

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

FÍSICA MÉDICA

PATRICK KWAN

**ESTUDOS SOBRE FLEXIBILIDADE MACROMOLECULAR DE PROTEÍNAS EM
SOLUÇÃO**

Botucatu

2023

PATRICK KWAN

**ESTUDOS SOBRE FLEXIBILIDADE MACROMOLECULAR DE PROTEÍNAS EM
SOLUÇÃO**

Monografia apresentada ao Instituto de
Biotecnologia, Universidade Estadual
Paulista “Júlio de Mesquita Filho” –
UNESP, Campus Botucatu, para
obtenção do título de Bacharel em
Física Médica.

Orientador: Prof. Dr. Mario de
Oliveira Neto

Botucatu

2023

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Kwan, Patrick.

Estudos sobre flexibilidade macromolecular de proteínas em solução / Patrick Kwan. - Botucatu, 2023

Trabalho de conclusão de curso (bacharelado - Física Médica) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Mario de Oliveira Neto

Capes: 20901003

1. Substâncias macromoleculares. 2. Proteínas. 3. Vírus Chikungunya. 4. Estrutura molecular.

Palavras-chave: Flexibilidade macromolecular; Proteínas flexíveis; SAXS; nsp2.

Patrick Kwan

**ESTUDOS SOBRE FLEXIBILIDADE MACROMOLECULAR DE PROTEÍNAS EM
SOLUÇÃO**

Trabalho de conclusão do curso de Física Médica apresentado ao Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Botucatu, para a obtenção de título de Bacharel em Física Médica.

Orientador: Prof. Dr. Mario de Oliveira Neto

Comissão examinadora:

Prof. Dr. Mario de Oliveira Neto
UNESP – Câmpus de Botucatu

Prof. Ass. Luis Antonio Justulin Junior
UNESP – Câmpus de Botucatu

Prof. Ass. Dr. Joel Mesa Hormaza
UNESP – Câmpus de Botucatu

Botucatu, 09 de janeiro de 2023

Dedico este trabalho aos meus familiares e a minha noiva, Ana.

“A persistência é o menor caminho do êxito” (Charles Chaplin)

Agradecimentos

Agradeço a Deus primeiramente por ter me dado forças para concluir esse trabalho. Agradeço aos meus pais, André e Cirlene, por todo o carinho, amor e apoio que sempre me deram. Agradeço também a minha noiva, Ana, por todo companheirismo, pelos conselhos, incentivos e por sempre me apoiar em todas as decisões da minha vida. Agradeço também aos meus irmãos, Bruna, Denis e Andrew, pelas risadas, alegrias e por acreditarem em mim.

Agradeço ao meu orientador, Mario de Oliveira Neto, por todos os ensinamentos, por acreditar no meu trabalho e me dar todo o auxílio desde o início até o final deste trabalho.

Sou grato a todos os meus amigos que me motivam a sempre dar o meu melhor. Agradeço a Pró-Reitora de Pesquisa (PROPe) da Unesp por conceder uma bolsa PIBIC Reitoria pelo trabalho feito.

“Quando tudo parecer dar errado em sua vida, lembre-se que o avião decola contra o vento, e não a favor dele” (Henry Ford)

Resumo

Neste trabalho é utilizado a estrutura molecular do vírus Chikungunya, mais precisamente a proteína não estrutural 2 (CHIKV nsP2), uma proteína multifuncional essencial para a replicação viral do vírus Chikungunya, como modelo para o uso da técnica de espalhamento de raios-X a baixo ângulo (SAXS) e para melhor entendimento de como o pacote de análise ATSAS pode impactar em uma análise sobre flexibilidade macromolecular. A CHIKV nsp2 possui uma região flexível em sua estrutura molecular, caracterizando funções específicas muito importantes, proporcionando adequação e conformação estrutural em determinadas situações de ligação e formação de complexos. Utilizando a ferramenta BUNCH reconstruiu-se a estrutura da solução de nsP2 com base nas estruturas cristalinas de nsP2h e nsP2p. O método de otimização de conjuntos (*ensemble optimization method* - EOM) permitiu a formulação de possíveis modelos estruturais da CHIKV nsp2 coerentes com a sequência de aminoácidos de suas regiões flexíveis e desordenadas, remetendo a uma previsão confiável do comportamento desses componentes em solução, atuando em suas funções específicas, possibilitando um estudo adequado e completo das características que conduzem seu rendimento funcional.

Palavras-Chave: SAXS. proteínas flexíveis. nsp2. flexibilidade macromolecular.

Abstract

In this work, the molecular structure of the Chikungunya virus, more precisely the non-structural protein 2 (CHIKV nsP2), a multifunctional protein essential for the viral replication of the Chikungunya virus, is used as a model for the use of the Small-angle X-ray Scattering technique and to better understand how the Data analysis software ATSAS can impact an analysis of macromolecular flexibility. CHIKV nsp2 has a flexible region in its molecular structure, characterizing very important specific functions, providing conformity and structural conformation in certain binding situations and complex formation. Using the BUNCH tool, the structure of the nsP2 solution was reconstructed based on the crystalline structures of nsP2h and nsP2p. The ensemble optimization method (EOM) allowed the formulation of possible embedded models of CHIKV nsp2 consistent with the amino acid sequence of its flexible and disordered regions, leading to a reliable prediction of the behavior of these components in solution, their functions specific, providing an adequate and complete study of the characteristics that lead to its functional performance.

Keywords: SAXS. flexible proteins. nsp2. macromolecular flexibility.

LISTA DE FIGURAS

- Figura 1:** Esquema da linha de SAXS do Laboratório Nacional de Luz Síncrotron. (ITRI et. al., 2015) 15
- Figura 2:** Esquema do vetor de espalhamento q (ITRI et. al. 2015). 16
- Figura 3:** Gráfico de Kratky de um espalhamento de dados de dispersão ilustrando mudanças no comportamento da curva para uma pouco flexível (esfera), partículas parcialmente flexíveis (esfera com espiral aleatória) e partículas completamente flexíveis (espiral aleatória) (RAMBO et.al., 2011). 17
- Figura 4:** Intensidades de espalhamento e função de distribuição de pares $p(r)$ para diferentes objetos geométricos (SVERGUN et. al., 2003). 18
- Figura 5:** Ajuste do BUNCH (linha azul claro) com os dados experimentais (círculos abertos). Função dos pares de distribuição de distância ($p(r)$) inserto a esquerda. 23
- Figura 6:** Gráfico de Kratky de um espalhamento de dados de dispersão ilustrando um comportamento da curva de uma proteína parcialmente flexível. 24
- Figura 7:** Ajuste do EOM propondo quatro conformações coexistentes em solução (linha vermelha) com os dados experimentais (círculos abertos). 25
- Figura 8:** *Pool* de modelos gerados aleatoriamente pelo EOM (barras cinzas) e frequência das conformações mais prováveis de raio de giro da CHIKV nsp2 em solução (linha cheia + pontos). 25
- Figura 9:** *Pool* de modelos gerados aleatoriamente pelo EOM (barras cinzas) e frequência das conformações mais prováveis de diâmetro máximo da CHIKV nsp2 em solução (linha cheia + pontos). 26
- Figura 10:** Ajuste do SREFLEX propondo cinco refinamentos (linha azul) com os dados experimentais (círculos abertos). 27

LISTA DE TABELAS

Tabela 1: Parâmetros de análise SAXS de nsP2.	23
Tabela 2: Parâmetros estruturais das quatro conformações mais prováveis utilizadas para descrever os dados da CHIKV nsp2.	26
Tabela 3: Parâmetros estruturais das cinco conformações refinadas da CHIKV nsp2.	27

LISTA DE ABREVIACÕES E SIGLAS

CHIKV - Chikungunya

CHIKV nsP2 - Proteína não estrutural 2 do vírus da Chikungunya

$D_{\text{máx}}$ - Diâmetro máximo

EOM - Método de otimização de conjuntos

$I(q)$ - Intensidade de espalhamento

NMA - Análise de modo normal

nsP2h - Proteína não estrutural 2 helicase do vírus da Chikungunya

nsP2p - Proteína não estrutural 2 protease do vírus da Chikungunya

NTPase - Nucleosídeo-trifosfatase

PDB - *Protein Data Bank*

$p(r)$ - Função de distribuição de pares

q - Vetor de espalhamento

R_g - Raio de giro

RNA - Ácido ribonucleico

RTPase - RNA trifosfatase

SAS - *Small-Angle Scattering*

SASBDB - *Small Angle Scattering Biological Data Bank*

SAXS - Espalhamento de raios-X a baixo ângulo

SAXSMoW 2.0 - *SAXS Molecular Weight*

SREFLEX - *SAXS REFinement through FLEXibility*

SUMÁRIO

1. Introdução	13
1.1. Flexibilidade de uma proteína	13
1.2. Espalhamento de raios-X a baixo ângulo.	14
1.2.1. Vetor de Espalhamento (q)	15
1.2.2. Intensidade de Espalhamento $I(q)$	16
1.2.3. Kratky	16
2. Objetivos	19
2.1. Objetivo geral	19
2.2. Objetivos específicos	19
2.3. Chikungunya e a proteína não estrutural 2 (nsp2)	19
3. Materiais e métodos	20
3.1. EOM	20
3.2. BUNCH	21
3.3. SREFLEX	21
4. Resultados e discussões	22
5. Conclusão	28
6. Referências bibliográficas	28

1. Introdução

Muitas das proteínas possuem nelas regiões flexíveis que apresentam funções específicas, manifestando adaptações e variabilidades que somente os complexos flexíveis podem manifestar. Atualmente, estudos são realizados para entender e eventualmente prever o comportamento de uma proteína flexível.

Conseguir determinar a estrutura terciária e quaternária das proteínas em solução é um passo importante na caracterização completa de uma proteína. Utilizando espalhamento de raios-X a baixo ângulo (*small angle X-ray scattering* - SAXS), é possível estudar detalhadamente a estrutura geral das proteínas. A determinação direta do envelope molecular, a predição de estruturas quaternárias em oligômeros, a verificação de transições conformacionais que ocorrem por interação com ligantes, assim como, a visualização de processos de enovelamento e desenovelamento protéico (SVERGUN *et al.*, 2003) podem ser realizadas com SAXS (SVERGUN *et al.*, 1996) (ARES *et al.*, 2018). Também é possível analisar a flexibilidade macromolecular em solução através de análises como a análise de Kratky, Porod-Debye (RAMBO *et al.*, 2011), das ferramentas *ensemble optimization method* (EOM) (TRIA *et al.*, 2015), BUNCH (PETOUKHOV *et al.*, 2005) e *SAXS REFinement through FLEXibility* (SREFLEX) (PANJKOVICH & SVERGUN, 2016).

1.1. Flexibilidade de uma proteína

A flexibilidade em larga escala dentro de uma proteína multidomínio desempenha um papel importante em sua função biológica. Por exemplo, uma proteína de múltiplos domínios consiste em dois ou mais domínios conectados por ligantes flexíveis (EKMAN *et al.*, 2005) (LEVITT, 2009) que determinam a extensão dos movimentos entre domínios e, além disso, levam a transições conformacionais funcionalmente relevantes em larga escala. A determinação da estrutura dos ligantes flexíveis contendo proteínas de múltiplos domínios é experimentalmente difícil. Pode ser bastante desafiador usar a cristalografia de raios X para resolver uma estrutura com múltiplas conformações, uma vez que esta metodologia é mais aplicável a uma proteína bem dobrada com um único estado dominante (EKMAN *et al.*, 2005) (LEVITT, 2009).

A cristalografia, técnica consolidada para estudo estrutural de proteínas, não é capaz de analisar as regiões desordenadas pois não é realizada com a proteína em solução e, portanto, é capaz apenas de detectar tais regiões flexíveis pela indicação da ausência de densidade eletrônica nesses locais da estrutura proteica (DYSON & WRIGHT, 2005).

As regiões flexíveis encontradas em proteínas intrinsecamente desordenadas funcionam, muitas vezes, para transcrição em estado ativo da proteína, e podem ser classificadas segundo a composição e disposição dos aminoácidos. Estudos de genomas completos apontam que proteínas com regiões estruturais flexíveis são predominantes em organismos mais complexos, e a quantidade dessas proteínas nos organismos é totalmente proporcional com o grau de complexidade genômica dos mesmos. Além disso, o estudo aprofundado do banco de dados indica uma correlação evidente entre a repetição de regiões estruturais e proteínas intrinsecamente não estruturadas, demonstrando uma provável evolução por expansão repetida.

Além disso, são muitas as proteínas modulares eucarióticas que possuem domínios globulares independentes conectados por uma região flexível, que variam em comprimento e composição de aminoácidos polares, os quais retêm uma tendência a maiores extensões dessa região conectora dos glóbulos. Proteínas de regiões flexíveis mais curtas podem ser encontradas em bactérias com função reguladora.

A região flexível das proteínas modulares assume a função principal de "liberar" os domínios conectados para realizar uma ligação que demanda adequação ou reformulação estrutural para ser concretizada. Estudos recentes também evidenciam a presença de proteínas modulares em associações com outras proteínas e com moléculas de RNA, demonstrando uma possível função de reconhecimento e preparação para outra ligação. É válido ressaltar também que várias proteínas intrinsecamente desordenadas passam por estados semi-ordenados, de acordo com a ligação que será realizado, e a sequência de aminoácidos da região flexível pode proporcionar certa previsão sobre possíveis posições e formatos de sua parte desordenada (DYSON & WRIGHT, 2005).

Recentemente os estudos na área de proteínas intrinsecamente desordenadas têm avançado através de técnicas que permitem a observação do comportamento da flexibilidade dessas proteínas modulares em solução. Nesse sentido, SAXS vêm sendo utilizado como uma importante técnica na caracterização desses complexos sistemas biológicos.

1.2. Espalhamento de raios-X a baixo ângulo.

O espalhamento de raios-X a baixo ângulo (SAXS) é muito utilizado para análise das informações estruturais de heterogeneidades de densidade eletrônica com dimensões características de 1 a 100 nm (GLATTER & KRATKY, 1982). Com ela é possível obter informações como raio de giro, volume, superfície molecular e estado de oligomerização (SVERGUN & STUHRMANN, 1991). Abordagens teóricas e computacionais foram desenvolvidas para o tratamento e interpretação de dados de SAXS, permitindo uma análise

detalhada da estrutura global de proteínas (FRANKE *et al.*, 2017), a predição de estruturas quaternárias (SVERGUN *et al.*, 1998) (ARES *et al.*, 2018), a verificação de possíveis diferenças entre as estruturas no seu estado cristalino e em solução, assim como a determinação relativa dos domínios de proteínas.

A técnica vem demonstrando muita validade de análise por tratar de informações dos sistemas em solução, sem a necessidade de cristalografia (apenas como referência para a geração de possíveis modelos) e sem restrições exigentes de tamanho. Nos experimentos realizados com o uso de SAXS, as proteínas são colocadas em solução-tampão e expostas a raios-X de alto fluxo (exposição em fluxo minimiza danos por radiação às proteínas) num intervalo de tempo muito curto, milissegundos. Desse modo, o espalhamento ocasionado pela proteína é "subtraído" do espalhamento total considerando a solução tampão, e a própria análise de uma média de espalhamento por parte das partículas de proteína permite qualificar a ordem ou desordem estrutural das mesmas (MALABY *et al.*, 2015).

De uma maneira geral, o espalhamento consiste de um feixe de raios-X, que incide inicialmente num monocromador, definindo seu comprimento de onda. O feixe atravessa então um sistema de colimação e atinge a amostra. Com isso, ocorre um espalhamento dessa solução atingida e os fótons de raios-X atingem um detector, que pode ser uma placa de imagem ou um detector eletrônico (Figura 1).

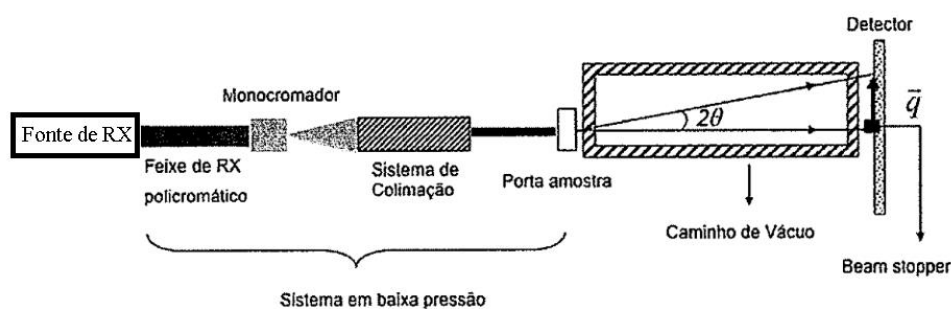


Figura 1: Esquema da linha de SAXS do Laboratório Nacional de Luz Síncrotron. (ITRI *et al.*, 2015)

1.2.1. Vetor de Espalhamento (q)

O vetor de espalhamento (q) é a diferença vetorial entre o feixe de raios-X que surge da amostra, mas não interage com ela, sendo o feixe de raios X que é espalhado para e atinge o detector em um determinada posição (Figura 2) (ITRI *et al.* 2015).

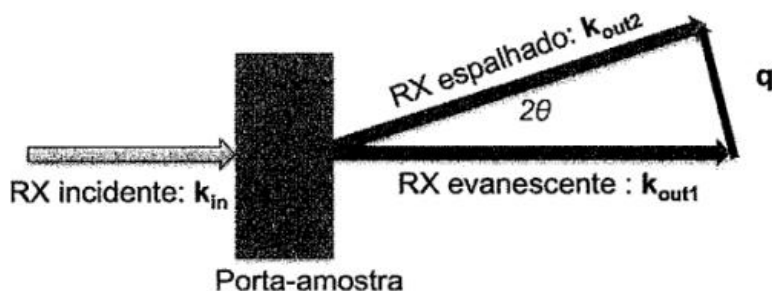


Figura 2: Esquema do vetor de espalhamento q (ITRI et. al. 2015).

1.2.2. Intensidade de Espalhamento $I(q)$

A intensidade de espalhamento $I(q)$ produzida por um material é uma função do ângulo de espalhamento e do vetor de espalhamento (q), e suas características estão relacionadas a densidade eletrônica $\rho(r)$ por uma transformada de Fourier. (A. CRAIEVICH, 2005).

1.2.3. Kratky

As curvas visualizadas no gráfico de Kratky (intensidade de espalhamento $I(q) \cdot q$ versus vetor de espalhamento (q)), mostrado na Figura 3, apresentam informações relevantes sobre a estrutura da proteína. A representação esperada de macromoléculas globulares não desordenadas (maior rigidez) é a de um pico (como um sino) decaindo em pequenas ondas ao longo da curva, remetendo ao comportamento assintótico em ângulos de alta dispersão, e a desconformidade dessa representação anuncia a provável desordem estrutural (maior flexibilidade) (RAMBO *et.al.*, 2011) (HOPKINS *et.al.*, 2017).

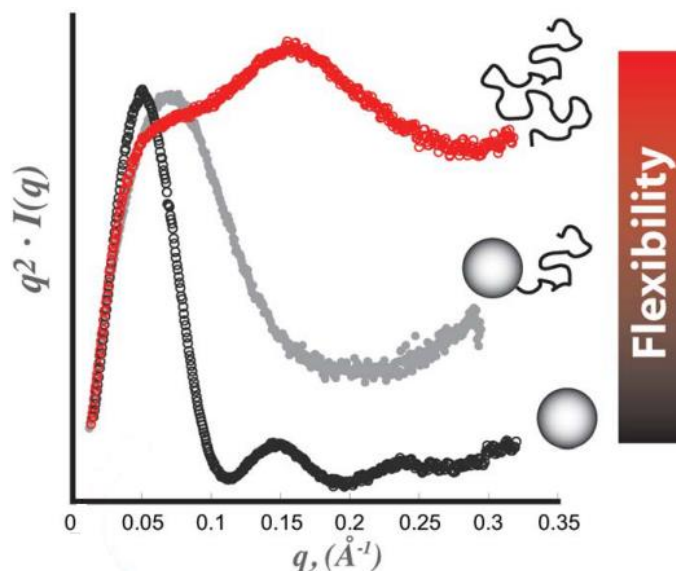


Figura 3: Gráfico de Kratky de um espalhamento de dados de dispersão ilustrando mudanças no comportamento da curva para uma pouco flexível (esfera), partículas parcialmente flexíveis (esfera com espiral aleatória) e partículas completamente flexíveis (espiral aleatória) (RAMBO et.al., 2011).

Os gráficos de Kratky podem avaliar qualitativamente a flexibilidade e/ou grau de desdobramento em amostras. As proteínas desdobradas (altamente flexíveis) devem ter um platô no gráfico de Kratky em q alto, enquanto as proteínas globulares compactas terão um pico em forma de sino (Gaussiano). Uma proteína parcialmente desdobrada (flexível) pode ter uma combinação de forma de sino e platô, ou um platô que decai lentamente para zero, assim indicado na Figura 3 (RAMBO *et.al.*, 2011).

Há também uma função de frequência de distâncias no espaço-real das moléculas ($p(r)$), calculada através de uma transformada inversa de Fourier da intensidade de espalhamento, $I(q)$. Ocorrendo uma descaracterização da curva da função $p(r)$ (forma de sino deslocado para direita ou esquerda, dependendo das características da forma das moléculas, mais alongadas ou de forma mais globular), evidencia-se também a desordem estrutural da proteína analisada, mostrada na Figura 4 (RAMBO *et.al.*, 2011).

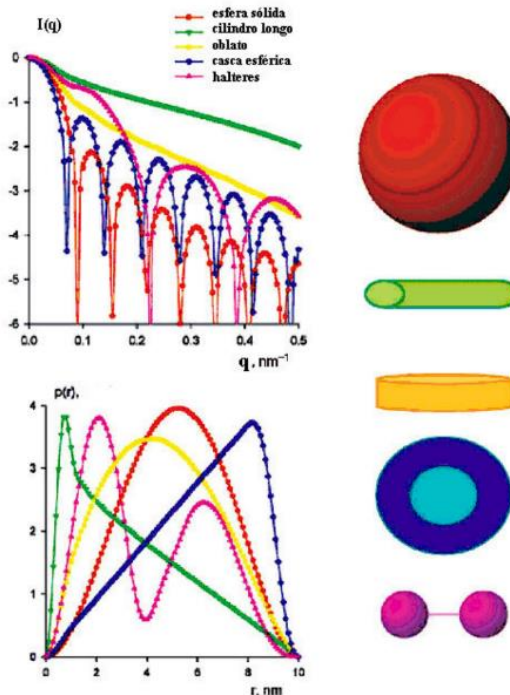


Figura 4: Intensidades de espalhamento e função de distribuição de pares $p(r)$ para diferentes objetos geométricos (SVERGUN et. al., 2003).

Dentre as ferramentas utilizadas em análise através da técnica de SAXS destaca-se, para o estudo de proteínas intrinsecamente desordenadas, o programa EOM, inserido no pacote de programas ATSAS, vem sendo uma estratégia pioneira na qualificação e demonstração estrutural de proteínas modulares. O programa realiza a análise em três etapas. Como a região flexível proporciona infinitas possibilidades de conformação estrutural das moléculas, a primeira etapa é responsável por gerar inúmeros possíveis modelos possíveis para os aminoácidos da região flexível em questão. Na segunda etapa, realiza-se o cálculo do perfil de espalhamento de cada uma das conformações previamente geradas. Na terceira e última etapa, o programa seleciona um conjunto de conformações que simbolizam várias outras conformações muito "semelhantes" geradas na primeira etapa (é empregado um algoritmo genérico que compara os parâmetros calculados, como o raio de giro (R_g) e a dimensão máxima ($D_{máx}$) da partícula), minimizando o número de possibilidades para análise. Findadas as três etapas, a análise dispõe de um conjunto de modelos estruturais prováveis para a proteína intrinsecamente desordenada, o que remete a um aumento na precisão de previsão nos estudos das proteínas modulares (TRIA et.al., 2015).

2. Objetivos

2.1. Objetivo geral

Este trabalho tem como objetivo estudar e verificar como a técnica de SAXS e os programas de análises podem auxiliar no estudo de proteínas que apresentam comportamentos complexos devido à presença de regiões flexíveis e intrinsecamente desestruturadas em sua sequência, para que eventualmente seja possível prever o comportamento de uma proteína flexível.

2.2. Objetivos específicos

Utilizar SAXS para analisar a estrutura da proteína não estrutural 2 (nsP2), proteína que desempenha vários papéis na replicação do RNA viral e nas interações vírus-hospedeiro (YEE-SONG LAW *et.al.*, 2021). O SAXS proporciona uma análise em cima de uma possível influência da região desenovelada no comportamento da proteína em solução.

Estudos estruturais de proteínas através de SAXS são de extrema importância para uma completa caracterização biofísica de sistemas biológicos. Objetiva-se o estudo das bases teóricas de SAXS, assim como a flexibilidade macromolecular, no qual visa-se compreender as novas descobertas feitas a respeito de proteínas flexíveis. Como a técnica de SAXS é sensível a mudanças conformacionais e de flexibilidade, estudos preliminares de quantificação de flexibilidade de uma proteína serão realizados.

O projeto tem como foco de estudo a proteína do vírus Chikungunya (CHIKV), referente ao código SASDK82 (<https://www.sasbdb.org/data/SASDK82>) da plataforma *Small Angle Scattering Biological Data Bank* (SASBDB).

2.3. Chikungunya e a proteína não estrutural 2 (nsp2)

O vírus da Chikungunya (CHIKV) é uma doença viral transmitida pelos mosquitos *Aedes aegypti* e *Aedes albopictus*.

A infecção por CHIKV geralmente causa febre, erupção cutânea e poliartrite em humanos (YEE-SONG LAW *et.al.*, 2021). A nsP2 é a maior proteína alfa viral madura que consiste em uma região de helicase N-terminal (nsP2h) e uma região de protease C-terminal (nsP2p) conectada através de um ligante flexível. A nsP2h exibe atividades de NTPase e RNA trifosfatase (RTPase) essenciais para a replicação de RNA viral. As helicases são alvos promissores de drogas antivirais porque suas atividades enzimáticas são essenciais para a replicação, transcrição e tradução do genoma viral. Neste caso, alguns inibidores também mostraram diminuir a replicação viral em culturas de células e modelos animais. A nsP2p exibe atividade de uma região de protease (enzimas que quebram ligações peptídicas entre os

aminoácidos das proteínas) de cisteína C-terminal, que é responsável pelo processamento de poliproteínas não estruturais. Até o momento, estruturas cristalinas das regiões N- e C-terminais de CHIKV nsP2 foram relatadas (YEE-SONG LAW *et.al.*, 2021), mas a estrutura tridimensional de nsP2 de comprimento total ainda não é conhecida.

Usando o método de espalhamento de raios-X de pequeno ângulo (SAXS), determinou-se que a nsP2 de comprimento total adota uma forma alongada por meio de um ligante interdomínio flexível. Estudos mostram que o processo de replicação viral pode tolerar a deleção de um resíduo ou a inserção de 10 resíduos flexíveis na região do ligante. A deleção de três ou cinco resíduos do ligante tornou-se letal para o vírus, o que pode ser devido a defeitos no processamento de poliproteínas e na montagem de complexos replicativos (YEE-SONG LAW *et.al.*, 2021).

3. Materiais e métodos

Os dados utilizados foram obtidos através do banco de dados SASBDB. A análise de Guinier do raio de giro foi conduzida usando a interface do programa Primus do pacote ATSAS (KONAREV *et.al.*, 2003). O ajuste dos dados experimentais e a avaliação da função de distribuição de pares de distância ($p(r)$), foram conduzidos utilizando o programa Gnom (SVERGUN, 1992). A análise do peso molecular foi calculada a partir dos dados SAXS usando o *webtool* SAXSMoW 2.0 (FISCHER *et.al.*, 2010) e o programa Crysol (SVERGUN *et.al.*, 1995) foi usado para simular o espalhamento e avaliar parâmetros estruturais a partir de estruturas cristalográficas.

Reconstruiu-se a estrutura da solução de nsP2 com base nas estruturas cristalinas de nsP2h (resíduos 1 a 463; PDB ID 6JIM) e nsP2p (resíduos 471 a 791; PDB ID 4ZTB), utilizando o programa BUNCH, estrutura no qual utilizou-se de modelo (YEE-SONG LAW *et.al.*, 2021). A modelagem de conjuntos para proteínas flexíveis e o ajuste da curva experimental foram conduzidos usando o método de otimização de conjuntos em sua versão 2.0 (EOM 2.0) (TRIA *et.al.*, 2015).

3.1. EOM

O programa *ensemble optimization method* (EOM) é um programa que ajusta a curva experimental de SAXS com a curva de espalhamento teórica de um conjunto de possíveis conformações de uma mesma proteína, gerando modelos estruturais baseados nas mais prováveis estruturas e parâmetros para uma proteína flexível em solução (TRIA *et.al.*, 2015). EOM é um programa que ajusta uma intensidade de espalhamento teórica média derivada de um conjunto de conformações para dados SAXS experimentais.

É uma abordagem popular para descrever dados de espalhamento de raios-X de pequeno ângulo (SAXS) de proteínas altamente desordenadas. O algoritmo EOM seleciona subconjuntos de estados coexistentes de grandes conjuntos de conformações aleatórias para ajustar os dados SAXS. O objetivo de utilizar o EOM é criar modelos semelhantes para que possam ser analisados e assim, tentar prever o tipo de comportamento que a flexibilidade da proteína pode exercer (SAGAR, A *et.al.*, 2021).

3.2. BUNCH

Realiza modelagem de proteínas de múltiplos domínios contra dados SAXS usando um corpo rígido combinado e abordagem de modelagem *ab initio*, permite a determinação da estrutura de domínio tridimensional de proteínas com base em vários conjuntos de dados de espalhamento de mutantes de exclusão quando a(s) estrutura(s) de domínios individuais estão disponíveis (PETOUKHOV *et.al.*, 2005).

O programa é principalmente orientado para proteínas de cadeia simples ou grupos simétricos com porção assimétrica contendo uma cadeia polipeptídica. A princípio, o programa BUNCH também pode ser usado para modelar complexos macromoleculares consistindo em várias subunidades quando nem todas as estruturas de subunidades são conhecidas. Nesse caso, não apenas o *loop* ausente dentro de uma única subunidade pode ser reconstruído, mas também a forma da subunidade ausente pode ser recuperada. Como os ligantes são frequentemente fortemente hidratados, foi adicionada ao BUNCH a possibilidade de aumentar a amplitude da porção de resíduos de *Dummins* resíduos representando loops ausentes em até 50%, o que permite que a água de ligação seja levada em consideração (PETOUKHOV *et.al.*, 2005).

As configurações dos *loops* fornecidos pelo BUNCH refletem a conformação média (geralmente flexível) dos *loops* e podem ser efetivamente usadas como um indicador do volume ocupado pelos *loops*, em vez de sua representação da verdadeira estrutura terciária (PETOUKHOV *et.al.*, 2005).

3.3. SREFLEX

O programa SREFLEX usa análise de modo normal (NMA) no espaço cartesiano para estimar a flexibilidade de modelos de biomacromoléculas de alta resolução e melhorar sua concordância com dados experimentais de raios-X de baixo ângulo e espalhamento de nêutrons (SAXS e SANS).

O método começa com uma certa conformação estrutural e um perfil SAS (*small-angle scattering*) correspondente que é relativamente inconsistente. A estrutura é automaticamente dividida em pseudo domínios com base na entrada do usuário ou na dinâmica de proteínas

prevista pela NMA. O algoritmo funciona em camadas, primeiro detectando grandes rearranjos seguidos por movimentos menores e mais localizados.

A saída consiste em um conjunto de modelos estruturais representando possíveis mudanças conformacionais que melhoram a consistência com o perfil experimental do SAS (PANJKOVICH & SVERGUN, 2016).

4. Resultados e discussões

Para compreender os programas de análise de SAXS, o estudo foi baseado na reconstrução da estrutura da solução de nsP2 com base nas estruturas cristalinas de nsP2h e nsP2p utilizando a ferramenta BUNCH. A análise de Guinier do raio de giro foi conduzida usando a interface do programa Primus do pacote ATSAS (KONAREV *et.al.*, 2003).

Parâmetros estruturais obtidos por SAXS da CHIKV nsp2 foram calculados pelo ajuste linear Guinier e através da função $p(r)$. O raio de giro (R_g) da nsP2 foi de 39.113 Å (Figura 5; Tabela 1), e a dimensão máxima da partícula ($D_{máx}$) foi de 123,3 Å (Figura 5; Tabela 1). O perfil da $p(r)$ da nsP2 atingiu o seu pico mais alto em aproximadamente 33 Å (Tabela 1), seguido por uma cauda longa estendida com um alto valor de r , sugerindo que a nsP2 adota uma conformação alongada e parcialmente dobrada em solução. Através do gráfico de Kratky da nsP2, obtido utilizando a *webtool* SAXS MoW 2.0 (FISCHER *et.al.*, 2010) (Figura 6), viu-se um pico em um valor de r mais baixo, seguido por uma descida lenta para a linha de base, indicando que em solução, o nsP2 de comprimento total é flexível.

O BUNCH pode, em princípio, ser usado para a modelagem de complexos macromoleculares constituídos por várias subunidades, quando nem todas as estruturas das subunidades são conhecidas. Portanto, reconstruiu-se a estrutura da solução de nsP2 com base nas estruturas cristalinas de nsP2h (resíduos 1 a 463; PDB ID 6JIM) e nsP2p (resíduos 471 a 791; PDB ID 4ZTB) usando BUNCH. O ligante ausente, não conhecido, entre os resíduos 464 a 470, foi representado como resíduos fictícios, gerados pelo próprio BUNCH (YEE-SONG LAW *et.al.*, 2021).

O ajuste estrutural calculado a partir do modelo BUNCH está em forte concordância com os dados experimentais (Figura 5).figura

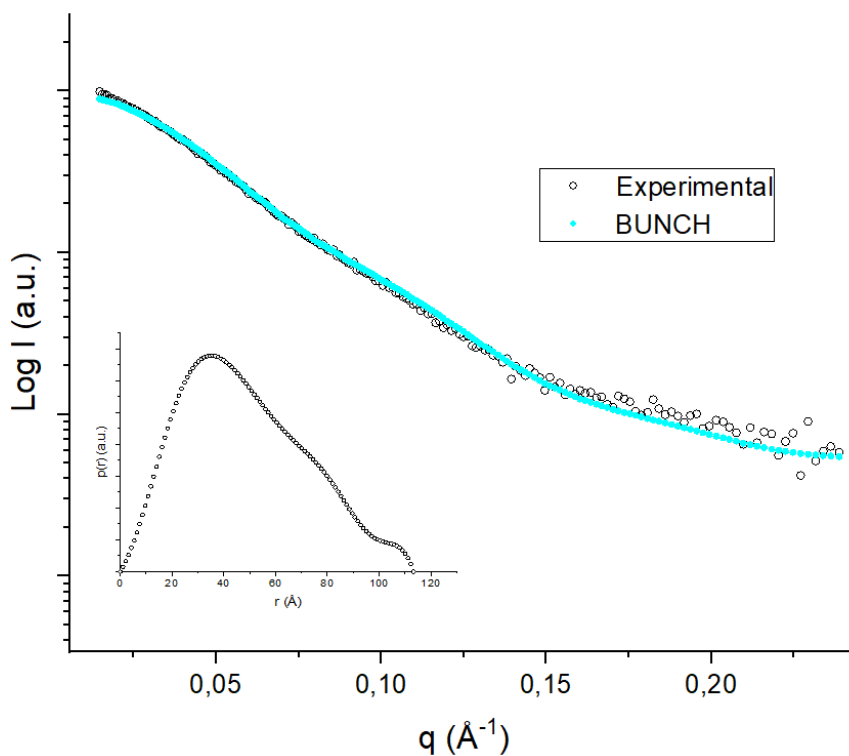


Figura 5: Ajuste do BUNCH (linha azul claro) com os dados experimentais (círculos abertos). Função dos pares de distribuição de distância ($p(r)$) inserto a esquerda.

Tabela 1: Parâmetros de análise SAXS de nsP2.

Modelo	R_g	$D_{máx}$	Pico da $p(r)$	Massa Mol.
SASDK82	39,2 Å	122 Å	32 Å	90 kDa
Reconstrução	39,113Å	123,3 Å	33 Å	95.0 kDa

Com a análise do gráfico de Kratky (Figura 6), percebe-se em solução uma molécula com perfil flexível.

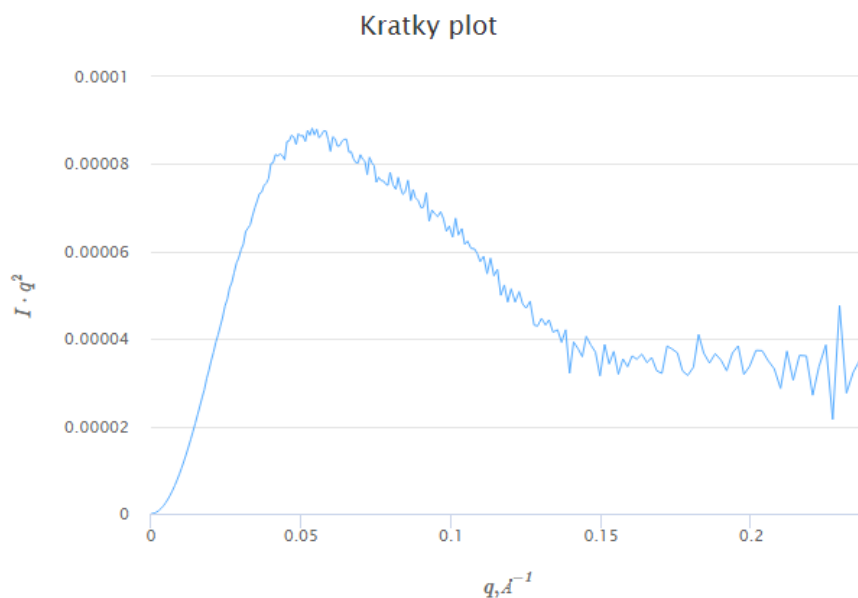


Figura 6: Gráfico de Kratky de um espalhamento de dados de dispersão ilustrando um comportamento da curva de uma proteína parcialmente flexível.

Através dos valores obtidos na reconstrução da estrutura feita com a ferramenta BUNCH, comparando-os com os parâmetros depositados no banco de dados da proteína da CHIKV nsp2 (Tabela 1) (SASDK82 - <https://www.sasbdb.org/data/SASDK82/>), consegue-se ver forte concordância entre os resultados alcançados.

Utilizou-se o programa EOM a fim de buscar modelos possíveis para a molécula em questão, gerando uma série de modelos e filtrando os modelos que com mais frequência apareciam, retirando assim os quatro modelos que mais se assemelham com os parâmetros fornecidos da CHIKV nsp2.

A partir do conjunto dos quatro modelos gerou-se uma curva de espalhamento, no qual foi visto uma ótima conformação, tornando os dados obtidos mais próximos da realidade (Figura 7; Tabela 2). Foram feitos ajustes dos dados experimentais obtendo a quantidade de vezes, em porcentagem, dos modelos mais frequentes, assim como os parâmetros estruturais (Figura 8 e Figura 9), mostrados na Tabela 2.

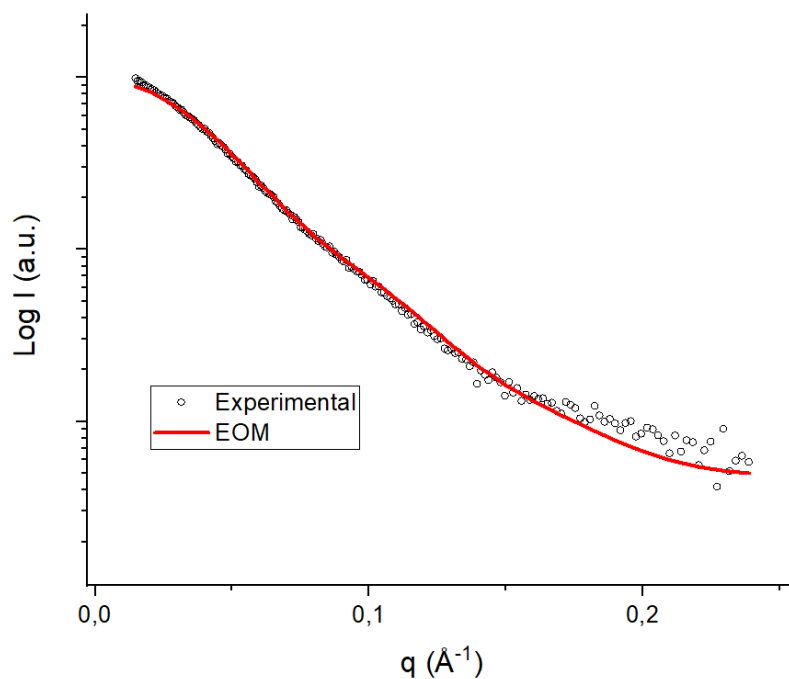


Figura 7: Ajuste do EOM propondo quatro conformações coexistentes em solução (linha vermelha) com os dados experimentais (círculos abertos).

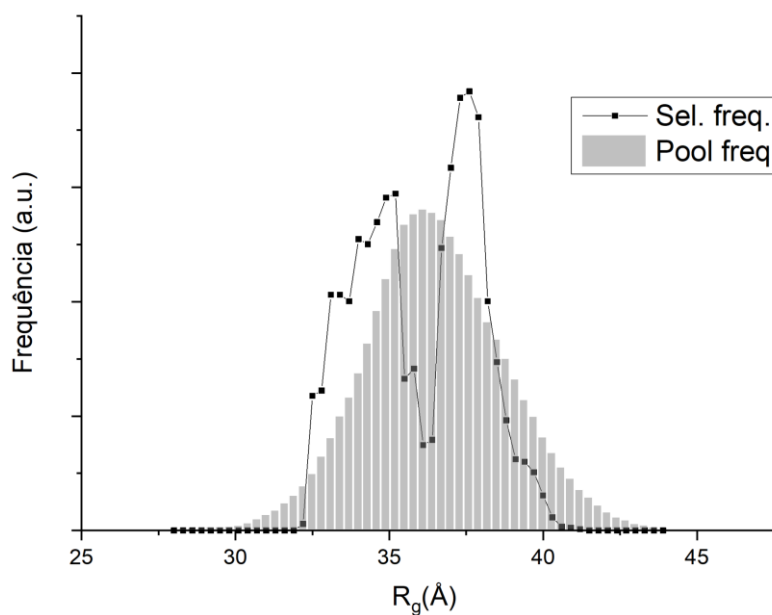


Figura 8: *Pool* de modelos gerados aleatoriamente pelo EOM (barras cinzas) e frequência das conformações mais prováveis de raio de giro da CHIKV nsp2 em solução (linha cheia + pontos).

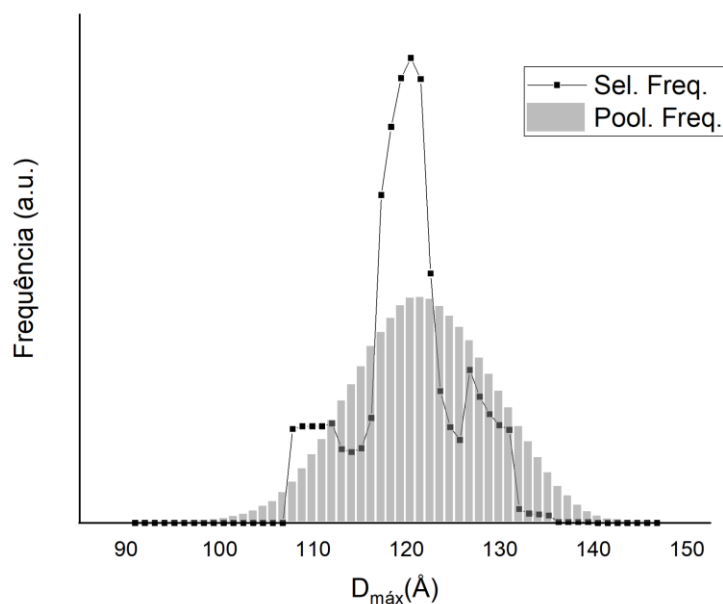


Figura 9: *Pool* de modelos gerados aleatoriamente pelo EOM (barras cinzas) e frequência das conformações mais prováveis de diâmetro máximo da CHIKV nsp2 em solução (linha cheia + pontos).

Tabela 2: Parâmetros estruturais das quatro conformações mais prováveis utilizadas para descrever os dados da CHIKV nsp2.

Modelo	Rg	Dmax	Frequência (%)
1	37,57	120,42	43%
2	33,12	110,38	29%
3	35,26	120,2	14%
4	38,26	120,54	14%

Por fim, utilizou-se o programa SREFLEX para refinamento da curva reconstruída. Inicialmente são gerados cinco modelos, uma curva de espalhamento (Figura 10) e os parâmetros adquiridos, que estão destacados na Tabela 3.

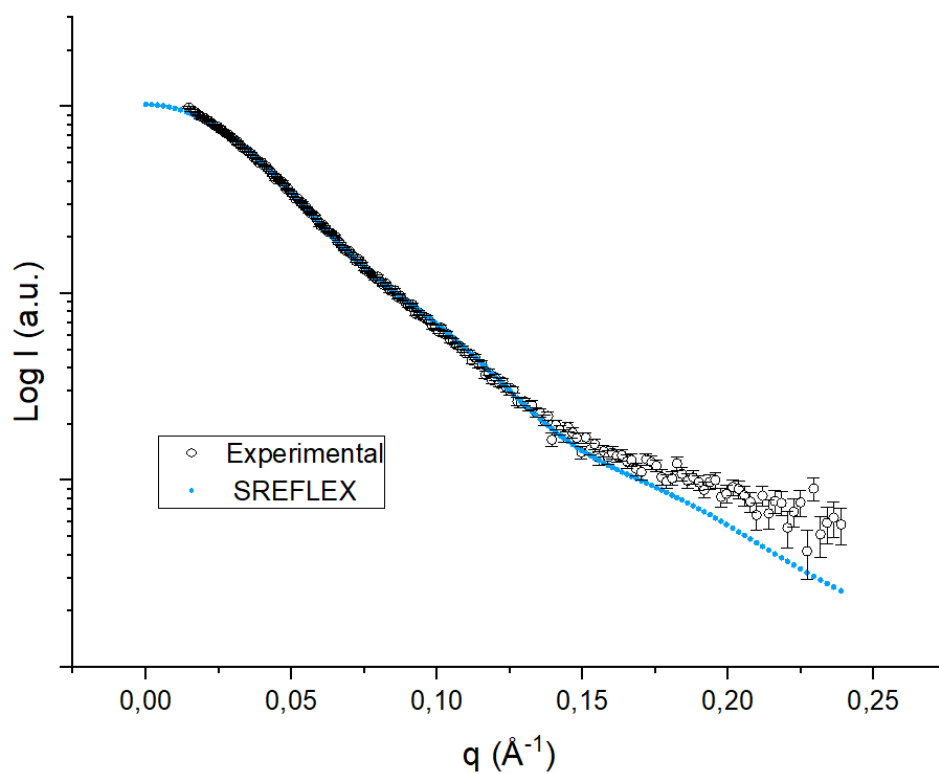


Figura 10: Ajuste do SREFLEX propondo cinco refinamentos (linha azul) com os dados experimentais (círculos abertos).

Tabela 3: Parâmetros estruturais das cinco conformações refinadas da CHIKV nsp2.

Modelo	Chi2	R _g
1	3,78	37,44 Å
2	3,87	37,04 Å
3	4,07	37,86 Å
4	4,1	36,88 Å
5	4,16	37,64 Å
SASDK82	4,03	38,2 Å

Quando se tem um modelo completo que não concorda com os dados SAXS, pode-se usar o SREFLEX, que usa a análise de modo normal (NMA) para classificar e alterar o modelo

de maneira flexível até que ele se ajuste aos dados. O SREFLEX fornece vários modelos e os respectivos ajustes aos dados. Também temos com detalhes qual o modelo possui conflitos e qual o modelo que se ajusta melhor. Entretanto, o SREFLEX funciona apenas para proteínas e funciona melhor com proteínas menores, não suportando simetria. Além disso, deve-se ter o modelo inicial que representa a proteína em sua forma completa.

Foi visto que através do uso do SREFLEX, obteve-se um Chi2 melhor do que o modelo inicial, mas demonstra-se uma curva não concordante com o experimental, mostrando que neste caso o SREFLEX não se enquadra bem, isso se dá pois a proteína se apresenta em sua forma reconstruída e não em sua forma completa, que é o foco do programa SREFLEX.

Nesses casos, o SREFLEX pode melhorar um pouco a consistência do perfil SAXS, mas não se espera encontrar uma solução adequada. No entanto, não é uma limitação do SREFLEX, mas sim uma limitação inerente das abordagens de refinamento baseadas em SAXS.

A abordagem apresentada neste trabalho visa auxiliar nos casos em que apenas um estado conformacional é conhecido da estrutura ou modelo cristalográfico e este modelo não corresponde aos dados SAXS correspondentes. Assim, o SREFLEX apresenta estruturas cristalográficas e dados experimentais SAXS em relativo desacordo.

5. Conclusão

Feitos tais estudos, percebe-se que através das análises de SAXS é possível analisar e entender o comportamento de uma estrutura flexível em solução. Com os modelos do EOM viu-se uma forma de prever o posicionamento e o comportamento da nsp2 em solução. Através dos modelos cristalográficos, gerados pelo EOM, foi possível perceber uma não intelecção entre as estruturas nsP2p e nsP2h, demonstrando que o ligante interdomínio flexível é o maior responsável pela replicação viral.

Em resumo, a compreensão do papel da flexibilidade de uma proteína é de extrema importância e é o início para compreender como uma proteína pode se adaptar e apresentar variabilidade em solução e que, através dos programas de análise SAXS, é possível compreender e prever como tal proteína pode se comportar em solução.

6. Referências bibliográficas

- ARES, E. A.; DE OLIVEIRA NETO, M.; POLIKARPOV, Igor. Biochemical characterization and low-resolution SAXS structure of two-domain endoglucanase BICel9 from *Bacillus licheniformis*. APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, v. 103, p. 1-

13, 2018.

- A. CRAIEVICH, "Small Angle X-Ray Scattering of Nanostructured Materials", chapter 8, cB 10, 2005.
- SAGAR, A., JEFFRIES, C.M., PETOUKHOV, M.V., SVERGUN, D.I., BERNADO P.. Comment on the Optimal Parameters to Derive Intrinsically Disordered Protein Conformational Ensembles from Small-Angle X-ray Scattering Data Using the Ensemble Optimization Method *J. Chem. Theory Comput.* 17(4), 2014-2021, 2021.
- EKMAN, D., BJÖRKLUND, A.K., ELOFSSON, A. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol.*, v. 348, p. 231–243, 2005.
- FRANKE, D.; PETOUKHOV, M.V.; KONAREV, P.V.; PANJKOVICH, A.; TUUKKANEN, A.; MERTENS, H.D.T.; KIKHNEY, A.G.; HAJIZADEH, N.R.; FRANKLIN, J.M.; JEFFRIES, C.M. AND SVERGUN, D.I. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Cryst.*, v. 50, p. 1212–1225, 2017.
- GLATTER, O. & KRATKY O. *Small Angle X-ray scattering*. Academic Press London, v. 36, p. 296-296, 1982.
- ITRI R., MORANDÉ, E. S., BARBOSA, L. R. S. Espalhamento de raios-X a baixos ângulos (SAXS) aplicado ao estudo de proteínas não interagentes em solução. Estratégias y avances em el estudio de toxinas de interés para la biomedicina, Benemérita Universidad Autonoma de Pueblam. v.1, p.171-190, e1, 2015.
- HOPKINS, J. B., GILLILAN, R. E., AND SKOU, S. *Journal of Applied Crystallography*. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis 50, 1545-1553, 2017.
- LEVITT, M. Nature of the protein universe. *Proc. Natl. Acad. Sci. USA.*; v. 106, p. 11079–11084, 2009.
- MANALASTAS-CANTOS, K., KONAREV, P. V., HAJIZADEH, N. R., KIKHNEY, A. G., PETOUKHOV, M. V., MOLODENSKIY, D. S., PANJKOVICH, A., MERTENS, H. D. T., GRUZINOV, A., BORGES, C., JEFFRIES, C. M., SVERGUN, D. I. & FRANKE, D. J. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Appl. Cryst.* v. 54, p. 343-355, 2021.
- PANJKOVICH, A. & SVERGUN, D. I. (2016). Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Phys. Chem. Chem. Phys.* v. 18, p. 5707–5719, 2016.

- PETOUKHOV, M.V. AND SVERGUN, D.I. Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, 89, 1237-1250, 2005.
- PIADOV, V. ; ARES, E.A.; DE OLIVEIRA NETO, M.; A. F. CRAIEVICH ; POLIKARPOV, I. . SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *PROTEIN SCIENCE*, v. 28, p. 454-463, 2019.
- RAMBO, R. P., & TAINER, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, 95(8), 559–571, 2011.
- SOARES, J.L., EDITORA SCIPIONE & TORTORA, GERARD J. *Microbiologia*. 10. ed. – Porto Alegre: Artmed, 2012.
- SVERGUN, D.I. & STUHRMANN, H.B. New developments in direct shape determination from small-angle scattering 1.Theory and model calculations. *ActaCrystallogr. A*, v. 47, p. 736-744, 1991.
- SVERGUN, D.I.; VOLKOV, V. V.; KOZIN, M. B.; STUHRMANN H. B. New developments in direct shape determination from small-angle scattering. 2. uniqueness. *Acta Crystallographica A*, v. 52, p. 419-426, 1996.
- SVERGUN, D.I.; ALDAG, I.; SIECK, T.; ALTENDORF, K.; KOCH, M. H. J.; KANE, D.J.; KOZIN, M.B.; GRUBER G. A Model of the Quaternary Structure of the Escherichia coli F1 ATPase from X-Ray Solution Scattering and Evidence for Structural Changes in the Delta Subunit during ATP Hydrolysis. *Biophysical Journal*, v. 75, p. 2212–2219, 1998.
- SVERGUN, D.I.; KOCH, M. H. J. Small-angle scattering studies of biological macromolecules in solution. *Reports Progress Physics*, v. 66, p. 1735 – 1782, 2003.
- YEE-SONG LAW; SAINAN WANG; YAW BIA TAN; ORION SHIH; AGE UTT; WEI YANG GOH; BING-JUN LIAN; MING WEI CHEN; U-SER JENG; ANDRES MERITS; DAHAI LUO. Interdomain Flexibility of Chikungunya Virus nsP2 Helicase-Protease Differentially Influences Viral RNA Replication and Infectivity. *Journal of Virology*, 95(6):e01470-20, 2021.