Universidade de São Paulo
Instituto de Física

Estudos estruturais de proteínas em solução por SAXS utilizando luz síncrotron

Hannes Fischer

Orientador: Prof. Dr. Aldo Felix Craievich

Co-Orientador: Prof. Dr. Igor Polikarpov

Tese apresentada ao Instituto de Física da
Universidade de São Paulo para obtenção
do título de Doutor em Ciências

Banca examinadora:

Prof. Dr.	Aldo Felix Craievich	IF-USP
Profa. Dra.	Iris Concepcion Linares de Torriani	IFI-UNICAMP
Profa. Dra.	Rosangela Itri	IF-USP
Prof. Dr.	Sylvio Roberto Accioly Canuto	IF-USP
Profa. Dra.	Yvonne Primerano Mascarenhas	IFSC-USP

São Paulo
2005

SBI-IFUSP

305M810T4348

Prof. Armando Corbani Ferraz
Presidente da Comissão de Pós Graduação

del. 11/04/05.

539.758
F529.2
D

Universidade de São Paulo
Instituto de Física

Estudos estruturais de proteínas em solução por

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Fischer, Hannes

Estudos estruturais de proteínas em solução por SAXS
utilizando luz síncrotron. São Paulo - 2005.

Tese (Doutoramento) – Universidade de São Paulo
Orientador: Prof. Dr. Aldo Felix Craievich

Instituto de Física – Departamento de Física Aplicada
Área de Concentração: Física

Unitermos:

1. Física experimental;
2. Macromolécula;
3. Fenômenos bioquímicos;
4. Química supramolecular.

USP/IF/SBI-041/2005

Sou grato....

Em especial a Deus que durante a tese colocou a amada Danielle na minha vida e nos deu um bendito filho. Como minha esposa tornou doce minha vida!

A meus pais e irmãos que sempre estiveram ao meu lado dando o estímulo e o ânimo necessários para que esse trabalho fosse desenvolvido com capricho e dedicação.

A Deus que me deu a oportunidade de conhecer o Prof. Aldo Craevich e por ter sido seu orientando. Excelente profissional e incentivador!

Ao Igor Polikarpov, meu coorientador que me estimulou a estudar macromoléculas em solução desde o início e que, de certa forma, é responsável pelos resultados deste trabalho.

Aos professores do IFSC, particularmente ao Richard Garratt e Glaucius Oliva, pela oportunidade que me deram de permanecer no IFSC por mais de 2 anos desfrutando de todos os recursos do mesmo, principalmente os computacionais.

Aos meus colegas de mestrado e doutorado com os quais passei preciosos momentos e por escrevermos vários artigos juntos. Gostaria de citar em especial o Stefano Trapani, com quem dividi a mesma sala por alguns anos e que sempre esteve pronto para me ajudar e citar, da mesma forma, Hamilton Napolitano, Frederico Ferreira, Fábio Macedo, Sandra Martha, Ricardo Aparício, Júlio Borges e Ricardo Cabrera.

Agradeço aos coordenadores das linhas de SAXS do LNLS, de forma especial ao Guinther Kellermann por sua ajuda e dedicação.

Ao pessoal do grupo de cristalografia do IFUSP e amigos que sempre me hospedaram em suas casas quando já não morava mais aqui. Gostaria de lembrar em especial do José Mario Prison que nunca disse não para me receber, o Luca, Antonio Carlos, Luciano, Evaldo e Valdir.

As secretárias do IFUSP e IFSC, em especial a Lia (Maria Mavília) que sempre estiveram prontas para me ajudar.

Também gostaria de agradecer de forma especial a FAPESP pelo suporte financeiro e ao revisor por ela indicado que sempre contribuiu positivamente com o trabalho.

Minha gratidão também a EMBRACO S.A. que permitiu minha ausência por períodos consecutivos para que eu pudesse terminar a redação da tese.

Resumo

Os objetivos da tese são, investigar (i) as estruturas terciária e quaternária de diversas proteínas em solução e (ii) as mudanças conformacionais de proteínas induzidas por ligantes e/ou temperatura, variando-se a concentração dos mesmos, etc. A principal técnica utilizada foi a de espalhamento de raios X a baixos ângulos (no inglês SAXS de small-angle X-ray scattering), associada a outras metodologias como cristalografia e modelagem de proteínas.

As contribuições dos estudos realizados de várias proteínas de interesse científico foram as seguintes: (i) propôs-se pela primeira vez um modelo de um receptor nuclear (RXR) composto por dois domínios, um de ligação ao ligante e outro ao DNA; (ii) elucidou-se o estado configuracional da fosfo-enol-piruvato carboxoquinase em solução; (iii) foi determinado o efeito de agregação e mudança conformacional induzido por ligantes na isoenzima fosfofrutoquinase e (iv) determinou-se a estrutura em solução da interleucina humana 22 e propôs-se modelos de interação com seus receptores.

Além disso, (i) foi desenvolvido um método para determinar o estado oligomérico de uma proteína em solução, em quaisquer condições de tampão e concentração, utilizando medidas de SAXS em escala relativa, e (ii) mostrou-se que a densidade das proteínas é uma função do seu peso molecular, contrariando conceitos clássicos, tendo as proteínas pequenas (< 20kDa) uma densidade mais elevada que as maiores.

Abstract

The purpose of this thesis is to investigate (i) the tertiary and quaternary structure of several proteins in solution and (ii) conformational changes that such proteins undergo when they bind to ligands or temperature, pH, etc. conditions are varied. The main experimental technique utilized in this study was Small Angle X-Ray Scattering (SAXS) associated to other techniques and approaches like protein crystallography and homology modeling.

The main contributions related to this work are: (i) for the first time a model for a nuclear receptor (RXR) containing two domains was proposed, one DNA binding domain and one ligand binding domain; (ii) the oligomeric state of the phosphor-enol-pyruvate carboxylase in solution was determined; (iii) the conformational changes induced by ligands to the isoenzyme phosphofructokinase was determined and (iv) the solution structure of human interleukin revealed and a model with its receptor proposed.

Additionally, (i) a method for determining the oligomeric state of a protein in solution under any buffer conditions and concentration using SAXS in relative scale is proposed and (ii) it was also shown that protein density is a molecular weight dependent function, where small proteins with less than 20kDa have in average a larger density than bigger ones.

Sumário

1	<i>Proteínas em solução: Estrutura e função</i>	1 - 1
1.1	Introdução	1
1.2	Estruturas das proteínas	3
1.2.1	Estrutura primária	3
1.2.2	Estrutura secundária	4
1.2.3	Estrutura terciária	5
1.2.4	Estrutura quaternária	6
1.3	Proteínas estudadas nesta tese	7
1.4	Objetivos e organização da tese.	8
2	<i>A técnica de SAXS</i>	2 - 1
2.1	SAXS como ferramenta para elucidação do envelope molecular de macromoléculas em solução	1
2.1.1	Espalhamento de raios X a baixos ângulos por sistemas coloidais	4
2.1.2	Informação contida em uma curva de SAXS	13
2.2	Cálculo de intensidade de SAXS a partir das coordenadas atômicas	13
2.3	Modelagem de configurações espaciais de domínios de proteínas utilizando dados de cristalografia de proteínas	16
2.4	Reconstrução <i>ab-initio</i> da forma da proteína a partir de resultados de SAXS	16
2.4.1	DAMMIN (Dummy atom modelling)	16
2.4.2	Funções de envelope	19
2.5	Arranjo experimental	20
2.6	Condições experimentais	21
2.6.1	Efeitos instrumentais ("Smearing")	22
2.6.2	Efeito da espessura da amostra na intensidade de espalhamento	23

3	<i>Determinação do estado oligomérico de macromoléculas em solução a partir de medidas de saxs em escala relativa</i>	3 - 1
3.1	Introdução	1
3.2	Efeito do valor máximo do módulo do vetor de onda utilizado na determinação do volume do objeto espalhador esférico	4
3.3	Efeito do valor máximo do módulo do vetor de espalhamento ($q_{\max}$) utilizado na determinação do volume de 1148 proteínas obtidas do PDB	7
3.4	Cálculo do volume a partir de dados experimentais de SAXS de várias proteínas em solução e conclusões	10
3.5	Dependência da densidade de proteínas em solução em função do peso molecular	21
4	<i>Estudos estruturais de receptores nucleares</i>	4 - 1
4.1	Introdução	1
4.2	Envelope de baixa resolução do receptor de retinoide X (RXR)	2
4.2.1	Mecanismos de funcionamento dos receptores nucleares	2
4.2.2	O RXR pode atuar sinergisticamente	3
4.3	Resultados	5
4.3.1	Medidas de SAXS dos dímeros e tetrâmeros de hRXR α DBD-LBD	5
4.3.2	Determinação ab initio do envelope da proteína	9
4.3.3	Estrutura do dímero e do tetrâmero de hRXR α DBD-LBD	14
5	<i>Estudos estruturais de outras proteínas em solução</i>	5 - 1
5.1	Introdução	1
5.2	Phosphoenolpyruvate carboxykinase (PEPCK)	2
5.2.1	Motivação do estudo da proteína PEPCK em solução	2
5.2.2	Resultados experimentais, análise de dados e conclusões	5
5.3	Mudanças conformacionais e estados de agregação na proteína PFK-2	9
5.3.1	Introdução e objetivos	9
5.3.2	Resultados experimentais, análise de dados e conclusões	11
5.4	Estrutura da interleucina humana 22 (hIL-22) em solução	20
5.4.1	Introdução e objetivos	20
5.4.2	Resultados, discussões e conclusões das medidas de SAXS e DLS da hIL-22 em solução	20
6.	<i>Conclusões e sugestões para trabalhos futuros.</i>	6-1

Abreviações utilizadas na tese

- $\langle \rangle_{\Omega}$ – Média 3D, também chamada promediação espacial
- D_{max} – dimensão máxima da partícula espalhadora
- $\gamma_0(r)$ – Função característica de Porod e representa a probabilidade de encontrar elemento de densidade eletrônica a uma distância r de outro elemento de densidade eletrônica.
- PDB – Brookhaven Protein Data Bank
- $p(r)$ – Função de distribuição de distâncias.
- SAXS – Espalhamento de raios X a baixos ângulos (abreviação do inglês – Small-Angle X-ray Scattering)
- XRD – Difração de raios X a altos ângulos (abreviação do inglês – X-Ray Diffraction)
-

1	<i>Proteínas em solução: Estrutura e função</i>	1 - 1
1.1	Introdução	1
1.2	Estruturas das proteínas	3
1.2.1	Estrutura primária	3
1.2.2	Estrutura secundária	4
1.2.3	Estrutura terciária	5
1.2.4	Estrutura quaternária	6
1.3	Proteínas estudadas nesta tese	7
1.4	Objetivos e organização da tese.	8

1 PROTEÍNAS EM SOLUÇÃO: ESTRUTURA E FUNÇÃO

1.1 Introdução

Organismos vivos são extremamente complicados e organizados. Os organismos que podemos ver a olho nu são em sua maioria compostos de muitas células e geralmente de vários tipos. Estas células por sua vez possuem estruturas sub celulares ou organelas, que são compostas por um grande número de macromoléculas, as quais, por si só, apresentam uma estrutura tridimensional complexa e intrigante, mesmo sendo compostas por unidades mais simples como os aminoácidos e açúcares, os quais são compostos basicamente por alguns elementos químicos (Figura 1). Vale lembrar que a conformação de uma macromolécula se deve às interações químicas das unidades monoméricas básicas que a compõe. A estrutura biológica sempre tem uma função associada à mesma, desempenhando um papel fundamental na existência de qualquer organismo vivo.

A determinação da seqüência de aminoácidos da insulina por Frederick Sanger, em 1953, foi um marco na bioquímica porque mostrou pela primeira vez que uma proteína tem uma seqüência de aminoácidos precisamente definida. Isto estimulou outros cientistas a fazerem estudos de seqüência de uma ampla variedade de proteínas. Hoje, milhares de proteínas já foram seqüenciadas.

Uma série de estudos marcantes revelou que as seqüências de aminoácidos das proteínas são geneticamente determinadas. A seqüência de nucleotídeos no DNA especifica

uma sequência complementar de nucleotídeos no RNA que, por sua vez, especifica a sequência de aminoácidos de uma proteína. Cada um dos 20 aminoácidos é codificado por uma ou mais sequências específicas de três nucleotídeos.

As proteínas exercem inúmeras funções bioquímicas como, catálise, transporte, armazenamento, proteção, exploração, regulação, construção de outras proteínas, mobilidade celular, etc., e são agrupadas geralmente como sendo, fibrosas, globulares ou de membrana. Partindo do pressuposto que a função da proteína, que é controlada por agentes fisiológicos, está associada com o arranjo de seus elementos de estrutura, pesquisadores têm investigado estes aspectos de arquitetura protéica e o movimento de domínios por muitos anos [1]. Vários estudos demonstraram que muitas mudanças conformacionais em proteínas podem ser caracterizadas, esquematicamente, como movimentos de corpo rígido de segmentos de estrutura preservada [2,3]. No decorrer da tese, mostrar-se-ão algumas aplicações de movimentos de corpo rígido estudados pela técnica de espalhamento de raios X a baixos ângulos (SAXS).

Hoje, conhecemos aproximadamente 30 mil estruturas de proteínas e estima-se que em 2020 conheceremos mais de 1 milhão de estruturas. A grande questão é como essas proteínas interagem umas com as outras, ou seja, como funcionam estes complexos biológicos. Para caminhar nesta direção muitos esforços têm sido feitos, e temos confiança que a técnica de SAXS contribuirá significativamente nesta questão. Outras técnicas muito promissoras são: microscopia eletrônica, difração de proteínas por lasers de raios X e técnicas espectroscópicas como NMR. Devemos lembrar também que o poder computacional em 2020 deverá ser, pelo menos, 100 vezes maior que o de hoje. Aliando as técnicas experimentais aos métodos computacionais (bancos de dados sofisticados, simulações de dinâmica molecular, docking, etc.), poderemos, num futuro próximo, ter à disposição bancos de dados de complexos biológicos e entender como funcionam, com isto compreendendo melhor como interagir com eles

Figura 1 – Ilustração da complexidade de uma célula

de forma ativa e obter resultados para tratamentos de doenças pelo menos 100 vezes mais rápido que hoje.

1.2 Estruturas das proteínas

1.2.1 Estrutura primária

As proteínas são as macromoléculas mais abundantes em um organismo e participam de todos os processos biológicos. Fazem parte de todas as estruturas celulares, atuam como, catalisadores, hormônios, antibióticos e anticorpos e são os principais constituintes de penas, cabelos, unhas, músculos, chifres, invólucros virais, clara de ovo, sedas, teias de aranha e muitos outros materiais biológicos.

No entanto, as proteínas são mais que uma reunião monótona de átomos – estes átomos estão organizados em unidades, constituindo aminoácidos. Existem apenas 20 aminoácidos diferentes compondo as proteínas. As proteínas são, portanto, polímeros de aminoácidos.

Na Figura 2 a está representado um fragmento de proteína, constituído de apenas 5 aminoácidos (peptídeos), que são, da esquerda para a direita, Arginina, Prolina, Alanina, Tirosina e Serina, respectivamente. Para facilitar as ligações químicas vamos usar uma representação plana (Figura 2b) deste fragmento ao invés de sua representação tridimensional (Figura 2a).

Figura 2 – (a) Representação em 3D de um peptídeo composto por 5 aminoácidos e (b) Representação plana

Observe que há uma regularidade na estrutura do peptídeo. Os átomos de nitrogênio são sempre intercalados por dois de carbono, formando um "esqueleto". A ligação entre os aminoácidos (ligação peptídica) é estabelecida por uma reação entre o grupo carboxílico de um aminoácido e o grupo amino de outro. Quando os aminoácidos polimerizam, o primeiro

tem seu grupo amina livre e é conhecido como o N terminal, sendo que o último tem seu grupo carboxila livre sendo assim conhecido como o C terminal. A ligação peptídica une dois aminoácidos resultando na formação de um dipeptídeo com a perda de uma molécula de água, como podemos ver na ilustração abaixo:

Podemos notar que todos os aminoácidos possuem uma estrutura idêntica, ou seja, um grupo amina, outro carboxílico e uma cadeia lateral. Assim, além da regularidade de dois átomos de carbono entre um de nitrogênio, também se observa que ao átomo de nitrogênio liga-se um átomo de hidrogênio e ao átomo de carbono, ligado diretamente ao nitrogênio, há um átomo de oxigênio. Estes átomos de hidrogênio e de oxigênio são muito importantes para a estruturação e estabilização de uma proteína, promovendo, por exemplo, pontes de hidrogênio.

O que mantém e determina a estrutura de uma proteína e as interações entre elas, na maioria das vezes, são forças de interação bem fracas (menores que 30kJ/mol como pontes de hidrogênio, forças de van der Waals, interações iônicas e hidrofóbicas). Ligações covalentes (fortes) mantêm átomos unidos permitindo a formação de moléculas, porém as não covalentes (fracas) acontecem em nível inter ou intramolecular.

1.2.2 Estrutura secundária

As restrições geométricas (distâncias máximas entre ligações covalentes, ângulos restritos que as ligações podem fazer, raio de van der Waals) e químicas (ligações de hidrogênio, atrações/repulsões iônicas, interações hidrofóbicas, etc.) impostas aos aminoácidos, determinam a forma com que aminoácidos se arranjam localmente, dando origem a estruturas como folhas beta e alfa hélices, por exemplo, conhecidas como sendo elementos da estrutura secundária da proteína.

Figura 3 – Exemplo de estrutura secundária. Esquerda: Folha beta paralela. Direita: Folha alfa.[4]

1.2.3 Estrutura terciária

A estrutura terciária de uma proteína é basicamente o enovelamento global de uma proteína. Ou seja, os aminoácidos se agrupam em estruturas secundárias, que por sua vez, por intermédio de aminoácidos sem aparente estrutura secundária (“loops”), formam uma estrutura terciária.

Talvez a diferença básica entre estrutura terciária e secundária seja que na primeira as interações são de maior distância enquanto que na segunda as interações são de curta distância. Isto é, não é incomum o 1º e o 100º aminoácido interagirem formando uma ligação estável. Outra diferença é que a estrutura terciária não é regular, sendo que a secundária é geralmente bem regular.

Geralmente a conformação adotada por uma proteína é globular, pois nestas condições a relação entre a superfície e o volume da mesma é minimizada, diminuindo as interações da proteína com o seu ambiente.

Figura 4 – Exemplo de estrutura terciária (ilustrado o segundo domínio da fosforilase).[4]

1.2.4 Estrutura quaternária

Muitas proteínas consistem de duas ou mais cadeias polipeptídicas com estrutura terciária bem definida, consideradas subunidades da proteína. Esta organização de subunidades de cadeias polipeptídicas independentes constitui a estrutura quaternária de uma proteína.

As ligações dominantes nas organizações estruturais de segunda, terceira e quarta ordem são predominantemente não covalentes, sendo que somente a primeira é basicamente covalente e toda a informação que uma proteína tem (como se enovelar, por exemplo) está contida na sequência da estrutura primária da mesma.

Algumas das vantagens de estruturas quaternárias são, maior estabilidade, eficiência genética (é necessário muito menos DNA para gerar um oligômero constituído por subunidades iguais do que diferentes de mesma massa), junção de sítios catalíticos (o monômero por si só pode não funcionar como um sítio de ação enzimático completo, porém um oligômero teria esta capacidade) e cooperatividade.

Figura 5 – Exemplo de estrutura quaternária. Hemoglobina, composta por 4 subunidades ($\alpha_1, \alpha_2, \beta_1, \beta_2$). Em (a) temos a hemoglobina não ligada ao oxigênio e em (b) a mesma ligada. No centro ilustrando a mudança conformacional induzida por essa ligação ao oxigênio em 2 subunidades (α_2, β_1). [4].

O reconhecimento biomolecular geralmente ocorre devido a complementaridade entre o ligante e a proteína. Estas ligações são em geral fracas o que leva a uma interação dinâmica em nível molecular entre proteínas, hormônios, receptores, etc... Existem muitas vantagens no fato de ligações fracas serem responsáveis pela dinâmica molecular, porém também desvantagens, uma delas seria a de que nós não sobreviveríamos por muito tempo

a temperaturas muito maiores ou menores que 25°C. A temperaturas acima de 40°C a maioria das proteínas perde sua função, pois as interações fracas pelas quais é formada não subsistem mais, desnaturando.

1.3 Proteínas estudadas nesta tese

Estudamos quatro proteínas nesta tese. Para termos uma idéia quais foram e o que foi elucidado, resumimos as suas funções e o tipo de estudo realizado na tabela abaixo:

Proteína	Função	Aplicabilidade	Tipo de estudo
Domínio de ligação ao ligante e DNA do receptor de ácido X retinóico (RXR)	Mediam a ação de hormônios sendo ativados por ligantes e moduladores da expressão de genes específicos	Oncologia, dermatologia, obesidade, visão, etc...	Envelope molecular do receptor composto por dois domínios.
Fosfo-enol-piruvato carboxiquinase (PEPCK)	Enzima-chave no metabolismo de açúcares do parasita <i>Trypanosoma cruzi</i>	Combate à doença de Chagas	Estado configuracional da mesma em solução
Isoenzima fosfofrutoquinase (PFK-2)	Enzima-chave no metabolismo de açúcares do parasita <i>Escherichia coli</i>	Pode ser útil no combate a doenças causadas por parasitas.	Regulação alostérica associada à agregação de proteínas e inibição enzimática.
Interleucina humana 22 (hIL-22)	Secretadas por leucócitos que afetam o crescimento e a função de outros leucócitos	Regulação de respostas inflamatórias	Estrutura em solução e modelos de interação da mesma com seu receptor

A atividade de muitas proteínas induz mudanças conformacionais associadas a movimentos de domínios em torno de uma dobradiça que envolve geralmente domínios rígidos que se deslocam ou giram em torno de uma dobradiça flexível (como no caso da PEPCK e PFK-2). Geralmente parte-se da estrutura cristalográfica já conhecida de dois estados conformacionais (PEPCK), um ligado e outro não, tenta-se encontrar as “dobradiças” desta proteína e propor modelos de funcionamento da mesma baseados nos eixos de rotação quando este se liga, por exemplo, ao ligante, e é submetida então, a um pH diferente ou é aquecida.

Quando só um estado conformacional é conhecido por cristalografia ou modelagem (PFK-2), tenta-se identificar as dobradiças comparando a estrutura com modelos de baixa resolução, obtidos por espalhamento de raios X de baixa resolução (SAXS), quando possível, ou com modelos teóricos. Em geral, a estrutura em nível atômico da proteína obtida por difração de raios X (DRX), é conhecida em apenas uma das condições de pH

e/ou temperatura e/ou aditivos. Os resultados da técnica de SAXS nos permitiram acompanhar as mudanças conformacionais resultantes das variações das condições do solvente.

A partir dos dados cristalográficos (obtidos pela técnica de difração de monocristais), pode-se determinar a curva de intensidade de espalhamento de baixo ângulo da proteína para as soluções em condições similares e compará-la com as intensidades experimentais. A estrutura das proteínas em outras condições (diferentes ou similares da cristalizada) corresponderá a modificações conformacionais a serem determinadas (PEPCK e PFK-2). Essa estrutura, com variações de forma, foi determinada deslocando e/ou rotando alguns domínios e subdomínios da estrutura inicial, até se obter um ajuste razoável entre as funções de intensidade de espalhamento $I(q)$ calculada e experimental.

Em casos de não se possuir a estrutura cristalográfica da proteína em estudo, pode-se tentar localizar domínios e eventuais mudanças que ocorram na mesma. Em outros casos propomos modelos de estrutura quaternária ou terciária quando somente alguns fragmentos da proteína eram conhecidos.

1.4 Objetivos e organização da tese.

Para um estudo detalhado da estrutura de proteínas em diferentes condições e meios são em geral utilizadas diversas técnicas complementares, notadamente XRD, SAXS, espalhamento de luz, fluorescência, ressonância magnética nuclear (NMR), dicroísmo circular (CD), etc.

Nesta tese de doutorado, a pesquisa é concentrada na aplicação da técnica de SAXS, a várias proteínas diferentes, colaborando com outros pesquisadores que estudam a mesma proteína com outras técnicas para assim chegar a uma conclusão clara sobre as estruturas das várias proteínas e ajudar a entender as respectivas funções das mesmas.

Decidimos estudar a estrutura de baixa resolução da várias proteínas como o receptor de ácido retinóico (RXR), fosfo-enol-piruvato carboxiquinase (PEPCK), isoenzima fosfofrutoquinase (PFK-2) e interleucina humana 22 (hIL-22) realizando experimentos de SAXS e colaborando com outros pesquisadores que utilizaram a técnica de XRD, entre outras.

A realização e a análise de dados das experiências de SAXS devem ser feitas de forma criteriosa para minimizar erros sistemáticos. Para isso, vários aspectos instrumentais relacionados com a análise de dados devem ser bem estudados e dominados. Alguns procedimentos associados precisaram ainda ser aprimorados para tornar a técnica de SAXS mais precisa e eficiente.

Com base no que foi exposto, o texto da tese foi organizado da seguinte forma:

Capítulo 1: Apresentam-se os aspectos relevantes da estrutura, conformação e atividade das proteínas;

Capítulo 2: Descrevem-se os aspectos básicos do espalhamento de raios X a baixos ângulos, tanto do ponto de vista experimental (instrumentos) quanto dos procedimentos de análise dos resultados (programas);

Capítulo 3: Apresentam-se dois tópicos originais que visam aumentar as possibilidades de uso e aprimorar conceitos estabelecidos associados à técnica de SAXS, notadamente (i) um procedimento para a determinação do estado oligomérico e peso molecular de proteínas em solução, a partir de determinação de intensidade de SAXS em escala relativa, e (ii) uma análise nova mostrando a dependência da densidade média de proteínas em solução com seu peso molecular;

Capítulo 4: São apresentados resultados de um estudo por SAXS da estrutura de baixa resolução dos receptores nucleares de ácido retinóico (RXR) em solução;

Capítulo 5: São apresentados resultados de estudos por SAXS das estruturas de baixa resolução de outras proteínas em solução: fosfo-enol-piruvato carboxiquinase (PEPCK), isoenzima fosfofrutoquinase (PFK-2) e interleucina humana 22 (hIL-22).

Capítulo 6: São mencionadas as conclusões gerais e citadas as referências dos artigos publicados, e em preparação, relacionados com a tese. Mencionam-se também as referências de publicações de outros estudos de proteínas em solução, realizados em paralelo, durante o desenvolvimento da tese.

¹ Lesk A M – Protein Architecture. *Oxford University Press*, New York, (1991).

² Gerstein M, Lesk AM and Chothia C – Structural mechanisms for domain movements in proteins. *Biochemistry*, **33**:6739 – 6749 (1994).

³ Wriggers W and Schulten K – Protein Domain Movements: Detection of Rigid Domains and Visualization of Hinges in Comparisons of Atomic Coordinates. *Proteins: Structure, Function, and Genetics*, **29**: 1-14 (1997).

⁴ Garrett RH and Grisham CM – Biochemistry. *Holt Rinehart & Winston*, 2nd edition (1999).

2	A técnica de SAXS	2 - 1
2.1	SAXS como ferramenta para elucidação do envelope molecular de macromoléculas em solução	1
2.1.1	Espalhamento de raios X a baixos ângulos por sistemas coloidais	4
2.1.2	Informação contida em uma curva de SAXS	13
2.2	Cálculo de intensidade de SAXS a partir das coordenadas atômicas	13
2.3	Modelagem de configurações espaciais de domínios de proteínas utilizando dados de cristalografia de proteínas	16
2.4	Reconstrução <i>ab-initio</i> da forma da proteína a partir de resultados de SAXS	16
2.4.1	DAMMIN (Dummy atom modelling)	16
2.4.2	Funções de envelope	19
2.5	Arranjo experimental	20
2.6	Condições experimentais	21
2.6.1	Efeitos instrumentais ("Smearing")	22
2.6.2	Efeito da espessura da amostra na intensidade de espalhamento	23

2 A TÉCNICA DE SAXS

2.1 SAXS como ferramenta para elucidação do envelope molecular de macromoléculas em solução

SAXS é uma técnica que nos fornece informações estruturais de flutuações ou heterogeneidades da densidade eletrônica numa matrizⁱ com dimensões características da ordem de 10 a 1000Å. Numerosos tipos de materiais apresentam heterogeneidades nesta escala: ligas metálicas, copolímeros, colóides, micelas, micro-emulsões, materiais

ⁱ Quando uma matriz como, por exemplo água, contém dissolvido nela partículas que tem uma densidade menor ou maior que a água, são estas variações que são possíveis de detectar por SAXS, ou seja, é necessário que haja um contraste de densidade eletrônica compatível com o comprimento de onda utilizado e os ângulos de espalhamento utilizados.

nanoporosos, cristais líquidos e soluções de macromoléculas biológicas (proteínas, ribossomo, DNA, vírus) [1, 2,3].

Como a técnica de SAXS é sensível a heterogeneidades em escala maior que a atômica, em muitos casos as macromoléculas espalhadoras podem ser descritas como tendo uma densidade eletrônica constante e única, sem considerar as variações da densidade eletrônica em escala atômica. No caso de nucleoproteínas, lipoproteínas, micelas, etc. não é possível considerar o sistema como tendo somente uma densidade eletrônica média, pois neste caso a estrutura não é internamente homogênea. Neste caso deve-se descrever a macromolécula por vários níveis de densidade eletrônica diferentes.

As proteínas podem estar muito diluídas na solução (solução ideal) ou interagindo, possuindo assim uma certa correlação espacial que influencia a intensidade espalhada. A forma das curvas de SAXS contém informações sobre o tamanho e a forma de partículas e também sobre a correlação espacial das mesmas. Como trabalharemos geralmente com sistemas de macromoléculas em soluções diluídas, não entraremos em muito detalhe quanto a interferências das ondas espalhadas causadas pela correlação espacial.

Com a técnica de SAXS podem-se obter modelos de baixa resolução de partículas (desde que monodispersas). Nesta tese focaremos partículas espalhadoras biológicas, mais precisamente, macromoléculas em solução em estado diluído (com correlação espacial desprezível). Neste campo é possível em muitos casos responder questões funcionais fundamentais. E o mais importante é que o método permite estudar partículas em sua forma nativa em solução, ou seja, em condições bem próximas às fisiológicas. Assim pode-se estudar mudanças estruturais em respostas a estímulos externos fornecendo informações preciosas na relação existente entre mudanças estruturais e o funcionamento destes sistemas.

Uma comparação entre técnicas utilizadas no estudo de macromoléculas com SAXS demonstra que apesar da baixa resolução inerente da técnica ($\sim 10 - 20 \text{ \AA}$), o método é em geral uma ferramenta muito útil para se estudar a estrutura, cinética de transformações estruturais e interações de macromoléculas biológicas em solução.

Tendências atuais apontam para uma fase de estudo integrado de sistemas biológicos, prometendo entender sistemas de diferentes níveis de organização biológica. Neste aspecto, SAXS e a microscopia eletrônica proverão o esqueleto estrutural para a análise de estruturas complexas no qual modelos de alta resolução de cristalografia e NMR podem ser ajustados.

As pesquisas de macromoléculas biológicas em solução por SAXS fornecem informações sobre a forma externa das mesmas. Esta técnica pode ser aplicada para estudar moléculas e arranjos de moléculas (oligômeros) de dimensões compreendidas entre 10-1000Å.

A intensidade de espalhamento produzida por um sistema monodisperso diluído de macromoléculas em solução é proporcional a amplitude de espalhamento da macromolécula promediado para todas as orientações:

$$I(q) \propto \langle A^2(\vec{q}) \rangle_{\Omega}$$

onde $|\vec{q}| = \frac{4\pi}{\lambda} \sin\theta$ é o módulo do vetor de espalhamento, λ é o comprimento de onda e 2θ é o ângulo de espalhamento. A partir do conhecimento de $I(q)$ podem ser calculados parâmetros estruturais tais como o raio de giro da macromolécula (R_g), volume (V), peso molecular (MW) e a função de distribuição de distâncias ou $p(r)$ das partículas espalhadoras. A $p(r)$ nos fornece uma idéia de frequência com que cordas de comprimento r encontram-se dentro de uma partícula e é muito sensível a assimetria geral e estrutura dos domínios dentro da partícula (veja Figura 4).

A grande vantagem da técnica de SAXS é que com ela pode-se estudar tanto proteínas muito pequenas ($< 10\text{kDa}$) a complexos de proteínas tão grandes ou maiores que vírus e ribossomos ($> \text{dezenas de MDa}$). Esta faixa não é possível de varrer com NMR ou microscopia eletrônica criogênica. A tendência indica que SAXS será uma técnica ainda muito utilizada na próxima década.

Tabela 1 – Vantagens e limitações das principais técnicas utilizadas no estudo de macromoléculas em solução [2]

Método	Amostras	Vantagens	Desvantagens
Cristalografia de proteínas (PCr)	Cristais	- Altíssima resolução (até 1 Å) revelando detalhes finos da estrutura atômica	- Precisa-se de cristais - Partes flexíveis podem não ser vistas. - Estrutura pode ser afetada por forças de empacotamento cristalino.
NMR	Soluções concentradas (~ 5 -10mg/mL)	- Alta resolução (2-3 Å) em solução	- Dificilmente possível de se estudar macromoléculas com MM maior que 50 kDa
Crio-ME	Soluções congeladas diluídas (< 1mg/mL)	- Pouquíssima quantidade de amostra - Visualização direta da forma e simetria da partícula	- Dificilmente possível de se estudar macromoléculas com MM menores que 200 kDa
SAXS/SANS	Soluções concentradas (~1 a ~100mg/mL)	- Análise da estrutura, cinéticas e interações em condições próximas as fisiológicas. - Estudo de misturas e sistemas fora do equilíbrio.	- Baixa resolução (~10-20 Å) - Necessita às vezes de informações adicionais para resolver ambiguidades na construção de modelos.

		- Ampla faixa de pesos moleculares (poucos kDa a centenas de MDa)	
Espalhamento de luz estático, dinâmico e ultracentrifugação	Soluções diluídas (< 1mg/mL)	- Não destrutivo - Pouquíssima quantidade de amostra - Medida relativamente fácil	- Baixa resolução (> 20 Å)

2.1.1 Espalhamento de raios X a baixos ângulos por sistemas coloidais

As ondas eletromagnéticas associadas aos raios X incidentes provocam oscilações dos elétrons dos átomos componentes da amostra irradiada. Quando a energia dos fótons incidentes é bem maior que a energia de ligação dos elétrons, estes se comportam como se fossem livres. As partículas livres carregadas oscilantes produzem ondas espalhadas coerentes com a onda incidente em todas as direções com intensidade definida pela equação de Thompson. Para um feixe de intensidade I_0 , na direção definida no plano de polarização do feixe temos:

$$I_{par} = I_0 r_e^2 \cos^2 2\theta$$

onde 2θ é o ângulo entre o feixe incidente e o feixe espalhado, r_e é o raio clássico do elétron ($2,82 \times 10^{-13}$ cm). Em direção definida no plano perpendicular ao de polarização do feixe temos:

$$I_{per} = I_0 r_e^2.$$

O espalhamento dos raios X pelos elétrons é muito maior que pelos prótons porque sua massa é muito menor. Por isso somente o espalhamento pelos elétrons é considerado.

Além do espalhamento dado pela equação de Thompson que é elástico, isto é, espalhamento sem mudança no comprimento de onda incidente, ou melhor, sem que haja troca de energia entre o fóton incidente e o elétron, existe também o espalhamento inelástico ou Compton. Para ângulos pequenos (que é o caso de SAXS) o espalhamento inelástico é pequeno e pode ser desprezado.

Seja $\rho(\vec{r})$ a densidade eletrônica num elemento de volume dV_r da amostra, $\vec{s}_0$ um vetor de módulo $1/\lambda$ (λ é comprimento de onda dos raios X) com a direção do feixe incidente e $\vec{s}$ o correspondente ao feixe espalhado (Figura 1). Se definirmos o vetor $\vec{q}$ como indicado na Figura 1, por $\vec{q} = 2\pi(\vec{s} - \vec{s}_0)$ o seu módulo resulta:

$$q = |\vec{q}| = \frac{4\pi \sin \theta}{\lambda}$$

Para ângulos pequenos podemos aproximar: $q \approx \frac{2\pi(2\theta)}{\lambda}$

Geralmente para experiências de SAXS em solução, 2θ é $\leq 10^\circ$ (175 mrad) para $\lambda \approx 1.5 \text{ \AA}$.

A diferença de fase entre uma onda espalhada por um elétron na origem e outro num elemento de volume $dV_{\vec{r}}$ localizado em $\vec{r}_{\vec{r}}$ é dada por $\vec{q} \cdot \vec{r}_{\vec{r}}$ (ver Figura 1).

Se observamos o espalhamento da amostra a uma distância muito maior que o tamanho da mesma a amplitude da radiação espalhada $dA(\vec{q})$ por cada elemento de volume $dV_{\vec{r}}$ é igual a:

$$dA(\vec{q}) = \rho(\vec{r}) e^{-i\vec{q} \cdot \vec{r}} dV_{\vec{r}}$$

onde $\rho(\vec{r})dV_{\vec{r}}$ é o número de elétrons no elemento de volume $dV_{\vec{r}}$ e $e^{-i\vec{q} \cdot \vec{r}}$ leva em conta a diferença de fase.

Assim, o espalhamento total pela amostra será:

$$A(\vec{q}) = \int_{V_{\vec{r}}} \rho(\vec{r}) e^{-i\vec{q} \cdot \vec{r}} dV_{\vec{r}}$$

Equação 2-1

Figura 1 – Representação vetorial do espalhamento por uma partícula de volume V e densidade eletrônica $\rho(\vec{r})$.

Note que para $q=0$,

$$A(0) = \int_{V_F} \rho(\vec{r}) dV_F = m$$

onde m é o número de elétrons no volume V_F . A intensidade espalhada é dada por

$$I(\vec{q}) = A(\vec{q}) \cdot A^*(\vec{q}) = \int_{V_F} \int_{V_F''} \rho(\vec{r}') \rho(\vec{r}'') e^{-i(\vec{r}' - \vec{r}'') \cdot \vec{q}} dV_F' dV_F'' \quad \text{Equação 2-2}$$

onde $A^*(\vec{q})$ é o complexo conjugado de $A(\vec{q})$. Seja $V_F \gamma(\vec{r})$ a função de autocorrelação da densidade eletrônica $\rho(\vec{r})$ da partícula, então:

$$\gamma(\vec{r}) = \frac{1}{V_F} \int_{V_F'} \rho(\vec{r}') \rho(\vec{r}' + \vec{r}) dV_F' \quad \text{Equação 2-3}$$

No caso de proteínas em solução, o espalhamento se origina da diferença de contraste eletrônico entre a proteína e o solvente (que é constante e definido como sendo ρ_0). Redefinindo, temos que:

$$\Delta\rho(\vec{r}) = \rho(\vec{r}) - \rho_0 \text{ e } A(0) = \int_{V_F} \Delta\rho(\vec{r}) dV_F = m$$

Da equação 2 e 3 temos que:

$$I(\vec{q}) = \int_{V_F} \gamma(\vec{r}) e^{-i\vec{r} \cdot \vec{q}} dV_F \text{ onde } \vec{r} = \vec{r}' - \vec{r}'' \quad \text{Equação 2-4}$$

Nas fórmulas acima supomos uma amostra com densidade eletrônica arbitrária definida por $\rho(\vec{r})$. Vamos agora focalizar o caso de proteínas de densidade eletrônica ρ_1 imersas em uma solução de densidade constante ρ_0 como a esquematizada na Figura 2. Neste caso a função de auto correlação pode ser simplesmente escrita como:

$$\gamma(\vec{r}) = \gamma_0(\vec{r})(\rho_1 - \rho_0)^2$$

onde $\gamma_0(\vec{r})$ é a função característica de Porod e é definida como

$$\gamma_0(\bar{r}) = \frac{\overline{V(\bar{r})}}{V}$$

onde $\overline{V(\bar{r})}$ é o volume em comum da partícula deslocada por $\bar{r}$ promediado para todas as direções de $\bar{r}$. A promediação orientacional de $\gamma_0(\bar{r})$ dado por $\gamma_0(r) = \langle \gamma_0(\bar{r}) \rangle_{\Omega}$ está relacionada com a função de distribuição de distâncias $p(r)$ por:

$$p(r) = 4\pi \cdot r^2 \gamma_0(r)$$

onde $\gamma_0(r)$ representa a probabilidade de encontrar elemento de densidade eletrônica a uma distância r de um elemento (Figura 3a). Se integrarmos essa probabilidade sobre a superfície de uma esfera de raio r e sobre o volume V temos a função de distribuição de distâncias como definida na equação acima. Uma boa análise da $p(r)$ pode-nos fornecer uma idéia do tamanho e a forma da proteína.

Na Figura 3a à direita está graficado a função $\gamma_0(r)$ correspondente a proteína representada à esquerda. Na Figura 3b a direita está representada a $p(r)$ para a mesma proteína. A partir de $r = D_{max}$ a função $p(r)$ vale zero, ou seja D_{max} é o máximo diâmetro da molécula.

Figura 2 – Exemplo de contraste no caso de partículas em solução ($\Delta\rho$).

No caso de proteínas dissolvidas numa solução próxima às condições fisiológicas (soluções predominantemente compostas por água, contendo diferentes sais ou açúcares em pequenas quantidades), a densidade eletrônica média do solvente é de $\rho_0 = 0,335 \text{ e}^-/\text{\AA}^3$ (água). A densidade média das proteínas é considerada aproximadamente igual a $\rho_1 = 0,43$

$\text{e}^-/\text{\AA}^3$ [4]. Veremos no capítulo 3 que esse valor é maior para proteínas de peso molecular inferior a 20kDa.

As macromoléculas biológicas são essencialmente compostas por átomos leves (H, C, N, O, S, no caso de proteínas e, além disso, P no caso de ácidos nucleicos) e como elas estão em soluções predominantemente de água há pouca diferença na densidade eletrônica (ou no contraste) entre a molécula e o solvente. Consequentemente, estes tipos de soluções espalham muito pouca radiação. Por exemplo, uma solução de 1mg/mL de uma proteína de 80kDa espalha da ordem de 4 fótons para cada 100.000 incidentes, o que torna muito atrativo o uso de feixes de raios X produzidos por fontes síncrotron que possuem um fluxo de fótons muito maior que aparelhos de raios X convencionais [4].

Figura 3: (a) Função de autocorrelação e correspondente média espacial (direita); (b) função de distribuição de distâncias.

No caso de espalhamento de partículas em solução a amostra é isotrópica ou seja, as partículas estão em movimento browniano e podem assumir todas orientações possíveis em

relação ao feixe incidente. Assim, temos acesso experimental somente a média esférica para todas orientações da macromolécula, ou seja:

$$I(q) = \left\langle \int_{V_F} V_F \gamma(\vec{r}) e^{-i\vec{r}\vec{q}} dV_F \right\rangle_{\Omega}$$

onde $\langle \rangle_{\Omega}$ significa a média esférica (média orientacional). Após promediar todas possíveis orientações com relação ao feixe incidente, as expressões 3 e 4 se reduzem a:

$$I(q) = 4\pi \int_0^{\infty} p(r) \frac{\sin rq}{rq} dr \quad \text{Equação 2-5a}$$

Vejamos a ilustração da Figura 4 onde está representada a função $p(r)$ para vários objetos, uma esfera de raio 25Å, duas esferas juntas de mesmo raio (25Å cada) e um cilindro de raio 20Å e comprimento de 120Å. Desta forma se a proteína for esférica, a $p(r)$ da mesma deveria ser parecida com a da esfera representada na Figura 4, e se numa solução existem proteínas diméricas (sendo os monômeros aproximadamente esféricos), deveríamos encontrar uma $p(r)$ similar a das duas esferas juntas ilustradas na Figura 4.

Figura 4 – Ilustração da função $p(r)$ para diferentes objetos espalhadores: esfera simples ou proteína esférica com raio de 25Å, duas esferas ou um dímero de proteína composta por monômeros esféricos onde cada esfera tem um raio de 25Å e um cilindro de raio 20Å e comprimento de 120Å.

Um parâmetro estrutural interessante utilizado na análise de mudanças conformacionais de macromoléculas em solução é R_g definido da seguinte maneira

$$R_g^2 = \frac{\int r^2 dv}{v}$$

O raio de giro da macromolécula (R_g) é semelhante a um “raio de inércia” da mecânica clássica sendo a raiz quadrada da distância quadrática média r entre os elementos de volume e o centro de massa eletrônico do objeto espalhador. Pode ser demonstrado [1, 3,5] que o raio de giro pode ser calculado a partir do função $\gamma_0(r)$ da seguinte forma:

$$R_g^2 = \frac{\int_0^D \gamma_0(r) r^4 dr}{2 \int_0^D \gamma_0(r) r^2 dr}$$

Como podemos ver, no caso de um sistema composto por partículas espalhadoras de diversos tamanhos, as maiores terão um peso maior e contribuirão mais para um aumento no raio de giro.

O valor de R_g pode alternativamente ser obtido a partir da porção da curva de espalhamento a muito baixos q 's para a qual é válida a aproximação de Guinier [5]:

$$I(q) = N(\Delta\rho)^2 v^2 e^{-\frac{1}{3}q^2 R_g^2}, q \rightarrow 0 \quad \text{Equação 2-6}$$

A equação de Guinier como expressada na Equação 2-6 é válida para sistemas de partículas idênticas e diluídas na região de $q \rightarrow 0$ (Figura 5). Representando a intensidade em escala $\ln I$ versus q^2 , espera-se um comportamento linear com uma inclinação igual a $-\frac{1}{3}R_g^2$ (Figura 5).

Figura 5 – Equação de Guinier é válida para sistemas de partículas idênticas e diluídas na região de $q \rightarrow 0$. Gráfico acima em escala logarítmo natural versus q^2 .

O conhecimento do raio de giro pode ser muito útil na investigação estrutural de uma partícula monodispersa em solução, pois um aumento ou diminuição do mesmo provocado por qualquer variação nas condições da solução indicaria uma mudança nas posições relativas entre os átomos que compõem a partícula.

A função $p(r)$ contém mais informação que o raio de giro. Porém, em soluções de proteínas muito diluídas (menos de $\sim 1\text{mg/mL}$), tempos longos de exposição para diminuição dos erros estatísticos são necessários, o que nem sempre é possível. Neste caso (soluções super diluídas e tempos de exposição curtos) não é possível coletar curvas de intensidade estatisticamente significantes numa ampla faixa de q para se poder determinar a $p(r)$. Assim a curva experimental é estatisticamente confiável somente a pequenos ângulos ($q \rightarrow 0$) onde somente o R_g obtido a partir da equação de Guinier pode ser determinado.

Muitas vezes interessa o estudo de variações rápidas de configuração molecular e são necessárias medidas de intensidade de espalhamento em tempos curtos, por exemplo 1 segundo ou menos. Nestes casos (estudos resolvidos no tempo) apenas na região de baixos q a intensidade é suficiente (estatisticamente significativa). Pode-se assim determinar a evolução temporal do R_g a partir da equação de Guinier o que é as vezes suficiente para detectar e caracterizar variações conformacionais rápidasⁱⁱ.

Para ilustrar a relação entre a função $p(r)$ e a forma da partícula espalhadora, vejamos a Figura 6 onde foram selecionadas 10 macromoléculas diferentes, umas esferoidais como a F e a I, outras com domínios afastados uns dos outros como A e J, outras prolatas como C, D, E e H e uma oblata (B).

ⁱⁱ Em laboratórios de fonte síncrotron de terceira geração, foram realizadas medidas com intervalos de tempo menores de $1\mu\text{s}$.

Figura 6 – Diversas macromoléculas e suas respectivas funções de distribuição de distâncias.

2.1.2 Informação contida em uma curva de SAXS

Uma das grandes dificuldades em problemas inversosⁱⁱⁱ é obter informações, por exemplo, do envelope de uma proteína a partir de uma curva de SAXS. Depara-se frequentemente na seguinte questão: será que dois envelopes completamente diferentes podem ter curvas de espalhamento semelhantes dentro dos erros experimentais? No caso geral não existe solução única pois é impossível determinar analiticamente um envelope tridimensional a partir de uma curva de espalhamento unidimensional. Como veremos, uma curva de SAXS contém informação de baixa resolução ($> 10\text{\AA}$) e assim podemos extrair dela somente alguns poucos parâmetros estruturais.

Para termos uma idéia do que significam “poucos parâmetros” analizaremos o teorema de Shannon [6,7] sobre o conteúdo de informação numa função tipo transformada de Fourier. A intensidade espalhada pela proteína $I(q)$ é uma função analítica que é completamente definida por seus valores nos pontos de amostragem (canais de Shannon) $I(q_k)$, $q_k = k\pi / D_{\max}$, $k = 0, 1, \dots, \infty$, onde $D_{\max}$ é a dimensão máxima da partícula. Assim, se temos uma curva de espalhamento no intervalo $[q_{\min}, q_{\max}]$, o número de parâmetros estruturais que podem ser extraídos é aproximadamente igual ao número de canais de Shannon $N_s = (q_{\max} - q_{\min}) D_{\max} / \pi$. Ao mesmo tempo, a informação completa é contida num intervalo finito e contínuo da curva. Se medirmos uma curva de espalhamento com incrementos muito menores que a distância de amostragem $\pi / D_{\max}$ podemos em princípio extrapolar além do intervalo experimental nos fornecendo então mais que N_s parâmetros independentes. A qualidade desta extrapolação analítica (e por conseguinte desta informação adicional) depende dos erros experimentais que variam desde extrapolações excelentes em dados sem erro até extrapolações completamente erradas quando o erro é muito grande.

Assim podemos garantir a extração de aproximadamente $N_i \equiv N_s$ parâmetros independentes que geralmente são em torno de 10-30. À baixa resolução, a forma de uma partícula é convenientemente parametrizada usando a função de envelope introduzida por Stuhrmann[8] que utilizou expansão de multipolos.

2.2 Cálculo de intensidade de SAXS a partir das coordenadas atômicas

ⁱⁱⁱ Coletamos informações no espaço recíproco e gostaríamos de obter informações no espaço real, porém informações importantes como a fase não podem ser obtidas.

Comparações entre curvas experimentais de SAXS e curvas teóricas avaliadas a partir de estruturas com resolução atômica, obtidas por XRD, têm sido amplamente utilizadas para validar modelos teóricos, verificar diferenças estruturais entre proteínas e ácidos nucleicos no estado cristalino e em solução e prever estruturas quaternárias [9,10,11,12,13,14,15,16].

Figura 7 – Representação esquemática de uma proteína em solução.

Proteínas em solução podem ser esquematicamente representadas como ilustrado na Figura 7. A partícula com densidade $\rho_a(\vec{r})$ é cercada por um solvente de densidade eletrônica média ρ_o . A hidratação da proteína é representada por uma camada de espessura Δ e densidade ρ_b que pode diferir de ρ_o . A curva de SAXS de uma proteína em uma solução suficientemente diluída é proporcional ao espalhamento de um único objeto espalhador. No caso específico de proteínas, consideramos um modelo de 3 densidades eletrônicas como mostrado na Figura 7. Pode ser demonstrado [1, 17] que para este modelo a intensidade de espalhamento se escreve como:

$$I(q) = \left\langle \left| \rho_a A_a(\vec{q}) - \rho_o A_c(\vec{q}) + \delta\rho A_b(\vec{q}) \right|^2 \right\rangle_{\Omega} \quad \text{Equação 2-7}$$

onde $A_a(\vec{q})$ é a amplitude de espalhamento da proteína no vácuo, $A_c(\vec{q})$ é a amplitude espalhada pelo volume de solvente deslocado pela proteína (volume excluído), $\delta\rho A_b(\vec{q})$ é a amplitude espalhada pela camada de hidratação, $\delta\rho = \rho_b - \rho_o$ e $\langle \rangle_{\Omega}$ significa a média para todas as possíveis orientações (promediação orientacional) da proteína.

Utilizamos um programa chamado CRY SOL [17] para a determinação de curvas de SAXS para proteínas com estrutura atômica conhecida. O programa usa expansão de multipólos [18] e é utilizado para um rápido cálculo da promediação orientacional (Equação 2-7). Dadas as coordenadas atômicas (no formato do Brookhaven Protein Data Bank) o

programa gera a curva de SAXS da proteína, que pode ser comparada com resultados experimentais.

Quando a estrutura cristalográfica da proteína é conhecida, a amplitude $A_a(\vec{q})$ pode ser determinada. O único parâmetro desconhecido que precisa ser ajustado é o volume excluído da proteína, V , e a diferença entre a densidade eletrônica entre o solvente e a camada de hidratação, $\delta\rho$. Esses parâmetros e um fator de escala c , são obtidos num procedimento de ajuste utilizando o CRY SOL. Assim a intensidade de SAXS modelada $I(q)$, dada pela Equação 2-7, é ajustada aos dados experimentais $I_{exp}(q)$ minimizando o funcional:

$$\chi^2(\delta\rho) = \frac{1}{N-1} \sum_{i=1}^N \left[\frac{I_{exp}(q_i) - cI(q_i, \delta\rho)}{\sigma(q_i)} \right]^2$$

onde N é o número de pontos experimentais e $\sigma(q_i)$ são os erros de $I_{exp}(q)$.

Este programa tem sido utilizado amplamente no laboratório em que foi desenvolvido (EMBL, Hamburgo, Alemanha) assim como em outros laboratórios (LURE, França; HASYLAB, Alemanha e Stanford, EUA).

Um exemplo de cálculo da curva de espalhamento a partir das coordenadas atômicas para diversas proteínas obtidas do Brookhaven Protein Data Bank (PDB) é representado na Figura 8.

Figura 8 – Função $I(q)$ obtidas a partir das coordenadas atômicas usando o programa CRY SOL

2.3 Modelagem de configurações espaciais de domínios de proteínas utilizando dados de cristalografia de proteínas

Com a técnica de SAXS é possível caracterizar variações conformacionais associadas por exemplo, com rotações e translações de domínios de macromoléculas em solução. Quando dados de difração de raios X a alto ângulo de proteínas cristalinas, obtidos em certas condições, são combinados com dados de SAXS de proteínas em solução em diferentes condições (pH, ligantes, temperatura, etc...), informações mais precisas sobre eventuais mudanças conformacionais podem ser estudadas.

Utilizamos o programa CRY SOL [17] bem como outro chamado *Modelling of Atomic Structures and SHape Analysis* (MASSHA) [19], um programa interativo em 3D para visualizar e manipular estruturas atômicas e modelos de baixa resolução.

MASSHA está acoplado a diversos módulos computacionais que permitem realizar um refinamento de corpo rígido de uma forma interativa e automática entre os modelos e a curva experimental. Porém, nem sempre é possível fazer modelagens deste tipo, pois muitas vezes não se dispõe dos dados cristalográficos de alta resolução da proteína que se pretende estudar. Nestes casos parte-se para uma análise mais cuidadosa, utilizando a função $p(r)$ para inferir alguns resultados ou outros programas que reconstróem a forma da partícula espalhadora a partir da curva experimental como o DAMMIN[20], GASBOR[21].

2.4 Reconstrução *ab-initio* da forma da proteína a partir de resultados de SAXS

Como resultado da promediação orientacional uma considerável quantidade de informação é perdida num experimento de espalhamento de raios X em soluções. Ainda assim é possível extrair informações tridimensionais da partícula a partir da curva de SAXS unidimensional [20,21,22,23,24,25,28,33]. A seguir descreveremos dois procedimentos *ab-initio* utilizados nesta tese.

2.4.1 DAMMIN (*Dummy atom modelling*)

Nos casos em que não se dispõe da estrutura com resolução atômica, utilizamos o programa DAMMIN[20]. Este programa permite a determinação *ab initio* da estrutura de baixa resolução (que define a forma e volume da molécula) de uma macromolécula biológica em solução (ou qualquer outro espalhador, desde que todos sejam iguais, ou seja, um sistema monodisperso, e tenham uma densidade eletrônica constante). Para isto é

necessário uma boa curva de SAXS e utilizando um algoritmo matemático (conhecido por *simulated annealing*) encontra-se uma solução que ajuste os dados experimentais e ao mesmo tempo minimize a área interfacial (da partícula com o solvente).

Este método funciona muito bem para proteínas com peso acima de 30 KDa onde domina o espalhamento da forma à estrutura interna da proteína (visto que neste caso a densidade eletrônica da partícula é aproximadamente constante).

O programa DAMMIN funciona da seguinte maneira: Parte-se geralmente de uma esfera de diâmetro igual ao diâmetro máximo (D_{max}) da macromolécula. Esse diâmetro é determinado a partir da função $p(r)$. A seguir preenche-se essa esfera de raio $R=D_{max}/2$ com esferas menores de raio r_0 . Assim o número de esferas será $N \cong (R/r_0)^3$. Parte-se então de uma proteína modelo esférica (de raio R) composto por N esferas menores (ou "pseudopontos") de raio r_0 . Claro que outros modelos iniciais também podem ser utilizados (cilindros, prismas, etc.), porém neste caso é necessário ter-se uma melhor idéia do sistema em estudo.

Em seguida começa o processo de minimização de discrepâncias entre a curva de espalhamento gerado a partir do modelo e a curva experimental. Devido ao grande número de variáveis e à complexa natureza do problema foi utilizado um algoritmo de minimização chamado *simulated annealing*[26]. Pretende-se encontrar no fim do processo de minimização (da diferença global entre a curva experimental e a calculada a partir do modelo) uma forma o mais parecida possível da macromolécula em solução.

A idéia principal deste método é de realizar modificações aleatórias no espaço configuracional das N pseudoesferas encontrando sempre uma configuração de esferas que minimize a discrepância entre a curva de SAXS gerada do sistema de esferas e da curva de SAXS real. A discrepância também pode aumentar (mas com uma probabilidade menor), isto se deve ao método de minimização empregado que com esta abordagem diminui os riscos do programa encontrar uma solução intermediária que também se ajuste bem aos dados experimentais (a idéia toda é encontrar a melhor solução possível, tipo um mínimo global num processo de otimização de N variáveis). Quanto mais iterações, menos as pseudoesferas se movimentam e finalmente obtêm-se uma solução que ajuste de forma excelente os dados experimentais. Claro que esta solução não é única, pois outros modelos com ajustes parecidos levam a configurações de pseudoesferas parecidas. Neste método, é interessante se gerar pelo menos 20 modelos e depois fazer uma média entre os mais parecidos.

Para ilustrar como o método de minimização funciona, vamos tentar explicar passo a passo como ele funciona. Para isto definimos $f(X) = \chi^2 + \alpha P(X)$ como sendo a função de minimização e $\alpha > 0$ é o peso da função de compactação $P(X)$ responsável pela manutenção compacta da proteína. χ^2 é definido como:

$$\chi^2 = \frac{1}{n-1} \sum_{j=1}^n \left[\frac{c_{DAM} I_{DAM}(q_j) - c I_{exp}(q_j)}{\sigma(q_j)} \right]^2$$

onde $I_{exp}(q_j)$ é a intensidade especificada nos n pontos q_j e σ_j seu correspondente desvio experimental. c_{DAM} e c são coeficientes de escala.

O algoritmo implementado no programa DAMMIN executa então os seguintes passos:

- (i) Inicia-se a partir de uma configuração randômica (X_0) das pseudoesferas com uma grande liberdade de permuta e configurações aleatórias definida por T_0 , por exemplo: $T_0 = f(X_0)$ quando então:
- (ii) Seleciona-se uma pseudoesfera randomicamente e de uma forma aleatória muda-se a sua fase para solvente ou partícula – ou seja obtendo uma nova configuração X' – e calcula-se a diferença entre a curva teórica gerada nesta nova configuração em relação a antiga com a experimental $\Delta = f(X') - f(X)$,
- (iii) se $\Delta < 0$, a pseudoesfera é movido para X' ; se $\Delta > 0$ a pseudoesfera é movido com uma probabilidade $\exp(-\Delta/T)$. Depois o passo (ii) se repete para X' (se for aceito) ou X .
- (iv) A taxa de permuta das pseudoesferas é mantida constante para $100N$ reconfigurações ou $10N$ reconfigurações bem sucedidas. (onde N é o número total de esferas que compõe a esfera maior), após isto, diminui-se os graus de liberdade com o qual as pseudoesferas podem permutar ($T' = 0.9T$) que continua até que nenhuma melhora na $f(X)$ é observada.

Como a técnica de SAXS está limitada a uma resolução de no máximo $10\text{\AA}$, a resolução do envelope da proteína restaurada utilizando o método *simulated annealing* limita-se em torno deste valor ($10\text{\AA}$) quando se consegue curvas de espalhamento de alta qualidade, o que nem sempre é possível devido a problemas técnicos de ordem experimental ou estabilidade da proteína. Contornando as dificuldades consegue-se facilmente uma resolução de $20\text{-}30\text{\AA}$ que é suficiente para identificar inequivocamente em muitos casos a forma aproximada da proteína bem como o volume ocupado por subdomínios e grupos atômicos da proteína.

Versões melhoradas de DAMMIN foram propostas por Svergun e colaboradores [27]. Outros procedimentos *ab initio* para determinação da forma de macromoléculas em solução têm sido propostos por Vigil [24], Walther [22] e Chacon [28,29] que usa um algoritmo genético para produzir modelos descritos por esferas densamente empacotadas.

2.4.2 Funções de envelope

Para proteínas com alto peso molecular (> 50 kDa) a determinação do envelope molecular pode ser feita com o auxílio da teoria de multipólos [8, 8,18]. Este foi um dos primeiros e mais elegantes métodos analíticos desenvolvidos para reconstruir a forma de uma partícula. Neste método a densidade eletrônica da proteína é tida como uniforme e o envelope molecular é descrito apenas pela função $F(\omega)$ (Figura 7).

$$\rho(\vec{r}) = \begin{cases} 1 & 0 \leq r < F(\omega) \\ 0 & r \geq F(\omega) \end{cases}$$

Esta função relaciona-se com a intensidade ($I(s)$) da seguinte maneira:

$$I(q) = \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(q)|^2$$

onde,

$$A_{lm}(q) = i^l (2/\pi)^{1/2} \sum_{p=0}^{\infty} \left[\frac{(-1)^p}{2^p p! [2(l+p+1)]!!} (qr)^{l+2p} f_{lm}^{(l+2p+3)} / (l+2p+3) \right] q^{l+2p},$$

$$f_{lm}^q = \int [F(\omega)]^q Y_{lm}^*(\omega) d\omega,$$

$$F(\omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{lm}^0 Y_{lm}(\omega),$$

$Y_{lm}^*(\omega)$ são os esféricos harmônicos, L é a máxima resolução e $(r, \omega) = (r, \theta, \varphi)$ são as coordenadas esféricas. L define o número de parâmetros independentes N_p para o qual os envelopes de baixa resolução são comparáveis com o número de canais de Shannon. No caso geral $N_p = (L+1)^2 - 6$, ou seja, são necessários 10-20 parâmetros para $L=3-4$, e este número se reduz ainda mais caso a partícula seja simétrica.

Portanto a intensidade de espalhamento $I(q)$ fica expressa em função dos coeficientes f_{lm} os quais por sua vez descrevem o envelope molecular $F(\omega)$. Esses coeficientes podem ser encontrados minimizando-se o seguinte funcional

$$\Phi = \Phi(f_m) = \frac{\int_{q_{\min}}^{q_{\max}} [I_{\text{exp}}(q) - I(q, f_m)] W(q) dq}{\int_{q_{\min}}^{q_{\max}} I_{\text{exp}}(q) W(q) dq}$$

onde $I_{\text{exp}}(q)$ é a intensidade experimental e $W(s)$ é uma função peso.

Esta técnica permite obter uma visão global da proteína e, em casos favoráveis, observar mudanças estruturais provocadas por mudança nas condições de solvatação. Além disso, a função $F(\omega)$ pode servir como base para auxiliar na determinação do envelope molecular em cristais.

Nos dois modelos *ab-initio* descritos acima (DAMMIN e SASHA), assume-se que a proteína possa ser bem descrita por uma densidade eletrônica constante em todo seu volume. As flutuações de densidade eletrônica existentes na proteína acabam contribuindo mais em altos valores de q e para se eliminar estas contribuições uma constante da curva de espalhamento experimental é subtraída. Para baixos valores de q essa contribuição é constante e pode ser ignorada, o que acontece aqui, por isso não faz sentido ajustar curvas experimentais para altos q 's quando se assume um modelo de 2 densidades eletrônicas somente. Para considerar uma curva de SAXS completa, precisa-se levar em conta essas flutuações de densidade eletrônica.

2.5 Arranjo experimental

Em um experimento típico de SAXS a amostra é atravessada por um feixe de raios X monocromático. O feixe transmitido é absorvido por um anteparo estreito e fino (beam-stop semitransparente) e o espalhamento a ângulos baixos (até 10 graus) é registrado por um detector. Um esquema do arranjo experimental típico associado a uma fonte de radiação síncrotron está ilustrado na Figura 9.

Figura 9 – Arranjo experimental para obtenção de uma curva de SAXS.

Mediante a técnica de SAXS é possível determinar a intensidade de espalhamento produzida por qualquer material com heterogeneidades na densidade eletrônica, na faixa compreendida entre alguns Ångström (Å) e algumas centenas de Ångström. No caso de sistemas macroscopicamente isotrópicos (como no caso de proteínas em solução) essa intensidade é expressa em função do módulo do vetor posição no espaço recíproco q , sendo q relacionado com o ângulo de espalhamento e pela relação [3] $q = \frac{4\pi}{\lambda} \sin(\epsilon/2)$.

No caso de ângulos baixos q pode ser aproximado por $q = \frac{2\pi}{\lambda} \epsilon$. O ângulo ϵ (Figura 9b) é determinado a partir de D (distância amostra-detector) e Δx (distância entre uma posição genérica definida pelo detector sensível a posição x e a posição associada ao feixe direto x_0 , $\Delta x = x - x_0$), mediante:

$$\epsilon = \frac{\Delta x}{D}$$

2.6 Condições experimentais

2.6.1 Efeitos instrumentais (“Smearing”)

Todas as medidas de SAXS foram realizadas no LNLS, Campinas onde utilizamos a linha de SAS. Vários comprimentos de onda dos raios-X incidentes foram utilizados nas diversas experiências ($\lambda=1.38\text{\AA}$, $1.61\text{\AA}$ e $1.73\text{\AA}$) de acordo com a necessidade. Para aquisição de dados utilizamos um detector 1D sensível à posição (DSP). Com ele e com a ajuda de um multicanal adquirimos a curva de intensidade de espalhamento em função do número de canal. O espalhamento parasita do ar, janelas e das fendas foi subtraído da intensidade total espalhada. Houve necessidade de fazer um “desmearing”, isto é, corrigir o dados experimentais para corrigir os efeitos de borramento (“smearing”) introduzidos pela janela de entrada dos fótons do DSP (que era de 8mm).

O desborramento foi realizado utilizando o programa GNOM [30,31], porém não foi aplicado diretamente aos dados experimentais e sim sobre a calculada que ele compara com a experimental.

A Figura 10 ilustra o efeito de “smearing” causado por uma fenda de 8mm em comparação com uma fenda muito estreita (1mm). Utilizamos a fenda de 8mm, pois assim conseguíamos ter um fluxo de fótons 8 vezes maior. Já a ângulos maiores (acima de $0,05\text{\AA}^{-1}$) esse efeito quase desaparece, porém se não for realizado o “desmearing”, erros sistemáticos são introduzidos nos resultados finais comprometendo assim os resultados.

Figura 10 – Efeito do “smearing” causado nas curvas de intensidade em função do ângulo e o respectivo “desmearing” (linha contínua). Verificou-se que a curva na qual se realizou o “desmearing” corresponde a curva supondo não haver “smearing” (fenda de 1mm).

A distância entre amostra e o detector foi escolhida de modo a adquirir o espalhamento de 0.01 a 0.6 Å⁻¹. Dependendo do comprimento de onda utilizado, utilizaram-se distâncias entre 60 e 140cm. Também isolamos as amostras dentro de duas janelas de mica o que permitiu realizarmos as medidas com o mínimo de espalhamento parasita possível. Todas as curvas foram corrigidas por efeitos associados à não homogeneidade da resposta do detector.

2.6.2 Efeito da espessura da amostra na intensidade de espalhamento

A intensidade de espalhamento a baixo ângulo é proporcional à espessura da amostra (x) e a um fator que leva em consideração a absorção da mesma, como indicado pela equação seguinte:

$$I(x) = Ke^{-\mu_m \rho x}$$

Onde K é uma constante (independente de espessura) que depende da estrutura do material, μ_m é o coeficiente mássico de absorção e ρ é a densidade mássica da solução estudada (para soluções aquosas diluídas $\rho \sim 1\text{g/cm}^3$). O coeficiente depende do comprimento de onda e da composição atômica da solução.

O máximo da intensidade espalhada corresponde ao máximo da equação acima que ocorre quando $x = 1/\mu_m \rho$. Para os dois comprimentos de onda utilizados nas experiências de SAXS $\lambda = 1.38\text{\AA}$ e $1.61\text{\AA}$ e com a espessura da cela utilizada $x = 1\text{mm}$ temos que a intensidade espalhada é igual a $I_{\text{max}}(q)$ e $0.8I_{\text{max}}(q)$ respectivamente.

Durante as duas últimas décadas notáveis progressos têm sido alcançados no campo de instrumentação associado à técnica experimental de espalhamento de raios X a baixo ângulo. Utilizando feixes de raios X potentes, fornecidos por fontes de radiação síncrotron e registrando a intensidade de espalhamento com detectores de raios X sensíveis à posição (uni e bidimensionais), numerosas proteínas e outras moléculas biológicas em solução têm sido estudadas com sucesso [32].

Atualmente é possível realizar experiências de SAXS de boa qualidade com menos de 50µg de proteínas utilizando a linha de SAXS do LNLS (5-10 µL de solução a 2-10 mg de proteína/mL de solvente). Num gerador de raios X do tipo anodo rotatório, foi possível realizar medidas de SAXS com soluções contendo apenas 50µg (5mg/mL) ou seja, 10µL de solução protéica, com estatística razoável. Trabalhando com baixas concentrações (< 5mg/mL) pode-se evitar efeitos de agregação, bem como economizar na preparação de proteínas purificadas que às vezes pode ser bem trabalhoso e caro.

Acompanhando os avanços na área experimental, novas abordagens teóricas e computacionais foram desenvolvidas recentemente para tratamento e interpretação de dados de espalhamento a baixos ângulos, permitindo a análise detalhada da estrutura global das proteínas, como a da determinação direta do envelope molecular[33], predição de estruturas quaternárias em oligômeros[34] e verificação de possíveis diferenças entre as estruturas no seu estado cristalino e em solução[17].

- ¹ Feigin, LA and Svergun, DI – Structure analysis by Small-angle X-ray and Neutron Scattering, pp. 83-87, *Plenum Press*, New York (1987).
- ² Svergun, DI and Koch, MHJ – Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66** 1735-1782 (2003).
- ³ Glatter, O & Kratky, O Editors. – Small-Angle X-ray Scattering. *London: Academic Pres.* (1982).
- ⁴ Winick, H and Doniach, S Editors – An Overview of Synchrotron Radiation Research, in *Synchrotron Radiation Research*, Plenum Press, New York and London (1980).
- ⁵ Guinier, A and Fournet, G – Small-angle scattering of X-rays, *John Wiley & Sons*, New York (1955).
- ⁶ Shannon, CE and Weaver, W – The Mathematical Theory of Communication, *University of Illinois Press*, Urbana, IL (1949).
- ⁷ Taupin, D and Luzzati, V – Information content and retrieval in solution scattering studies. I. Degrees of freedom and data reduction. *J. Appl. Cryst.* **15**, 289-300 (1982).
- ⁸ Stuhmann, HB – Ein neues Verfahren zur Bestimmung der Oberflächenform und der inneren Struktur von gelösten globulären Proteinen aus Röntgenkleinwinkelmessungen. *Z. Phys. Chem. Frankfurt*, **72**, 177-198 (1970a).
- ⁹ Langridge, R, Marvin, DA, Seeds, WE, Wilson, HR, Hooper, CW, Wilkins, MHF and Hamilton, LD – The Molecular Configuration of Deoxyribonucleic Acid.: Molecular Models and their Fourier Transforms. *J. Mol. Biol.* **2**, 38-64 (1960).
- ¹⁰ Svergun, DI, Barberato, C, Koch, MHJ, Fetler, L and Vachette, P – Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins*, **27**, 110-117 (1997).
- ¹¹ Ninio, J, Luzzati, V and Yaniv, M – Comparative small-angle x-ray scattering studies on unacylated, acylated and cross-linked *Escherichia coli* transfer RNA I Val. *J. Mol. Biol.*, **71**, 217-229 (1972).
- ¹² Fedorov, BA, Ptitsyn, OB and Voronin, LA – X-Ray diffuse scattering of globular protein solutions: Consideration of the solvent influence. *FEBS lett.* **28**, 188-190 (1972).
- ¹³ Fedorov, BA and Denesyuk, AI – Large-angle X-ray diffuse scattering, a new method for investigating changes in the conformation of globular proteins in solutions. *J. Appl. Cryst.* **11**, 473-477 (1978).
- ¹⁴ Müller, JJ – Calculation of scattering curves for macromolecules in solution and comparison with results of methods using effective atomic scattering factors. *J. Appl. Cryst.* **16**, 74-82 (1983).
- ¹⁵ Pavlov, MY, Sinev, MA, Timchenko, AA and Ptitsyn, OB – A study of apo- and holo-forms of horse liver alcohol dehydrogenase in solution by diffuse x-ray scattering. *Biopolymers*, **25**, 1385-1397 (1986).
- ¹⁶ Grossmann JG, Abraham ZH, Adman ET, Neu M, Eady RR, Smith BE and Hasnain SS – X-ray scattering using synchrotron radiation shows nitrite reductase from *Achromobacter xylosoxidans* to be a trimer in solution. *Biochemistry*, **32**, 7360-7366 (1993).

-
- ¹⁷ Svergun, DI, Barberato, C and Koch, MH – CRY SOL- a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**:768-773 (1995).
- ¹⁸ Stuhrmann, HB – Interpretation of small-angle scattering of dilute solutions and gases. A representation of the structures related to a one-particle scattering functions. *Acta Crystallogr.* **A26**:297-306 (1970b).
- ¹⁹ Konarev, PV, Petoukhov, MV and Svergun, DI – MASSHA - a graphic system for rigid body modelling of macromolecular complexes against solution scattering data. *J. Appl. Crystallogr.* **34**, 527-532 (2001).
- ²⁰ Svergun, DI – Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**:2879-2886 (1999)
- ²¹ Svergun, DI, Petoukhov, MV and Koch, MHJ – Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946-2953 (2001).
- ²² Walther, D, Cohen, FE and Doniach, S – Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. *J. Appl. Crystallogr.* **33**, 350-363 (2000).
- ²³ Petoukhov, MV and Svergun, DI – New methods for domain structure determination of proteins from solution scattering data. *J. Appl. Cryst.* **36**, 540-544 (2003).
- ²⁴ Vigil, D, Gallagher, SC, Trewhella, J and Garcia, AE – Functional Dynamics of the Hydrophobic Cleft in the N-Domain of Calmodulin. *Biophys J* **80**, 2082-2092 (2001).
- ²⁵ Kozin MB. and Svergun DI – Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* **34**, 33-41 (2001).
- ²⁶ Ingber, L – Simulated annealing: practice versus theory. *Math. Computer Modelling.* **18**:29 –57 (1993).
- ²⁷ Heenan, RK – The FISH Reference Manual. Rutherford Appleton Laboratory, Didcot (2000).
- ²⁸ Chacon, P, Moran, F, Diaz, JF, Pantos, E and Andreu, JM – Low-Resolution Structures of Proteins in Solution Retrieved from X-Ray Scattering with a Genetic Algorithm. *Biophys J* **74**, 2760-2775 (1998).
- ²⁹ Chacon, P, Diaz, JF, Moran, F and Andreu, JM – Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *J Mol Biol* **299**(5), 1289-1302 (2000).
- ³⁰ Svergun, DI – Mathematical methods in small-angle scattering data analysis *J. Appl. Cryst.* **24**, 485-492 (1991).
- ³¹ Svergun, DI – Determination of the regularization parameter in indirect-transform methods using perceptual criteria *J. Appl. Cryst.* **25**, 495-503 (1992).
- ³² Koch, MHJ – Scattering of Non-Crystalline Systems. In "Handbook on Synchrotron Radiation", vol. 4, Eds. Ebashi S, Koch M and Rubenstein E. *Elsevier Science Publishers B.V.* (1991).
- ³³ Svergun, DI and Stuhrmann, HB – New developments in direct shape determination from small-angle scattering .1. Theory and model-calculations. *Acta Cryst.* **A47**:736-744 (1991)
- ³⁴ Svergun, DI – Solution scattering from biopolymers: advanced contrast variation data analysis. *Acta Crystallogr.* **A50**, 391-402 (1994).

3	<i>Determinação do estado oligomérico de macromoléculas em solução a partir de medidas de saxs em escala relativa</i>	3 - 1
3.1	Introdução	1
3.2	Efeito do valor máximo do módulo do vetor de onda utilizado na determinação do volume do objeto espalhador esférico	4
3.3	Efeito do valor máximo do módulo do vetor de espalhamento ($q_{\max}$) utilizado na determinação do volume de 1148 proteínas obtidas do PDB	7
3.4	Cálculo do volume a partir de dados experimentais de SAXS de várias proteínas em solução e conclusões	10
3.5	Dependência da densidade de proteínas em solução em função do peso molecular	21

3 DETERMINAÇÃO DO ESTADO OLIGOMÉRICO DE MACROMOLÉCULAS EM SOLUÇÃO A PARTIR DE MEDIDAS DE SAXS EM ESCALA RELATIVA

3.1 Introdução

O peso molecular de uma proteína pode ser determinado por SAXS mediante medidas de intensidades em escala absoluta ou mediante comparação de curvas experimentais com uma curva de espalhamento “padrão” de proteínas de peso molecular e concentração conhecidos.

O que propomos neste trabalho é obter uma estimativa precisa do estado oligomérico de uma macromolécula em solução, ou seja, determinar a sua massa. Isso pode ser feito muito mais precisamente com gel de electroforese e espectroscopia de massa, mas obter estimativa aproximada ou exata do oligômero de solução não pode ser feito com relativa facilidade por estas técnicas nas mesmas condições (concentração da proteína, tampão, etc).

Medir o peso molecular de uma proteína não é algo novo e tem sido descrito em alguns trabalhos utilizando a técnica de SAXS [1,2,3,4] bem como inúmeros outros

utilizando outras técnicas como eletroforese [5,6], espectroscopia de massa [7], espalhamento de luz [8,9], espectroscopia óptica [10], NMR [11] entre outras [12,13,14]. Abaixo uma breve comparação entre as técnicas utilizadas hoje em dia.

Técnica	Condições da medida	Acurária na medida
SAXS	Solução	5-15%
Espalhamento de Luz	Solução	10-50%
Espectroscopia óptica	Solução	2-30%
Espectrometria de massa	Gás ionizado	< 1%
Eletroforese	Gel	5-20%
NMR	Solução	< 1%

Neste trabalho visamos obter o peso molecular com precisão superior a 10% utilizando uma curva qualquer de SAXS em escala relativa. Discutiremos o problema bem como eventuais dificuldades experimentais e limitações. Esperamos com este método contribuir de alguma forma a resolver problemas de biologia estrutural e relacionados.

Quando estudamos proteínas em solução, elas nem sempre se comportam da mesma forma que numa coluna de purificação ou de exclusão de massa, ou num gel, submetidas a ensaios de cristalização ou em outra situação qualquer. Temos constatado inúmeras vezes proteínas se comportarem de modo completamente diferente quando em solução, principalmente no que diz respeito ao seu estado oligomérico (dependendo do pH, concentração, temperatura, etc... ela (dis)associa). Conhecer o estado oligomérico também é muitas vezes importante no que diz respeito a função e atividade biológica da proteína.

Não encontramos nenhum artigo onde se utiliza SAXS para determinar a massa molecular de macromoléculas em solução a não ser um interessante trabalho para determinação da massa molecular de polímeros usando um método de diluição múltipla [3,4]. Porém neste trabalho não encontramos uma abordagem para proteínas em solução e poucos exemplos sucedidos são apresentados.

Três condições básicas devem ser obedecidas para que a análise exposta a seguir seja válida:

- 1) O sistema deve ser estatisticamente isotrópico e monodisperso. Neste caso tanto faz se as partículas estão orientadas aleatoriamente ou se elas rotacionam em função do tempo.
- 2) Não existe correlação espacial entre as posições instantâneas das partículas. Ou seja, se trata de um sistema "diluído". O estar "diluído" depende do peso molecular (PM) das proteínas. Em caso de não saber se um certo sistema é diluído devem ser feitas várias curvas com diluições sucessivas e verificar se a forma da curva não é alterada. Se sim, deve-se sempre utilizar a mais diluída

estatisticamente significativa. Em caso corrente, para qualquer solução de proteínas uma concentração de 1mg/mL praticamente pode ser considerada como um sistema diluído.

- 3) As curvas devem estar livres de qualquer efeito de “borramento” (smearing) causado ou por um feixe de raios-X não pontual ou detector não pontual. Procedimentos de desmearing são cruciais para obtenção e estimativa correta da intensidade de espalhamento a ângulo zero.

A intensidade espalhada a ângulo zero $I(0)$ é dada por:

$$I(0) = (\Delta\rho)^2 \cdot v^2 \quad \text{Equação 1}$$

Outro parâmetro importante é a integral da intensidade em todo espaço recíproco, que está diretamente relacionada às flutuações médias da densidade eletrônica dos objetos espalhadores, e é definida como:

$$Q = \int_{q_{\min}}^{q_{\max}} I(q) \cdot q^2 dq \quad \text{Equação 2a}$$

onde $q_{\min} \rightarrow 0$ e $q_{\max} \rightarrow \infty$. Como $Q = 2\pi^2 \cdot \gamma(0) = 2\pi^2 \int_V \rho^2(r) dr = 2\pi^2 (\Delta\rho)^2 V$, temos:

$$Q = 2\pi^2 \cdot V \cdot (\Delta\rho)^2 \quad \text{Equação 2b}$$

Substituindo Equação 1 na Equação 2b temos que

$$V = \frac{2\pi^2 \cdot I(0)}{Q} \quad \text{Equação 3}$$

Sendo a Equação 3 válida para intensidades em escala relativa, não existe a necessidade de transformar as intensidades para uma escala absoluta. Por outro lado, o fator de contraste de densidade eletrônica também desapareceu quando substituímos a Equação 1 na Equação 2b. Assim podemos conhecer o volume do objeto espalhador simplesmente conhecendo $I(0)$ e a integral Q , ou seja, não necessitamos determinar qual o contraste e muito menos fazer medidas em escala absoluta.

Portanto, conhecendo $I(0)$ e Q , temos o volume da partícula espalhadora e, se sabemos a densidade média (ρ) da mesma, temos finalmente a massa:

$$M = \frac{2\pi^2 \cdot I(0) \cdot \rho}{Q}$$

Equação 4

Esta equação é geral e deveria valer para qualquer sistema de proteínas em solução se as três condições mencionadas no início desta seção fossem obedecidas. Assim eliminamos vários problemas como fazer medidas em escala absoluta, não precisamos saber a densidade do solvente e nem a concentração da amostra.

Precisamos sim estimar com precisão valores de $I(0)$ e Q o que nem sempre é muito fácil e podem estar sujeitos a erros, pois para determinar $I(0)$ temos que extrapolar os dados experimentais a ângulo zero e para determinar Q teríamos que extrapolar os dados para altos ângulos. Faremos particularmente uma análise detalhada do método quando são utilizadas curvas de SAXS somente num intervalo limitado de valores de q .

3.2 Efeito do valor máximo do módulo do vetor de onda utilizado na determinação do volume do objeto espalhador esférico

A primeira modelagem feita foi utilizando um modelo teórico de esferas duras de raios variando entre $1\text{\AA}$ e $360\text{\AA}$. Calculamos teoricamente a intensidade de espalhamento produzida por uma esfera destas utilizando a seguinte equação:

$$I(q) = (\Delta\rho^2)V^2 \left(3 \frac{\sin qR - qR \cos qR}{(qR)^3} \right)^2$$

Equação 3-5

Calculamos $I(0)$ e a integral Q definida na Equação 2a. Esta integral foi calculada nos limites entre $q_{\min} = 0$ e $q_{\max}$ variou-se entre $0.1\text{\AA}^{-1}$ e $1\text{\AA}^{-1}$. Em seguida graficamos a Equação 3 para esferas com volumes entre 10^4 - $10^6\text{\AA}^3$ (Figura 1).

Figura 1: Gráfico do resultado da Equação 3 para diferentes $q_{\max}$ em função do volume da esfera espalhadora. (a) intervalo do tamanho de partículas entre 1 a 10^9Å^3 e (b) ampliação da região cinza em a ilustrando o tamanho típico da maioria das proteínas (entre $4000\text{-}500000 \text{Å}^3$). Escalas em log-log para maximizar as diferenças.

Se tivéssemos integrado em todo espaço recíproco (Equação 2a) esperaríamos um comportamento linear se graficássemos a Equação 3, e de fato, quando integramos para valores de $q_{\max}$ maiores de 0.2Å^{-1} temos esse comportamento no intervalo de volume de 10^4Å^3 ou maiores. Porém se $q_{\max}$ for menor que 0.2Å^{-1} há um desvio deste valor, o que é perfeitamente compreensível visto que estamos ignorando parte do espalhamento no espaço recíproco. Esferas muito grandes, com volumes superiores a 10^5Å^3 apresentam um comportamento linear mesmo para $q_{\max}$ abaixo de 0.2Å^{-1} o que é justificável pelo fato de a maior parte do espalhamento se concentrar a baixíssimos ângulos, ou seja, q 's.

Outro resultado decorrente desta análise foi verificar qual o valor de $q_{\max}$ necessário para que mais de 98% de todo espalhamento no espaço recíproco seja integrado e isto está representado de forma esquemática na Figura 2.

A análise referente ao cálculos da integral Q até um certo valor de $q_{\max}$ (e não até infinito) é importante para proteínas em solução porque nestes casos a intensidade para q relativamente grandes pode ser extremamente pequena com os erros estatísticos muito elevados, tornando inviável qualquer extrapolação para obter Q até $q=\infty$.

Figura 2: Gráfico ilustrando o valor de $q_{\max}$ necessário para esferas com determinado raio de giro para diferentes completezas. (a) Intervalo de esferas com raio de giro entre 1 a 300 Å e (b) ênfase para valores de raio de giro entre 10 a 50 Å, que corresponde à maioria das macromoléculas existentes na natureza, ilustrado na caixa cinza em a.

Observando a Figura 2b podemos notar claramente que para uma completeza de 90%, ou seja, 90% de todo espalhamento produzido pela esfera espalhadora com R_g de 10 Å temos que integrar q entre 0 e 0.45 Å^{-1} . Se queremos uma completeza ainda maior, em torno de 98%, o intervalo de integração já aumenta de 0 e 0.82 Å^{-1} . A importância da completeza é crucial, pois se integrarmos somente até um valor de $q_{\max} = 0.1 \text{ Å}^{-1}$ temos uma completeza de somente 10% no caso da mesma esfera com raio de giro de 10 Å. Claro que para esferas maiores, com raio de giro de 30 Å, por exemplo, a completeza a $q_{\max} = 0.1 \text{ Å}^{-1}$ já sobe para 80%.

Como vimos na Figura 1b, existe um pronunciado desvio da linearidade da Equação 3 para esferas com R_g menores de 30 Å quando $q_{\max} = 0.1 \text{ Å}^{-1}$ ou menores de 17 Å quando $q_{\max} = 0.2 \text{ Å}^{-1}$. Já para intervalos de integração acima de $q_{\max} = 0.4 \text{ Å}^{-1}$ quase todo intervalo de tamanhos se comporta de forma linear, sugerindo ser este um valor excelente para ser usado como referência. De fato, se olharmos a Figura 2b, notamos que para valores de $q_{\max} = 0.4 \text{ Å}^{-1}$ temos uma completeza de 90%. Logo, supomos ser 90% o valor de completeza mínimo necessário para uma boa estimativa do volume da partícula sob análise. Logo, a completeza depende do tamanho da partícula em estudo e assumimos que a curva azul da Figura 2b (90% de completeza) corresponde a um valor ideal para calcularmos com precisão razoável o volume e conseqüentemente o peso molecular de proteínas ou qualquer

outro objeto espalhador. Resumindo, para a maioria das proteínas precisamos integrar até um valor de $q_{\max}$ entre $0.2-0.3\text{\AA}^{-1}$.

Para realizar uma análise mais refinada dos resultados discutidos anteriormente, selecionamos 1148 proteínas da base de dados do PDB [15], com peso molecular acima de 10 KDa e que foram depositados entre 1995 e 2002 aplicando um filtro de sequência por homologia de 50% evitando assim incluir proteínas redundantes na análise.

3.3 Efeito do valor máximo do módulo do vetor de espalhamento ($q_{\max}$) utilizado na determinação do volume de 1148 proteínas obtidas do PDB

As curvas teóricas de SAXS foram calculadas utilizando-se o programa CRY SOL [16] descrito no capítulo 2. Fizemos um gráfico da Equação 3 para diferentes valores limites de integração $q_{\max}$ como pode ser visto na Figura 3 e Figura 4. A única diferença entre estas figuras é o tipo de ajuste representado pela linha contínua. Ajustamos os dados obtidos pela integração correspondentes as intensidades produzidas das 1148 proteínas para diferentes $q_{\max}$ utilizando os dados teóricos da Figura 1 representado na Figura 3 e fizemos o ajuste linear dos dados (usando uma equação $y = A + Bx$) representado na Figura 4. Podemos observar claramente a grande dispersão dos pontos em torno dos ajustes para $q_{\max} = 0.1\text{\AA}^{-1}$, porém para ajustes para $q_{\max}$ acima de $0.2\text{\AA}^{-1}$ temos uma dispersão muito melhor como já previsto anteriormente. Com base nestes ajustes calculamos o desvio médio dos pontos da reta da Figura 4 para termos uma idéia da dispersão do tamanho ideal das proteínas do valor calculado utilizando uma simples reta e o resultado está ilustrado na Figura 5.

(a)

(b)

Figura 3: Gráfico ilustrando a razão entre $I(0)/Q$ em função do volume calculado teoricamente para diversos $q_{\max}$ para as 1148 proteínas. (a) Sobreposição dos dados teóricos de esferas duras da Figura 1, (b) em escala logarítmica.

(a)

(b)

Figura 4: Gráfico ilustrando a razão entre $I(0)/Q$ em função do volume calculado teoricamente para diversos $q_{\max}$ para as 1148 proteínas. (a) Ajuste linear dos dados, (b) mesmo gráfico em escala logarítmica.

Podemos observar que para valor de integração de $q_{\max}$ menores de $0.2\text{\AA}^{-1}$ temos uma dispersão maior de 13-16%, ou seja, erros cometidos poderão ser em média em torno de 14% e em alguns casos até 20-30%. Quando integramos até $q_{\max} = 0.2\text{\AA}^{-1}$ o erro já diminui para 12% e para $q_{\max} = 0.4\text{\AA}^{-1}$ temos um erro de menos de 8% (veja Figura 5).

Figura 5: (a) Distribuição do desvio em % do peso molecular ideal em função da integração de $q_{\max}$ e (b) Erro médio cometido para as 1148 proteínas em função de $q_{\max}$.

Figura 6: Coeficientes obtidos ajustando os dados do gráfico da Figura 4 com a equação da reta $y = A + Bx$ onde $y = I(0)/Q$ e $x = V$. (a) Coeficiente linear para proteínas alongadas (vermelho), todas juntas (preto) e esferoidais (verde) em função de $q_{\max}$ (b) Coeficiente angular em função de $q_{\max}$ para proteínas alongadas (vermelho), todas juntas (preto) e esferoidais (verde).

Fizemos mais algumas simulações para ver como os coeficientes A e B da reta variam caso utilizássemos somente proteínas alongadas, ou seja, mais alongadas numa direção do que na outra numa razão de 2:1 e no caso de proteínas esféricas (Figura 6). Observamos que existe um comportamento característico e que deve ser levado em conta para uma determinação mais precisa da massa molecular. Para descobrir se a proteína em estudo é esferoidal ou alongada, basta analisar a função de distribuição de distâncias (veja figura 2-6). Quando esta função é aproximadamente simétrica temos a indicação da proteína ser esferoidal e quando ela é assimétrica significa que a proteína é ou oblata (assimétrica deslocada para a direita) ou prolata (assimétrica deslocada para esquerda).

Se definirmos como $R_2 = R_{\max}$ e $R_1 = R[P(r)_{\max}]$, consideramos que a proteína é alongada se $R_2/R_1 > 3$. No caso de uma esfera essa razão vale 2.

3.4 Cálculo do volume a partir de dados experimentais de SAXS de várias proteínas em solução e conclusões

Agora relataremos a parte dos exemplos experimentais coletados por nós mostrando como podemos calcular o peso molecular somente com uma curva de SAXS em escala arbitrária e daremos um exemplo passo a passo sobre como determinar a massa molecular de uma proteína e finalmente mostraremos uma tabela onde analisamos vários casos de proteínas diferentes.

Partimos de uma curva de SAXS em escala arbitrária. Investigamos somente proteínas em soluções diluídasⁱ, caso contrário o método não se aplicaria e se tornaria inválido. Utilizamos como exemplo uma proteína chamada Jararagina [17]. Primeiramente extrapolamos os dados para $q = 0$, utilizamos para isto o programa GNOM [18,19]. Foi importante fazer um “desmearing” correto dos dados, ou seja, corrigir os dados dos efeitos de borramentos causados pela largura das fendas de entrada no detector. (Isto se aplica mesmo em configurações utilizadas em síncrotrons, que utilizam em geral um feixe quase pontual, mas um detector com fenda finita. Em detectores CCD modernos este efeito é praticamente desprezível) como descrito no capítulo 2.6.1. Os dados obtidos da Jararagina foram medidos na linha de SAXS do LNLS e também tiveram que ser corrigidas pela altura da fenda. Na Figura 7a temos a curva de espalhamento de SAXS em escala arbitrária e a intensidade extrapolada para $q=0$ foi normalizada em 1. Desta forma temos que $I(0)=1$, e na Figura 7b está o gráfico para podemos calcular a integral Q definida na Equação 2b.

Figura 7: (a) Intensidade versus módulo do vetor de espalhamento q da Jararagina, dados experimentais em quadrados e ajuste obtido pelo GNOM em linha

ⁱ Para que os efeitos de interferência dos raios X espalhados por cada proteína individual não interfiram com a vizinha introduzindo artefatos no padrão de espalhamento.

contínua. (b) Intensidade multiplicada por q^2 versus q também conhecido como (Kratky plot) necessário para cálculo da integral Q .

Para obtermos o volume teríamos que integrar o gráfico da Figura 7b de zero a infinito, porém não temos informações acima de $q=0.6\text{\AA}^{-1}$ o que nos impede de realizarmos essa integral, por este motivo vamos integrar este gráfico de 0 a $0.1\text{\AA}^{-1}$, $0.2\text{\AA}^{-1}$, $0.3\text{\AA}^{-1}$, $0.4\text{\AA}^{-1}$ e $0.5\text{\AA}^{-1}$ e analisar os resultados obtidos.

Obtivemos os seguintes valores para as integrais em Q representados na Tabela 1:

$q_{\max} (\text{\AA}^{-1})$	$Q (\times 10^{-5})$	$I(0)/Q$	Volume ($\text{\AA}^3$)	Massa (Da)	Razão
0.1	5.42	18447	159691	116574	0.81
0.2	8.60	11631	132336	96605	0.96
0.3	10.64	9402	121244	88508	1.06
0.4	11.93	8379	122498	89423	1.05
0.5	13.66	7318	130189	95038	0.99

Tabela 1: Integral Q da Figura 7b para diferentes valores de integração de $q_{\max}$, razão entre $I(0)=1$ e Q , volume obtido utilizando os valores de A e B da Figura 6, Massa M obtida utilizando o valor da densidade média de $1.37(3) \text{ g/cm}^3$ e finalmente a razão entre o peso molecular teórico de 93.7 KDa e o experimental, neste caso uma discrepância nula corresponderia a uma razão igual a 1.

Como pode ser visto na Tabela 1, foi possível calcular o volume da proteína a partir de uma integração de $q_{\max}$ com somente $0.1\text{\AA}^{-1}$, porém com um erro superior a 20% sendo que nos outros casos, quando utilizamos um valor de integração de $q_{\max}$ maior de $0.2\text{\AA}^{-1}$ temos um erro menor de 6%.

Repetimos este procedimento para outras 20 proteínas a fim avaliarmos melhor a capacidade e limitações da técnica. Na Tabela 2 apresentamos os valores da integral para 20 diferentes proteínas medidas pela técnica de SAXS para diferentes valores de $q_{\max}$. Não foi possível obter para todas proteínas valores da integral Q até $0.5\text{\AA}^{-1}$ devido ao espalhamento extremamente fraco acima de $0.2\text{\AA}^{-1}$ ou $0.3\text{\AA}^{-1}$. A fim de facilitar comparações normalizamos todas as intensidades a ângulo zero em um $[I(0)=1]$.

Proteína	Peso Molecular (KDa)		Q (x 10 ⁻⁵) para diferentes q _{max} em Å ⁻¹				
	Monômero	Oligômero	0.1	0.2	0.3	0.4	0.5
Lisozima	14.6		20.27	49.81			
RXR [20]	45	90	7.49	12.34			
EPSPS	44.5	89	8.22	12.59			
Prot1	50		10.92	17.73			
PEPCK [21]	58	116	5.63	8.25			
Alfa Galactosidade	48.76		11.59	17.00			
Troponina C	18.3		13.50	34.15			
Integrina	21.6		12.01	31.44	39.92		
Interleukina 22	16.84	33.68	9.68	19.02	30.07		
Interleukina 22	16.84	67.36	6.74	10.88	16.09		
PH	26.7	213.6	3.54	4.37	4.99		
GREP [22]	21	42	5.63	11.48	16.59	22.57	
DM43	32.39	64.78	6.85	11.82	16.19	20.36	26.71
Jararagina	46.85	93.7	5.42	8.60	10.64	11.93	13.66
PFK-2 sem ATP							
[23]	34	68	8.64	12.31	13.24	13.97	16.15
BSA	66		8.31	13.57	17.80	23.18	31.00
PFK-2 com ATP							
[23]	34	136	5.45	7.27	8.01	9.47	12.13
FMS	41	164	4.76	6.12	6.63	7.38	8.99
Prot4	17.5		19.72	43.54	51.19	62.89	81.75
RXR	45	180	3.42	5.84	8.12	10.65	13.29

Tabela 2: Integral (Q) para diferentes valores de integração de q_{max} e 20 diferentes proteínas reais medidas por SAXS, q_{max} coletado varia entre 0.2 e 0.5 Å⁻¹. Todas as curvas foram normalizadas tal que I(0)=1 a fim de resaltar a variação de Q em função da massa molecular. Peso molecular monômero significa qual o peso molecular do monômero e oligômero é o peso molecular total da proteína.

Depois que calculamos a integral, pudemos facilmente encontrar o volume das 20 diferentes proteínas utilizando a equação $y = A + Bx$ com coeficientes obtidos pelo gráfico 8 para diferentes valores de integração de q_{max}. Os resultados do volume estão representados na Tabela 3.

Como a densidade média das proteínas é em torno de 1.370(3) g/cm³ pudemos calcular o valor estimativo da massa das proteínas que estão representados na Tabela 4.

Proteína	Peso Molecular (KDa)		Volume ($\text{\AA}^3$) para diferentes q_{max} em $\text{\AA}^{-1}$				
	Monômero	Oligômero	0.1	0.2	0.3	0.4	0.5
Lisozima	14.6		17462	16728			
RXR	45	90	105473	90421			
EPSPS	44.5	89	93106	88503			
Prot1	50		61691	60649			
PEPCK	58	116	151562	139024			
Alfa Galactosidade	48.76		56104	63585			
Troponina C	18.3		43317	27857			
Integrina	21.6		52996	30907	27846		
Interleukina 22	16.84	33.68	73970	56033	38914		
Interleukina 22	16.84	67.36	121026	103641	77891		
PH	26.7	213.6	261371	268989	264243		
GREP	21	42	151708	97794	75356	62656	
DM43	32.39	64.78	118641	94790	77372	69979	65189
Jararagina	46.85	93.7	158843	133104	120902	122616	130189
PFK-2 sem ATP	34	68	86944	90671	95971	104042	109705
BSA	66		91797	81559	69856	60886	55784
PFK-2 com ATP	34	136	157760	158841	162460	155662	146986
FMS	41	164	185468	190213	197539	201102	199387
Prot4	17.5		18890	20220	20409	19521	19369
RXR	45	180	271804	199443	160207	137978	133922

Tabela 3: Volume obtido após aplicar a equação $y = A + Bx$ com coeficientes obtidos da Figura 6 para diferentes valores de integração de q_{max} e 20 diferentes proteínas reais medidas por SAXS, q_{max} coletado varia entre 0.2 e $0.5 \text{\AA}^{-1}$. Todas as curvas foram normalizadas tal que $I(0)=1$ a fim resaltar a variação de V em função da massa molecular. Peso molecular monômero significa qual o peso molecular do monômero e oligômero é o peso molecular total da proteína.

Finalmente calculamos a razão entre a massa real da proteína e a massa obtida por nosso método empírico e este resultado é apresentado na Tabela 5. Nesta tabela os valores correspondentes a massas encontradas para cada q_{max} com erro menor de 10% são representadas em verde, com erros entre 10-25% em laranja e com erros superiores a 25% em vermelho. Aparentemente temos uma subestimação da massa molecular o que pode ser devido a densidade eletrônica excessivamente alta. De fato, se modificarmos a densidade eletrônica e a reduzirmos para 1.25 g/cm^3 temos um acordo muito melhor entre o teórico e o experimental. Estes resultados estão apresentados na Tabela 6 (massa obtida) e Tabela 7 (razões entre real e experimental).

Se desconsideramos os dados de massas molecular obtidos para Q integrado somente até $q_{\max}=0.1\text{\AA}^{-1}$, podemos notar entre a Tabela 5 e Tabela 7 uma considerável melhora entre os dados reais e experimentais, sendo que os dados com erros superiores a 25% diminuindo de 11 para somente 5 pontos e os com erros entre 10-25% diminuindo de 24 para 21 e o aumento dos dados com erros inferiores a 10% de 15 para 24.

Proteína	Peso Molecular (KDa)		Massa (KDa) para diferentes $q_{\max}$ em $\text{\AA}^{-1}$				
	Monômero	Oligômero	0.1	0.2	0.3	0.4	0.5
lisozima	14.6		12.7	12.2			
RXR	45	90	77.0	66.0			
EPSPS	44.5	89	68.0	64.6			
Prot1	50		45.0	44.3			
PEPCK	58	116	110.6	101.5			
Alfa Galactosidade	48.76		41.0	46.4			
Troponina C	18.3		31.6	20.3			
Integrina	21.6		38.7	22.6	20.3		
Interleukina 22	16.84	33.68	54.0	40.9	28.4		
Interleukina 22	16.84	67.36	88.3	75.7	56.9		
PH	26.7	213.6	190.8	196.3	192.9		
GREP	21	42	110.7	71.4	55.0	45.7	
DM43	32.39	64.78	86.6	69.2	56.5	51.1	47.6
Jararagina	46.85	93.7	115.9	97.2	88.2	89.5	95.0
PFK-2 sem ATP	34	68	63.5	66.2	70.1	75.9	80.1
BSA	66		68.6	60.5	54.5	50.2	48.7
PFK-2 com ATP	34	136	115.2	115.9	118.6	113.6	107.3
FMS	41	164	135.4	138.8	144.2	146.8	145.5
Prot4	17.5		13.8	14.8	14.9	14.2	14.1
RXR	45	180	198.4	145.6	116.9	100.7	97.8

Tabela 4: Massa calculada supondo densidade média em torno de $1.370(3) \text{ g/cm}^3$ utilizando para isto resultados obtidos da tabela 3. Peso molecular monômero significa qual o peso molecular do monômero e oligômero é o peso molecular total da proteína.

Proteína	Peso Molecular (KDa)		Razão massas para diferentes $q_{\max}$ em $\text{\AA}^{-1}$				
	Monômero	Oligômero	0.1	0.2	0.3	0.4	0.5
lisozima	14.6	14.6	1.15	1.20			
RXR	45	90	1.17	1.36			
EPSPS	44.5	89	1.31	1.38			
Prot1	50	50	1.11	1.13			
PEPCK	58	116	1.05	1.14			
Alfa Galactosidade	48.76	48.76	1.19	1.05			
Troponina C	18.3	18.3	0.58	0.90			
Integrina	21.6	21.6	0.56	0.96	1.06		
Interleukina 22	16.84	33.68	0.62	0.82	1.19		
Interleukina 22	16.84	67.36	0.76	0.89	1.18		
PH	26.7	213.6	1.12	1.09	1.11		
GREP	21	42	0.38	0.59	0.76	0.92	
DM43	32.39	64.78	0.75	0.94	1.15	1.27	1.36
Jararagina	46.85	93.7	0.81	0.96	1.06	1.05	0.99
PFK-2 sem ATP	34	68	1.07	1.03	0.97	0.90	0.85
BSA	66	66	0.96	1.09	1.21	1.31	1.35
PFK-2 com ATP	34	136	1.18	1.17	1.15	1.20	1.27
FMS	41	164	1.21	1.18	1.14	1.12	1.13
Prot4	17.5	17.5	1.27	1.19	1.17	1.23	1.24
RXR	45	180	0.91	1.24	1.54	1.79	1.84

Tabela 5: Tabela mostrando a razão entre a massa molecular obtida pela curva de SAXS para diferentes valores de integração $q_{\max}$ e a massa molecular real da proteína indicada na coluna oligômeros. Em verde estão as razões com erro menor de 10%, em laranja entre 10-25% e em vermelho dados com mais de 25% de erro.

Proteína	Peso Molecular (KDa)		Massa (KDa) para diferentes $q_{\max}$ em $\text{\AA}^{-1}$				
	Monômero	Oligômero	0.1	0.2	0.3	0.4	0.5
lisozima	14.6	14.6	14.0	13.4			
RXR	45	90	84.4	72.3			
EPSPS	44.5	89	74.5	70.8			
Prot1	50	50	49.4	48.5			
PEPCK	58	116	121.2	111.2			
Alfa Galactosidade	48.76	48.76	44.9	50.9			
Troponina C	18.3	18.3	34.7	22.3			
Integrina	21.6	21.6	42.4	24.7	22.3		
Interleukina 22	16.84	33.68	59.2	44.8	31.1		
Interleukina 22	16.84	67.36	96.8	82.9	62.3		
PH	26.7	213.6	209.1	215.2	211.4		
GREP	21	42	121.4	78.2	60.3	50.1	
DM43	32.39	64.78	94.9	75.8	61.9	56.0	52.2
Jararagina	46.85	93.7	127.1	106.5	96.7	98.1	104.2
PFK-2 sem ATP	34	68	69.6	72.5	76.8	83.2	87.8
BSA	66	66	75.2	66.3	59.7	55.1	53.4
PFK-2 com ATP	34	136	126.2	127.1	130.0	124.5	117.6
FMS	41	164	148.4	152.2	158.0	160.9	159.5
Prot4	17.5	17.5	15.1	16.2	16.3	15.6	15.5
RXR	45	180	217.4	159.6	128.2	110.4	107.1

Tabela 6: Massa calculada supondo densidade média em torno de 1.25 g/cm^3 utilizando para isto resultados obtidos da tabela 3. Peso molecular monômero significa qual o peso molecular do monômero e oligômero é o peso molecular total da proteína.

Proteína	Peso Molecular (KDa)		Razão massas para diferentes $q_{\max}$ em $\text{\AA}^{-1}$				
	Monômero	Oligômero	0.1	0.2	0.3	0.4	0.5
lisozima	14.6	14.6	1.05	1.09			
RXR	45	90	1.07	1.24			
EPSPS	44.5	89	1.19	1.26			
Prot1	50	50	1.01	1.03			
PEPCK	58	116	0.96	1.04			
Alfa Galactosidade	48.76	48.76	1.09	0.96			
Troponina C	18.3	18.3	0.53	0.82			
Integrina	21.6	21.6	0.51	0.87	0.97		
Interleukina 22	16.84	33.68	0.57	0.75	1.08		
Interleukina 22	16.84	67.36	0.70	0.81	1.08		
PH	26.7	213.6	1.02	0.99	1.01		
GREP	21	42	0.35	0.54	0.70	0.84	
DM43	32.39	64.78	0.68	0.85	1.05	1.16	1.24
Jararagina	46.85	93.7	0.74	0.88	0.97	0.96	0.90
PFK-2 sem ATP	34	68	0.98	0.94	0.89	0.82	0.77
BSA	66	66	0.88	1.00	1.11	1.20	1.24
PFK-2 com ATP	34	136	1.08	1.07	1.05	1.09	1.16
FMS	41	164	1.11	1.08	1.04	1.02	1.03
Prot4	17.5	17.5	1.16	1.08	1.07	1.12	1.13
RXR	45	180	0.83	1.13	1.40	1.63	1.68

Tabela 7: Tabela mostrando a razão entre a massa molecular obtida pela curva de SAXS para diferentes valores de integração $q_{\max}$ e a massa molecular real da proteína indicada na coluna oligômeros. Em verde estão as razões com erro menor de 10%, em laranja entre 10-25% e em vermelho dados com mais de 25% de erro.

Estes resultados nos indicam que provavelmente a densidade das proteínas de fato é um pouco menor que a atualmente mais aceita que era em torno de 1.35 g/cm^3 . De fato, se fizermos uma simulação supondo que em média proteínas tem uma camada de $3 \text{\AA}$ de água hidratada em torno delas sendo a densidade das proteínas de 1.37 g/cm^3 e a densidade da água de 1 g/cm^3 , chegamos a conclusão que a densidade média das proteínas vistas por SAXS varia de 1.05 a 1.30 g/cm^3 como ilustrado na Figura 8.

Figura 8: Gráfico ilustrando como a densidade média da proteína varia em função do volume da mesma, utilizamos modelos simples de proteínas esféricas e cúbicas com os volumes indicados e utilizando duas densidades distintas de 1.37 g/cm^3 e de 1.42 g/cm^3 .

Se, por exemplo, utilizarmos a densidade obtida na Figura 8 para calcularmos a massa das proteínas da tabela 3 chegamos a valores muito mais próximos do valor ideal o que nos leva a concluir que uma boa estimativa da densidade média do volume visto por SAXS é fundamental para uma estimativa relativamente precisa e correta da massa da proteína.

De todas proteínas medidas, somente a GRPE e RXR, marcadas em negrito na tabela 7 e indicadas em azul na Figura 9 (dois pontos extremos), tiveram um desvio muito grande do valor esperado de massa molecular e isso explicamos da seguinte maneira. A GRPE (42KDa) é uma proteína extremamente alongada ($R_2/R_1 > 4.3$, portanto seria necessário uma medida a maiores q 's, no mínimo q_{max} em torno de $0.5\text{-}0.6 \text{\AA}^{-1}$, vemos uma tendência da razão se aproximar de 1 para valores crescentes da integração de q , porém quanto maior q , mais erros podemos cometer, pois somente um tratamento de dados perfeito poderia garantir uma análise sem muito erro para cálculos da integral acima de $0.3 \text{\AA}^{-1}$. Desta forma o método fica um pouco limitado para proteínas muito alongadas ou muito pequenas (menores de 5 kDa).

Quando medidos a RXR tetramérica (180KDa), precisamos de mais de 10mg/mL de proteína, o que significa que eventuais efeitos de interferência possam ter causado um decréscimo em $I(0)$ que implica numa subestimação do volume e por consequência da massa ou também poderia haver uma mistura de tetrâmeros com dímeros o que explicaria o aumento na razão, pois um valor da razão acima de 1 significa uma subestimação da massa molecular e uma razão abaixo de 1 significa uma superestimação da massa molecular, porém tudo indica que o método não funcionou para esta proteína em particular visto que o espalhamento da RXR foi comparada com o da BSA em $I(0)$ e sua massa foi consistente com um tetrâmero.

Em geral o método se mostrou extremamente útil para analisar os dados de SAXS. É mais uma ferramenta para assegurar se uma medida foi realizada da maneira correta, se existem interferências na curva de espalhamento (muitas vezes supõe-se que a solução é monodispersa, mas este método pode ajudar a

Figura 9: Distribuição das razões do peso molecular real com o obtido pela técnica proposta para todas as 20 proteínas estudadas. Sob o quadrado cinza estão todas as proteínas com erros menores de 10% e fora delas em vermelho proteínas medidas com somente $q_{\max}=0.2^{-1}$.

3.5 Dependência da densidade de proteínas em solução em função do peso molecular

A densidade das proteínas é uma quantidade muito importante. Por exemplo ela pode ser útil para o processo de faseamento em cristalografia de monocristais, ou seja a atribuição das fases às amplitudes cujos módulos são calculados a partir das intensidades observadas (reflexões de Bragg). Uma medida precisa da densidade das proteínas também é essencial na predição correta do volume da proteína a partir da seqüência primária de aminoácidos. Um valor preciso também facilitará a determinação do número de subunidades na célula unitária, sendo que por muito tempo tem-se acreditado que a densidade média das proteínas é em torno de 1.35 g/cm^3 derivado de medidas de volume específico parcial de proteínas por métodos hidrodinâmicos [24,25] e compressibilidade adiabática [26]. Este valor porém foi revisado por Andersson e Hovmöller [27] que mostraram a densidade média de 28 diferentes proteínas ser em torno de $1.22(2) \text{ g/cm}^3$.

Quillin e Matthews[28] sugerem que a densidade média das proteínas deva ser em torno de $1.43(3) \text{ g/cm}^3$ e citam o artigo de Anderson e Hovmöller justificando a razão da discrepância ao algoritmo utilizado. Porém ainda há certa controvérsia quanto à densidade média das proteínas. Por exemplo, fazendo uma análise detalhada dos dados obtidos por Quillin e Matthews, descobrimos que há uma dependência do tamanho da proteína com sua densidade como ilustrado na Figura 10. Percebe-se que para proteínas com peso molecular abaixo de 20 KDa a densidade é maior que $1.44(3) \text{ g/cm}^3$ chegando a mais de 1.5 g/cm^3 para proteínas abaixo de 10 KDa o que pode ser uma indicação que proteínas pequenas sejam mais rígidas ou pelo menos mais compactas que proteínas grandes (com menos cavidades). Proteínas maiores de 40 KDa tem uma densidade média de $1.410(6) \text{ g/cm}^3$.

Figura 10: Gráfico ilustrando a dependência de densidade em função da massa molecular da proteína de acordo com Quillin e Matthews[28].

¹ Glatter, O and Kratky, O – *Small-Angle X-ray Scattering*. London: Academic Pres (1982).

² Guinier, A and Fournet, G – *Small-Angle Scattering of X-Rays*. John Wiley & Sons, New York (1955).

³ Pleštil J – Molecular-weight determination by the small-angle scattering method without using absolute intensities and contrast. *Makromol. Chem.-Macromol. Symp.* **15**, 185-190 (1988).

⁴ Pleštil J, Pospíšil H, Ostanevich YM and Degovics G – Molecular-Weight Determination from Small-Angle Scattering Without Absolute Intensities: Advantages and Limitations. *J. Appl. Cryst.* **24**, 659 – 664 (1991).

⁵ Williams, JG and Gratzer, WB – Limitations of detergent-polyacrylamide gel electrophoresis method for molecular weight determination of proteins. *J. Chromatogr.* **57**(1): 121 (1971).

⁶ He, Y and Yeung, ES – Rapid determination of protein molecular weight by the Ferguson method and multiplexed capillary electrophoresis. *J. Proteome Research* **1**(3), 273-277 (2002).

⁷ Trauger SA, Webb W and Siuzdak G – Peptide and protein analysis with mass spectrometry. *Spectroscopy International Journal*, **16**(1): 15-28 (2002).

⁸ Beignon, M, Bohic, S, Le Guennec, M, Le Goff, D, Roger, P and Proutiere, A – Molecular weight determination of macromolecules with a new simplified and coherent light scattering method, *J. Mol. Struct.* **443**(1-3): 233-253 (1998).

⁹ Raczek, J – A general-method for the determination of molecular-weight distributions by dynamic light-scattering. *Eur. Polym. J.* **18**(10): 847-861 (1982).

¹⁰ Dolomatov, MY, Khashper, LM and Kuzmina, ZF – Spectroscopic method for determination of the average molecular-weight. *Chem. Technol. Fuels Oils-Engl*, **27**(7-8): 401-403 (1991).

- ¹¹ Hatada, K, Ute, K, Kashiya, M, Imanari, M – Direct determination of molecular-weight and its distribution by the absolute calibration method using the online GPC NMR. *Polym. J.* **22**(3): 218-222 (1990).
- ¹² Fields, TB – Rapid screening method for molecular weight determination of proteins, peptides and viruses. *Abstr. Pap. Amer. Chem. Soc.* **224**, 14 (2002).
- ¹³ Guan, Y, Huang, XD, Xiao, CP, Wan, GX, Zhang, WC, Peng, YX, Ji, PJ, Xu, J and Ye, ML – A new method for determination of molecular weight of macromonomer containing methylacryloyloxy end group. *Acta Polym. Sin.* **2**, 190-193 (2000).
- ¹⁴ Kubota, I and Tsugita, A – A new semi-empirical method for the determination of the subunit molecular-weight of a protein. *Eur. J. Biochem.* **106**(1): 263-268 (1980).
- ¹⁵ Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE – The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).
- ¹⁶ Svergun, DI, Barberato, C and Koch, MHJ – CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.* **28**, 768-773 (1995).
- ¹⁷ Neves-Ferreira AG, Perales J, Fox JW, Shannon JD, Makino DL, Garratt RC, Domont GB – Structural and functional analyses of DM43, a snake venom metalloproteinase inhibitor from *Didelphis marsupialis* serum. *J Biol Chem.* **277**(15):13129-37 (2002).
- ¹⁸ Svergun, DI – Mathematical methods in small-angle scattering data analysis *J. Appl. Cryst.* **24**, 485-492 (1991).
- ¹⁹ Svergun, DI – Determination of the regularization parameter in indirect-transform methods using perceptual criteria *J. Appl. Cryst.* **25**, 495-503 (1992).
- ²⁰ Fischer H, Dias SM, Santos MA, Alves AC, Zanchin N, Craievich AF, Apriletti JW, Baxter JD, Webb P, Neves FA, Ribeiro RC and Polikarpov I – Low resolution structures of the retinoid X receptor DNA-binding and ligand-binding domains revealed by synchrotron X-ray solution scattering. *J Biol Chem.* **278**(18):16030-8 (2003).
- ²¹ Trapani S, Linss J, Goldenberg S, Fischer H, Craievich AF and Oliva G – Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from *Trypanosoma cruzi* at 2 Å resolution. *J Mol Biol.* **313**(5):1059-72 (2001).
- ²² Borges JC, Fischer H, Craievich AF, Hansen LD and Ramos CH – Free human mitochondrial GrpE is a symmetric dimer in solution. *J Biol Chem.* **278**(37):35337-44 (2003).
- ²³ Cabrera R, Fischer H, Trapani S, Craievich AF, Garratt RC, Guixe V and Babul J – Domain motions and quaternary packing of phosphofructokinase-2 from *Escherichia coli* studied by small angle x-ray scattering and homology modeling. *J Biol Chem.* **278**(15):12913-9 (2003).
- ²⁴ Kuntz, I D Jr and Kauzmann, W – Hydration of proteins and polypeptides. *Adv. Protein Chem.* **28**, 239-345 (1974).
- ²⁵ Squire, PG and Himmel, ME – Hydrodynamics and protein hydration. *Arch. Biochem. Biophys.* **196**, 165-177 (1979).
- ²⁶ Gekko, K and Noguchi, H – Compressibility of globular proteins in water at 25°C. *J. Phys. Chem.* **83**, 2706-2714 (1979).
- ²⁷ Andersson KM and Hovmöller S – The average atomic volume and density of proteins. *Z. Kristallogr.* **213**(7-8): 369-373 (1998).
- ²⁸ Quillin ML and Matthews BW – Accurate calculation of the density of proteins. *Acta Cryst. D* **56**: 791-794 (2000).

4	<i>Estudos estruturais de receptores nucleares</i>	4 - 1
4.1	Introdução	1
4.2	Envelope de baixa resolução do receptor de retinoide X (RXR)	2
4.2.1	Mecanismos de funcionamento dos receptores nucleares	2
4.2.2	O RXR pode atuar sinergisticamente	3
4.3	Resultados	5
4.3.1	Medidas de SAXS dos dímeros e tetrâmeros de hRXR α DBD-LBD	5
4.3.2	Determinação ab initio do envelope da proteína	9
4.3.3	Estrutura do dímero e do tetrâmero de hRXR α DBD-LBD	14

4 ESTUDOS ESTRUTURAIS DE RECEPTORES NUCLEARES

4.1 Introdução

Os receptores nucleares são encontrados em todo o reino animal, abrangendo desde os insetos até o homem. Seqüenciamento do genoma humano confirmou que homem possui 48 receptores nucleares. A superfamília de receptores nucleares é constituída por cerca de 150 receptores, os quais mediam a ação dos hormônios lipofílicos (esteróides, retinóides, vitamina D e tireoidianos), de metabólitos endógenos (ácidos biliares e graxos, eicosanóides e oxiesterol), e de xenobióticos. A ação destes receptores se dá através da sua interação com a região promotora de genes alvo da ação hormonal.

Os receptores nucleares, como todos os membros da superfamília, são proteínas modulares, com três domínios principais. O domínio amino-terminal, que não é conservado entre os membros da superfamília, age estimulando a transcrição através da função de transativação AF-1. O DBD (DNA binding domain), situado na região central da proteína é responsável pela ligação do receptor ao DNA, enquanto o LBD (ligand binding domain) é responsável pela ligação do receptor ao hormônio. O LBD é essencial para a homodimerização e heterodimerização do receptor, além de sua associação com proteínas co-reguladoras, co-repressoras e co-ativadoras, que são essenciais para repressão ou ativação, respectivamente, da transcrição dos genes regulados pelos receptores.

No seu conjunto, os receptores nucleares estão envolvidos em praticamente todas as funções fisiológicas do organismo, e o grupo de agonistas e antagonistas (glucocorticóides, estrogênios, antiestrogênios, vitamina D, retinóides, androgênios, antiandrogênios e

tiroxina) que neles atuam constitui hoje a classe de fármacos mais utilizada na prática médica. Como exemplo, os estrogênios são utilizados para o tratamento da osteoporose, os anti-estrogênios para o tratamento de câncer de mama, os retinóides para o tratamento de leucemias, a vitamina D para a osteodistrofia renal e os glucocorticóides para o controle de rejeição de transplantes e de doenças imunológicas como a artrite reumatóide. Áreas tais como oncologia, dermatologia e obesidade são somente alguns dos poucos exemplos de aplicabilidade terapêutica.

Isto faz desta superfamília de proteínas um dos mais importantes alvos para o desenvolvimento de medicamentos, sendo o desenvolvimento de moléculas sintéticas com o intuito de modular, antagonizar ou bloquear uma expressão genética uma das principais metas das indústrias farmacêuticas.

4.2 Envelope de baixa resolução do receptor de retinoide X (RXR)

Retinóides são derivados da vitamina A e apresentam uma multiplicidade de efeitos biológicos nas células de mamíferos. Esses componentes atuam como peças chaves na diferenciação e crescimento celular e têm papel crítico em processos como a visão e a reprodução. Eles também são potentes inibidores da carcinogênese, sendo utilizados no tratamento e quimioprevenção de diversos tipos de cânceres. Os retinóides regulam a expressão de genes mediante a ação do receptor nuclear de ácido retinóico formas *all-trans* e *9-cis* (RARs) e do receptor específico para forma *9-cis* (RXR). Os receptores ligam seqüências curtas de DNA denominadas elementos responsivos do receptor do hormônio de ácido retinóico.

4.2.1 Mecanismos de funcionamento dos receptores nucleares

O receptor RXR, assim como todos os membros da superfamília de receptores nucleares, contém os domínios de ligação ao DNA (DBD) e de ligação ao ligante (LBD) (veja figura abaixo).

O mecanismo molecular geral de funcionamento dessas proteínas está sendo mostrado abaixo. À esquerda temos o LBD do receptor que, como acontece em diversos receptores, quando não está ligado ao hormônio se liga a proteínas corepressoras (vermelho). O DBD do receptor contata o DNA na posição que precede a sequência codificadora do gene regulado. No meio: A ligação do hormônio induz uma mudança conformacional que resulta na expulsão do corepressor. À direita: Um co-ativador liga-se ao receptor e dá início a transcrição. O receptor RXR quando homodimerizado, ao contrário de muitos receptores nucleares, não se liga ou se liga muito fracamente a corepressores.

4.2.2 O RXR pode atuar sinergisticamente

Uma série de receptores ligam-se ao DNA somente quando formam heterodímeros com RXR. Em tal situação, cada membro do dímero tem parte na regulação da via metabólica ativada. A figura abaixo é um esquema da ativação sinérgica do heterodímero RAR-RXR.

A: Ambos receptores estão na forma apo (ou sem ligante). Um repressor (tijolo) liga-se a RAR (branco) e impede a transcrição. B: RAR é ativado (hormônio em preto). O repressor é substituído por um coativador que permite a transcrição. C: RXR (cinza) é ativado, mas o repressor ainda permanece ligado ao RAR impedindo a transcrição. D: Ambos receptores recrutam um ativador, corroborando para uma resposta transcripcional mais efetiva.

A estrutura cristalográfica do receptor RXR completo, ou mesmo contendo somente os domínios DBD e LBD, ainda não foi determinada. De maneira a ganhar informação

sobre o envelope do receptor RXR humano isoforma α (hRXR α) na sua forma dimérica e tetramérica, e contendo somente os domínios DBD e LBD, realizamos medidas de SAXS desta proteína em solução. Para obtenção da proteína, realizamos superexpressão em sistema heterólogo (*E. Coli*) da proteína recombinante, tendo sido ela expressa com uma cauda de histidina e purificada por cromatografia.

Também com os dados coletados de espalhamento de raios X a baixos ângulos foi possível propor modelos da interface de interação do homodímero e homotetrâmero do RXR. Utilizamos as estruturas cristalográficas dos domínios isolados, já determinados experimentalmente, como modelos de partida e os comparamos com modelos obtidos *ab initio* dos dados de SAXS.

A coleta de dados de SAXS como usual foi realizada na linha de SAXS do LNLS (Campinas, Brasil) usando um detetor 1D sensível a posição [1]. Foram medidas as proteínas nas concentrações de 2, 3.1, 6.3 e 14mg/mL¹ e em um comprimento de onda de $\lambda=0.138$ nm para as distâncias amostra-detector de 632.5 e 1345.2mm cobrindo um intervalo do módulo do vetor de espalhamento entre $0.1 < q < 6 \text{ nm}^{-1}$ ($q=4\pi\sin\theta/\lambda$, onde 2θ é o ângulo de espalhamento). As curvas de espalhamento foram coletadas em intervalos de 30-60s para monitorar eventuais instabilidades do feixe e deterioração da amostra. Finalmente, os dados foram normalizados pela intensidade do feixe incidente e corrigidos pela curva de resposta do detetor.

O espalhamento do tampão foi subtraído e as curvas resultantes foram normalizadas pela concentração das proteínas em solução. A função de distribuição de distâncias $p(r)$ e o raio de giro R_g foram obtidos pelo programa de transformada indireta de Fourier GNOM [2]. As massas moleculares dos oligômeros de hRXR α em solução foram estimadas por comparação da intensidade $I(0)$ extrapolada da proteína de referência (albumina de soro bovino) que tem uma massa de 66 kDa. Antes de fazermos a reconstrução da forma das proteínas, subtraímos uma constante das curvas de espalhamento para que a região de altos ângulos decaísse com q^{-4} obedecendo à lei de Porod para partículas homogêneas [3]. O valor da constante é obtido automaticamente da parte externa da curva por uma aproximação linear nas coordenadas de $q^4 I(0)$ por q^4 pelo programa de reconstrução de forma chamado DAMMIN [4]. Este procedimento diminui a contribuição do espalhamento devido à estrutura interna da proteína e fornece uma aproximação do espalhamento da forma da mesma (redução do problema a duas densidades eletrônicas bem definidas, a do solvente e da partícula espalhadora). As heterogeneidades de densidade eletrônica de uma proteína contribuem principalmente para altos ângulos de espalhamento.

Como as $p(r)$'s não apresentavam nenhuma parte negativa, que é um indício da ausência de efeitos de interferência nas curvas de espalhamento produzido por correlações

¹ Entre 2-3mg/mL as proteínas eram predominantemente diméricas. Com 6mg/mL houve misturas de espécies diméricas e tetraméricas e acima de 6mg/mL as proteínas eram tetraméricas. Todas as proteínas foram preparadas pela Dr. Sandra Martha Gomes Dias.

espaciais, concluímos que as soluções estudadas foram suficientemente diluídas a fim dos efeitos de interferência não serem perceptíveis. O fator de estrutura $S(q)$ para partículas altamente anisotrópicas, como neste caso, que descreve correlações interpartículas deve ser próximo a 1 mesmo para altas concentrações de proteínas [5].

Utilizamos os domínios já conhecidos da hRXR α , o LBD e o DBD que foram obtidos do *Protein Data Bank* (PDB ID:1G1U e 2NLL) [6,7]. As posições relativas destes domínios (LBD e DBD) foram encontradas rotacionando-os respectivamente e sobrepondo-os da melhor forma possível com a estrutura de baixa-resolução obtida pelos procedimentos *ab initio* descritos no capítulo 2. Utilizamos o programa MASSHA [8] para este fim. Em seguida calculamos a curva de espalhamento, dimensão máxima (D_{max}) e raio de giro (R_g) do modelo de coordenadas atômicas com o auxílio do programa CRY SOL levando em consideração a camada de hidratação [9]. Finalmente utilizamos o programa SUPCOMB [10] para sobrepor os modelos de baixa resolução *ab initio* com o modelo cristalográfico.

4.3 Resultados

4.3.1 Medidas de SAXS dos dímeros e tetrâmeros de hRXR α DBD-LBD

As formas diméricas e tetraméricas do receptor hRXR α eluíram separadamente em cromatografia de troca iônica utilizando-se a resina Q-sepharose High Performance (Amersham) (amostras 1 e 2, referentes aos picos 1 e 2)[11], conforme avaliado por experimento de *Dynamic Light Scattering*, exclusão por tamanho e SAXS.

As proteínas destes picos foram analisadas por eletroforese do tipo SDS-PAGE, tendo sido verificado que seu peso molecular é consistente com o esperado para o hRXR α (contendo somente os domínios DBD e LBD), além de que a pureza foi visualmente avaliada como sendo superior a 95%. A fim de determinar o envelope das espécies oligoméricas presentes nestes picos realizamos experimentos de SAXS e os resultados podem ser vistos na Figura 1.

Figura 1: Curvas de espalhamento experimental e ajustes dos procedimentos de modelagem da hRRX α Δ AB. (a) dímero a 3.1mg/mL e (b) tetrâmero a 14mg/mL. Acima um gráfico em $\log I$ por q focando os ajustes para q altos. Abaixo o mesmo gráfico focando a região de baixos q 's em escala linear com um inserto ilustrando a validade da lei de Guinier ($\log I$ por q^2). (1) Dados experimentais; (2) Dados experimentais após subtração de uma constante como descrito no texto; (3) Intensidade de espalhamento resultante do modelo gerado pelo DAMMIN; (4) Intensidade de espalhamento gerado pelo modelo do envelope (SASHA).

Os dados estruturais obtidos por estes ajustes podem ser vistos na Tabela 1.

Tabela 1 – Principais dados experimentais e de modelagem

Parâmetro/Amostra	hRXR α dímero			hRXR α tetrâmero		
	Exp. ^a	Env. ^b	DAM ^c	Exp.	Env.	DAM ^c
Massa (kDa)	79(5)			185(10)		
D _{max} (nm)	11.5(5)	12.0	11.0(2)	19(1)	20	18(2)
R _g (nm)	3.38(5)	3.41	3.36(1)	5.27(4)	5.27	5.2(2)
V(nm ³)	105	151	119(5) ^e	246	338	285(3)
Parâmetros livres	6.7 ^d	13 ^e	1359	14.1 ^d	31 ^f	486
Discrepância χ		0.57	1.32(1)		0.99	1.30(2)
Resolução ^g (nm)	3.0	3.3	3.0	1.05	2	2.1

^a Exp., calculados a partir dos dados experimentais.

^b Env., parâmetros dos modelos de envelope (SASHA).

^c DAM, parâmetros da média de 20 modelos de DAMs.

^d Número de canais de Shannon $N_s = D_{max}(q_{max} - q_{min})/\pi$

^e Simetria 2 imposta

^f Simetria 222 imposta

^g Dada por $R = 2\pi/q_{max}$

Os oligômeros de hRXR α ΔAB das amostras 1 e 2 correspondem as massas moleculares de 79 e 185kDa, respectivamente. Isto foi determinado experimentalmente por SAXS comparando as intensidades extrapoladas para ângulo zero I(0) com as da albumina de soro bovino de 66kDa (veja tabela acima). Isto foi um indício que os picos 1 e 2 são respectivamente, dímeros e tetrâmeros de hRXR α ΔAB. Evidência adicional para o estado oligomérico das proteínas vem dos resultados da análise da técnica de cromatografia por exclusão de tamanho. Por esta técnica determinou-se que o raio de Stokes das proteínas foram de 4.48nm e 4.66nm para as proteínas do pico 1 nas concentrações de 5.6mg/ml e 12.3mg/ml respectivamente. O raio de Stokes correspondente à proteína da fração 2 varia entre 5.2 – 5.7nm para proteínas nas concentrações entre 3 – 11mg/ml.

Cromatogramas de exclusão de tamanho representativos foram ilustrados na Figura 2. Esses resultados concordam com os obtidos pelos envelopes de baixa resolução de SAXS. O valor calculado para o raio de Stokes para o dímero é de 4.15nm e 5.6nm para o tetrâmero [12]. Também fizemos experimentos de espalhamento de luz dinâmica das proteínas provenientes dos picos 1 e 2 nas concentrações de 1.2mg/ml e 7 mg/ml que resultaram respectivamente nos raios de Stokes de 4.42 e 4.3nm para o pico 1 e 5.13 e 5.42 para o pico 2. Considerando todas estas evidências experimentais, concluímos que os picos

1 e 2 representam na sua grande maioria populações diméricas e tetraméricas da hXR α Δ AB.

As medidas de SAXS das proteínas do pico 1 com concentrações entre 3 – 14mg/ml resultaram em curvas idênticas as proteínas do pico 2 de concentração similar. Isto sugere que as populações diméricas do pico 1 podem se converter em populações tetraméricas para concentrações maiores o que é consistente com a idéia anterior que os dímeros e tetrâmeros estão em equilíbrio dinâmico sendo o mesmo influenciado pela sua concentração [13,14].

Figura 2: Cromatogramas de exclusão de tamanho representativos dos picos 1 e 2 das frações de proteína obtidos pela coluna de Q-sefarose.

Os valores experimentais de $D_{\max}$ e R_g , 19 nm e 5.27 nm para o tetrâmero e 11.5 nm e 3.38 nm para o dímero, respectivamente, sugerem que a proteína nos dois picos é alongada (veja tabela 1). A própria $p(r)$ como pode ser visto na Figura 3 é típica de proteínas alongadas [15]. Pode ser observado claramente que o $D_{\max}$ do tetrâmero (19 nm) é 7.5 nm maior que o do dímero (11.5 nm), o que sugere o tetrâmero ser bem mais alongado que o dímero.

Figura 3: Função de distribuição de distâncias dos dímeros (triângulo) e tetrâmeros (círculos) da hRXR α Δ AB.

4.3.2 Determinação *ab initio* do envelope da proteína

A determinação da forma do dímero e tetrâmero do hRXR α foi realizado utilizando dois métodos *ab-initio*, um de expansão de múltiplos (SASHA) e outro usando DAMMIN como descrito no capítulo 2.

No primeiro método (SASHA), ajustamos uma função de envelope aos dados experimentais partindo de uma forma inicial esférica. O envelope foi representado por harmônicos esféricos com $L=4$ (13 parâmetros independentes) e $L=10$ (31 parâmetros independentes) para o dímero e tetrâmero, respectivamente. Estes valores decorrem da análise de parâmetros independentes que se pode extrair de uma curva de SAXS (canais de Shannon) e simetria imposta para o dímero ($N_s=6.7$ e simetria 2) e tetrâmero ($N_s=14.1$ e simetria 222) respectivamente.

A solução dos envelope de baixa resolução para os parâmetros acima utilizados é única e esta ilustrada na parte superior das Figura 4 e Figura 5.

No segundo método de reconstrução de forma *ab initio*, utilizamos o programa DAMMIN também impomos condições de simetria para diminuir o número de parâmetros livres e reduzir o tempo de processamento. Assim impomos um eixo de ordem 2 para o dímero e três eixos de ordem 2 para o tetrâmero.

Na reconstrução da forma tanto do dímero quanto tetrâmero do hRXR α partiu-se de uma esfera inicial com diâmetro máximo definido pela $p(r)$ sendo o $D_{\max} = 11.5$ nm para o dímero e $D_{\max} = 20.0$ nm para o tetrâmero. O raio das esferas que preencheram o espaço de busca inicial foi de $r_a = 0.375$ nm para o dímero e $r_a = 0.575$ nm para o tetrâmero o que resultou em $N_{\text{DAM}} = 2718$ para o dímero e $N_{\text{DAM}} = 3884$ para o tetrâmero. Como foi imposta simetria, o número de parâmetros livres final foi reduzido para $N_{\text{red}} = 2718/2 = 1359$ para o dímero e $N_{\text{red}} = 3884/8 = 486$ para o tetrâmero. Outra restrição imposta que reduziu ainda mais o número de parâmetros livres foi a imposição de forma prolata o que é fácil de observar na $p(r)$ (Figura 3) que é típico para estruturas alongadas tipo cilindros (veja exemplos C, E e H da figura 6 do capítulo 2).

Os modelos finais tanto do dímero quanto tetrâmero do hRXR α continham em torno de 395 ± 10 DAMs e 265 ± 5 DAMs respectivamente. Uma média de 20 modelos está representada na parte superior e inferior das Figura 4 e Figura 5. Utilizamos o programa SUPCOMB para fazer a média dos 20 modelos, e quando um modelo não era parecido com os outros era descartado. Uma representação de 5 modelos gerados de forma *ab initio* tanto para dímero como tetrâmero, se encontra nas Figura 6 e Figura 7, respectivamente.

As concentrações de proteínas utilizadas nas medidas de SAXS de 3.1 mg/ml (dímero de hRXR α) e 14 mg/ml (tetrâmero de hRXR α) permitiram obter modelos finais com resolução de 3.0 nm e 2.1 nm, respectivamente (Tabela 1). Esta resolução não permite uma determinação sem ambigüidade da posição espacial dos elementos de estrutura secundária, porém a posição relativa dos domínios cristalográficos foi obtida.

Figura 4: Envelope de baixa resolução do dímero do hRXR α Δ AB obtidos utilizando o programa SASHA (sólido) e DAMMIN (bolinhas) ajustando as curvas de espalhamento experimental. (em cima) dímeros ab initio e (baixo) modelo proposto. Espaçamento da grade é de 2nm. (b) e (c) é o mesmo modelo rotacionado 90° nos eixos z e y, respectivamente.

Figura 5: Envelope de baixa resolução do tetramero do hRXR α Δ AB obtidos utilizando o programa SASHA (sólido) e DAMMIN (bolinhas) ajustando as curvas de espalhamento experimental. (em cima) dímeros *ab initio* e (baixo) modelo proposto. Espaçamento da grade é de 2nm. (b) e (c) é o mesmo modelo rotacionado 90° nos eixos z e y, respectivamente.

A forma reconstruída pelo SASHA para o dímero está ilustrado na Figura 4 e o ajuste aos dados experimentais na Figura 1. Sobreposto temos um modelo derivado de

fragmentos cristalográficos (LBD e DBD) de alta resolução do hRXR α que puderam ser encaixados sem ambigüidade no envelope do dímero. Fizemos a mesma análise para o tetrâmero e o seu envelope restaurado tanto pelo SASHA quando DAMMIN está ilustrado na Figura 5 e o ajuste aos dados experimentais também está ilustrado na Figura 1.

Em relação ao tetrâmero do hRXR α Δ AB, foi possível encaixar um tetrâmero de hRXR α LBD e quatro fragmentos do hRXR α DBD. Para encontrar um modelo para o tetrâmero, ajustamos dois dímeros, rotacionando e deslocando-os em torno de um eixo de simetria, variando a posição relativa entre os DBDs e LBDs. Para isto foi utilizado o programa MASSHA até encontrarmos o melhor ajuste ao modelo *ab initio*.

Calculamos o raio de Stokes para o dímero e tetrâmero destes modelos propostos que resultaram em 4.15nm e 5.6nm, respectivamente [12].

Figura 6: 5 exemplos de envelopes de baixa resolução do dímero do hRXR α Δ AB obtidos utilizando o programa DAMMIN (bolinhas) e a média dos mesmos na última coluna. Segunda e terceira linha mesmo modelos rotacionados em 90° nos eixos z e y, respectivamente.

Figura 7: 5 exemplos de envelopes de baixa resolução do tetrâmero do hRXR α Δ AB obtidos utilizando o programa DAMMIN (bolinhas) e a média dos mesmos na última coluna. Segunda e terceira linha mesmo modelos rotacionados em 90° nos eixos z e y, respectivamente.

4.3.3 Estrutura do dímero e do tetrâmero de hRXR α DBD-LBD

Os modelos de SAXS obtidos pelos métodos *ab initio* (veja Figura 4) indicam que o dímero tem um formato de “U”, com dois domínios claramente identificáveis, ou seja o LBD e o DBD. Os dois LBDs formam a base do “U” e cada LBD está conectado ao longo de uma extensa α -hélice a seu respectivo DBD que está no extremo do “U” (veja Figura 4 segunda linha)

Os LBDs dimerizam pela mesma interface descrita na estrutura cristalográfica do hRXR α [7,16] envolvendo contatos entre as α -hélices H9 e H10. As duas α -hélices que conectam os LBDs aos DBDs e o ângulo formado entre eles é de aproximadamente 10° gerando uma distância entre DBDs da ordem de 2nm que é comparável com os sítios de ligação do elemento responsivo DR1 (3nm). Porém a orientação dos DBDs é face a face, e para que eles pudessem se ligar ao DR1 um teria que rodar 180°. Assumindo que esta rotação pode ocorrer, o modelo é consistente com que se sabe sobre o papel do hRXR α Δ AB homodímero se ligar a seu elemento responsivo de DNA (DR1).

O tetrâmero do hRXR α Δ AB tem uma forma oblata e mais alongada com uma forma tipo um "X" (veja Figura 5). Tal como no dímero, foi possível localizar os domínios correspondentes aos DBDs e LBDs. Porém neste caso, o LBD é um tetrâmero localizado no centro da molécula e os quatros DBDs estão conectados ao mesmo. A posição relativa dos LBDs é consistente com resultados cristalográficos anteriores e estudos de SAXS que mostram que o tetrâmero é composto por um dímero de dímeros onde os LBDs de um dímero estão num arranjo de topo a topo com o outro dímero. No ajuste dos dados cristalográficos com os modelos *ab initio* permitimos a rotação relativa entre os homodímeros em torno de seu eixo de simetria e também o deslocamento entre eles. Este modelo é consistente com a idéia já discutida para o dímero que eles utilizam a interface composta pelas α -hélices H9 e H10. A superfície de tetramerização também é consistente com artigos previamente publicados (onde existe a interação entre as α -hélices H3-H3, H11-H11 e H12 com a região hidrofóbica H3-H5 do dímero adjacente [17]). Afirmar mais que isto não é possível devida a baixa resolução obtida pelos modelos de SAXS, porém percebe-se uma nítida diferença entre o dímero e o tetrâmero (veja segunda linha da Figura 4 e Figura 5). As distâncias entre os DBDs de cada dímero do tetrâmero são de 6nm, isto sugere que os DBDs do tetrâmero teriam dificuldade de se ligarem ao elemento responsivo DR1, a não ser que sofressem uma mudança conformacional significativa. cremos esta ser uma razão pelo qual o hRXR α Δ AB tetrâmero não se ligar muito bem ao DR1.

Certamente não podemos afirmar que esta é a única solução que os domínios individuais podem assumir. Outras combinações seriam possíveis com ajustes de mesma qualidade, porém esta solução deve estar bem próxima da realidade. As dimensões máximas obtidas da $p(r)$ impõe que as proteínas sejam alongadas, impedindo que se posicionem os domínios DBD e LBD da hRXR α em outras posições. Além do mais, o raio de Stokes obtido pelo modelo de baixa resolução proposto, concorda também com dados obtidos pela técnica de espalhamento de luz dinâmico e cromatografia por exclusão de tamanho.

Estes resultados de baixa resolução representam os primeiros modelos para um receptor nuclear contendo tanto o LBD como DBD e a comparação entre os modelos do hRXR α Δ AB tetramérico e dimérico nos respondem algumas questões relacionadas ao

porque de uma forma se ligar melhor com um elemento responsivo que a outra. Como já discutido acima, pelo fato dos DBDs do dímero estarem mais próximos uns dos outros, eles podem se ligar ao DR1 sem se submeterem a mudanças conformacionais muito grandes e de fato, os homodímeros ligam-se preferencialmente ao DR1. Claro que um dos DBDs teria que rotacionar em torno de 180° para essa ligação ocorrer. Alguns fatos levam-nos a crer que os LBDs influenciam o posicionamento dos DBDs. De um lado, a afinidade de ligação do homodímero de RXR a elementos DR1 é aumentada pelo hormônio 9-*cis*, ou seja, a mudança conformacional induzida pela ligação ao ligante deve de alguma maneira melhorar o ajuste entre o homodímero e o DR1.

Em contraste ao dímero que adota uma forma tipo um “U”, o tetrâmero com sua forma de um “X” e os DBDs espaçados em mais de 6nm dificilmente se ligaria ao DR1 a não ser que sofresse uma mudança conformacional significativa. Podemos ver que o arranjo dos elementos de estrutura terciária é sensível a alterações na estrutura quaternária.

O equilíbrio dinâmico entre os oligômeros de RXR é influenciado pela concentração do receptor, ligante e DNA [13, 14]. Desta forma, é plausível que o armazenamento do RXR em forma tetramérica *in vivo* seja um mecanismo para a célula armazenar o excesso de receptores até que funções fisiológicas que exijam homodímeros ou heterodímeros de RXR sejam necessárias.

O mecanismo estrutural responsável pela inativação dos receptores, revelado por estudos cristalográficos do tetrâmero do LBD, envolve a interação da hélice H12 dos monômeros de um dímero com as regiões das hélices H3-H5 dos monômeros do dímero adjacente. Esta interação compete pela interação de coativadores, a qual ocorre na mesma região.

Enquanto nossos estudos revelaram um envelope em baixa resolução do tetrâmero do hRXR α Δ AB o qual corresponde ao esperado por estudos cristalográficos e biológicos, remanesce a possibilidade do RXR formar tetrâmeros biologicamente ativos em uma configuração ainda diferente. Porém, nosso estudo mostrou pela primeira vez um modelo da hRXR α Δ AB de baixa resolução que contém mais de um domínio. Além disto, nossos estudos oferecem uma sugestão do porque do dímero deste receptor possuir maior afinidade de ligação que o tetrâmero a elementos do tipo DR1. Estes resultados indicam que elementos de estrutura quaternária podem influenciar a estrutura terciária, ao mesmo tempo em que sugere que alterações em um dos domínios (LBD) pode levar a mudanças no outro domínio (DBD), alterando suas propriedades de interação.

¹ Kellermann, G, Vicentin, F, Tamura E, Rocha, M, Tolentino, H, Barbosa, A, *et al.* – The small-angle X-ray scattering beamline at the Brazilian synchrotron light laboratory *J. Appl. Crystallog.* 30:880-883 (1997).

- ² Svergun, DI – Determination of the regularization parameter in indirect transform methods using perceptual criteria *J. Appl. Crystallog.* **25**:495-503 (1992).
- ³ Porod, G – in Small-angle X-ray Scattering (Glatter, O and Kratky, O, Eds) pp. 17-51, *Academic Press*, London (1982).
- ⁴ Svergun, DI – Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**:2879-2886 (1999).
- ⁵ Guinier, A and Fournet, G – Small-angle scattering of X-rays, *John Wiley & Sons*, NY (1955).
- ⁶ Rastinejad, F, Wagner, T, Zhao, Q and Khorasanizadeh, S – Structure of the RXR-RAR DNA binding complex on the retinoid acid response element DR1. *EMBO J.* **19**: 1045-1054 (2000).
- ⁷ Bourguet, W, Ruff, M, Chambon, P, Gronemeyer, H and Moras, D – Crystal structure of the ligand-binding domain of the human nuclear receptor RXR. *Nature* **375**:377-382 (1995).
- ⁸ Konarev, PV, Petoukhov, MV and Svergun, DI – MASSHA- a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. *J. Appl. Cryst.* **34**: 527-532 (2001).
- ⁹ Svergun, DI, Barberato, C and Koch, MH – CRY SOL- a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**:768-773 (1995).
- ¹⁰ Kozin MB and Svergun DI – Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* **34**, 33-41 (2001).
- ¹¹ Fischer H, Dias SM, Santos MA, Alves AC, Zanchin N, Craievich AF, Apriletti JW, Baxter JD, Webb P, Neves FA, Ribeiro RC, Polikarpov I – Low resolution structures of the retinoid X receptor DNA-binding and ligand-binding domains revealed by synchrotron X-ray solution scattering. *J Biol Chem.* **278**(18):16030-8 (2003).
- ¹² Garcia de la Torre, J, Huertas, ML and Carrasco, B – Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719-730 (2000).
- ¹³ Kersten, S, Pan, L, Chambon, P, Gronemeyer, H and Noy, N – Role of ligand in retinoid signaling. 9-cis-retinoic acid modulates oligomeric state of the retinoid X receptor. *Biochemistry* **34**:13717-18721 (1995).
- ¹⁴ Chen, ZP, Iyer, J, Bourguet, W, Held, P, Mioskowski, C, Lebeau, L, Chambon, P and Gronemeyer, H – Ligand- and DNA- induced dissociation of RXR tetramers. *J. Mol. Biol.* **275**:55-65 (1998).
- ¹⁵ Feigin, LA and Svergun, DI – Structure analysis by Small-angle X-ray and Neutron Scattering, pp. 83-87, Plenum Press, New York (1987).
- ¹⁶ Egea, PF, Mitschler, A, Rochel, N, Ruff, M, Chambon, P and Moras, D – Crystal structure of the human RXR alpha ligand-binding domain bound to its natural ligand: 9-cis retinoic acid. *EMBO J.* **19**:2592-2601 (2000).
- ¹⁷ Gampe, RT, Jr, Montana, VG, Lambert, MH, Wisely, GB, Milburn, MV and Xu, HE – Structural basis for autorepression of retinoid X receptor by tetramer formation and the AF-2 helix. *Genes Dev.* **14**:2229-2241 (2000).

5	<i>Estudos estruturais de outras proteínas em solução</i>	5 - 1
5.1	Introdução	1
5.2	Phosphoenolpyruvate carboxykinase (PEPCK)	2
5.2.1	Motivação do estudo da proteína PEPCK em solução	2
5.2.2	Resultados experimentais, análise de dados e conclusões	5
5.3	Mudanças conformacionais e estados de agregação na proteína PFK-2	9
5.3.1	Introdução e objetivos	9
5.3.2	Resultados experimentais, análise de dados e conclusões	11
5.4	Estrutura da interleucina humana 22 (hIL-22) em solução	20
5.4.1	Introdução e objetivos	20
5.4.2	Resultados, discussões e conclusões das medidas de SAXS e DLS da hIL-22 em solução	20

5 ESTUDOS ESTRUTURAIS DE OUTRAS PROTEÍNAS EM SOLUÇÃO

5.1 Introdução

Neste capítulo serão descritos alguns estudos utilizando SAXS levando em conta a estrutura cristalográfica já conhecida da proteína, propondo-se modelos de ação quando a mesma se liga a um ligante.

A Phospho-enol-pyruvate carboxykinase (PEPCK), é uma enzima chave no metabolismo de açúcares do parasita *Trypanosoma cruzi* e tem sido considerada como um bom alvo para o desenvolvimento de uma nova droga anti-chagásica [1]. A estrutura cristalina recém resolvida da PEPCK do *T. cruzi* mostrou que seu enovelamento é muito similar ao enovelamento da PEPCK monomérica da *E. coli*, a únicaⁱ outra PEPCK com estrutura cristalina conhecida [2]. Porém, a análise dos contatos entre os monômeros da estrutura cristalina não forneceu uma informação clara sobre o estado atual da proteína em solução em condições fisiológicas.

ⁱ Única até meados de 2000, em 2001 um grupo publicou a estrutura da PEPCK humana a qual foi fundamental para estudos de afinidade de drogas computacional.

Os resultados de SAXS mostraram que a PEPCK da *T. cruzi* em solução é dimérica. Uma análise mais detalhada ainda mostrou que os dímeros estão arranjados de forma muito parecida a um arranjo específico dos monômeros (dímero tipo I) no estado cristalino. Realizaram-se também refinamentos de corpo rígido que indicaram a estrutura interna dos monômeros adotar uma conformação mais aberta, em contraste com a cristalina, que é fechada e ainda verificou-se que a posição e orientação relativa dos dois monômeros difere da do estado cristalino.

Descrever-se-á também a isoenzima fosfofrutoquinase (PFK-2), que é uma proteína dimérica da *E. coli* que é inibida pelo MgATP que se liga ao sítio alostérico e promove a tetramerização da proteína. Esta enzima é particularmente interessante para estudar o efeito de agregação molecular e também o de ligação de ligantes, pois ela somente existe como dímero ou tetrâmero e suas propriedades catalíticas e de ligação são bem conhecidas. Estudou-se a cinética de formação de tetrâmeros e se propôs um modelo da proteína na presença de Frutose e MgATP. Também mostrou-se que a temperatura leva a dissociação do tetrâmero.

Finalmente foram realizados alguns estudos com a interleucina humana 22 (hIL-22), que ilustram a importância da técnica de SAXS pois explica como dois monômeros de hIL-22 interagem e se empacotam formando uma estrutura dimérica muito parecida com outra interleucina já conhecida, a hIL-10. Os resultados mostram claramente como os dímeros se empacotam, sendo esse empacotamento completamente diferente do proposto no artigo de Nagem *et al.*[10]. Também descrever-se-á como a hIL-22 interage com seu receptor e propôs-se um modelo para o tetrâmero.

5.2 Phosphoenolpyruvate carboxykinase (PEPCK)

5.2.1 Motivação do estudo da proteína PEPCK em solução

A doença de Chagas humana (Descoberta por Carlos Chagas em 1909) acontece do sul dos Estados Unidos até a Patagônia, onde põe sob risco mais de 60 milhões de pessoas de 18 países americanos e afeta cerca de 18 milhões de indivíduos. Transmitida principalmente por um inseto sugador de sangue, o triatomíneo vulgarmente conhecido por "barbeiro", "chupão" ou "fincão", pode também veicular-se de homem a homem através de transfusão de sangue, por via placentária, por transplantes de órgãos, por acidentes em laboratório e por outras vias excepcionais, como a oral. A doença humana transcorre numa fase inicial aguda, caracterizada por febre, muitos parasitas circulantes e poucas semanas de duração, seguindo-se uma fase crônica, afebril, com poucos parasitas no sangue e causadora de importantes alterações cardíacas em cerca

de 20-30% dos casos e digestivas (danos principalmente no esôfago e intestino grosso) em aproximadamente 10% dos pacientes.ⁱⁱ

Ao picar uma pessoa infectada pelo parasita, geralmente à noite e na região da face, o barbeiro torna-se portador dos tripanossomos, que se reproduzem em seu intestino. Ao picar outro indivíduo sadio, o inseto defeca e elimina suas fezes contaminadas. A vítima, ao coçar o local da picada, espalha as fezes do mosquito sobre o ferimento. Dessa maneira, os parasitas penetram nas células da pele, atingindo a circulação sanguínea. Nessa etapa, chamada de fase aguda, não há manifestação de sintomas, na maioria dos casos. Quando ocorrem, a vítima apresenta forte reação local à picada e febre alta (veja Figura 1)ⁱⁱⁱ

Figura 1: Ciclo da transmissão da doença de Chagas pelo “barbeiro”.

Se não for diagnosticada na fase aguda, quando ainda tem cura, a doença evolui para a forma crônica. Os tripanossomos instalam-se nos músculos humanos, especialmente no coração. Ao atingir e destruir fibras musculares, provocam insuficiência e arritmia cardíaca, que podem levar à morte. O sistema digestivo também pode ser afetado.

A prevenção consiste no saneamento básico, no combate ao agente transmissor e na melhoria das condições de habitação, já que o inseto costuma se abrigar nas frestas de paredes de barro ou madeira. No Brasil, os números de caso de internação causados pela doença vêm decrescendo. Segundo o Ministério da Saúde, foram 1.836 em 1990, 1.232 em

ⁱⁱ Texto extraído de <http://www.vertentes.com.br/chagas/index.htm>

ⁱⁱⁱ Texto e ilustração extraída de <http://www.netium.com.br/lcpa/Index.htm>

1994 e 966 em 1995. Não há estimativa do total de casos existentes por se tratar de uma doença que, no início, é assintomática.

O protozoário, causador da doença de Chagas, cataboliza a glucose com número reduzido de catabolitos, resumidamente succinato e L-alanina, mesmo sob condições aeróbicas ("fermentação aeróbica"), devido a dois fatores: ineficiência relativa da cadeia respiratória para a reoxidação do NADH durante o catabolismo ativo da glucose e devido à falta da dehidrogenase NADH e do sítio I de fosforilação [2].

A Phospho-enol-pyruvate carboxykinase (PEPCK), é uma enzima chave no metabolismo de açúcares do parasita *Trypanosoma cruzi* e tem sido considerada como um bom alvo para o desenvolvimento de uma nova droga anti-chagásica [1]. A estrutura cristalina recém resolvida da PEPCK do *T. cruzi* mostrou que seu enovelamento que é muito similar ao enovelamento da PEPCK monomérica da *E. coli*, a única outra PEPCK com estrutura cristalina conhecida até então [2]. Porém, a análise dos contatos entre os monômeros da estrutura cristalina não forneceu uma informação clara sobre o estado atual da proteína em solução em condições fisiológicas.

Esta enzima foi cristalizada por Stefano Trapani que resolveu a estrutura atômica da mesma [2]. A PEPCK cristaliza sob condições específicas no sistema cristalino ortorômbico ($P2_12_12_1$) com dimensões da cela unitária de $a=66\text{\AA}$, $b=107,6\text{\AA}$ e $c=179,1\text{\AA}$ com $\alpha=\beta=\gamma=90^\circ$. A unidade assimétrica contém 2 monômeros da PEPCK. Considerando as operações de simetria desse sistema cristalino, temos 8 monômeros por cela unitária como ilustrado na Figura 2 (somente 4 monômeros são vistos por motivos de clareza).

Figura 2: A PEPCK da *T. cruzi* cristalizou no sistema cristalino ortorômbico ($P2_12_12_1$) com dimensões da cela unitária de $a=66\text{\AA}$, $b=107,6\text{\AA}$ e $c=179,1\text{\AA}$ com $\alpha=\beta=\gamma=90^\circ$. A unidade assimétrica contém 2 monômeros da PEPCK. Considerando as operações de simetria desse sistema cristalino, temos 8 monômeros por cela unitária (somente 4 monômeros são vistos por motivos de clareza).

O problema apresentado foi descobrir se, em solução, a PEPCK se encontra numa configuração monomérica ou dimérica e, se fosse dimérica, na configuração AB (dímero II) ou BC (dímero I) como ilustrado na figura acima.

5.2.2 Resultados experimentais, análise de dados e conclusões

A amostra de PEPCK estudada estava concentrada a 8.0mg/mL numa solução contendo 1M NaCl e 20mM KH_2PO_4 a pH 6,7. A quantidade total da amostra utilizada na experiência foi 120 μL . A amostra foi centrifugada por 3 minutos em torno de 20000g e como nenhuma precipitação foi observada realizaram-se as experiências. As amostras foram encapsuladas numa cela de mica (espessura de 1mm).

A experiência de SAXS também foi realizada no LNLS. Realizaram-se três medidas sucessivas (cada uma com duração de 10min) para checar eventuais danos causados pela radiação, que por sua vez não foram observados.

Realizaram-se medidas de SAXS para determinar qual a configuração da PEPCK em solução (monomérica ou dimérica) e verificar eventuais mudanças conformacionais (movimentos de domínios ou orientações relativas) da proteína em solução com a determinada cristalograficamente.

As curvas teóricas de SAXS correspondentes aos monômeros e dímeros I (BC) e II (AB) da PEPCK em solução foram determinadas utilizando o programa CRY SOL. A Figura 3 mostra a curva experimental junto com o melhor ajuste das curvas teóricas. A qualidade do ajuste (medido pelo parâmetro χ) é claramente melhor para o dímero I. Os valores do χ diminuem de 8,02 no caso do monômero, para 2,76 no caso do dímero tipo II ao valor de 1,32 para o dímero tipo I. Devido a alta concentração de sal na solução, calculamos e fixamos a densidade eletrônica do solvente em $0,347 \text{ e}^-/\text{\AA}^3$. O melhor ajuste dos dados teóricos e experimentais para o dímero I resultaram num volume deslocado de $151.000 \text{\AA}^3$ e num $\delta\rho$ de $0,09 \text{ e}^-/\text{\AA}^3$ o que corresponde a uma camada de hidratação de 0,29 gramas de água por grama de proteína (0,29w/w).

Figura 3: Dados experimentais de SAXS da PEPCK da *T. cruzi* em solução e curvas teóricas calculadas para três modelos distintos da enzima baseado em sua estrutura cristalina em diferentes configurações: (i) modelo monomérico, (ii) dimérico I (BC) e (iii) dimérico II (AB). Ao lado esquerdo os três modelos estão separados para fácil visualização.

Monômero fechado da PEPCK da *T. cruzi* determinado cristalograficamente

Nenhuma investigação conclusiva sobre mudanças conformacionais induzidas por ligantes foram realizadas até o momento com a PEPCK da *T. cruzi*. A estrutura cristalográfica da *T. cruzi* PEPCK determinada por Trapani [2] não contém o substrato, apresenta porém um ion de sulfato (talvez um fosfato, o mapa de sensidade eletrônica não permite distinguir entre as duas moléculas) na região onde o fosfato se liga, o que confere a PEPCK uma configuração “fechada”, pois comparações com a PEPCK da *E. coli* mostram que na presença do substrato o monômero é “fechado” e na ausência “aberto”. Desta forma deve-se esperar que a PEPCK da *T. cruzi* também deveria apresentar mudanças conformacionais grandes induzidas pelo substrato.

Uma vez estabelecido a PEPCK ser um dímero do tipo I em solução, investigaram-se detalhes adicionais da conformação interna dos monômeros, isto é, tentou-se estabelecer se dois domínios internos do monômero estão numa configuração “aberta” ou “fechada”. Os resultados cristalográficos demonstraram a estrutura ser “fechada” e no local em torno do qual os domínios se abrem e fecham (o sítio

Modelo do monômero da PEPCK *T. cruzi* aberto

ativo) foi encontrado uma molécula de sulfato ou fosfato que provavelmente induziu o fechamento. Estudou-se esta enzima por SAXS em solução sem o íon sulfato, e íons fosfato se encontravam na solução a concentrações moderadas.

Modelos teóricos de conformações “fechadas” e “abertas” foram para a PEPCK da *T. cruzi* foram obtidos utilizando-se os códigos da PEPCK da *E. coli* (PDB 1AQ2 e 1OEN) como modelo. A configuração “fechada” da PEPCK do *T. cruzi* foi determinada cristalograficamente e serviu como ponto de partida quando então pequenos incrementos na rotação relativa entre os domínios foi empregada. Os dois eixos de rotação e os resíduos envolvidos nos dois C-terminais foram determinados utilizando-se o programa DynDom [3] na PEPCK da *E. coli* nas duas conformações (aberta e fechada). Em seguida sobrepueram-se duas cópias dos monômeros da PEPCK da *E. coli* no dímero da *T. cruzi*. Seis modelos abertos parcialmente foram modelados simplesmente movendo os resíduos de 5 a 35° (em passos de 5°) em torno dos dois eixos. Em seguida utilizando estes modelos, ajustaram-se as curvas teóricas obtidas a partir destes modelos aos dados experimentais utilizando o CRY SOL e o melhor ajuste (χ) foi obtido em torno de 20-25° como pode ser visto na Figura 4.

Finalmente uma modelagem de corpo rígido, onde a posição e orientação relativa dos monômeros é variada, foi realizada utilizando o programa MASSHA para cada um dos seis ângulos de rotação dos domínios internos (Figura 4). Permitiu-se variar a posição relativa de um dos monômeros em torno de $\pm 2\text{\AA}$ de sua posição original (três passos em cada direção x, y e z). A orientação relativa de um dos monômeros foi variada também num cone de 10°. As simulações computacionais demonstraram que essas variações levam a adicionais reduções do χ como pode ser visto na Figura 4.

Esses resultados indicam que a estrutura da PEPCK do *T. cruzi* em solução difere da de sua estrutura cristalina em dois pontos:

- (i) No cristal a conformação interna do monômero é “fechada” enquanto que em solução é aberta, isto é, devido à ausência de uma molécula que pudesse “fechar” o monômero. Desta forma espera-se encontrar futuramente uma estrutura cristalina “aberta” se for possível cristalizar a PEPCK sem a presença de substrato;
- (ii) A posição e orientação relativa dos monômeros que formam o dímero são ligeiramente diferentes que a da do estado cristalino.

Figura 4: Valores das discrepâncias χ correspondentes ao melhor ajuste dos dados experimentais aos teóricos para diferentes ângulos de abertura entre os domínios intramoleculares (quadrados). Os resultados de um refinamento de corpo rígido de pequenos deslocamentos relativos entre os monômeros também está representado e apresenta um ajuste ainda melhor (bolinhas).

Antes de iniciar a análise dos resultados obtidos pela PEPCK em solução utilizou-se o programa DAMMIN que demonstrou ser bastante eficiente ao resolver o envelope molecular a uma resolução de 25Å da PEPCK como pode ser visto na Figura 5.

Desta forma concluiu-se que por SAXS que a PEPCK da T. cruzi em solução é dimérica. Uma análise mais detalhada ainda mostrou que os dímeros estão arranjados de forma muito parecida ao dímero tipo I no estado cristalino. Os refinamentos de corpo rígido indicaram que a estrutura interna dos monômeros adota uma conformação aberta, em contraste com a cristalina que é fechada e ainda verificou-se que a posição e orientação relativa dos dois monômeros diferem da do estado cristalino.

No estudo desta proteína, a técnica de SAXS demonstrou sua utilização em auxiliar cristalógrafos de proteína na determinação da estrutura em solução quando existem ambigüidades ou dificuldades de se obterem cristais em diferentes estados conformacionais.

Figura 5: Lado esquerdo: ajuste final do modelo restaurado pelo programa DAMMIN em relação aos dados experimentais. Lado direito: modelo restaurado (bolinhas) e modelo cristalográfico sobreposto.

5.3 Mudanças conformacionais e estados de agregação na proteína PFK-2

5.3.1 Introdução e objetivos

Um equilíbrio entre diversos estados oligoméricos é encontrado em numerosas proteínas em solução. Um deles é aquele no qual um ligante se associa a uma enzima promovendo uma mudança nas propriedades catalíticas da mesma acompanhada por um processo de associação-dissociação. Este é o caso da isoenzima fosfofrutoquinase (PFK-2), uma proteína dimérica da *E. coli* que é inibida pelo MgATP que se liga ao sítio alostérico^{iv} e promove a tetramerização da proteína [4,5]. Esta enzima é importante na quebra da glicose, particularmente ela reduz a Frutose-6-P para Frutose-1,6 bis fosfato como ilustrado na Figura 6. Esta enzima é particularmente interessante para estudar o efeito de agregação molecular e também o de ligação de ligantes, pois ela somente existe como dímero ou tetrâmero e suas propriedades catalíticas e de ligação são bem conhecidas [6,7,8].

^{iv} Sítio alostérico não é o sítio ativo e sim outro sítio ("allo" significa outro).

Figura 6: Caminho metabólico da quebra de glicose em piruvato. A PFK é acionada no passo 3. No lado direito da figura temos a reação que ela cataliza. [9]

Estudos por gelfiltração demonstraram que na ausência de MgATP a PFK-2 elui como dímero num largo intervalo de concentração da proteína (0,05-30mg/mL), mas logo que MgATP é adicionado a concentrações crescentes ela tetrameriza. Porém neste sistema, a técnica de gelfiltração é de natureza somente qualitativa, pois quando o MgATP é introduzido, ele interfere na absorção da amostra fazendo com que seja impossível quantificar a fração de dímeros/tetrâmeros em solução para diferentes concentrações de MgATP.

Os objetivos deste estudo são determinar por SAXS o estado de agregação da PFK-2 em função da concentração de MgATP e as mudanças conformacionais associadas quando esta proteína se liga a Frutose-6-P. Desta forma teremos informações bem mais completas que as fornecidas por técnicas espectroscópicas porque as medidas de SAXS não são afetadas por problemas de superposição de espectros de absorção óptica, pois são somente sensíveis à mudanças em volume de forma da proteína.

5.3.2 Resultados experimentais, análise de dados e conclusões

A dependência do estado de agregação da PFK-2 em função da concentração do ATP foi estudada para duas concentrações de proteínas: 2mg/mL e 4.2mg/mL numa solução tampão contendo Tris HCl (pH 7.6) MgCl₂ 5mM, DTT 2mM e 5% de glicerol.

Várias medidas a diferentes frações molares de MgATP/PFK-2 foram realizadas. O volume total de cada amostra medida não ultrapassou 150µL. As amostras foram centrifugadas por 3 minutos a 20000g e nenhuma precipitação foi observada. A amostra também foi encapsulada numa janela de 1mm de espessura de mica. Todas medidas de SAXS foram realizadas na linha de SAXS do LNLS em Campinas. O comprimento de onda incidente foi de $\lambda=1.488 \text{ \AA}$ (borda do Ni). Utilizou-se um detector sensível a posição para registrar a intensidade espalhada em função do módulo do vetor de espalhamento q . Realizou-se também o desborramento dos dados como descrito na seção 2.6.1 do capítulo 2. A distância entre a amostra-detector foi ajustada de tal modo a cobrir um intervalo em q de 0.01 a $0.6 \text{ \AA}^{-1}$. Todo caminho do feixe foi evacuado a fim de reduzir ao máximo o espalhamento parasita produzido pelo ar. Em geral foram coletados entre 5-30 exposições de 60-180s a fim de acompanhar eventuais danos que a amostra pudesse sofrer sob ação dos raios X bem como monitorar eventuais mudanças de posição no próprio feixe incidente.

5.3.2.1 PFK-2 a 2mg/mL

A curva de espalhamento da PFK2 a 2mg/mL com concentrações de MgATP variando entre 0 to 540µM está ilustrado na Figura 7. Nesta figura a intensidade foi normalizada a concentração equivalente ($c=1 \text{ mg/ml}$). Nessas condições a intensidade normalizada é proporcional ao volume da molécula:

$$\frac{I(q)}{c} \propto v$$

Equação 1

Figura 7: Curvas de SAXS da PFK-2 a 2mg/mL e concentrações crescentes de MgATP.

Observando a Figura 7 podemos perceber mudanças nas curvas de espalhamento quando o MgATP é adicionado à solução. Observa-se em particular uma forte variação para concentrações em torno de 200 μM . A baixa concentração de MgATP, a intensidade de espalhamento extrapolada a ângulo zero, $I(0)$, é próxima de 1, enquanto que para concentrações elevadas $I(0)$ é próximo de 2. Lembrando a equação 1, o resultado acima mostra que, se em solução diluída a baixa concentração de MgATP a proteína se encontra em estado dimérico com volume v , a alta concentração ela forma tetrâmeros com volume $2v$. As funções de distribuição de distâncias das curvas da Figura 7 estão ilustradas na Figura 8.

Figura 8: Distribuição de distâncias – $p(r)$ – como função da fração molar de MgATP na PFK-2. Os dados foram normalizados para 1 para efeitos de comparação.

Observando as funções $p(r)$ e tendo em conta a análise quantitativa da seção 2.2.1 e a Figura 6 desta mesma seção, podemos identificar uma transição clara entre uma estrutura alongada (prolata) para uma estrutura achatada (oblata). Esta transição ocorre para concentrações de MgATP entre $180\mu\text{M}$ (3.15^v) e $240\mu\text{M}$ (4.20). Isto significa que com $180\mu\text{M}$ (3.15) MgATP tem-se somente dímeros em solução e com $240\mu\text{M}$ (4.20) observam-se somente tetrâmeros. Isto é uma indicação que a transição entre os estados não ligado e ligado ocorre num intervalo de concentração relativamente pequeno. Os resultados sugerem ser necessários quase 4 MgATP por 2 dímeros para que o tetrâmero possa ser formado, caso contrário, o equilíbrio entre as duas espécies oligoméricas é fortemente deslocada para dímeros.

Na Figura 9, onde graficou-se a variação da intensidade à ângulo zero $I(0)$ em função da concentração de MgATP, o mesmo fato é mostrado de forma mais clara, como $I(0)/c$ a altas concentrações de MgATP é aproximadamente igual a $2I(0)$ a baixas concentrações.

Figura 9: Ilustração da transição dímero $\rightarrow$ tetrâmero em função da fração molar de MgATP na PFK-2. As linhas tracejadas ilustram as espécies presentes na solução.

^v Os números entre parênteses indicam a fração molar, ou seja a quantidade de moléculas de ATP por molécula da proteína. Um tetrâmero da PFK têm 4 moléculas idênticas, logo no caso em questão existiriam 3.15 moléculas de ATP por tetrâmero ou por 2 dímeros ou 4 monômeros de PFK.

Figura 10: Ilustração do raio de giro na transição dímero → tetrâmero em função da fração molar de MgATP na solução de PFK-2. As linhas tracejadas ilustram as espécies presentes na solução. O ajuste nos gráficos de Guinier ($\ln I(q)$ versus q^2) para obter o raio de giro foi realizado entre $0.03 < q < 0.06 \text{ \AA}^{-1}$.

Percebe-se na Figura 9 um deslocamento do equilíbrio entre as espécies para tetrâmeros quando a fração molar fica em torno de 4 MgATP para cada 2 dímeros. O gráfico da variação do raio de giro do objeto espalhador em função da concentração de MgATP da Figura 10 também ilustra esta transição. Novamente se observa uma transição entre 180-240 μM ou seja 3.15-4.20 MgATP para cada monômero de PFK-2. Todos esses resultados indicam que a transição ocorre para uma concentração molar da proteína muito próxima à concentração molar do ligante, ou seja em torno de 4 moléculas de MgATP para cada tetrâmero de PFK-2.

A fim de obter curvas de SAXS com estatística razoável para valores de q maiores, $q = 0.3 \text{ \AA}^{-1}$, realizou-se uma nova série de experiências para uma concentração de proteína maior (4.2 mg/mL). Isto permitiu determinar com maior detalhe as funções $I(q)$ para a região de concentração de ligante (MgATP) compreendida entre 3.33 e 4.17 mol de MgATP por mol de PFK-2 e caracterizar assim mais precisamente a região de transição.

5.3.2.2 PFK-2 a 4.2mg/mL

A Figura 11 ilustra a função $p(r)$ de sistemas concentrados para diferentes concentrações MgATP na região de transição. Percebe-se que as curvas são estatisticamente melhores que no caso diluído (Figura 8). Agora pode-se observar que, quando temos uma concentração de MgATP superior a 3.33 vezes a razão molar de PFK-2, existem espécies tetraméricas em solução, sendo isto sugerido pelo ombro que aparece entre 60-80Å. Assim, com uma razão molar de 3.44 tem-se claramente uma certa porcentagem de tetrâmeros. Para valores maiores que 3.44 nota-se um claro aumento na fração de tetrâmeros e a 4.17 tem-se o equilíbrio totalmente deslocado para a forma tetramérica.

Figura 11: Distribuição de distâncias – $p(r)$ – como função da fração molar de MgATP na PFK-2. Os dados foram normalizados para 1 para efeitos de comparação.

A fim de estimar com maior precisão a fração de espécies oligoméricas em solução, utilizou-se uma combinação linear de curvas de intensidade de espalhamento de dois estados (dímero e tetrâmero). Os resultados estão representados na Figura 12. Todas as curvas foram obtidas com concentrações de proteína da ordem de 4.2mg/mL.

Figura 12: Ajuste entre dados experimentais (pontos) e combinação linear dois estados (dimérico e tetramérico) em função da concentração de MgATP. Proteína a 4.2mg/mL.

Desta análise conseguiu-se extrair a fração de tetrameros em solução como função de MgATP como ilustra a Figura 13.

Figura 13: Porcentagem de espécies tetraméricas em função da concentração de MgATP resultante do ajuste da combinação linear de dois estados como indicado na Figura 12.

5.3.2.3 Mudança conformacional induzida pelo MgATP e frutose-6-P

Outro objetivo deste estudo foi o de observar eventuais mudanças conformacionais internas a nível de estrutura terciária induzidas pela ligação de MgATP (além do efeito de tetramerização neste caso) e frutose-6-P à PFK-2. Para isto coletou-se dados com qualidade estatística aceitável até $q=0.55\text{\AA}^{-1}$ utilizando 3 diferentes concentrações de proteínas 2, 4.2 e 30mg/mL para a PFK-2 com alta concentração de MgATP e sem MgATP. Para a PFK com frutose-6-P utilizaram-se concentrações de 5 e 25mg/mL com alta concentração de frutose e sem MgATP. As curvas correspondentes as diversas concentrações de proteínas foram utilizadas para definir as funções $I(q)/c$ nas diversas faixas de q (baixas concentrações de proteína para baixos q 's e altas concentrações para altos q 's). Os resultados são mostrados na Figura 14.

Figura 14: Espalhamento da PFK-2 sem ligantes (quadrados), ligada ao MgATP formando tetrâmeros (círculos) e ligada a frutose-6-P (triângulos).

Este resultado demonstra uma mudança conformacional induzida quando PFK-2 se liga à frutose-6-P bem como a MgATP. Percebem-se diferenças entre as curvas em todo

intervalo de q , mesmo no caso da PFK-2 ligado a frutose-6-P em comparação com a não ligada.

Para a proteína PFK-2 quando a enzima está ligada ao MgATP pode-se notar um acentuado vale da função $I(q)$ em $q = 0.2\text{\AA}^{-1}$ ilustrando que além de ocorrer a tetramerização (evidenciada pelo aumento de intensidade para $q < 0.08\text{\AA}^{-1}$) há uma mudança conformacional a nível de organização terciário (movimento de domínios dos monômeros) quando o MgATP se liga no sítio alostérico. No caso da frutose-6-P se ligar a PFK-2 no sítio ativo as diferenças entre as funções $I(q)$ sugerem, como veremos, também uma mudança conformacional.

Para ilustrar a mudança conformacional de forma ainda mais clara, analisamos as funções $p(r)$ representadas na Figura 15.

Figura 15: Função distribuição de distâncias da PFK-2 em tampão sem ligantes (curva azul), ligada ao MgATP (preta) formando tetrâmeros e ligada a frutose-6-P (vermelha).

Observando as diferenças entre as curvas da Figura 15, percebemos que para a frutose-6-P ligada a PFK-2, a função $p(r)$ é inferior à da enzima nativa para distâncias r elevadas, o que é um forte indício de uma compactação da molécula.

Para determinar a estrutura e arranjo dos domínios em baixa resolução, propôs-se um modelo da estrutura cristalográfica da PFK obtido por modelagem por homologia utilizando-se como modelo as estruturas “abertas” e “fechadas” da riboquinase (códigos PDB 1RKA e 1RKD, respectivamente). Como estes PDBs contêm somente um monômero

da riboquinase, a segunda subunidade monomérica da riboquinase foi obtido aplicando-se uma operação de simetria de ordem 2 apropriada. Utilizou-se o programa DynDom [3] para encontrar os resíduos nas regiões flexíveis entre os domínios móveis. Desta forma variando a abertura dos domínios e a posição relativa entre os mesmos obtiveram-se modelos para a PFK holo (ou seja, não ligada a nenhum ligante), ligada a frutose-6-P e MgATP como ilustrado na Figura 16.

Figura 16: Modelos estruturais baseados em modelagem por homologia e SAXS da PFK (a) não ligada, (b) ligada a frutose-6-P e ao (c) MgATP. Também ilustrado o envelope obtido pelo método *ab-initio* utilizando o DAMMIN.

Os resultados demonstraram que quando o MgATP se liga ao sítio alostérico, ocorrem duas coisas simultaneamente: uma mudança na estrutura terciária e quaternária. Também estudou-se o processo de tetramerização em função da concentração de MgATP. A tetramerização foi observada somente quando mais de 3 moléculas de MgATP estavam presentes para cada 2 enzimas diméricas da PFK-2.

Quando a molécula de açúcar (frutose-6-P) se liga a PFK no sítio alostérico ocorre um fechamento do sítio ativo que promove posteriormente uma maior afinidade do MgATP à PFK (passo entre (a) e (b) da figura 16). Quando porém o MgATP se liga ao sítio alostérico, a PFK tetrameriza e esconde o sítio ativo, consequentemente impedindo a entrada de frutose-6-P ao sítio ativo (passo entre (a) e (c) da figura 16).

Finalmente quando nada está ligado a proteína, ela se encontra numa forma mais “aberta” como observado pela análise das medidas de SAXS (estado (a) da figura 16).

5.4 Estrutura da interleucina humana 22 (hIL-22) em solução

5.4.1 Introdução e objetivos

As interleucinas são proteínas secretadas por leucócitos que afetam o crescimento e a função de outros leucócitos. Existem várias espécies distintas de interleucinas, mas apenas a IL-2 tem sido empregada no tratamento clínico do câncer, as demais encontram-se em fase de ensaios clínicos. A interleucina humana 22 (hIL-22) também conhecida como fator de indução das células T é uma citocina que tem estrutura similar a IL-10. Estudos prévios mostraram que a hIL-22 induz reações de fase aguda “in vivo” e “in vitro”. Isto sugere que esta citocina contribui na regulação de respostas inflamatórias. O peso molecular desta proteína é de 16.8 kDa, sendo constituída por 147 aminoácidos.

Nagem *et al.*[10] resolveu a estrutura cristalográfica da hIL-22 porém não conseguiu explicar como a proteína interage com o receptor, pois a proteína cristalizou de forma que torna impossível essa interação ou seja, no dímero descrito no artigo deles, a região do sítio ativo é ocupada por outro monômero.

Elucidou-se a estrutura da hIL-22 em solução mediante um estudo por SAXS e propuseram-se modelos de interação da proteína com seu receptor.

5.4.2 Resultados, discussões e conclusões das medidas de SAXS e DLS da hIL-22 em solução

O estado oligomérico das proteínas foi estudado tanto por SAXS como por espalhamento dinâmico de luz (DLS). As medidas de DLS foram realizadas num equipamento da Protein Solutions, o DynaPro MS200 a 4°C. Para verificar o estado oligomérico em função da concentração, fizeram-se medidas de 0.32, 0.63, 1.25, 2.50 e 5.00 mg/mL numa solução de 10 mM de Tris-HCl e 10 mM NaCl num pH 8.0.

O raio de Stokes da hIL-22 medido variou entre 2.9nm a 3.5nm das proteínas a baixas a altas concentrações respectivamente. Estes resultados concordam com os raios de Stokes obtidos pelos modelos obtidos pelos resultados de SAXS tanto para dímero (2.9nm) quanto para o tetrâmero (3.3nm). Para determinar o raio de Stokes dos modelos de baixa resolução de SAXS utilizou-se um programa chamado Hydropro [11].

Os resultados experimentais de $D_{\max}$ e R_g , 11 nm e 3.25 nm para o tetrâmero e 8.0 nm e 2.74 nm para o dímero, respectivamente, sugerem que a proteína é globular nos dois estados oligoméricos. A função de distribuição de distâncias $p(r)$ da Figura 18 ilustra perfis típicos de proteínas globulares prolatas [12].

Figura 17: Espalhamento da hIL-22 sem ligantes. (a) dímero e (b) tetrâmero. No inserto está representado o gráfico de Guinier ($\log I$ versus q^2) deixando claro que a amostra é monodispersa não apresentando agregação. (1) dados experimentais; (2) dados experimentais subtraídos por uma constante para forçar 2 densidade eletrônicas que foi utilizado como entrada do programa DAMMIN; (3) ajuste do DAMMIN e (4) Ajuste utilizando estruturas dos monômeros de hIL-22 conhecidos.

Figura 18: Função distribuição de distâncias da hIL-22 tanto do tetrâmero (quadrados preenchidos) e dímero (quadrados ocos).

Os oligômeros da hIL-22 se comportam como dímeros para concentrações de até 2.5mg/mL e tetrâmeros para concentrações superiores a 5mg/mL como pôde ser observado pela determinação experimental de suas massas moleculares de 32(3) e 63(5) kDa, respectivamente. As massas moleculares foram obtidas pela técnica de SAXS tendo como referência uma solução de albumina de soro bovino (veja tabela abaixo).

5.4.2.1 Modelos *ab initio* utilizando DAMMIN

Utilizou-se o programa DAMMIN para reconstruir a forma tanto do dímero como do tetrâmero. Para restringir o número de parâmetros livres, assumiu-se um eixo de simetria de ordem 2 para o dímero e três eixos de ordem 2 para o tetrâmero. O fato de se utilizar restrições de simetria reduz o número de parâmetros livres e conseqüentemente o tempo de processamento computacional.

Na reconstrução da forma tanto do dímero quanto tetrâmero da hIL-22 partiu-se de uma esfera inicial com diâmetro máximo definido pela $p(r)$ sendo o $D_{\max} = 9.0$ nm para o dímero e $D_{\max} = 11.0$ nm para o tetrâmero. O raio das esferas que preencheram o espaço de busca inicial foi de $r_a = 0.275$ nm para o dímero e $r_a = 0.325$ nm para o tetrâmero o que resultou em $N_{\text{DAM}} = 2550$ para o dímero e $N_{\text{DAM}} = 3559$ para o tetrâmero. Como foi imposta simetria, o número de parâmetros livres final foi reduzido para $N_{\text{red}} = 2550/2 = 1275$ para o dímero e $N_{\text{red}} = 3559/4 = 890$ para o tetrâmero.

Os modelos finais tanto do dímero quanto tetrâmero da hIL-22 continham em torno de 760 ± 20 DAMs e 690 ± 20 DAMs respectivamente. Uma média de 20 modelos *ab-initio* sobreposto ao modelo obtido utilizando a estrutura cristalográfica está representada na Figura 19 e Figura 20. Utilizou-se o programa SUPCOMB para fazer a média dos 20 modelos, e quando um modelo não era parecido com os outros era descartado.

Os principais dados experimentais e os dados decorrentes de modelagem estão representados na Tabela 1.

Figura 19: Modelos do tetrâmero da hIL-22 obtidos por modelagem *ab initio* utilizando o programa DAMMIN e dados cristalográficos disponíveis (utilizando o programa MASSHA). (a) Superposição do envelope obtido pelo DAMMIN (malha em azul) proveniente de uma média de 20 modelos e do modelo que melhor ajustou os dados experimentais obtido a partir dos dados cristalográficos da hIL-22. (b) e (c) são os mesmos modelos rotacionados em sentido anti-horário por 90° em torno do eixo z e y, respectivamente. As figuras sobrepostas foram feitas utilizando o software PyMOL (Delano Scientific, San Carlos, C.A., <http://www.pymol.org>).

5.4.2.2 Estrutura do dímero e tetrâmero da hIL-22 obtida por SAXS

Utilizou-se a estrutura já conhecida da hIL-22 que foi obtida do *Protein Data Bank* (PDB ID:1M4R)[10]. As posições relativas dos monômeros foram encontradas rotacionando-os respectivamente até encontrar-se um bom acordo entre a curva de espalhamento teórico do modelo e os dados experimentais. Utilizou-se o programa MASSHA para este fim.

Tabela 1 – Principais dados experimentais e de modelagem

Parâmetro/Amostra	hIL-22 dimérica			hIL-22 tetramérica		
	Exp. ^a	Mod. ^b	DAM ^c	Exp. ^a	Mod. ^b	DAM ^c
Massa (kDa)	32(3)	33.7		63(5)	67.4	
D _{max} (nm)	8.0(5)	8.2	8.2(1)	11.0(5)	10.5	10.8(1)
R _g (nm)	2.74(6)	2.83	2.78(2)	3.25(4)	3.32	3.32(2)
V(nm ³)		44			88	
Parâmetros livres	7.6 ^d	6 ^e	1275 ^e	12.3 ^d	6 ^e	890 ^f
Discrepância χ		0.83	0.81		0.75	0.73
Resolução (nm)	2.1	4.2	4.2	1.8	4.2	1.8

^a Exp., calculados a partir dos dados experimentais.

^b Modelo teórico, parâmetros dos modelos cristalográficos (MASSHA).

^c DAM, parâmetros da média de 20 modelos de DAMs.

^d Número de canais de Shannon $N_s = D_{max}(q_{max} - q_{min})/\pi$

^e Simetria 2 imposta

^f Simetria 222 imposta

O modelo dimérico da hIL-22 obtido por SAXS, está ilustrado na Figura 19. Nota-se que o dímero tem uma estrutura tipo um V aberto com duas sub-estruturas claramente visíveis que correspondem ao homodímero da hIL-22. Esta estrutura é muito parecida com o da interleucina humana 10 (hIL-10) [10,13,14,15]. Também modelou-se a interação da hIL-22 com seu receptor. Utilizou-se para este fim o receptor α da interleucina humana 10 (hIL-10R α) como pode ser visto na Figura 21. (Utilizou-se a estrutura da hIL-10R α pois esta já se encontrava no banco de dados do PDB).

A interface de dimerização da hIL-22 é praticamente a mesma da hIL-10, e neste caso acredita-se que esta interface é composta pelos resíduos CYS132-GLY136 que fazem parte do loop entre as hélices D e E de acordo com a nomenclatura adotada por Nagen *et al.*[10].

Da mesma forma como se obteve o dímero da hIL-22, propôs-se um modelo do tetrâmero. A forma do mesmo é um pouco oblata e com formato de um X deslocado (Figura 20). Como o dímero, é possível identificar claramente sub-estruturas que correspondem aos domínios individuais do tetrâmero. Também propôs-se um modelo da hIL-22 ligada a 4 seus receptores hIL-10R α (Figura 22). Isto sugere que em solução a hIL-22 possa se ligar simultaneamente a 4 receptores.

Comparando o empacotamento da hIL-22 com outras proteínas, encontramos uma grande similaridade dela com a estrutura da interferon γ (hIF γ) como descrito por Randal e Kossiakoff [16]. O homodímeros de Randal e Kossiakoff da hIF γ cristalizam de uma forma tal que um empacotamento tetramérico pode ser observado semelhante ao observado neste estudo, sugerindo que para altas concentrações, a hIF γ também se comporte como um tetrâmero se ligando a 2 receptores ao invés de 4 como no caso da hIL-22.

Figura 20: Modelos do tetrâmero da hIL-22 obtidos por modelagem *ab initio* utilizando o programa DAMMIN e dados cristalográficos disponíveis (utilizando o programa MASSHA). (a) Superposição do envelope obtido pelo DAMMIN (malha em azul) proveniente de uma média de 20 modelos e do modelo que melhor ajustou os dados experimentais obtido a partir dos dados cristalográficos da hIL-22. (b) e (c) são os mesmos modelos rotacionados em sentido anti-horário por 90° em torno do eixo z e y, respectivamente. As figuras sobrepostas foram feitas utilizando o software PyMOL (Delano Scientific, San Carlos, C.A., <http://www.pymol.org>).

Figura 21: Modelo estrutural do dímero da hIL-22 (parte central) ligada a seu receptor – hIL-10R α (parte exterior). (a) e (b) são os mesmos modelos rotacionados em sentido horário por 90° em torno do eixo x, respectivamente. As figuras sobrepostas foram feitas utilizando o software PyMOL (Delano Scientific, San Carlos, C.A., <http://www.pymol.org>).

Figura 22: Modelo estrutural do tetrâmero da hIL-22 (parte central) ligada a 4 receptor – hIL-10R α (parte exterior). (a) e (b) são os mesmos modelos rotacionados em sentido horário por 90° em torno do eixo x, respectivamente. As figuras sobrepostas foram feitas utilizando o software PyMOL (Delano Scientific, San Carlos, C.A., <http://www.pymol.org>).

Em nosso estudo mostramos a importância da técnica de SAXS porque ela ilustra como dois monômeros de hIL-22 interagem e se empacotam formando uma estrutura dimérica muito parecida com a hIL-10. Os resultados ilustram claramente como os dímeros se empacotam, sendo esse empacotamento completamente diferente do proposto no artigo de Nagem *et al.*[10]. Também conseguimos descrever como a hIL-22 interage com seu receptor e propomos um modelo do tetrâmero encontrado em solução em concentrações

acima de 5mg/mL. Finalmente chegamos a um modelo do tetrâmero e propomos que o mesmo possa ser ligado simultaneamente a 4 receptores, ou seja, uma ligação de 4:4.

- ¹ Linss J, Goldenberg S, Urbina JA and Amzel LM – Cloning and characterization of the gene encoding ATP-dependent phospho-enol-pyruvate carboxykinase In *Trypanosoma-Cruzi* - comparison of primary and predicted secondary structure with host CTP-dependent enzyme, *Gene* **136** (1-2): 69-77 (1993).
- ² Trapani S, Linss J, Goldenberg S, Fischer H, Craievich AF and Oliva G – Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from *Trypanosoma cruzi* at 2 Å resolution. *J Mol Biol.* **313**(5):1059-72 (2001).
- ³ Hayward S and Berendsen HJC – Systematic analysis of domain motions in protein from conformational change: new results on citrate synthase and T4 lysozyme. *Prot. Struc. Func. Genet.* **30**: 144-154 (1998).
- ⁴ Babul, J – Phosphofructokinases from *Escherichia coli*. Purification and characterization of the nonallosteric isozyme. *J. Biol. Chem.* **253**, 4350-4355 (1978).
- ⁵ Guixé, V, Rodríguez, PH, and Babul, J – Ligand-Induced Conformational Transitions in *Escherichia coli* Phosphofructokinase 2: Evidence for an Allosteric Site for MgATP²⁻. *Biochemistry* **37**(38), 13269-13275 (1998).
- ⁶ Cabrera R, Guixé V, Alfaro J, Rodríguez PH and Babul J – Ligand-dependent structural changes and limited proteolysis of *Escherichia coli* phosphofructokinase-2. *Archives of Biochemistry and Biophysics* **406** (2): 289-295 (2002).
- ⁷ Guixé, V and Babul, J – Effect of ATP on phosphofructokinase-2 from *Escherichia coli*. A mutant enzyme altered in the allosteric site for MgATP. *J. Biol. Chem.* **260**, 11001-11005 (1985).
- ⁸ Cabrera R, Fischer H, Trapani S, Craievich AF, Garratt RC, Guixé V and Babul J – Domain motions and quaternary packing of phosphofructokinase-2 from *Escherichia coli* studied by small angle x-ray scattering and homology modeling. *J Biol Chem.* **278**(15):12913-9 (2003).
- ⁹ Garrett RH and Grisham CM. *Biochemistry. Holt Rinehart & Winston*, 2nd edition (1999).
- ¹⁰ Nagem RA, Colau D, Dumoutier L, Renauld JC, Ogata C and Polikarpov I – Crystal structure of recombinant human interleukin-22. *Structure* **10**(8):1051-62 (2002).
- ¹¹ Garcia de la Torre, J, Huertas, ML and Carrasco, B – Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719-730 (2000).
- ¹² Feigin, LA and Svergun, DI – “Structure analysis by Small-angle X-ray and Neutron Scattering”, pp. 83-87, *Plenum Press*, New York (1987).
- ¹³ Zdanov, A, Schalk-Hihi, C, Gustchina, A, Tsang, M, Weatherbee, J and Wlodawer, A – Interleukin-10: crystal structure reveals the functional dimer with an unexpected topological similarity to interferon gamma. *Structure* **3**, 591-601 (1995).
- ¹⁴ Zdanov, A, Schalk-Hihi, C and Wlodawer, A – Crystal structure of human interleukin-10 at 1.6 resolution and a model of a complex with its soluble receptor. *Protein Sci.* **5**, 1955-1962 (1996).
- ¹⁵ Walter, MR and Nagabhushan, TL – Crystal-structure of interleukin-10 reveals an interferon gamma-like fold. *Biochemistry* **34**, 12118-12125 (1995).
- ¹⁶ Randal, M and Kossiakoff AA – The Structure and Activity of a Monomeric Interferon- γ : α -Chain Receptor Signaling Complex. *Structure*, **9**, 155-163 (2001).

6. *Conclusões e sugestões para trabalhos futuros.* 6-1

6. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS.

As principais conclusões desta tese se referem à elucidação das estruturas terciária e quaternária de algumas proteínas em solução - importantes pelas suas funções biológicas - mediante a técnica de espalhamento de raios X a baixos ângulos (SAXS). Foram obtidos modelos da estrutura de baixa resolução dessas proteínas visando contribuir para a compreensão das respectivas funções das mesmas.

Foram estudadas nesta tese as estruturas das seguintes proteínas:

- i. Domínio de ligação ao ligante e DNA do receptor de ácido X retinóico humano (hRXR α Δ AB);
- ii. Enzima Fostó-enol-piruvato carboxoquinase (PEPCK);
- iii. Isoenzima fosfofrutoquinase (PFK-2);
- iv. Interleucina humana 22 (hIL-22).

Para todas as proteínas listadas acima foi possível estabelecer modelos estruturais plausíveis. As conclusões mais relevantes associadas a cada uma delas são as seguintes:

i) ***Domínios de ligação ao ligante e ao DNA do receptor de ácido X retinóico (RXR).***

Os resultados de SAXS permitiram determinar, pela primeira vez, modelos estruturais de baixa resolução para um receptor nuclear contendo, tanto o domínio de ligação ao ligante (LBD), quanto o domínio de ligação ao DNA (DBD). Mais especificamente, a comparação entre os modelos estruturais que obtivemos para as formas tetraméricas e diméricas do receptor (hRXR α Δ AB), nos permitiu obter conclusões relevantes, entre elas, uma explicação plausível sobre a razão pela qual uma das formas (a dimérica) se liga com maior afinidade com um elemento responsivo (DNA) que a outra (a tetramérica). Descobrimos que os DBDs no dímero estão mais próximos uns dos outros que nos tetrâmeros, favorecendo a ligação do receptor ao elemento responsivo de DNA (DR1), sem sofrer mudanças conformacionais acentuadas. De fato, os homodímeros ligam-se preferencialmente ao DR1.

Concluimos também que, em contraste com o dímero que adota uma forma do tipo "U", o tetrâmero exibe uma forma do tipo "X". No tetrâmero a distância entre os DBDs é

de aproximadamente 6nm e, nestas condições, o DBD dificilmente se ligaria ao DR1, a não ser que sofresse uma mudança conformacional significativa. Destes resultados se conclui que o arranjo dos elementos de estrutura terciária é claramente sensível a alterações na estrutura quaternária.

ii) Enzima fosfo-enol-piruvato carboxoquinase (PEPCK).

Para esta proteína em solução, foi estabelecido um modelo estrutural baseado nos resultados de SAXS, complementados com outros de difração de raios X de monocristais. Nossos resultados confirmaram que a proteína PEPCK em solução é dimérica. Por outro lado, conseguimos estabelecer a conformação dos dímeros, que demonstramos estar arranjados em forma de "V". Essa estrutura é plausível porque sua configuração permite que os ligantes tenham fácil acesso ao sítio ativo. Finalmente descobrimos que, em solução, a posição e orientação relativa dos monômeros da proteína é ligeiramente diferente que no estado cristalino.

iii) Isoenzima fosfofrutoquinase (PFK-2).

Neste caso a proteína em solução foi estudada por SAXS, e os resultados foram comparados com modelos de estrutura de alta resolução obtidos por modelagem e por homologia. Determinamos o estado de agregação da PFK-2 em função da concentração de MgATP e as mudanças conformacionais associadas quando a Frutose-6-P e/ou MgATP se ligam à proteína, comparando com a proteína nativa.

Quando a PFK-2 não está associada a ligantes (APO) se encontra num estado dimérico "aberto". Demonstramos que, quando o MgATP se liga a sítio alostérico, ocorrem simultaneamente mudanças na estrutura terciária e quaternária da proteína: o monômero se fecha adotando uma configuração mais compacta, ocluindo assim o sítio ativo, e também passa de uma configuração dimérica a uma tetramérica.

Quando o ligante frutose-6-P se associa à PFK-2, a proteína continua dimérica como no estado APO, mas os domínios se fecham tornando-a mais compacta.

O efeito de compactação ou "fechamento" da estrutura induzida pelo MgATP e também a tetramerização do PFK-2, que nós estabelecemos, permite explicar o fato conhecido da frutose-6-P ter pouca afinidade com a PFK-2 associada ao MgATP.

Por outro lado, o efeito de compactação induzido pela ligação da frutose-6-P permitiu entender porque a PFK-2 ligada à frutose-6-P tem maior afinidade pelo MgATP que na forma APO.

Finalmente estudamos o estado de associação da PFK-2 saturada por MgATP em função da temperatura. Mostramos que a proteína mantém uma configuração tetramérica

de 4°C até 40°C. Acima desta temperatura a proteína se dissocia formando dímeros, perdendo assim sua atividade.

iv) **Interleucina humana 22 (hIL-22).**

Utilizando a técnica de SAXS, complementada por dados cristalográficos, mostramos que a proteína hIL-22 em solução possui uma configuração completamente diferente da que adota em estado cristalino, assumindo forma de "V" bem aberto com os sítios ativos expostos e acessíveis, em lugar de uma forma de eixos maiores paralelos com os sítios ativos voltados para o interior. A forma em "V" com os sítios ativos localizados no lado exterior, permite explicar porque os receptores podem de fato ligar à hIL-22. Também demonstramos que, em soluções de alta concentração, a hIL-22 adota uma configuração tetramérica com formato de "X" deslocado e, em baixas concentrações, possui uma configuração dimérica.

Em decorrência deste trabalho foram feitos avanços no conhecimento de aspectos básicos associados às proteínas em solução, como um estudo inédito da dependência da densidade de uma macromolécula com seu peso molecular e um procedimento novo para a determinação do estado oligomérico de uma proteína em solução por medidas de SAXS em escala relativa.

O conjunto dos resultados obtidos nesta tese permite concluir que a técnica de SAXS, quando utilizada de forma cuidadosa e tendo em conta todos os fatores que podem reduzir a qualidade das medidas, é provavelmente única para determinação ab-initio do envelope de macromoléculas em solução, o estabelecimento de modelos estruturais quando complementada com dados cristalográficos e o estudo de variações conformacionais e agregação de macromoléculas em função de diferentes parâmetros físicos e/ou químicos. A técnica de SAXS utilizando fontes clássicas ou fontes síncrotron permite, na prática, obter resultados de boa qualidade para macromoléculas em solução diluídas com peso molecular entre 1kDa e 1TDa.

Alguns aspectos das diversas pesquisas realizadas nesta tese não foram ainda esclarecidos. Novos estudos pertinentes - sugestões para futuros trabalhos - são os seguintes:

- i) *No caso do receptor de ácido X retinóico (RXR), seria interessante realizar estudos da proteína como um todo, incluindo o N terminal e o elemento responsivo (DNA) para entender como ela se liga ao DNA. Também seria interessante cristalizar o complexo, obter a estrutura de alta resolução e compará-la com os resultados de SAXS.*

- ii) Para a enzima fosto-enol-piruvato carboxoquinase (PEPCK) propomos estudá-la na ausência e na presença de seus ligantes para estabelecer as mudanças conformacionais induzidas.
- iii) Sugerimos resolver a estrutura de alta resolução da Isoenzima fosfofrutoquinase (PFK-2) para determinar quais são os aminoácidos envolvidos nas mudanças conformacionais da mesma.
- iv) Uma proposta interessante é também estudar a interleucina humana 22 (hIL-22) ligada a seus receptores para verificar a quantos receptores o tetrâmero se liga.
- v) Sugerimos fazer um estudo mais aprofundado quanto a dependência da densidade com o peso molecular da proteína, levando em conta centenas de proteínas diferentes (ou seja, medir a densidade experimentalmente destas proteínas) e verificar se a equação por nós encontrada ainda é válida.
- vi) Finalmente, no que se refere à metodologia da obtenção do estado oligomérico de proteínas em solução, sugerimos realizar um estudo similar ao feito nesta tese, considerando um número bem maior de proteínas de estrutura conhecida, para verificar a validade geral do procedimento.

Os artigos científicos decorrentes das pesquisas incluídas nesta tese estão listados a seguir:

- Fischer H, Napolitano H, Oliva G and Craievich A – A new method for obtaining proteins volume and molecular mass based on Small Angle X-ray Scattering (em preparação).
- Fischer H, Polikarpov I and Craievich A – Average protein density is a molecular-weight-dependent function - *Protein Sci.* **10**(13):2825-8 (2004).
- Fischer H, Dias SM, Santos MA, Alves AC, Zanchin N, Craievich AF, Apriletti JW, Baxter JD, Webb P, Neves FA, Ribeiro RC, Polikarpov I. – Low-resolution structures of the human retinoid X receptor DNA-binding and ligand-binding domains revealed by synchrotron X-ray solution scattering – *J. Biol. Chem.* **278** (18):16030-8 (2003).
- Trapani S, Linss J, Goldenberg S, Fischer H, Craievich AF, Oliva G – Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from *Trypanosoma cruzi* at 2 Å resolution. *J Mol Biol.* **313**(5):1059-72 (2001).
- Cabrera R, Fischer H, Trapani S, Craievich AF, Garratt RC, Guixe V and Babul J – Domain motions and quaternary packing of phosphofructokinase-2 from *Escherichia coli* studied by small angle x-ray scattering and homology modeling. – *J Biol Chem.* **278**(15):12913-9 (2003).
- Fischer H, Craievich AF and Polikarpov I – Low resolution structures of human interleukin-22 dimer and tetramer as revealed by synchrotron X-ray solution scattering (em preparação)

Além de uma participação predominante nas pesquisas incluídas nesta tese, realizamos em paralelo vários trabalhos em colaboração com diversos cientistas sobre aplicação da técnica de SAXS a várias outras proteínas em solução. Os artigos decorrentes destas colaborações são os seguintes:

- Borges JC, Fischer H, Craievich AF and Ramos CHI – Low-resolution structural study of two human Hsp40 chaperones in solution. HJA1 from subfamily A and HJB4 from subfamily B, have different quaternary structures (aceito *Journal of Biological Chemistry*, Janeiro de 2005).
- Bugs MR, Forato LA, Bortoleto-Bugs RK, Fischer H, Mascarenhas YP, Ward RJ, Colnago LA – Spectroscopic characterization and structural modeling of prolamin from maize and pearl millet - *Eur. Biophys. J. Biophys. Lett.* **33**(4):335-43 (2004).
- Forato LA, Doriguetto AC, Fischer H, Mascarenhas YP, Craievich AF, Colnago LA – Conformation of the Z19 prolamin by FTIR, NMR, and SAXS – *J. Agric. Food Chem.* **52**(8):2382-5 (2004).
- Santos SF, Zanette D, Fischer H and Itri R – A Systematic Study of Bovine Serum Albumin (BSA) and Sodium Dodecyl Sulfate (SDS) Interactions by Surface Tension and Small Angle X-Ray Scattering – *J. Colloid Interface Sci.*, **262**(2):400-8 (2003).
- Borges JC, Fischer H, Craievich AF, Hansen LD and Ramos CH – Free human mitochondrial GrpE is a symmetric dimer in solution. *J Biol Chem.* **278**(37):35337-44 (2003).
- Aparicio R, Fischer H, Scott DJ, Verschueren KHG, Kulminskaya AA, Eneiskaya EV, Neustroev KN, Craievich AF, Golubev AM and Igor Polikarpov – Structural Insights into the beta-Mannosidase from *T. reesei* Obtained by Synchrotron Small-Angle X-ray Solution Scattering Enhanced by X-ray Crystallography – *Biochemistry*, **41**(30): 9370-9375 (2002).
- Ferreira FM, Aparicio R, Fischer H, Craievich AF, Oliva G – Homology Model of the N-acetylglucosamine 6-phosphate Deacetylase from *Escherichia coli* Obtained from Solution SAXS Data (em preparação).