Calm. Ca ²⁺			
K14–S39	S18–K95		M1–K76	K22–K116	K76–K116	
K14–K76	K31–K78	M1–K95	M1–K78	K31–K95	K78–K95	
K14–K78	K31–S82	S18–K116	S18–K22	K31–S102	K78–S102	K95–K116
K14–S82	K31–K116	K22–S102	K22–K31	S39–K95	K78–K116	S102–K116
K14–S102	S39–K78		K22–K76	K76–K78	S82–K95	
S18–K76	K78–S82		K22–K78	K76–K95	S82–K116	
S18–K78			K22–K95	K76–S102	K95–S102	

Ao todo, são 36 cross-links para o experimento da Calmodulina livre e 26 para o da Calmodulina complexada com cálcio.

Quando se compara a análise dos dados feita pelo grupo experimental [51] com a Tabela 4.4, percebem-se algumas diferenças como consequência do tratamento posterior. Para a experimentalista, considerando a qualidade dos espectros e apenas cross-links intramoleculares, apenas 2 cross-links são exclusivos da Calmodulina livre e todos os encontrados para Calmodulina complexada com cálcio são comuns com a livre. Os cross-links exclusivos são: K76–K116 e S82–K95. Constatando a presença de ambos os cross-links nos dados brutos de ambos os experimentos, conclui-se que a qualidade de espectro desses cross-links para a forma complexada foi discriminatória para a experimentalista, tornando essas conexões particulares da Calmodulina livre em suas considerações finais.³ Já para os cross-links comuns entre os dois experimentos, foram utilizados os dados das espécies relacionadas aos cross-links K76–K78, K22–K31 e K95–S102 para a quantificação e diferenciação das duas conformações [51].

Importante ressaltar que, tendo em vista o objetivo de comparação dos cross-links validados nas diferentes espécies da Calmodulina, a busca de compatibilidade com uma estrutura sempre leva em consideração a união dos conjuntos apresentados

³O parâmetro para a seleção destes cross-links não é descrito na referência

na Tabela 4.4, mesmo que experimentalmente para uma dada espécie não tenham sido identificados todos os cross-links testados.

4.2.1 Cross-links compatíveis com as estruturas de referência

A Tabela 4.5 mostra as ligações cruzadas validadas nas estruturas de referência para a Calmodulina livre (30 estruturas) e Calmodulina complexada com cálcio (1 estrutura).

Table 4.5: Cross-links validados nas respectivas estruturas de referência para a Calmodulina livre (30 conformações de RMN) e Calmodulina complexada com cálcio (1 conformação cristalográfica)

Calmodulina livre	Calmodulina Ca^{2+}
M1–K76	K76–K116
M1–K78	K78–S82
M1–K95	K78–K116
S18–K22	S82–K116
K22–K31	K95–S102
K76–K78	

Calmodulina livre

Para a Calmodulina livre, que possui uma estrutura obtida por RMN em solução e para a qual foram propostas 30 conformações, a Figura 4.15 mostra a fração destas conformações que satisfaz cada cross-link.

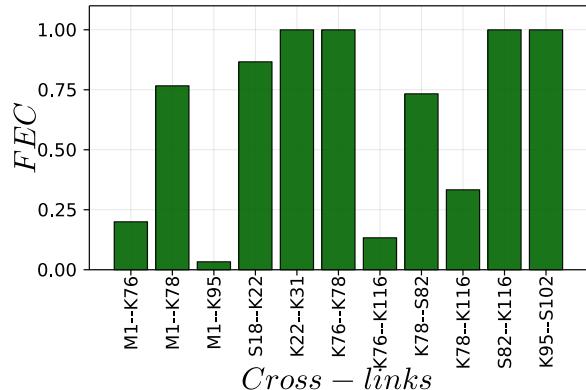


Figura 4.15: Fração de Estruturas Compatíveis para cada cross-link validado nas estruturas de referência, Calmodulina livre (30 conformações).

A Figura 4.16 mostra a representação dos cross-links listados na Tabela 4.5 em algumas das estruturas do conjunto da Calmodulina livre em que são compatíveis. São compatíveis com as possíveis estruturas da Calmodulina livre 11 ligações cruzadas. Dentre estas, a ligação cruzada M1–K95, não foi reportada nos dados experimentais desta espécie, então é contemplado 28% do conjunto experimental mais 1 cross-link.

Ao analisar a disposição de alguns dos cross-links na Figura 4.16, pode-se facilmente identificar a perturbação nestes promovida pela complexação com íons cálcio e, consequentemente, formação da α -hélice entre os domínios. São eles: K78–K116 (Amarelo em 4.16a), M1–K95 (Azul em 4.16b) e K76–K116 (Verde em 4.16c). Uma vez que estes cross-links ligam os dois domínios da proteína, certamente serão impactados pelo aumento da distância entre essas duas porções mediante a formação da hélice.

Na Figura 4.19 são observados os 6 cross-links validados nessa conformação, porém acredita-se que a restrição M1–K78 não se torne tão relevante dentro da simulação devido à potencial mobilidade do N-terminal. Diferentemente de K18–K22, K22–K31 e K95–S102 que se localizam próximas aos sítios de coordenação com o metal e K76–K78 e K78–S82 que ligam resíduos da mesma α -hélice.

Sendo assim, além da confirmação dos comportamentos sugeridos acima, a dinâmica molecular da proteína possibilitará o entendimento dos cross-links que não apresentam dependência direta com o aumento de distância entre os domínios promovido pela mudança de conformação. São eles: M1–K76 e S82–K116.

Outro ponto de interesse nos dados da dinâmica é a possível validação da restrição S82–K95, mencionada no trabalho do grupo experimental, que não foi validada nas estruturas de referência.

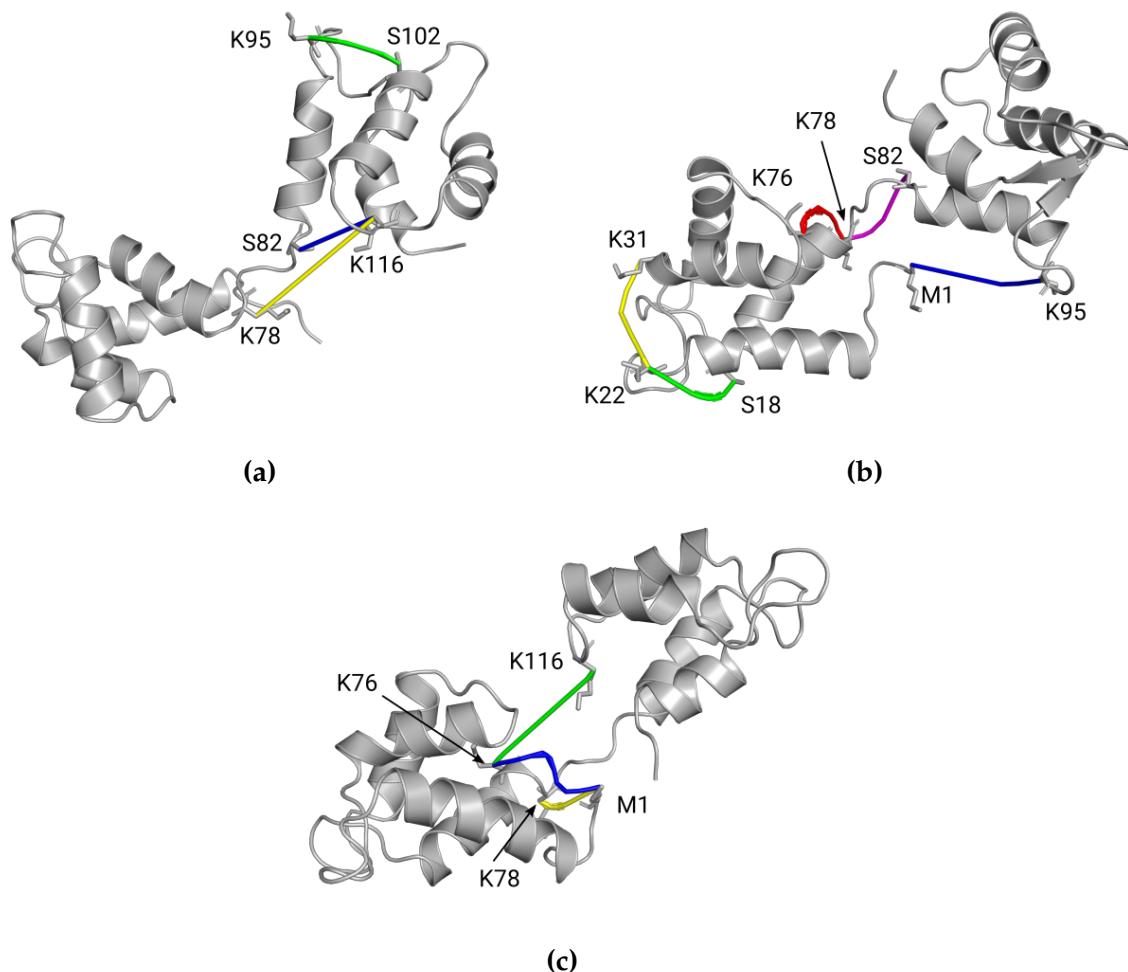


Figura 4.16: Representação de cross-links nas estruturas de referência - Calmodulina livre. **(a)** Conformação 7 - Destacado em azul o cross-link S82–K116, amarelo o K78–K116 e verde o K95–S102. **(b)** Conformação 25 - Destacado em azul o cross-link M1–K95, amarelo o K22–K31, verde o S18–K22, vermelho o K76–K78 e magenta o K78–S82. **(c)** Conformação 29 - Destacado em azul o cross-link M1–K76, amarelo o M1–K78 e verde o K76–K116.

A Figura 4.17 mostra a PVES dos cross-links compatíveis com as estruturas das simulações *all-atom* da Calmodulina livre.

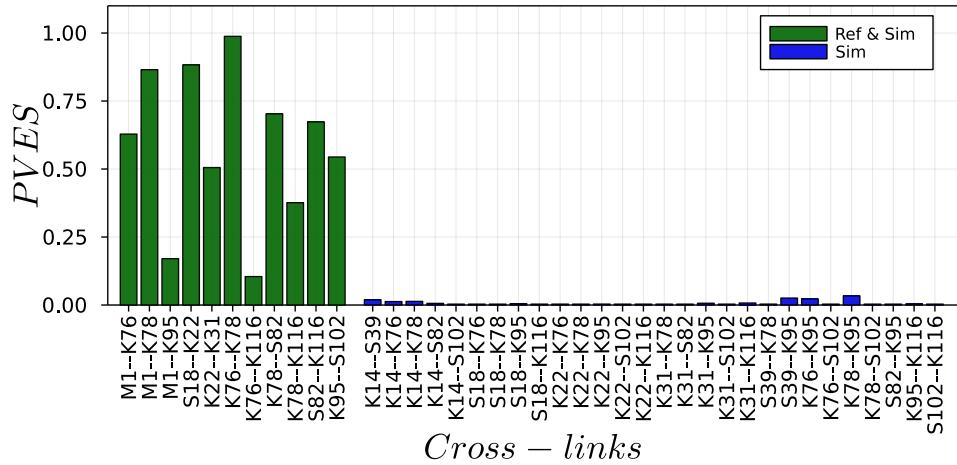


Figura 4.17: PVES para cada cross-link compatível com as estruturas das simulações *all-atom* da Calmodulina livre.

Com o auxílio da Figura 4.16, podemos identificar na Figura 4.17 que os cross-links compatíveis com as estruturas das simulações *all-atom* que apresentam PVES mais distantes de 1 são aqueles que conectam resíduos de domínios diferentes. São eles: M1–K95, K76–K116 e K78–K116.

Os casos M1–K76 e K22–K31 são um pouco mais intrigantes e não há razão aparente para a baixa PVES. Ao analisarmos as distribuições de PVES em função da distância entre os resíduos para cada cross-link deste grupo compatível com as estruturas de referência, Figura 4.18, obtemos novas informações.

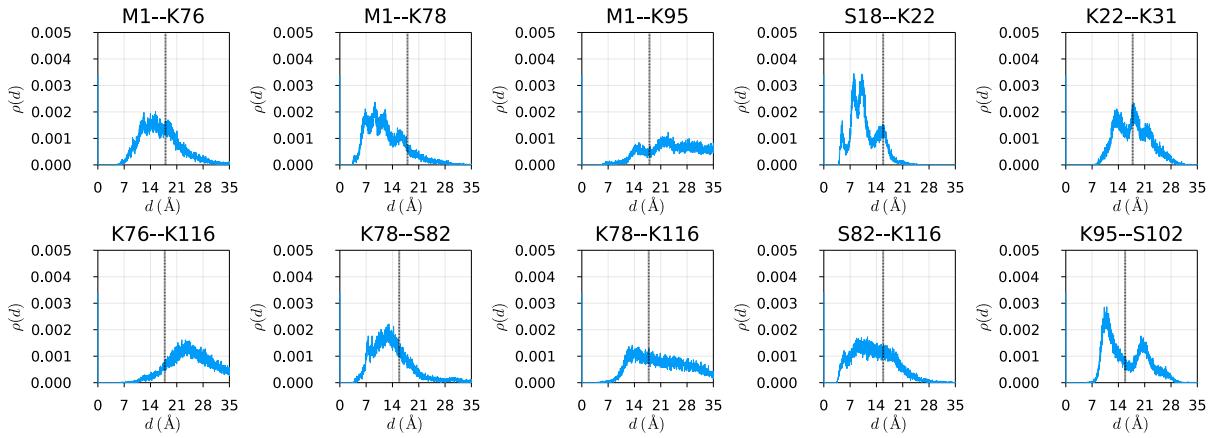


Figura 4.18: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links compatíveis com as estruturas de referência. - Calmodulina livre, *all-atom*. A referência pontilhada é o limite estatístico.

Por conta do maior número de estruturas iniciais (1 simulação de 200 ns para cada uma das 30 estrutura de referência), podemos perceber picos nas distribuições, indicando múltiplas regiões de maior densidade de estruturas. O cross-link S18–K22 é um exemplo de distribuição com esta característica.

Neste conjunto de cross-links temos distribuições predominantemente largas que indicam a grande variação das distâncias encontradas entre seus respectivos resíduos ao longo das simulações. Sendo as ligações cruzadas M1–K76 e K22–K31 praticamente centralizadas no limite estatístico e significativamente simétricas, a PVES próxima de 0.5 é justificada.

Vemos também que a ligação S82–K95, reportada nos dados experimentais mas não validada nas estruturas de referência foi contemplada pela compatibilidade com as estruturas da simulação.

Calmodulina complexada com cálcio

A Figura 4.19 mostra a representação dos cross-links listados na Tabela 4.5 na estrutura cristalográfica da Calmodulina complexada com cálcio. São compatíveis com a estrutura cristalográfica da Calmodulina complexada com cálcio 6 ligações cruzadas. Dentro destas, a ligação cruzada K78–S82 não foi reportada nos dados experimentais

para esta espécie. Com estes números, é contemplado 19% do conjunto experimental mais 1 cross-link.

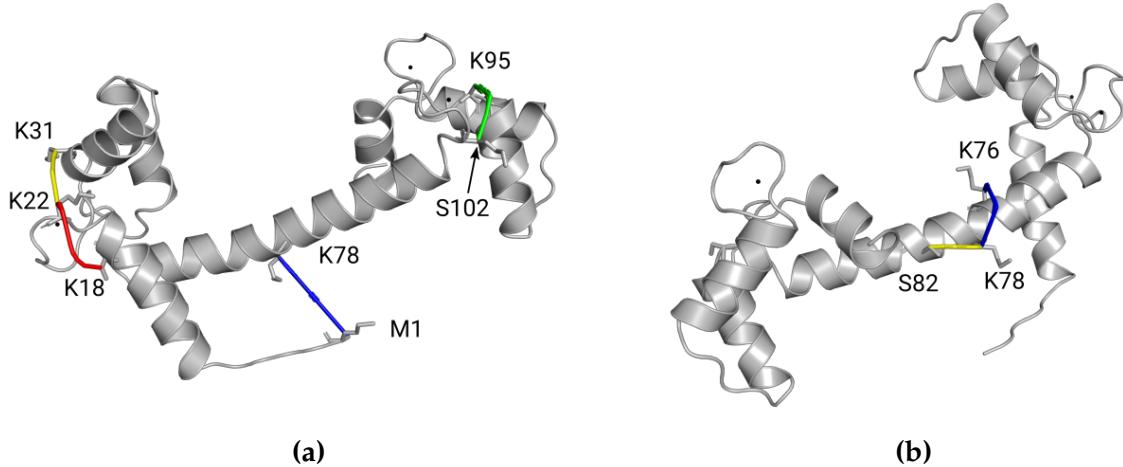


Figura 4.19: Representação de cross-links na estrutura de referência - Calmodulina complexada com cálcio. **(a)** Destacado em azul o cross-link M1–K78, em amarelo o K22–K31, em verde o K95–S102 e em vermelho o S18–K22. **(b)** Destacado em azul o cross-link K76–K78 e em amarelo o K78–S82.

A Figura 4.20 mostra a PVES dos cross-links compatíveis com as estruturas da simulação *all-atom* da Calmodulina complexada com cálcio.

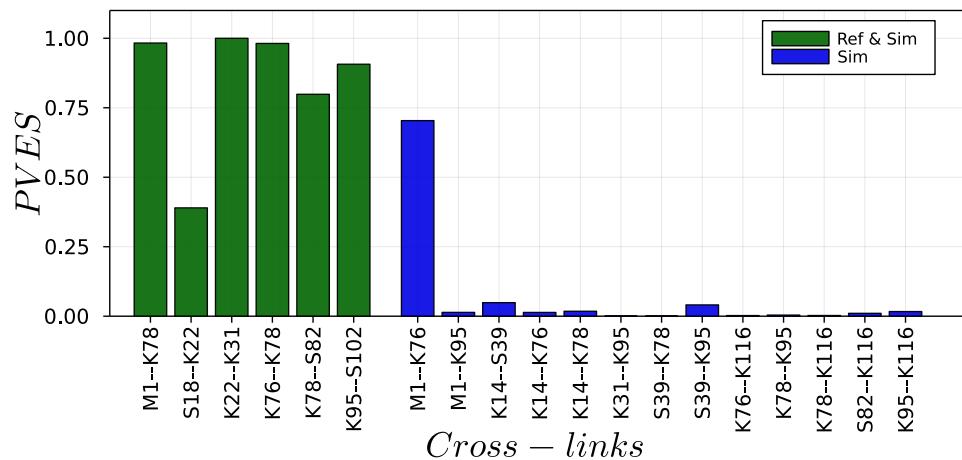


Figura 4.20: PVES para cada cross-link compatível com as estruturas das simulações *all-atom* da Calmodulina complexada com cálcio.

Situadas nos sítios de complexação com cálcio (Figura 4.19a), as ligações cruzadas K22–K31 e K95–S102 possuem PVES próximos de 1. Entretanto, para o cross-link S18–K22, que também se localiza nesta região, a PVES é significativamente menor. A Figura 4.21, que apresenta as distribuições da PVES para os cross-links em questão, deixa evidente que, apesar de apresentar uma curva muito próxima a forma de uma Gaussiana, as distâncias mais comuns encontradas para a distância entre o resíduo de serina 18 (S18) e lisina 22 (K22) são muito próximas do limite estatístico de validação.

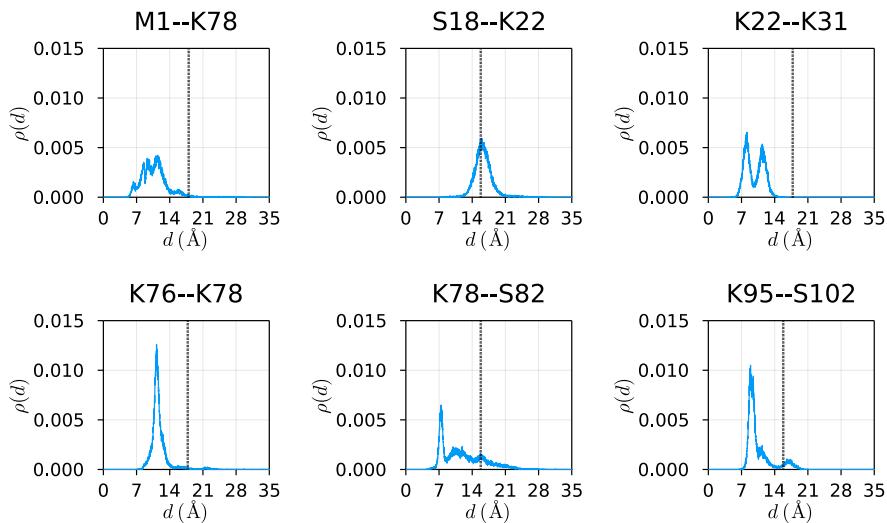


Figura 4.21: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links compatíveis com a estrutura cristalográfica. - Calmodulina complexada com cálcio, *all-atom*. A referência pontilhada é o limite estatístico.

Na Figura 4.21 podemos observar também a mobilidade do N-terminal através da distribuição correspondente à ligação cruzada M1–K78 e a distribuição estreita correspondente ao cross-link K76–K78 localizado na α -hélice inter-domínio.

Por conta da grande variação da distância apresentada pela distribuição o cross-link K78–S82 aparenta indicar uma perda da estrutura secundária, uma vez que a ligação cruzada se localiza em uma α -hélice. Todavia, a distância destas distribuições é a topológica. Quando consultamos a distribuição por distância euclidiana para essa ligação cruzada (Figura II.4), pode-se confirmar que este não é o caso.

4.2.2 Cross-links não compatíveis com a estrutura de referência

A Tabela 4.6 mostra as ligações cruzadas validadas nas estruturas das simulações para a Calmodulina livre e Calmodulina complexada com cálcio (barras azuis nas Figuras 4.17 e 4.20).

Table 4.6: Cross-links compatíveis com estruturas das simulações *all-atom* mas não validados nas estruturas de referência para a Calmodulina livre (30 estruturas de RMN) e Calmodulina complexada com cálcio (1 estrutura cristalográfica)

Calmodulina livre				Calmodulina Ca^{2+}	
K14–S39	S18–K95	K31–K78	K76–K95	M1–K76	S39–K95
K14–K76	S18–K116	K31–S82	K76–S102	M1–K95	K76–K116
K14–K78	K22–K76	K31–K95	K78–K95	K14–S39	K78–K95
K14–S82	K22–K78	K31–S102	K78–S102	K14–K76	K78–K116
K14–S102	K22–K95	K31–K116	S82–K95	K14–K78	S82–K116
S18–K76	K22–S102	S39–K78	K95–K116	K31–K95	K95–K116
S18–K78	K22–K116	S39–K95	S102–K116		S39–K78

Calmodulina livre

As ligações cruzadas reportadas na Tabela 4.6 para a Calmodulina livre, são aquelas representadas em azul na Figura 4.17. Destas 28 ligações cruzadas, 2 cross-links não estão no conjunto experimental. Os cross-links S18–K116 e K22–S102 juntos com o M1–K95 validado nas estruturas de referência somam 3 dados excedentes. Assim, 100% do conjunto experimental e mais 3 cross-links são compatíveis com as estruturas das simulações.

Com valores de PVES muito menores que os cross-links compatíveis com as estruturas de referência (Figura 4.17), este conjunto indica uma especificidade significativa para as estruturas em que são validados. Considerando a abordagem da simulação de todas as 30 estruturas para a Calmodulina livre, é razoável se pensar que os cross-links que não são compatíveis com nenhuma estrutura de referência sejam

validados em estruturas pouco prováveis, uma vez que o conjunto de referência já apresenta uma variedade conformacional grande. Por isto os valores muito menores de PVES.

A Figura 4.22 mostra as distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos 2 cross-links compatíveis com as estruturas da simulação *all-atom*, mas não validados nas estruturas de referência, que não foram reportados nos dados experimentais da espécie da Calmodulina livre.

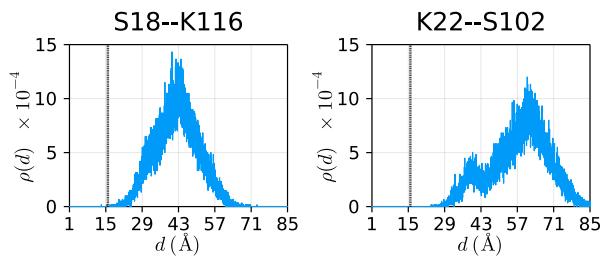


Figura 4.22: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links S18–K116 e K22–S102. - Calmodulina livre, *all-atom*. A referência pontilhada é o limite estatístico.

As ligações cruzadas das distribuições acima conectam domínios diferentes da proteína. Apesar de serem validados em algumas das estruturas da simulação, vemos nestas distribuições a raridade das ocorrências de compatibilidade. Isto se deve ao fato de que o conjunto de estruturas mais provável apresenta distâncias topológicas entre os resíduos em torno de 43 Å para a ligação S18–K116 e de 63 Å para a ligação K22–S102.

Simulações coarse-grained. As simulações *coarse-grained* para a Calmodulina livre não validaram nenhum novo cross-link que já não estivesse validado pelas simulações *all-atom*.

Calmodulina complexada com cálcio

As ligações cruzadas reportadas na Tabela 4.6 para a Calmodulina complexada com cálcio, são aquelas representadas em azul na Figura 4.20. Destas 13 ligações

cruzadas, 4 não foram reportadas nos dados experimentais para a espécie complexada com cálcio. São estas: K14–K76, K14–S39, K14–K78 e S39–K78. Com estas novas validações, 54% do conjunto experimental para a espécie é contemplado e mais 5 cross-links.

A Figura 4.23 mostra as distribuições das PVESs em função da distância topológica entre os resíduos dos 4 cross-links compatíveis com as estruturas da simulação *all-atom*, mas não validados na estrutura de referência, que não foram reportados nos dados experimentais.

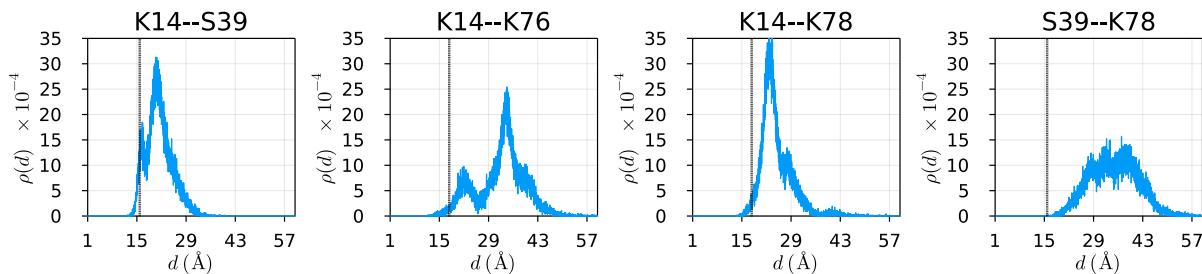


Figura 4.23: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links K14–S39, K14–K76, K14–K78 e S39–K78. - Calmodulina complexada com cálcio, *all-atom*. A referência pontilhada é o limite estatístico.

As distribuições da Figura 4.23 nos mostram que as distâncias mais prováveis entre os resíduos são deslocadas a direita do limite estatístico de validação. Ao verificar a distância topológica para estas ligações na estrutura de referência, a não validação é justificada pois os resíduos apresentam distâncias semelhantes aos picos da distribuição de densidade. Estas distâncias são: 19.6 Å para K14–S39, 16.9 Å para K14–K76, 19.8 Å para K14–K78 e 27.1 Å para S39–K78.

Ao retomar a análise da Figura 4.20, vemos que o cross-link M1–K76 se sobressai das demais ligações cruzadas não compatíveis com a estrutura cristalográfica. Apesar de não ser validado, como o cross-link M1–K78 é compatível com a estrutura, é rationalizada a PVES calculada para a ligação envolvendo o resíduo de lisina 76 (K76) argumentando a proximidade dos resíduos K76 e K78 na sequência.

A Figura 4.24 mostra as distribuições das PVESs em função da distância topológica entre os resíduos dos 2 cross-links mencionados. Ambos compatíveis com as

estruturas da simulação *all-atom*, mas apenas M1–K78 validado na estrutura cristalográfica.

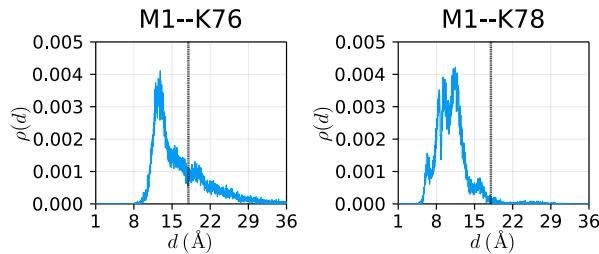


Figura 4.24: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links M1–K76 e M1–K78. - Calmodulina complexada com cálcio, *all-atom*. A referência pontilhada é o limite estatístico.

Com auxílio da Figura 4.24, podemos constatar a semelhança das duas distribuições para resíduos tão próximos como o K76 e K78. Por conta do impedimento estérico que o resíduo K76 sofre por cadeias laterais próximas, mesmo apresentando uma distância euclidiana menor que o limite estatístico, o cross-link M1–K76 não é validado na estrutura cristalográfica.⁴ Porém, um caminho acessível ao solvente é possibilitado ao longo da simulação, por isso a compatibilidade significativa com as estruturas da simulação.

Simulações coarse-grained. As estruturas da simulação *coarse-grained* foram compatíveis com 16 cross-links. Todos os validados pela estrutura cristalográfica (6) e outros 10. Sendo que destes outros, 5 também foram compatíveis com as estruturas da simulação *all-atom* e 5 são cross-links que ainda não haviam sido validados para a Calmodulina complexada com cálcio. Dentro do conjunto de novos cross-links, os S18–K76 e S18–K78 não foram reportados nos dados experimentais para a espécie complexada com cálcio.

Com o acréscimo destes novos 5, um total de 24 cross-links foram compatíveis com as estruturas das simulações com a Calmodulina complexada com cálcio, 65% do conjunto experimental para a espécie e mais 8 cross-links.

⁴Concluiu-se sobre a acessibilidade do resíduo através da distribuição das PVESs em função da distância euclidiana na Figura II.5.

A Figura 4.25 apresenta a PVES para os cross-links compatíveis com as estruturas da simulação *coarse-grained*.

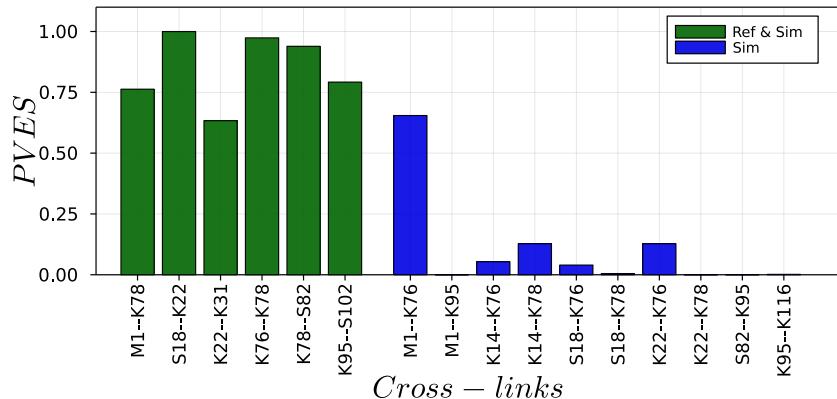


Figura 4.25: PVES para cada cross-link compatível com as estruturas das simulações *coarse-grained* da Calmodulina complexada com cálcio.

Nesta simulação, a ligação cruzada M1–K76 também foi validada em proporção significativa nas estruturas devido a proximidade do resíduos, semelhantemente às simulações *all-atom*.

4.2.3 Cross-links não compatíveis com as estruturas das simulações

Calmodulina livre

Para a espécie Calmodulina livre, as estruturas das simulações foram compatíveis com todo o conjunto experimental.

Calmodulina complexada com cálcio

Não foram compatíveis com a estrutura cristalográfica ou as estruturas das simulações 9 cross-links, 35% do conjunto experimental. A Tabela 4.7 lista os cross-links do conjunto experimental que não foram compatíveis com nenhuma estrutura das simulações.

Table 4.7: Cross-links do conjunto experimental não compatíveis com as estruturas de referência nem com as das simulações para a Calmodulina complexada com cálcio.

Calmodulina Ca ²⁺	
S18–K116	K76–K95
K22–K95	K76–S102
K22–S102	K78–S102
K22–K116	S102–K116
K31–S102	

Um detalhe pertinente é que estas ligações cruzadas são todas, com exceção da S102–K116, inter-domínio. Uma vez que a complexação com cálcio proporciona a formação de uma α -hélice entre os dois domínios, há uma promoção do afastamento dessas duas regiões que pode impedir ação do ALC. Todavia, estas ligações foram reportadas nos dados experimentais. Se considerarmos os dados das simulações como referência, os mesmos sugerem a presença de ligações cruzadas oriundas da espécie Calmodulina livre nos dados experimentais para espécie complexada com cálcio.

4.2.4 Agrupamento por comportamento das ligações cruzadas em diferentes condições biológicas/experimentais

Nesta secção faremos um agrupamento por comportamento com os dados dos experimentos para a Calmodulina livre e Calmodulina complexada com cálcio. Este agrupamento tem o intuito de identificar ligações cruzadas preferenciais nas diferentes condições da Calmodulina.

Fazendo a compilação dos dados dos experimentos da Calmodulina livre e Calmodulina complexada com cálcio, o software identificou 2 clusters. Como para a calmodulina temos somente 2 espécies/condições experimentais, os cross-links identificados no cluster 1 são aqueles que, através do XIC experimental, mostraram um aumento da detecção de seus precursores da condição 1 (Calmodulina livre) para a

2 (Calmodulina complexada com cálcio) e para o cluster 2 o contrário. A Figura 4.26 mostra esse comportamento em gráfico.

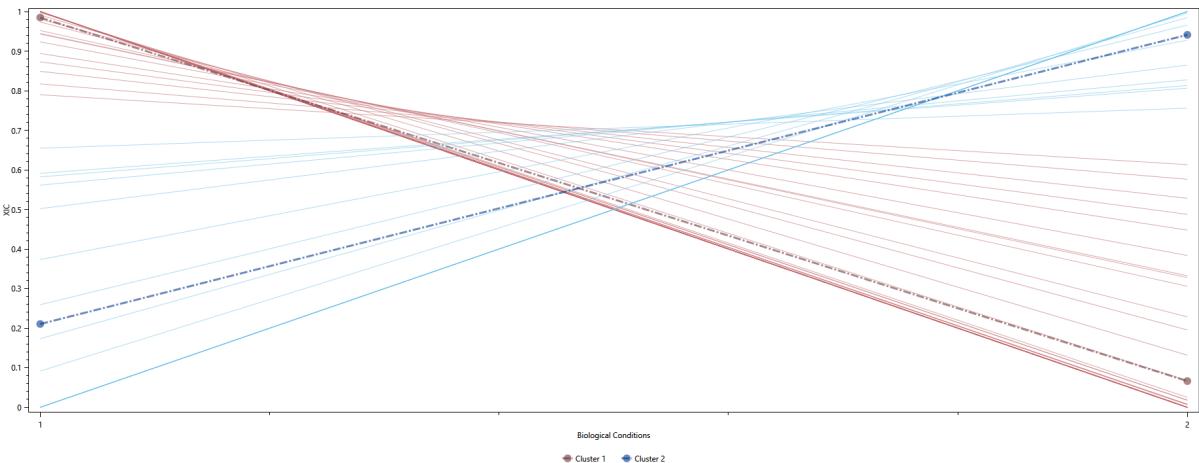


Figura 4.26: Representação do comportamento dos clusters identificados pelo software QuinXL. Onde a condição biológica 1 é a ausência de íons cálcio e a 2 a presença de íons cálcio. Os Clusters 1 e 2 são representados pelas cores vermelho e azul, respectivamente.

A Figura 4.27 mostra quais são os cross-links de cada grupo na sequencia de aminoácidos da proteína.

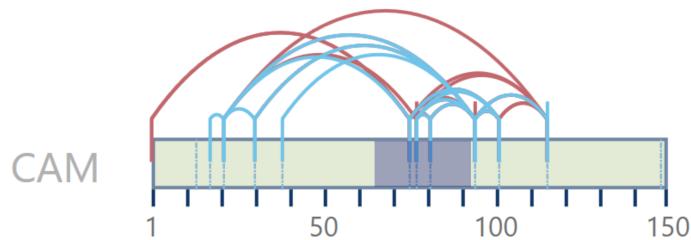


Figura 4.27: Representação dos cross-links na sequência da proteína coloridos em função do cluster identificado. Os Clusters 1 e 2 são representados pelas cores vermelho e azul, respectivamente. Destaque em azul para a parte da sequência na qual é formada a α -hélice inter-domínio.

Ao identificar os domínios da proteína como as porções antes e depois do destaque azul na Figura 4.27, podemos observar que a clusterização não mostra uma relação de proibição de ligação inter-domínio ou de preferência intra-domínio em prol da complexação com cálcio.

A Tabela 4.8 lista os cross-links que foram compatíveis com estruturas das simulações *all-atom* para ambas as espécies (Calmodulina livre e Calmodulina complexada com cálcio) agrupadas pelo cluster a qual elas pertencem.

Table 4.8: Cross-links compatíveis com as estruturas das simulações *all-atom* comuns entre Calmodulina livre e Calmodulina complexada com cálcio agrupados pelos clusters em que eles se encontram na identificação do software QuinXL.

Cluster 1	Cluster 2
M1–K76	K78–S82
K31–K95	K78–K116
K76–K78	K95–S102
K76–K116	K95–K116
	S18–K22
	K22–K31
	S39–K95
	K78–K95

Sendo assim, os cross-links do Cluster 1 deveriam aparecer com maior intensidade na condição 1 (ausência de íons cálcio) e os do Cluster 2 na condição 2 (presença de íons cálcio). Entretanto, quando levamos em consideração estes comportamentos para os crosslinks da Tabela 4.8 comparando a PVES para as diferentes simulações, não temos resultados conclusivos.

Para o Cluster 1 as ligações cruzadas K76–K116, K31–K95, K78–K116 e K76–K78 seguem a tendência apontada para o grupo pelo QuinXL, porém para os cross-links M1–K76, K78–S82, K95–S102 e K95–K116 o comportamento é o inverso do que o esperado para o grupo segundo os dados do software.

Para o Cluster 2, o mesmo comportamento dos dados é observado. Metade do conjunto é condizente com a interpretação mediante os dados do programa e metade não. Os cross-links K22–K31 e S39–K95 apresentam um aumento da PVES da condição 1 para a 2 e os cross-links S18–K95 uma diminuição.

Como visão geral do procedimento, acredito que as incompatibilidades com os dados das simulações são consequências das incompatibilidades dos dados experimentais anteriormente apontadas. O agrupamento por comportamento em diferentes condições faz sentido estruturalmente. Uma vez que observamos nas simulações a

maior compatibilidade de ligações cruzadas inter-domínio para a Calmodulina livre, é plausível que um comportamento deste tipo possa ser identificado pelo método utilizado pelo software. Todavia, a natureza dos dados experimentais é essencial para uma boa representação da realidade estrutural da proteína.

Capítulo 5

Considerações Finais

A técnica de Espectrometria de Massas de Cross-linking é uma poderosa ferramenta que consiste na ligação covalente entre dois resíduos proteicos e identificação dessa conexão. As aplicações são diversas e coletam informações estruturais inter e intramoleculares possibilitando entendimento de interações enzimáticas [54], estruturas terciárias e quaternárias de proteínas, e, mais recentemente, assistência a modelagens computacionais [55].

As análises realizadas com a SalBIII deixam claro a importância da dinâmica proteica para entender a satisfação de restrições que não ocorrem nas estruturas de referência. Justificando 42 ligações cruzadas experimentais não compatíveis com a estrutura cristalográfica através do conjunto de conformações da simulação *all-atom* e 47 da simulação *coarse-grained*, um estudo como esse pode direcionar a melhor compreensão dos dados experimentais e até a seleção de restrições para aplicações desse tipo de informação. É de significativa relevância a informação de que 69 cross-links não foram satisfeitos, mesmo depois uma simulações com 5 μ s e amostragem *coarse-grained*. A interpretação estatística das distância entre resíduos participantes de cross-links ao longo da simulação nos ajudaram a interpretar o motivo da ocorrência e intensidade dos cross-links. Ainda, podem trazer informações conformacionais detalhadas, como indica o gráfico de densidade de conformações em função da distância topológica entre os resíduos. Como o da ligação cruzada E62–E86 na Figura 4.5, por exemplo.

A adição de íons cálcio à Calmodulina induz a formação da α -hélice inter-domínio que impede diretamente a satisfação de 3 cross-links. Todavia, através da dinâmica foi possível verificar a compatibilidade de uma ligação cruzada com parte significativa das estruturas da simulação, M1–K76. Este cross-link era compatível com parte das estruturas da Calmodulina livre, mas não havia sido validado para a estrutura cristalográfica da Calmodulina complexada com cálcio. A clusterização por comportamento das ligações cruzadas em função da condição biológica/experimental com base nas ligações inter e intra-domínio não geraram relações compatíveis com os dados das simulações.

É relevante ressaltar também que foram encontradas correlações positivas entre a qualidade dos espectros e a porcentagem do conjunto filtrado que é compatível com a estrutura cristalográfica e com as estruturas da simulação.

Finalmente, conclui-se que, assim como a dinâmica dos linkers [56], a dinâmica de conformação da proteína é de significante relevância para a análise de dados de cross-linking e, considerando a magnitude das moléculas estudadas, simulações *coarse-grained* são ferramentas úteis quando interpretadas de forma coerente em análise conformacional qualitativa.

Appendix A

Análises preliminares dos campos de força

A.1 Campo de força

Duas simulações *coarse-grained* foram produzidas com a SalBIII através do software GROMACS, 1 μ s com o campo de força SIRAH e 1 μ s com o campo de força Martini, ambos seguindo os procedimentos sugeridos pelos respectivos desenvolvedores.

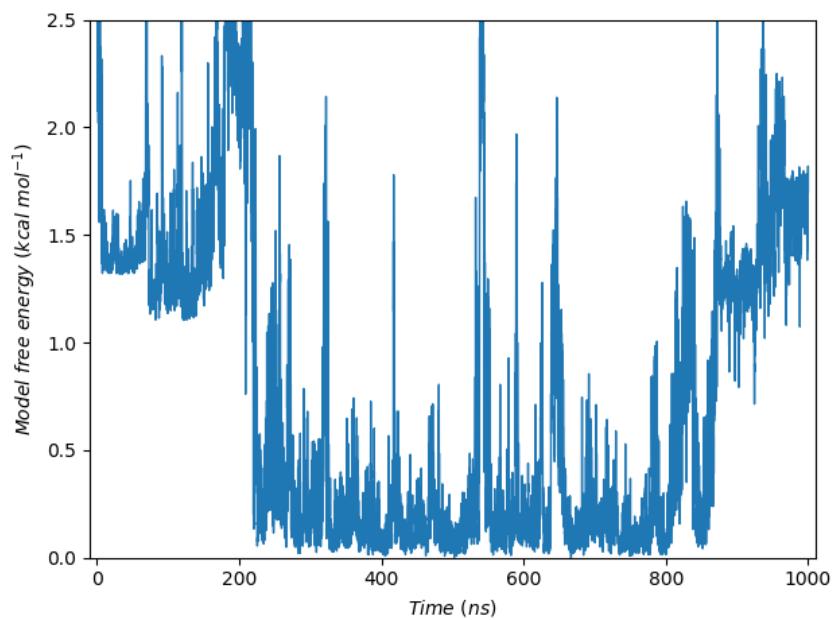


Figura A.1: Gráfico da energia livre da conformação pelo tempo de simulação com o campo de força SIRAH - 1 μ s

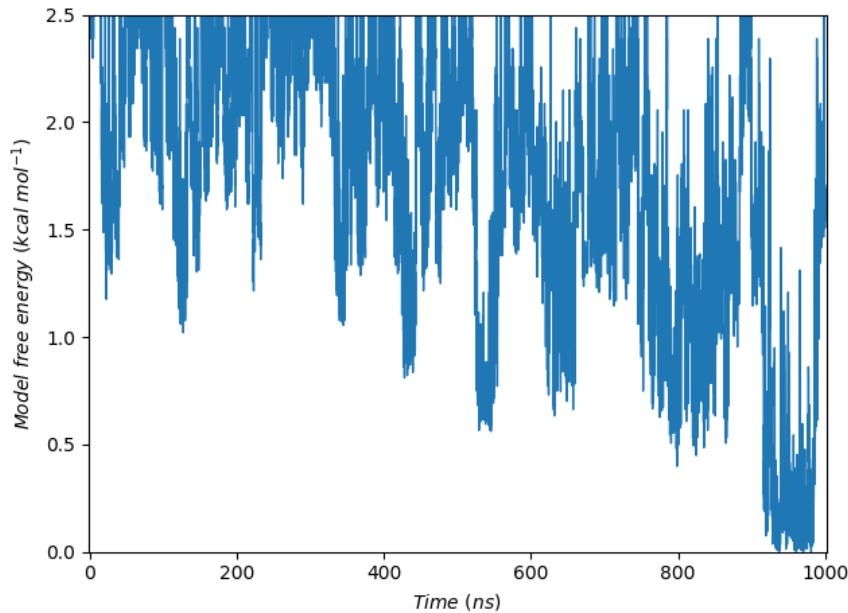


Figura A.2: Gráfico da energia livre da conformação pelo tempo de simulação com o campo de força Martini

O campo de força a ser escolhido deveria apresentar um comportamento energético compatível com uma amostragem termodinâmica de conformações, ou seja, o padrão energético da simulação deveria seguir um regime natural de enovelamento e desnaturação da proteína. Entretanto, os resultados obtidos, representados pelas Figuras A.1 e A.2, não apresentaram diferenças significativas para a escolha de um em detrimento do outro. Apesar dos resultados não apontarem um campo de força como o melhor ou mais adequado no momento, os procedimentos utilizados para a obtenção dos resultados já os diferenciam de certa maneira. Percebeu-se que ambos procedimentos são muito flexíveis, permitindo assim a manipulação quase livre do processo de produção da dinâmica. Todavia, as ferramentas de remapeamento dos campos de força apresentaram diferenças em sua performance. Depois da separação das frames da trajetória, o remapeamento é feito para cada uma das conformações. Um programa para a execução deste procedimento foi escrito usando a ferramenta de remapeamento de cada um dos campos de força e, nesse cenário, ao analisar os resultados deste de cada remapeamento, percebeu-se que a ferramenta de do campo

de força Martini não foi tão eficiente quanto a do SIRAH. Em números, em 10.000 conformações para cada trajetória, 108 erros de remapeamento foram encontrados para o procedimento com a ferramenta do Martini e nenhum para o do SIRAH. Com isso, uma preferencia, por questões práticas, para/com o campo de força SIRAH surgiu apesar dos resultados apresentados acima não serem determinantes.

A.2 Coarse-grained vs All-Atom

O desvio quadrático médio (RMSD) é uma representação da movimentação da molécula simulada. Afim de comparar o comportamento da molécula nas simulações dos três campos de força (*All-atom* - Amber e *Coarse-grained* - SIRAH e Martini), construíram-se as representações da função RMSD de cada um deles.

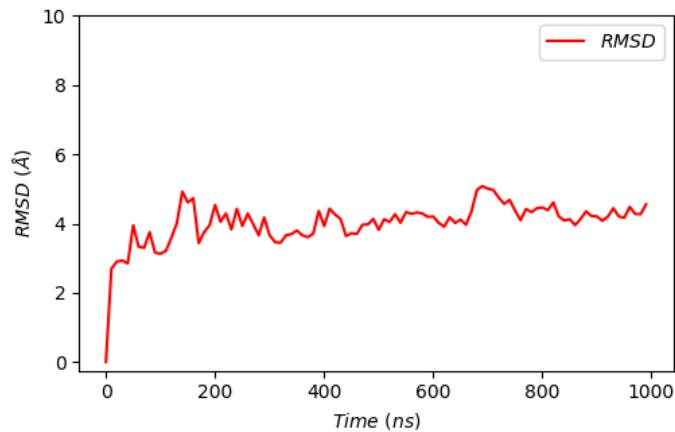


Figura A.3: Desvio quadrático médio (RMSD) - 1 μ s, Amber

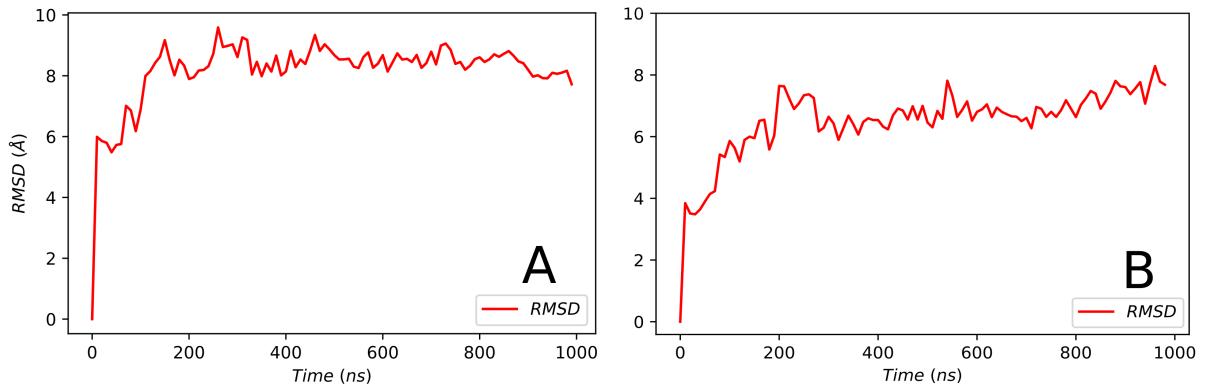


Figura A.4: Desvio quadrático médio (RMSD) - 1 μ s, Martini (A) SIRAH (B)

Comparando a Figura A.3 e a Figura A.4, pudemos perceber que os valores de RMSD da última são maiores que a da primeira. Devido ao número reduzido de graus de liberdade, a consideração de apenas interações de curta distância, a superfície de potencial mais suavizada e a passos de tempo de integração maiores, os campos de força coarse-grained demonstram uma velocidade de 2 a 5 ordens de grandeza maior do que os campos de força all-atom [57], isto é, há amostragem de mais conformações para um mesmo intervalo de tempo de simulação.

Estudos de comparação entre campos de força coarse-grained e all-atom mostram as diferenças dos valores de propriedades termodinâmicas. Entre dados obtidos experimentalmente, com campo de força coarse-grained e com campo de força all-atom. Com isso, apontam que apesar dos campos de força coarse-grained serem aproximações significativas, ainda há espaço para aprimoramento nesse quesito [58]. Quando se comparam a flexibilidade da cadeia e a média de propriedades de longo alcance a conclusão é que os campos de força coarse-grained produzem uma boa descrição dos dados apresentados pelos campos de força all-atom [59].

Para detalhar a movimentação da estrutura, um mapeamento da movimentação dos resíduos foi feita através da função de flutuação quadrática média (RMSF) [60]. Uma vez em que o RMSD é o desvio quadrático médio das posições de todos os átomos, como um só corpo, ao longo do tempo, o RMSF é a média do desvio de um

átomo, ou grupo de átomos, relativo a uma estrutura de referência.

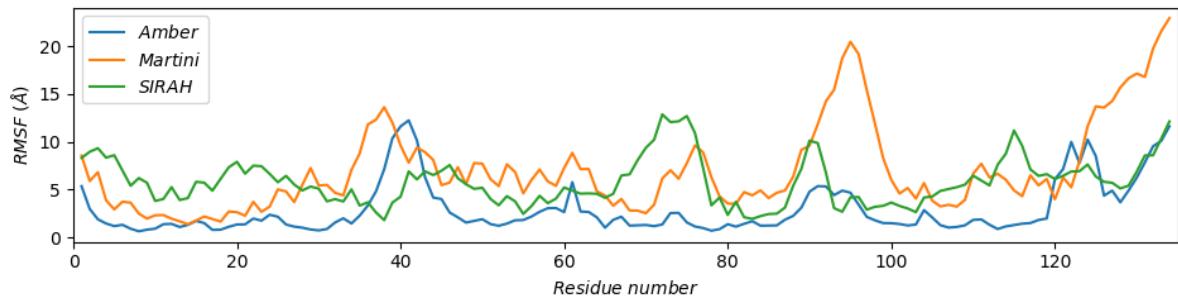


Figura A.5: Flutuação quadrática média (RMSF) dos carbonos- α dos três campos de força

Tomando a flutuação dos resíduos com o campo de força Amber como comparação para as outras flutuações, encontramos valores para a correlação de Pearson de 0.64 para o campo de força Martini e 0.18 para o SIRAH. Todavia, analisando o RMSF da trajetória produzida com o campo de força Amber, obtivemos a Figura A.6.

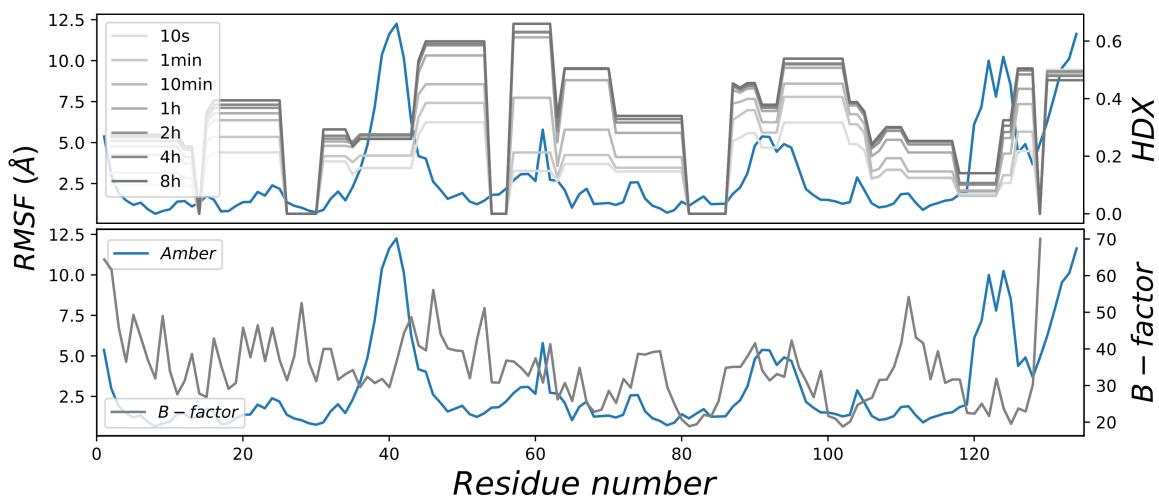


Figura A.6: Flutuação quadrática média (RMSF) dos carbonos- α em comparação a troca de hidrogênio-deutério (HDX) e ao B-fator - Amber.

O experimento de troca hidrogênio-deutério (HDX) consiste em, através do solvente rico em deutério, realizar a troca de hidrogênios da proteína com deutérios do solvente e analisar os produtos da reação, via espectrômetro de massas, para diferentes

tempos de reação. Assim, determinam-se as regiões da proteína que são mais e menos expostas ao solvente. Entretanto, não vemos qualquer sinal de correlação entre os dados de HDX e RMSF.

O B-factor é um parâmetro que reflete a precisão do experimento de cristalografia de raio X, sendo ele maior quando a precisão da posição do átomo/resíduo correspondente é menor. Contudo, para a correlação entre as duas curvas, obtivemos $r = -0.02$, mostrando que os dados B-fator também não são correlatos ao RMSF.

Finalmente, por questões práticas mencionadas na secção anterior e pela contato mais direto que o grupo tem com os desenvolvedores, escolhemos o campo de força SIRAH para utilizar no restante do estudo.

A.3 RMSD

Com o intuito de entendermos mais detalhadamente a movimentação da proteína, produzimos um gráfico, através dos dados obtidos com o MDLovoFit [60]. Neste gráfico, produzido utilizando 100 conformações igualmente espaçadas na simulação *all-atom* com o campo de força *Amber*, verificou-se o valor mínimo obtido pela função RMSD em função da porcentagem de átomos considerados (ϕ) no alinhamento das conformações.

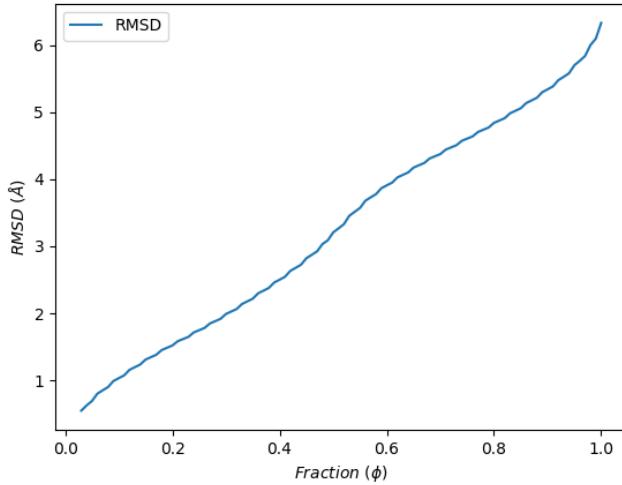


Figura A.7: RMSD em função da fração selecionada da proteína

Segundo a Figura A.7, a função RMSD segue um regime quase linear em relação a fração de átomos da proteína considerada para o alinhamento. Com isso, percebemos que os valores mínimos de RMSD acima de 4.5 Å são oriundos de aproximadamente 30% dos átomos. Assim, na Figura A.8, mostramos a dependência do desvio quadrático médio com o tempo utilizando, arbitrariamente, a fração (ϕ) de 0.7 como comparação à consideração de todos os átomos ($\phi = 1$) da proteína.

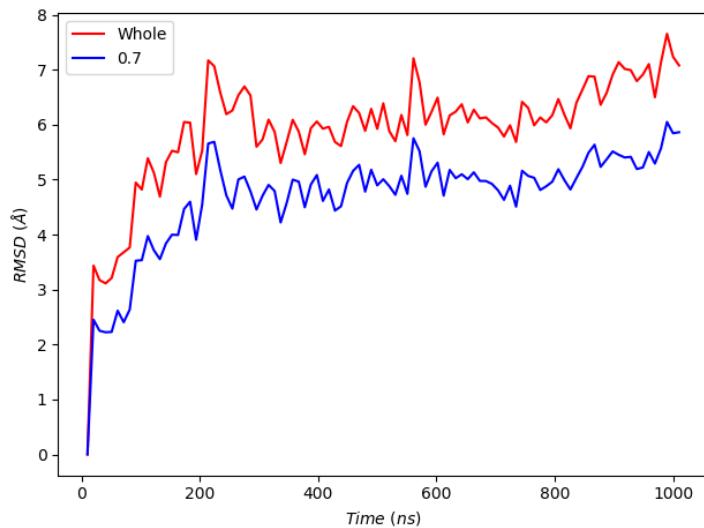


Figura A.8: Desvio quadrático médio com $\phi = 1$ e $\phi = 0.7$

Para melhor representar o gráfico da Figura A.8, foi feita uma representação gráfica de sobreposição das 100 conformações em que a fração selecionada da proteína é ilustrada na cor azul e o restante da molécula em vermelho.

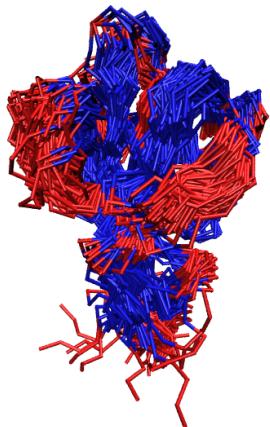


Figura A.9: Sobreposição de conformações da simulação *all-atom* de $1 \mu\text{s}$

Unindo as informações, vemos claramente na Figura A.9 uma significativa movimentação do N-terminal na parte inferior da figura, um dos motivos pelos quais os valores de RMSD são mais altos para maiores valores de ϕ .

I**Anexo A**

Arquivos mdp utilizados na produção das simulações.

em1_CGPROT.mdp

```
; VARIOUS PREPROCESSING OPTIONS
title      = EMIN
cpp        = /lib/cpp
include    =
define     = -DGN_GO

; RUN CONTROL PARAMETERS
integrator = steep
nsteps     = 5000

; Langevin Dynamics Options
; None (see manual)

; ENERGY MINIMIZATION OPTIONS
emtol      = 1.0
emstep     = 0.05
; nstcgsteep = 500
nbfgscorr  = 10

; OUTPUT CONTROL OPTIONS
nstxout    = 50
nstvout    = 50
nstfout    = 0
nstlog     = 50
nstenergy   = 50
nstxtcout  = 50
xtc-precision = 1000
xtc-grps    =
energygrps = 

; NEIGHBORSEARCHING PARAMETERS
nstlist    = 10
ns_type    = grid
pbc       = xyz
rlist     = 1.2

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype = PME
rcoulomb    = 1.2
vdwtype     = Cut-off
rvdw        = 1.2

optimize_fft = yes
fourierspacing = 0.2
```

em2_CGPROT.mdp

```
; VARIOUS PREPROCESSING OPTIONS
title      = EMIN
cpp        = /lib/cpp
include    =
define     = 

; RUN CONTROL PARAMETERS
integrator = steep
nsteps     = 5000

; Langevin Dynamics Options
; None (see manual)

; ENERGY MINIMIZATION OPTIONS
emtol      = 1.0
emstep     = 0.05
; nstcgsteep = 500
nbfgscorr  = 10

; OUTPUT CONTROL OPTIONS
nstxout    = 50
nstvout    = 50
nstfout    = 0
nstlog     = 50
nstenergy   = 50
nstxtcout  = 50
xtc-precision = 1000
xtc-grps    =
energygrps = 

; NEIGHBORSEARCHING PARAMETERS
nstlist    = 10
ns_type    = grid
pbc       = xyz
rlist     = 1.2

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype = PME
rcoulomb    = 1.2
vdwtype     = Cut-off
rvdw        = 1.2

optimize_fft = yes
fourierspacing = 0.2
```

eq1_CGPROT.mdp

```
; GENERAL INPUT FLAGS
title      = 5ns equilibration NVT
cpp        = /usr/bin/cpp
define     = -DPOSRES

; RUN FLAGS
integrator = md
dt          = 0.020
nsteps      = 250000
pbc         = xyz

; NON-BONDED INTERACTIONS FLAGS
ns_type    = grid
nstlist    = 10
rlist      = 1.2
coulombtype = PME
rcoulomb   = 1.2
vdwtype    = Cut-off
rvdw       = 1.2

optimize_fft = yes
fourierspacing = 0.2

; TEMPERATURE COUPLING FLAGS
tcoupl     = V-rescale
tc-grps    = Protein SIRAH-Solvent
tau_t      = 1.0 1.0
ref_t      = 300 300

; INITIAL STEP VELOCITIES FLAGS
gen_vel    = no
gen_temp   = 300
gen_seed   = -1

; PRESSURE COUPLING FLAGS
Pcoupl    = no

; OUTPUT FLAGS
nstxout   = 0
nstvout   = 0
nstlog    = 5000
nstenergy = 0
nstxtcout = 5000
```

eq2_CGPROT.mdp

```
; GENERAL INPUT FLAGS
title      = 25ns equilibration NVT
cpp        = /usr/bin/cpp
define     = -DGN_GO_SOFT

; RUN FLAGS
integrator = md
dt          = 0.020
nsteps      = 1250000
pbc         = xyz

; NON-BONDED INTERACTIONS FLAGS
ns_type    = grid
nstlist    = 10
rlist      = 1.2
coulombtype = PME
rcoulomb   = 1.2
vdwtype    = Cut-off
rvdw       = 1.2

optimize_fft = yes
fourierspacing = 0.2

; TEMPERATURE COUPLING FLAGS
tcoupl     = V-rescale
tc-grps    = Protein SIRAH-Solvent
tau_t      = 1.0 1.0
ref_t      = 300 300

; INITIAL STEP VELOCITIES FLAGS
gen_vel    = no
gen_temp   = 300
gen_seed   = -1

; PRESSURE COUPLING FLAGS
Pcoupl    = no

; OUTPUT FLAGS
nstxout   = 0
nstvout   = 0
nstlog    = 5000
nstenergy = 0
nstxtcout = 5000
```

```

md_CGPROT.mdp
(para 5 μs, nsteps = 250000000)
                                nstlog      = 5000
                                nstenergy   = 0
                                nstxtcout   = 5000

; GENERAL INPUT FLAGS
title          = 1us production NPT
cpp            = /usr/bin/cpp
define         =

; RUN FLAGS
integrator    = md
dt             = 0.020
nsteps         = 50000000
pbc            = xyz

; NON-BONDED INTERACTIONS FLAGS
ns_type        = grid
nstlist         = 10
rlist           = 1.2
coulombtype   = PME
rcoulomb       = 1.2
vdwtype        = Cut-off
rvdw            = 1.2

optimize_fft   = yes
fourierspacing = 0.2

; TEMPERATURE COUPLING FLAGS
tcoupl         = V-rescale
tc-grps        = Protein SIRAH-Solvent
tau_t          = 2.0 2.0
ref_t          = 300 300

; INITIAL STEP VELOCITIES FLAGS
gen_vel        = no
gen_temp       = 300
gen_seed       = -1

; PRESSURE COUPLING FLAGS
Pcoupl         = Parrinello-Rahman
pcoupltype    = isotropic
tau_p          = 8.0
compressibility = 4.5e-5
ref_p          = 1.0
refcoord-scaling = com

; OUTPUT FLAGS
nstxout        = 0
nstvout        = 0

```

II

Anexo B

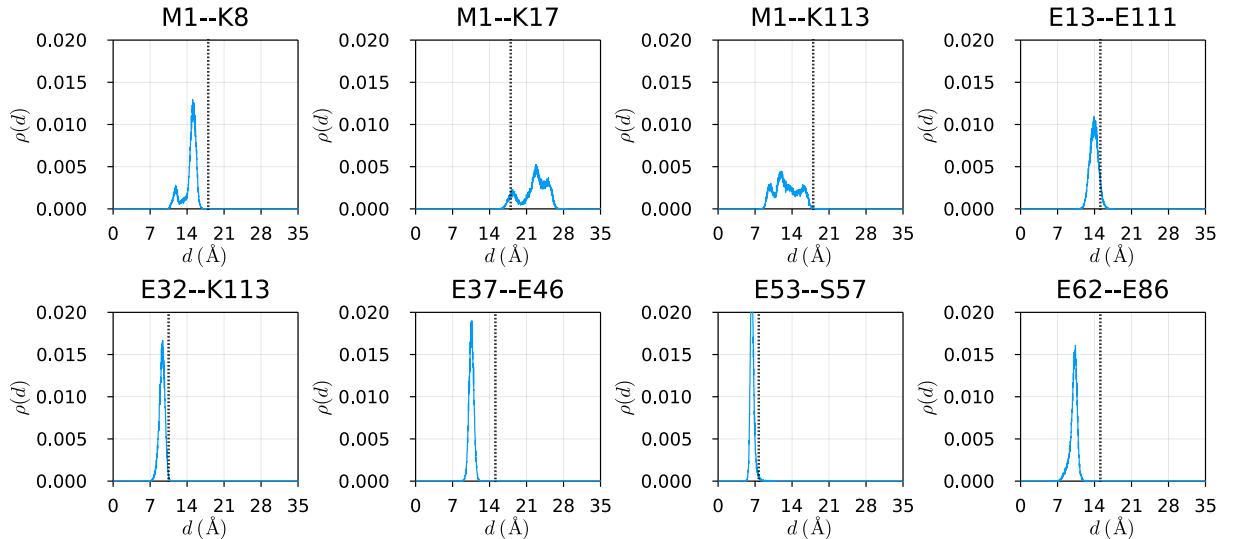


Figura II.1: Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links dà Figura 4.4, a referência pontilhada é o limite estatístico.

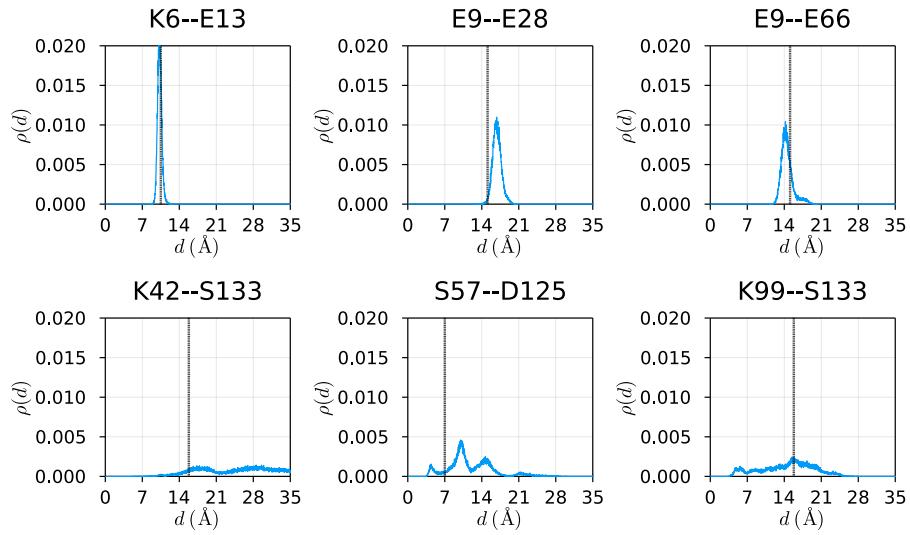


Figura II.2: Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links dà Figura 4.8, a referência pontilhada é o limite estatístico.

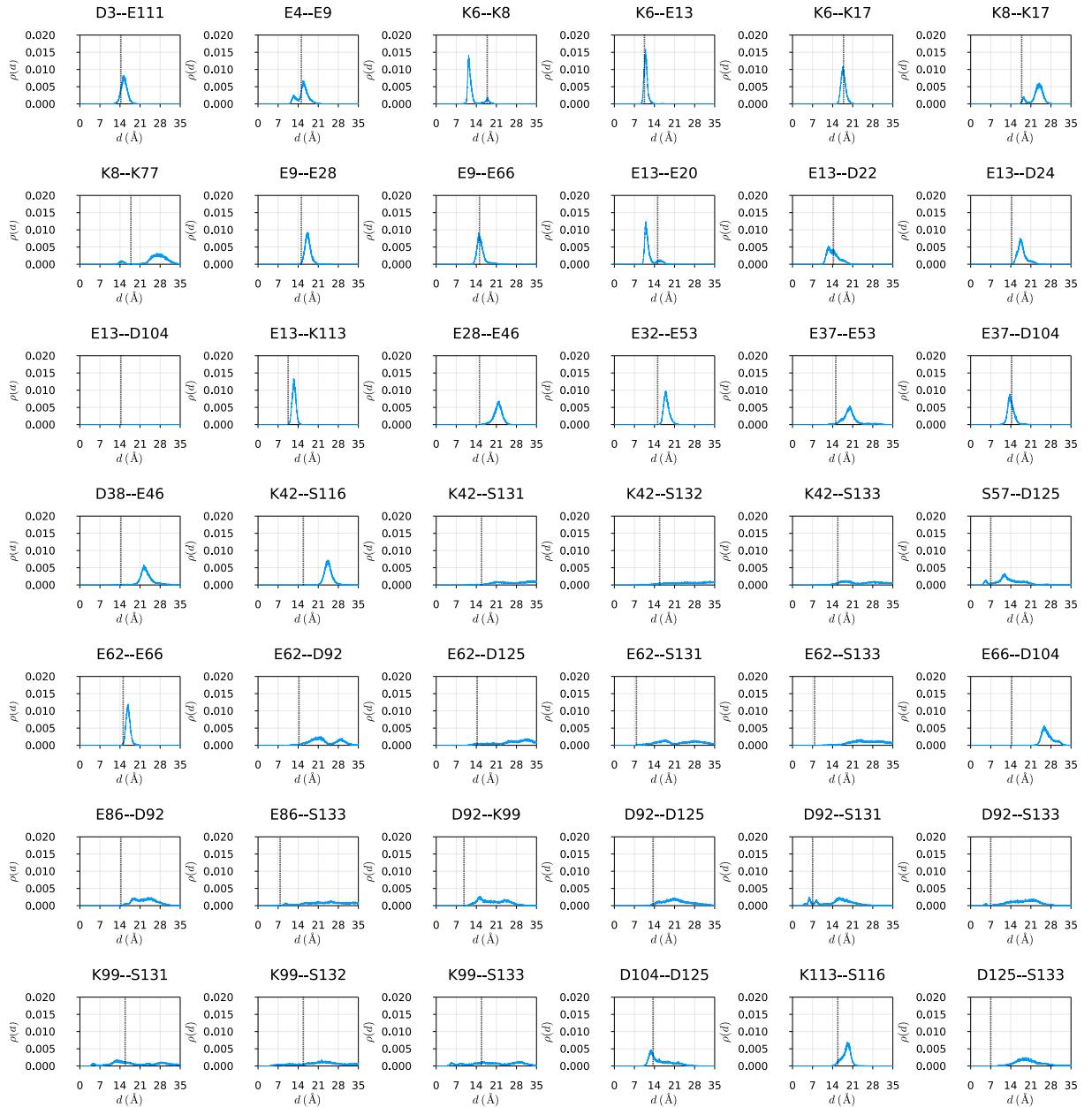


Figura II.3: Distribuições da densidade de estruturas em função da distância topológica entre os resíduos dos cross-links dà Figura 4.6, a referência pontilhada é o limite estatístico.

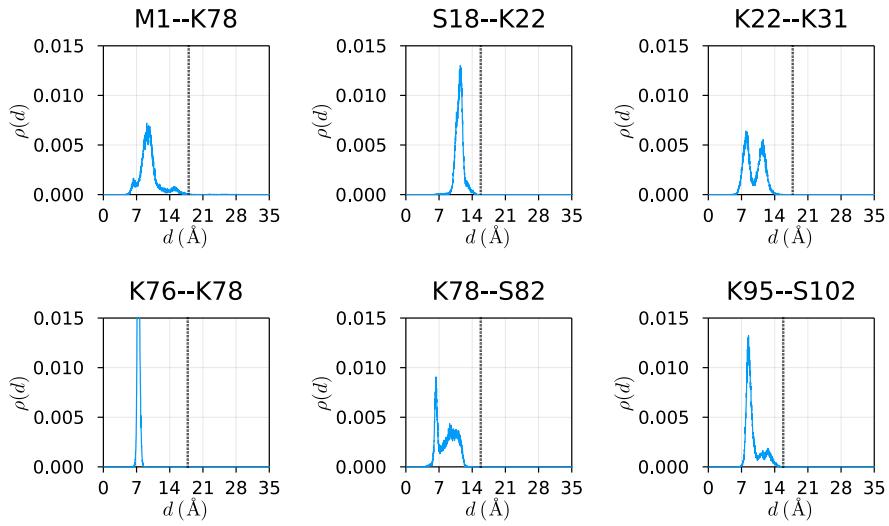


Figura II.4: Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links compatíveis com a estrutura cristalográfica.- , *all-atom*. A referência pontilhada é o limite estatístico.

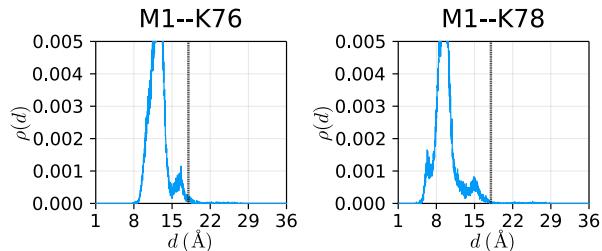


Figura II.5: Distribuições da densidade de estruturas em função da distância euclidiana entre os resíduos dos cross-links M1-K76 e M1-K78.- , *all-atom*. A referência pontilhada é o limite estatístico.

Bibliography

- [1] Chu, F.; Thornton, D. T.; Nguyen, H. T. Chemical cross-linking in the structural analysis of protein assemblies. *Methods* **2018**, *144*, 53–63.
- [2] Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64–71.
- [3] Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **1988**, *60*, 2299–2301.
- [4] Griffiths, J. A Brief History of Mass Spectrometry. *Analytical Chemistry* **2008**, *80*, 5678–5683.
- [5] Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics* **1993**, *17*, 75–86.
- [6] Maurer, M. M.; Donohoe, G. C.; Valentine, S. J. Advances in ion mobility-mass spectrometry instrumentation and techniques for characterizing structural heterogeneity. *Analyst* **2015**, *140*, 6782–6798.
- [7] Sinz, A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Review of Proteomics* **2014**, *11*, 733–743.
- [8] Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrometry Reviews* **2006**, *25*, 663–682.

- [9] Trimpin, S.; Keune, S.; Räder, H. J.; Müllen, K. Solvent-free MALDI-MS: Developmental improvements in the reliability and the potential of MALDI in the analysis of synthetic polymers and giant organic molecules. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 661–671.
- [10] Lima, D. B.; de Lima, T. B.; Balbuena, T. S.; Neves-Ferreira, A. G. C.; Barbosa, V. C.; Gozzo, F. C.; Carvalho, P. C. SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis. *Journal of Proteomics* **2015**, *129*, 51–55.
- [11] Leitner, A.; Joachimiak, L. A.; Unverdorben, P.; Walzthoeni, T.; Frydman, J.; Förster, F.; Aebersold, R. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proceedings of the National Academy of Sciences* **2014**, *111*, 9455–9460.
- [12] Fioramonte, M.; De Jesus, H. C. R.; Ferrari, A. J. R.; Lima, D. B.; Drekenner, R. L.; Correia, C. R. D.; Oliveira, L. G.; Neves-Ferreira, A. G. D. C.; Carvalho, P. C.; Gozzo, F. C. XPlex: An Effective, Multiplex Cross-Linking Chemistry for Acidic Residues. *Analytical Chemistry* **2018**, *90*, 6043–6050.
- [13] Gomes, A. F.; Gozzo, F. C. Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI-and ESI-MS/MS. *Journal of mass spectrometry* **2010**, *45*, 892–899.
- [14] Tang, X.; Bruce, J. E. A new cross-linking strategy: protein interaction reporter (PIR) technology for protein–protein interaction studies. *Molecular biosystems* **2010**, *6*, 939–947.
- [15] Ferrari, Állan J. R., Modelagem da estrutura de proteínas baseada em restrições de distância obtidas por ligação cruzada e espectrometria de massas. Ph.D. thesis, Universidade Estadual de Campinas, 2019.

- [16] Ferrari, A. J. R.; Clasen, M. A.; Kurt, L.; Carvalho, P. C.; Gozzo, F. C.; Martínez, L. TopoLink: evaluation of structural models using chemical crosslinking distance constraints. *Bioinformatics* **2019**, *1*–2.
- [17] Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein science* **2014**, *23*, 747–759.
- [18] Jacobsen, R. B.; Sale, K. L.; Ayson, M. J.; Novak, P.; Hong, J.; Lane, P.; Wood, N. L.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. S. Structure and dynamics of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry. *Protein science* **2006**, *15*, 1303–1317.
- [19] Müller, F.; Rappsilber, J. A protocol for studying structural dynamics of proteins by quantitative crosslinking mass spectrometry and data-independent acquisition. *Journal of Proteomics* **2020**, *218*, 103721.
- [20] Chen, Z. A.; Rappsilber, J. Protein dynamics in solution by quantitative crosslinking/mass spectrometry. *Trends in biochemical sciences* **2018**, *43*, 908–920.
- [21] Iacobucci, C.; Götze, M.; Sinz, A. Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Current Opinion in Biotechnology* **2020**, *63*, 48–53, Nanobiotechnology Systems Biology.
- [22] Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of structural biology* **2011**, *173*, 530–540.
- [23] Ziemianowicz, D. S.; Saltzberg, D.; Pells, T.; Crowder, D. A.; Schräder, C.; Hepburn, M.; Sali, A.; Schriemer, D. C. IMProv: A Resource for Cross-link-Driven Structure Modeling that Accommodates Protein Dynamics. *Molecular & Cellular Proteomics* **2021**, *20*.

- [24] Joo, K.; Heo, S.; Joung, I.; Hong, S. H.; Lee, S. J.; Lee, J. Data-assisted protein structure modeling by global optimization in CASP12. *Proteins: Structure, Function, and Bioinformatics* **2018**, *86*, 240–246.
- [25] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A., et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *1*.
- [26] van Gunsteren, W. F.; Mark, A. E. Validation of molecular dynamics simulation. *The Journal of chemical physics* **1998**, *108*, 6109–6116.
- [27] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- [28] Khuntawee, W.; Sutthibutpong, T.; Phongphanphanee, S.; Karttunen, M.; Wong-Ekkabut, J. Molecular dynamics study of natural rubber–fullerene composites: connecting microscopic properties to macroscopic behavior. *Physical Chemistry Chemical Physics* **2019**, *21*, 19403–19413.
- [29] Martínez, L.; Shimizu, S. Molecular Interpretation of Preferential Interactions in Protein Solvation: A Solvent-Shell Perspective by Means of Minimum-Distance Distribution Functions. *Journal of Chemical Theory and Computation* **2017**, *13*, 6358–6372.
- [30] Andrew, R. L. Molecular modeling principles and applications. *2nd, editor.: Pearson Education Limited 2001*,
- [31] McMillan, S. The leapfrog integrator. http://www.physics.drexel.edu/~steve/Courses/Comp_Phys/Integrators/leapfrog/.
- [32] Cai, W. Ewald Notes. http://micro.stanford.edu/mediawiki/images/4/46/Ewald_notes.pdf.

- [33] Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.
- [34] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995**, *103*, 8577–8593.
- [35] Huang, S. T.; Cambanis, S. Stochastic and multiple Wiener integrals for Gaussian processes. *The Annals of Probability* **1978**, *5*, 585–614.
- [36] Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.
- [37] Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **1983**, *50*, 1055–1076.
- [38] Namba, A. M.; Silva, V. B. d.; Silva, C. H. T. P. d. Dinâmico: teoria e aplicações no planejamento de fármacos. *EcláQuâmica* **2008**, *33*, 13 – 24.
- [39] Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **1988**, *110*, 1657–1666.
- [40] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations*; 2003.
- [41] Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *Journal of Chemical Theory and Computation* **2015**, *11*, 723–739.

- [42] Tieleman, D. P.; Marrink, S.-J.; Kandasamy, S. K.; Periole, X.; Monticelli, L.; Larson, R. G. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation* **2008**, *4*, 819–834.
- [43] Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chemical Reviews* **2016**, *116*, 7898–7936.
- [44] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B* **2007**, *111*, 7812–7824.
- [45] Machado, M. R.; Barrera, E. E.; Klein, F.; Sónora, M.; Silva, S. The SIRAH force field 2.0: Altius, Fortius, Citius. **2018**, 1–28.
- [46] Luhavaya, H.; Dias, M. V.; Williams, S. R.; Hong, H.; De Oliveira, L. G.; Leadlay, P. F. Enzymology of Pyran Ring A Formation in Salinomycin Biosynthesis. *Angewandte Chemie - International Edition* **2015**, *54*, 13622–13625.
- [47] Zhang, M.; Tanaka, T.; Ikura, M. Calcium-Induced Conformational Transition Revealed by the Solution Structure of Apo Calmodulin. *Nature Structural Biology* **1995**, *2*, 758–767.
- [48] Babu, Y.; Bugg, C. E.; Cook, W. J. Structure of calmodulin refined at 2.2 Å resolution. *Journal of Molecular Biology* **1988**, *204*, 191–204.
- [49] Stevens, F. C. Calmodulin: an introduction. *Canadian Journal of Biochemistry and Cell Biology* **1983**, *61*, 906–910, PMID: 6313166.
- [50] Chin, D.; Means, A. R. Calmodulin: a prototypical calcium sensor. *Trends in Cell Biology* **2000**, *10*, 322–328.
- [51] dos Santos, Luana O., Uma nova abordagem para estudo de mudança conformacional de proteínas utilizando ligação cruzada quantitativa associada à espectrometria de massas. Ph.D. thesis, Universidade Estadual de Campinas, 2019.

- [52] Kurt, L. U.; Clasen, M. A.; Santos, M. D.; Lyra, E. S.; Santos, L. O.; Ramos, C. H.; Lima, D. B.; Gozzo, F. C.; Carvalho, P. C. Characterizing protein conformers by cross-linking mass spectrometry and pattern recognition. *Bioinformatics* **2021**,
- [53] Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007**, *126*, 014101.
- [54] Mintseris, J.; Gygi, S. P. High-density chemical cross-linking for modeling protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **2020**, *117*, 93–102.
- [55] Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski-Wills, J. C.; Nilges, M.; Cramer, P.; Rappsilber, J. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO Journal* **2010**, *29*, 717–726.
- [56] Gong, Z.; Ye, S.-X.; Nie, Z.-F.; Tang, C. The Conformational Preference of Chemical Cross-linkers Determines the Cross-linking Probability of Reactive Protein Residues. *The Journal of Physical Chemistry B* **2020**,
- [57] The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 225–248.
- [58] Baron, R.; De Vries, A. H.; Hünenberger, P. H.; Van Gunsteren, W. F. Comparison of atomic-level and coarse-grained models for liquid hydrocarbons from molecular dynamics configurational entropy estimates. *Journal of Physical Chemistry B* **2006**, *110*, 8464–8473.
- [59] Baron, R.; Trzesniak, D.; De Vries, A. H.; Elsener, A.; Marrink, S. J.; Van Gunsteren, W. F. Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models. *ChemPhysChem* **2007**, *8*, 452–461.

- [60] Martínez, L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS ONE* **2015**, *10*, 1–10.