

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Química

Departamento de Físico-Química

Eduardo Kiyota

Mestrado

**Expressão, Purificação e Caracterização Estrutural dos
Fatores de Transcrição bZIP
SCF12 e SCF5 de Cana-de-Açúcar**

Orientador Prof. Dr. Ricardo Aparicio (IQ/Unicamp)

Co-orientador: Prof. Dr. Marcelo Menossi Teixeira (IB/Unicamp)

Campinas, agosto de 2008

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE
QUÍMICA DA UNICAMP**

K658e	<p>Kiyota, Eduardo. Expressão, purificação e caracterização estrutural dos fatores de transcrição bZIP SCF12 e SCF5 de cana-de-açúcar / Eduardo Kiyota. -- Campinas, SP: [s.n], 2008.</p> <p style="text-align: center;">Orientador: Ricardo Aparício.</p> <p style="text-align: center;">Co-orientador: Marcelo Menossi Teixeira.</p> <p style="text-align: center;">Dissertação - Universidade Estadual de Campinas, Instituto de Química.</p> <p style="text-align: center;">1. Fatores de transcrição bZIP. 2. Cana-de-açúcar. 3. Cristalografia de proteínas. 4. SAXS. I. Aparício, Ricardo. II. Teixeira, Marcelo Menossi. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.</p>
-------	--

Título em inglês: Expression, purification and structural characterization of the sugarcane bZIP transcription factors SCF12 and SCF5

Palavras-chaves em inglês: bZIP transcription factors, Sugarcane, Protein crystallography, SAXS

Área de concentração: Físico-Química

Titulação: Mestre em Química na área de Físico-Química

Banca examinadora: Prof. Dr. Ricardo Aparício (orientador), Prof. Dr. João Alexandre Ribeiro Gonçalves Barbosa (LNLS-Campinas), Profa. Dra. Ljubica Tasic (IQ-UNICAMP)

Data de defesa: 08/08/2008

Este trabalho é dedicado à minha família:

ao meu querido pai Mamoru, por todo seu amor, amizade, por toda a sua força, por mostrar que a vida deve ser vivida com honestidade, com decência, com dignidade; à minha querida mãe Adelina, por todo seu amor, carinho, dedicação, e por sua coragem e fibra que demonstrou diante dos obstáculos que superamos até este momento; aos meus queridos irmãos, Fabricio e Juliano, por toda nossa amizade e pelo companheirismo; e ao meu querido irmão Rodrigo, que tão cedo partiu e provocou uma profunda dor. Porém, sua breve passagem por nossas vidas nos fez uma família mais forte, mais unida. O tempo passou, a dor diminuiu, e restou uma enorme saudade...

“quando alguém evolui, evolui tudo à sua volta”

Paulo Coelho

Agradecimentos:

Ao meu orientador, professor Ricardo Aparicio, pela oportunidade e pelos ensinamentos;

Ao meu co-orientador, professor Marcelo Menossi, pela confiança e pelos ensinamentos;

Ao Prof. Dr. João Alexandre Ribeiro Gonçalves Barbosa, à Profa Dra. Ljubica Tasic, à Profa. Dra. Anete Pereira de Souza, ao Prof. Dr. Munir Salomão Skaf e à Profa. Dra. Juliana de Maria Félix, pela participação na banca;

Ao meu amigo Marcelo Leite dos Santos, pela grande ajuda durante a realização deste trabalho;

Ao Dr. Paulo S. Schlögl, meu amigo, pelas conversas, pelos ensinamentos, por todo o apoio;

Ao Daniel Razzo, por seu enorme auxílio nos experimentos de CD;

Ao pessoal da CPG-IQ, pelo suporte;

À minha amiga Aline, pela amizade, pelas conversas;

Ao pessoal que trabalha/trabalhou comigo no CBMEG, Sandra e Tânia (meninas da secretaria); Seu Chico; Márcio; ao meu amigo Wilson (Virsão); às minhas amigas Dani, Kiara e Ana (pelas conversas, pelos cafés); aos meus amigos Mário, Tati, Thaís e Guilherme (muitas conversas e almoços); ao pessoal do LGF e ao pessoal do antigo e saudoso Laboratório de Genética de Plantas;

Aos meus amigos da panela99, Almir, Aninha, Américo, Kátia, Fer, Viana, Vianinha, Rodrigo Indaia, Tháta, Warti, Danilo;

Aos meus amigos de Salto, Messias, Eliandra, Jackson, Edna, Nal, Rosângela, Gersão, Fran, Willian Hebe, Carlão, Micheli, Buizzo, Maria Eugênia, James, etc., pelos churrascos, pelas conversas;

Aos meus grandes amigos David Figueira e Paulo Gomes, pela amizade e por toda ajuda durante esses anos todos de convivência, desde que cheguei à Unicamp;

À minha querida família.

Eduardo Kiyota
Curriculum Vitae

Dados Pessoais

Nome Eduardo Kiyota
Nascimento 20/02/1977 - Salto/SP - Brasil

Formação Acadêmica/Titulação

1999 - 2003 Graduação em Química.
Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil

Atuação profissional

1. Universidade Estadual de Campinas - UNICAMP

Vínculo institucional

1998 - Atual Vínculo: Servidor público , Enquadramento funcional: Técnico químico , Carga horária: 40, Regime: Integral

Atividades

03/1999 - Atual Outra atividade técnico-científica, Cbmeg, Genética de Plantas
Especificação:
análise de proteínas

03/2005 - Atual Outra atividade técnico-científica, Cbmeg
Especificação:
biologia molecular e estrutural

Produção em C, T & A

Produção bibliográfica

Artigos completos publicados em periódicos

1. Kiyota, Eduardo, de Sousa, Sylvia Morais, dos Santos, Marcelo Leite, da Costa Lima, Aline, Menossi, Marcelo, Yunes, José Andrés, Aparicio, Ricardo
Crystallization and preliminary X-ray diffraction analysis of maize aldose reductase. Acta Crystallographica. Series F. , v.63, p.990 - 992, 2007.

Eventos

Participação em eventos

1. Apresentação de Poster / Painel no(a) **XVIII Reunião Anual de Usuários do LNLS**, 2008. (Outra)
Clonagem, Expressão, Purificação e Caracterização Estrutural da UGPase de Cana-de-Açúcar.
2. Apresentação de Poster / Painel no(a) **XXXVII Reunião Anual da SBBq**, 2008. (Congresso)
PRELIMINARY CRYSTALLOGRAPHIC ANALYSIS OF MAIZE ALDOSE.
3. Apresentação de Poster / Painel no(a) **XVIII Reunião Anual de Usuários do LNLS**, 2008. (Outra)
Structural characterization of bZIP SCF5 transcription factor from sugarcane by Small-Angle X-ray Scattering (SAXS)..
4. Apresentação de Poster / Painel no(a) **XXXVII Reunião Anual da SBBq**, 2008. (Congresso)
STRUCTURAL STUDIES OF UGPase FROM SUGARCANE.
5. Apresentação de Poster / Painel no(a) **XV Congresso Interno de Iniciação Científica**, 2007. (Congresso)
Estudos Estruturais da UGPase de Cana-de-Açúcar.
6. Apresentação de Poster / Painel no(a) **XXXVI Reunião Anual da SBBq**, 2007. (Congresso)
PRELIMINARY CRYSTALLIZATION STUDIES OF MAIZE ALDOSE REDUCTASE.
7. Apresentação de Poster / Painel no(a) **XXXVI Reunião Anual da SBBq**, 2007. (Congresso)
SMALL-ANGLE X-RAY SCATTERING (SAXS) STUDIES OF A BASIC LEUCINE.
8. Apresentação de Poster / Painel no(a) **XVII Reunião Anual de Usuários do LNLS**, 2007. (Outra)
Structural studies of two bZIP (SCF12 and SCF5) transcription factors from sugarcane.
9. Apresentação de Poster / Painel no(a) **XXXV Reunião Anual da SBBq**, 2006. (Congresso)
Cloning, expression and purification of a novel bZIP (SCF12) protein from sugarcane.
10. Apresentação de Poster / Painel no(a) **50o Congresso Brasileiro de Genética**, 2004. (Congresso)
Maize plants expressing both Proprotein Convertase 1 (PC1) and human Proinsulin: producing human insulin in maize kernels.
11. Apresentação de Poster / Painel no(a) **XXXII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular**, 2003. (Congresso)
Production of human growth hormone in transgenic maize seeds.
12. Apresentação de Poster / Painel no(a) **XXXI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular**, 2002. (Congresso)
Production of Human Pro-Insulin in Transgenic Maize.

Resumo

Os fatores de transcrição do tipo bZIP estão presentes em organismos eucariotos e estão envolvidos na regulação da expressão gênica e no controle de muitos processos intracelulares. Esses fatores se ligam a seqüências específicas no DNA e são capazes de reconhecer seqüências reguladoras no promotor de um gene. As bZIPs são caracterizadas por uma região conservada rica em resíduos de aminoácidos básicos, e um zíper de leucinas, que possui repetições de uma seqüência de aminoácidos hidrofóbicos onde há uma leucina que ocupa a mesma posição a cada 7 resíduos. Estudos estruturais com bZIPs mostraram que essas proteínas enovelam-se na forma de uma extensa hélice- α e são capazes de formar dímeros através de um arranjo do tipo *coiled-coil*. Neste trabalho, a parte correspondente à região básica e ao zíper de leucinas de duas bZIPs, SCF5 e SCF12 de cana-de-açúcar, pertencentes a sub-famílias diferentes, foram clonadas, expressas e purificadas para estudos estruturais. O DNA correspondente à SCF12 foi clonado em pET28a e a proteína recombinante foi produzida em *E. coli* BL21 (DE3) pRil. A SCF12 purificada por cromatografia de afinidade (IMAC) teve sua estrutura secundária caracterizada por dicroísmo circular. A SCF5, clonada em pET3C e expressa em *E. coli* BL21 (DE3) pLysS foi purificada por cromatografia de troca catiônica. Cristais de um complexo da proteína ligada a uma seqüência de DNA de 24 pares de bases foram obtidos mas não exibiram qualidade suficiente para permitir a determinação da estrutura cristalográfica. Entretanto, foi possível obter um modelo do complexo a partir de experimentos de espalhamento de Raios X a baixos ângulos (SAXS, do inglês *Small Angle X-Ray Scattering*) em solução, e interpretá-lo à luz de estruturas de homólogas já conhecidas.

Abstract

The bZIP transcription factors are present in eukaryotic organisms and are involved in the regulation of gene expression and many intracellular processes. These factors bind specific DNA sequences and are able to recognize regulatory sequences of a gene promoter. The bZIPs are characterized by a conserved region rich in basic amino acid residues as well as by having the leucine zipper region, which possess a sequence of hydrophobic residues where there are leucines every seventh amino acids. Structural studies have shown that bZIP-folding is alpha-helical and these proteins are capable of dimer formation via coiled-coil arrangement. In this work, the basic region and the leucine zipper of two sugarcane bZIPs, SCF12 and SCF5, belonged to two different bZIP-families were cloned, expressed and purified for structural studies. The corresponding SCF12 DNA was cloned into pET28a expression vector and the protein was produced in *E. coli* BL21 (DE3) pRil cells. SCF12 protein was purified by affinity chromatography (IMAC) and had its secondary structure characterized by CD. SCF5, cloned into pET3c and expressed in *E. coli* BL21 (DE3) pLysS was purified by cation exchange chromatography. Crystals of a complex formed by SCF5 protein and a 24-base-pair DNA sequence were obtained but unfortunately with quality insufficient for crystallographic structure determination. However, it was possible to obtain a model of the analyzed complex applying Small Angle X-ray Scattering (SAXS) technique by protein homologous structure comparison.

Índice Geral

Índice de Tabelas	xiii
Índice de Figuras	xiv
Lista de abreviações	xvi
1- Introdução	1
1.1- Cana-de-açúcar	1
1.2- Proteínas	2
1.3- Fatores de transcrição	5
1.4- bZIPs em plantas	7
1.5- Técnicas para caracterização estrutural de proteínas	10
1.5.1- Dicroísmo circular (CD)	10
1.5.2- Espalhamento dinâmico de luz	15
1.5.3- Espalhamento de Raios X a Baixos Ângulos (SAXS)	17
1.5.4- Cristalização de proteínas	21
1.5.4.1- Métodos de cristalização	21
1.5.4.2- Método de difusão de vapor: técnica da gota suspensa	23
2- Objetivos	25
3- Materiais e métodos	26
3.1- Proteína SCF12	26
3.1.1- Ensaio de expressão da proteína SCF12 em <i>E. coli</i>	27
3.1.1.1- Influência da cepa de <i>E. coli</i> utilizada na produção da SCF12	27
3.1.1.2- Influência da temperatura na indução	28
3.1.2- Análise das amostras em gel de poliacrilamida (SDS-PAGE)	28
3.1.3- Expressão da proteína em grande escala	29
3.1.4- Purificação	29
3.1.5- Dicroísmo circular	30
3.2- Proteína SCF5	30
3.2.1- Purificação por troca catiônica	31
3.2.2- Reação de ligação da SCF5 com a fita de DNA	31
3.2.3- Espalhamento dinâmico de luz (DLS)	32
3.2.4- Medidas de SAXS (<i>Small-Angle X-ray Scattering</i>)	32
3.2.5- Estudo da estabilidade térmica da proteína SCF5 por Dicroísmo Circular	33
3.2.6- Ensaio de cristalização	33
4- Resultados	39
4.1- Proteína SCF12	39

4.1.1 Clonagem e expressão	39
4.1.2- Purificação	40
4.1.3- Análise de dicroísmo circular (CD)	41
4.2- Proteína SCF5	43
4.2.1 Purificação por troca catiônica	43
4.2.2- Estudo da estabilidade térmica por CD	45
4.2.3- Espalhamento Dinâmico de Luz (DLS)	48
4.2.4- Experimento de SAXS (<i>Small-Angle X-ray Scattering</i>)	49
4.2.5- Ensaio de cristalização	55
5- Discussão	59
5.1- SCF12	59
5.2- SCF5	61
6- trabalhos adicionais	64
6.1-UGPase	64
6.2-Aldose Redutase (AR)	64
7- Conclusão e Perspectivas	66
Anexo 1	67
8- Referências	71

Índice de Tabelas

Tabela 1- Descrição das soluções contidas no <i>Crystallization Basic Kit for Proteins</i> (Sigma).....	35
Tabela 2- Descrição das soluções contidas no <i>Crystallization Extension Kit for Proteins</i> (Sigma).....	36
Tabela 3- Predição da estrutura secundária da proteína SCF12 usando o programa CONTIN (<i>Dichroweb internet service</i>).....	42
Tabela 4: Amostras submetidas à análise por SAXS.....	50
Tabela 5- Raios de giro calculados para os modelos ligados ao DNA.....	52

Índice de Figuras

Figura 1- Estrutura geral de um aminoácido e a ligação peptídica.....	4
Figura 2- Esquema das estruturas primária, secundária, terciária e quaternária de uma proteína.....	5
Figura 3- Representação da bZIP GCN4 ligada ao sítio de DNA ATF/CREB (entrada PDB 1DGC).....	8
Figura 4- Representação dos heptatos de duas bZIPs interagindo entre si.....	8
Figura 5- Representação de uma onda plana linearmente polarizada.....	11
Figura 6- Representação da onda eletromagnética.....	11
Figura 7- Ilustração esquemática da luz circularmente polarizada.....	12
Figura 8- Luz circularmente polarizada e o efeito de CD.....	14
Figura 9- Espectros característicos de CD para os diferentes tipos de estrutura secundária regular.....	15
Figura 10- a) representação esquemática do experimento de SAXS. b) representação vetorial do espalhamento de Raios X por uma amostra.....	18
Figura 11- Ilustração da função $\rho(r)$ para diferentes formas de partículas.....	20
Figura 12- Porção de um típico cristal de proteína.....	22
Figura 13- Representação de um experimento de cristalização de proteína pela técnica da gota suspensa.....	24
Figura 14- Fotografia de uma placa de cristalização usada nos experimentos para cristalização da proteína SCF5/DNA.....	34
Figura 15- Esquema das placas de cristalização montadas após resultado dos testes iniciais de varredura de condições.....	37
Figura 16- Nova abertura de condições.....	38
Figura 17- Seqüência expressa de aminoácidos da proteína SCF12.....	39
Figura 18- Análise da expressão da proteína de 14,3 kDa SCF12, produzida em <i>E. coli</i> BL21 (DE3) p-Ril e pLys-s.....	40
Figura 19- Expressão e purificação da proteína SCF12.....	41
Figura 20 - Espectro de dicroísmo circular obtido para proteína SCF12.....	42
Figura 21- Seqüência expressa de aminoácidos da proteína SCF5.....	43
Figura 22- Perfil cromatográfico para a SCF5.....	44
Figura 23- Análise por SDS-PAGE 15% das frações da proteína SCF5 obtidas da cromatografia de troca catiônica.....	45
Figura 24- Espectros de CD (elipticidade θ vs λ) em diferentes temperaturas durante o processo de aquecimento da SCF5 de 0 a 90°C.....	46
Figura 25- Espectros de CD (elipticidade θ vs λ) em diferentes temperaturas durante o processo de resfriamento da SCF5 de 80 a 20°C.....	46

Figura 26- Estabilidade térmica da proteína CSF5.....	47
Figura 27- DLS.....	48
Figura 28- a) Representação da estrutura do complexo 1GD2-1L1M computacionalmente preparada. b) Comparação entre as curvas de SAXS experimental da bZIP SCF5/DNA (quadrados com barras de erro) e calculada com o programa CRY SOL para o modelo cristalográfico preparado (linha verde).....	51
Figura 29- a) Curva de espalhamento da proteína SCF5 ligada a DNA (quadrados com barras de erro) e curva ajustada pelo programa GNOM (linha azul). Também são apresentados: b) a função de distribuição de distâncias calculada pelo GNOM e c) o gráfico de Guinier.....	53
Figura 30- Modelo de baixa resolução para o complexo SCF5/DNA obtido pelo programa GASBOR.....	56
Figura 31- Testes iniciais de cristalização: separação de fases.....	57
Figura 32- Fotos dos cristais obtidos pelo refinamento da condição 22 do <i>Crystallization Basic Kit for Proteins</i>	57
Figura 33- Fotos dos cristais obtidos na placa 5.....	58

Lista de abreviações

APS: *Ammonium Persulfate*

blast: *Basic Local Alignment Search Tool*

BSA: *Bovine Serum Albumin*

CD: *Circular Dichroism*

cDNA: DNA complementar

DLS: *Dynamic Light Scattering*

DMF: Dimetil Formamida

DMSO: Dimetilsulfóxido

EDTA: *Ethylenediaminetetraacetic Acid*

HEPES: *4-2-hydroxyethyl-1-piperazineethanesulfonic acid*

IMAC - *Imobilized Metal Affinity Chromatography*

IPTG: Isopropil β -D-1-thiogalactopiranosideo

MES: *2-N-morpholino ethanesulfonic acid*

MRE: *mean residue ellipticity*

mRNA: RNA mensageiro

NADP: Nicotinamida Adenina Dinucleótido Fosfato

NADPH: Nicotinamida Adenina Dinucleótido Fosfato forma reduzida

PEG: Polietileno Glicol

Rg: Raio de Giro

rRNA: RNA ribossomal

SAXS: *Small Angle X-ray Scattering*

SDS-PAGE: *Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis*

SUCEST: *Sugarcane EST Genome Project*

TEMED: *N,N,N',N'-Tetramethylethylenediamine*

tRNA: RNA transportador

1- Introdução

1.1- Cana-de-açúcar

A cana-de-açúcar é uma planta que pertence ao gênero *Saccharum L.*. Há pelo menos seis espécies do gênero, sendo a cana-de-açúcar cultivada um híbrido multiespecífico, recebendo a designação *Saccharum spp.* As espécies de cana-de-açúcar são provenientes do Sudeste Asiático. A planta é a principal matéria-prima para a fabricação do açúcar e álcool (etanol).

A cana é umas das mais antigas culturas sob cultivo, sendo um importante fator na economia e cultura do Brasil desde o início da época colonial. O cultivo da cana-de-açúcar gera uma grande quantia de recursos financeiros ao país, principalmente através da produção de açúcar, álcool anidro (aditivo da gasolina) e o álcool hidratado (combustível para veículos) (Pessoa *et al.*, 2005).

Em 1975, o governo brasileiro lançou o Programa Nacional do Álcool (PROALCOOL) para desenvolver a pesquisa e os meios de produção de álcool como combustível. Houve também o desenvolvimento do primeiro carro movido a etanol. As usinas de açúcar passaram produzir paralelamente o etanol. Atualmente, o Brasil é o maior produtor de cana-de-açúcar do planeta, seguido por Índia e Austrália. Ainda assim, produz mais que o dobro que o segundo colocado (Marris, 2006).

Cerca de 55% da cana-de-açúcar processada vai para a produção de etanol e dos 45% restantes é produzido açúcar. No período de 2003-2004, o país produziu 14,4 bilhões de litros de etanol e 24,2 milhões de toneladas de açúcar. O custo médio de produção de açúcar no Brasil é US\$ 180 por tonelada, enquanto que na Austrália, sai por US\$ 385 por tonelada. Na Europa, que utiliza beterraba, o preço subsidiado de uma tonelada de açúcar alcança US\$ 710 (Pessoa *et al.*, 2005).

Com a implementação dos carros bi-combustível e devido aos altos preços atingidos pelo barril do petróleo no mercado internacional, o mercado sucro-álcooleiro encontra-se em crescimento e há expectativa de crescimento maior

caso os países importadores de petróleo passem a utilizar o etanol como combustível.

Uma das maneiras de atender a crescente demanda por álcool combustível é a utilização de variedades melhoradas de cana-de-açúcar. Uma das ferramentas para isto é a biologia molecular e engenharia genética. Desde o lançamento do projeto SUCEST (*Sugarcane Expressed Sequence Tag Project*) em 1998, aproximadamente 90% dos genes da cana-de-açúcar foram seqüenciados e pesquisas estão sendo desenvolvidas com genes envolvidos em transdução de sinais, resposta a patógenos, acúmulo de açúcar, tolerância a alumínio, ciclo celular, entre outros (Pessoa *et al.*, 2005).

Compreendendo a função de seus genes, espera-se obter variedades transgênicas de cana-de-açúcar ainda mais adequadas ao cultivo e ao processamento, atendendo melhor as necessidades do mercado. Assim, este trabalho tem como objetivo a caracterização estrutural de duas proteínas de cana-de-açúcar do tipo bZIP, envolvidas em processos de regulação da expressão gênica.

1.2- Proteínas

As proteínas são polímeros formados por moléculas menores, os aminoácidos. A análise de um grande número de proteínas, provenientes de um grande número de fontes mostrou que elas são formadas por 20 aminoácidos-padrão. Os aminoácidos comuns são conhecidos como α -aminoácidos porque possuem um grupo amino primário ($-\text{NH}_2$) e um grupo carboxílico ($-\text{COOH}$) ligado ao mesmo carbono, conhecido como carbono- α (Figura 1). A única exceção é a prolina que possui um grupo amino secundário ($-\text{NH}-$). Esses dois grupos ligados ao carbono- α ionizam-se em solução, conferindo aos aminoácidos algumas propriedades características. Os valores de pK dos grupos α -carboxílicos dos 20 aminoácidos estão em torno de 2,2 ao passo que os valores de pK para os grupos α -amino estão próximos a 9,4. Em pH fisiológico (pH $\sim 7,4$) o grupo α -amino está na forma protonada ($-\text{NH}_3^+$) e o grupo α -carboxílico está na forma de base

conjugada (-COO-). Um aminoácido pode, portanto, agir como um ácido ou como uma base. Outra questão é o tipo de grupo R que o aminoácido possui. Basicamente os grupos R são classificados como: (1) grupos R apolares; (2) grupos R polares não carregados e (3) grupos R polares carregados. As características dos grupos R influenciam fortemente as propriedades do aminoácido (Voet *et al.*, 2000).

A polimerização dos aminoácidos para formar as cadeias protéicas corresponde à reação de condensação entre os aminoácidos, onde o carbono do grupo carboxílico de um aminoácido se liga ao nitrogênio do grupo amino do outro aminoácido, resultando uma ligação peptídica do tipo amida e a liberação de uma molécula de água. Essa ligação é denominada ligação peptídica (Figura 1). Polímeros compostos de 2 e 3 aminoácidos são chamados de dipeptídeo e tripeptídeo, respectivamente. Quando o número é maior que 3 temos os oligopeptídeos e se o número de aminoácidos for maior que 10 o polímero é denominado polipeptídeo. Na prática, são chamados apenas de peptídeos. Após serem incorporadas em um peptídeo, os aminoácidos são chamados de resíduos de aminoácidos. Os peptídeos são polímeros lineares, ou seja, cada resíduo participa de duas ligações peptídicas e liga-se a seus vizinhos de forma cabeça-cauda, em vez de formar cadeias ramificadas. Os resíduos das extremidades da cadeia fazem apenas uma ligação peptídica. O resíduo com o grupo amino livre é chamado de amino-terminal ou N-terminal. O resíduo com o grupo carboxílico livre é denominado carboxi-terminal ou C-terminal.

As proteínas estão no centro da ação nos processos biológicos. Praticamente todas as transformações que definem o metabolismo celular são mediadas pela catálise protéica. As proteínas exercem também funções regulatórias, controlando as condições intracelulares e extracelulares e mandando informações para outros componentes da célula. Além disso, as proteínas são componentes estruturais essenciais das células. Apesar da grande importância, existe ainda uma infinidade de proteínas cuja função ainda não é muito bem definida ou é completamente desconhecida. Uma peça chave para decifrar a função de uma dada proteína é o entendimento de sua estrutura.

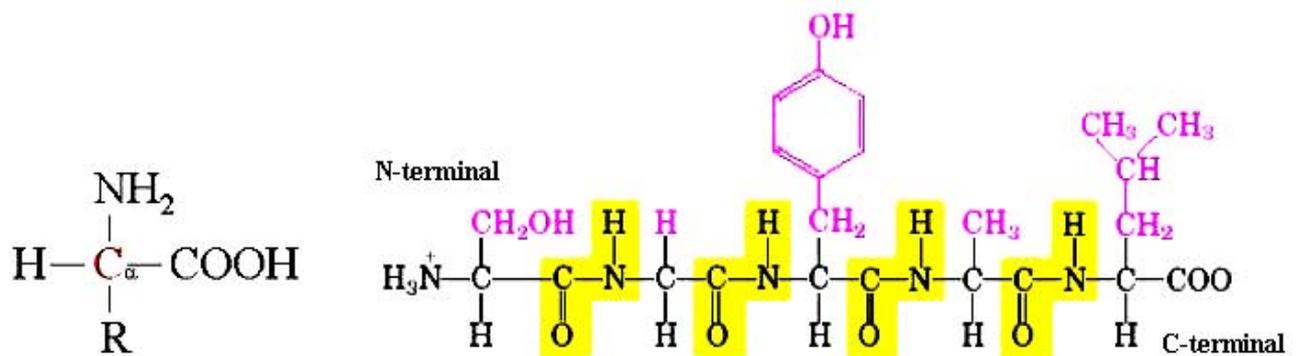


Figura 1- Estrutura geral de um aminoácido e a ligação peptídica. A Figura da esquerda representa a estrutura geral de um aminoácido. Os grupos R diferenciam os 20 aminoácidos-padrão. Na Figura da direita, é ilustrado um peptídeo constituído por 5 aminoácidos. Em vermelho estão os grupos R e as ligações peptídicas estão sombreadas em amarelo (adaptado de Lehninger *et al.*, 2005).

As proteínas são moléculas que contêm uma ou mais cadeias polipeptídicas e podem ser descritas de acordo com níveis de organização, denominados estrutura primária, secundária, terciária e quaternária. A estrutura primária consiste na seqüência de aminoácidos da sua cadeia polipeptídica. Os elementos de estrutura secundária correspondem às hélices- α , folhas- β e voltas (*turn*). Estes elementos são formados devido ao arranjo local dos átomos orientados pelas restrições geométricas (distâncias, ângulos, raios de van der Waals, etc.) e químicas (ligações de hidrogênio, atrações/repulsões iônicas, interações hidrofóbicas, etc.). Esses elementos de estrutura secundária passam a interagir e se organizar espacialmente. O enovelamento destas estruturas secundárias corresponde à estrutura terciária e, quando a proteína corresponde a duas ou mais cadeias polipeptídicas com estrutura terciária definidas (subunidades), o arranjo espacial destas subunidades independentes corresponde à estrutura quaternária da proteína. Na Figura 2 são mostrados exemplos de estrutura primária, secundária, terciária e quaternária de proteína. A estrutura tridimensional de uma proteína, ou seja, o arranjo espacial dos elementos que a compõem, e sua função estão intimamente ligadas (Stryer, 1995; Voet *et al.*, 2000).

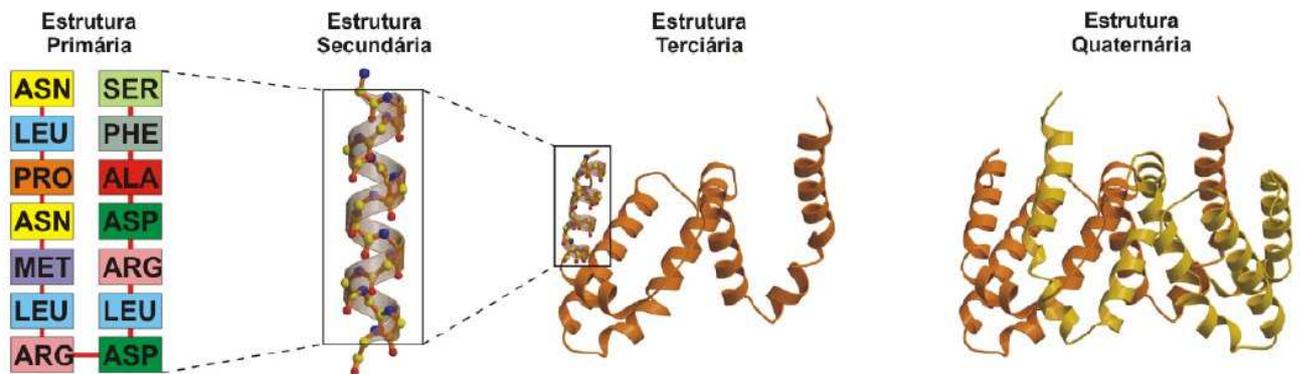


Figura 2- Esquema das estruturas primária, secundária, terciária e quaternária de uma proteína. (adaptado de Lehninger *et al.*, 2000).

1.3- Fatores de transcrição

Os fatores de transcrição regulam a expressão de genes e estão envolvidos no controle de muitos processos intracelulares. Estes fatores se ligam a seqüências específicas de DNA e reconhecem seqüências reguladoras no promotor de um gene, sendo capazes de modular a transcrição através da interação com componentes basais da maquinaria transcrricional (Holstege *et al.*, 1999).

Muitos desses fatores podem ser agrupados em uma pequena quantidade de diferentes famílias de proteínas de acordo com o tipo de domínio de ligação ao DNA que cada uma possui (Jakoby *et al.*, 2002).

Organismos eucariotos possuem três tipos de enzimas responsáveis pela síntese de moléculas de RNA: as RNA polimerases I, II e III. A RNA polimerase I está envolvida na transcrição dos genes de RNA ribossomal (rRNA) enquanto que a RNA polimerase III atua na transcrição de RNAs transportadores (tRNA), RNA ribossomal 5S e alguns outros pequenos RNAs nucleares (Kornberg, 1999).

A RNA polimerase II é responsável pela transcrição de genes que codificam proteínas, sintetizando, então, os denominados RNAs mensageiros (mRNA). Apesar de ser essencial para a transcrição, essa enzima não atua sozinha.

Existem várias outras proteínas que interagem com o DNA e entre si, formando grandes complexos protéicos responsáveis pela regulação da síntese do mRNA, e ainda outras enzimas que interagem com a própria RNA polimerase II, formando um complexo protéico denominado holoenzima.

Algumas dessas proteínas são necessárias para a transcrição de qualquer tipo de gene e são denominadas fatores gerais de transcrição. Outras, denominadas reguladores (ativadores e repressores), regulam a taxa de transcrição de um grupo de genes ou um único gene (Lee *et al.*, 2000). Então, quando esses fatores gerais de transcrição ligam-se de maneira específica, formando um determinado arranjo com o DNA-alvo, a holoenzima se liga a este complexo, formando o complexo de pré-iniciação que pode iniciar a transcrição (Kornberg, 1999).

Os membros da família bZIP (*basic leucine zipper*) são reguladores transcricionais e são encontrados em eucariotos, sendo caracterizados por uma região conservada, rica em resíduos de aminoácidos básicos, que se liga ao DNA-alvo e um zíper de leucina, responsável pela interação entre bZIPs.

As propriedades estruturais das proteínas do tipo bZIP foram verificadas através de estudos cristalográficos realizados com a GCN4, um ativador transcricional de fungos que pertence a esta classe de proteínas (O' Shea *et al.*, 1991; Keller *et al.*, 1995). Foi demonstrado que os monômeros destas proteínas são α -helicoidais e, no domínio zíper de leucina, as hélices são torcidas, produzindo uma estrutura do tipo "*coiled-coil*" (Figura 3). Como ilustra a Figura 4, este zíper de leucina consiste de várias repetições de um heptato (posições "a" até "g") formado por resíduos hidrofóbicos onde um resíduo de leucina ocupa sempre a mesma posição de cada repetição (posição "d"), numa extensão de 30 a 40 resíduos. O tipo de resíduo de aminoácido que ocupa determinada posição no heptato, especialmente nas posições "e" e "g" e nas posições "a" e "d", proporciona as interações e o correto direcionamento entre as α -hélices, determinando a especificidade da associação de subunidades iguais (homodímeros) ou diferentes (heterodímeros) (Lupas, 1996). Nas posições "a" e "d" estão situados aminoácidos hidrofóbicos e as posições "e" e "g" são ocupados

por aminoácidos carregados (Hurst, 1995). A interação entre a GCN4 com o DNA alvo se faz pelo contato de 13 aminoácidos com os fosfatos das ligações fosfodiéster presentes entre as bases. A inclinação das hélices- α (região básica) advindas do zíper de leucinas dimerizado, semelhantes a uma pinça, permite que os aminoácidos encostem-se à fenda maior (*major groove*) do DNA alvo (Figura 3), permitindo interações do tipo van der Waals e ligações de hidrogênio entre a proteína e os fosfatos da molécula de DNA (König and Richmond, 1993).

Algumas bZIPs, como a GCN4, podem formar apenas homodímeros, enquanto outras, como o fator de transcrição humano c-Fos (*proto-oncogene protein c-fos*) formam heterodímeros. Existem, porém, bZIPs que podem formar tanto homodímeros quanto heterodímeros, como é o caso da proteína humana c-Jun (*proto-oncogene c-jun*) (Smeal *et al.*, 1989; Kouzarides and Ziff, 1989).

O processo de dimerização pode regular a interação entre uma bZIP e o complexo de pré-iniciação ou ainda, modular a interação com o DNA, constituindo um mecanismo de controle da expressão gênica (Chiu *et al.*, 1989).

1.4- bZIPs em plantas

Em geral, a maioria dos genes de plantas são regulados no nível da transcrição e destes, a maior parte é controlada por elementos flanqueadores 5' de seqüências de DNA, presentes nas regiões promotoras. Análises moleculares de genes que respondem a diversos estímulos em plantas identificaram *cis-acting elements*, ou seja, seqüências de DNA que regulam a expressão de genes no mesmo cromossomo, que possuem o motivo ACGT comum (Foster *et al.*, 1994). As bZIPs de plantas se ligam preferencialmente em seqüências de DNA que possuem o motivo ACGT (Izawa *et al.*, 1993).

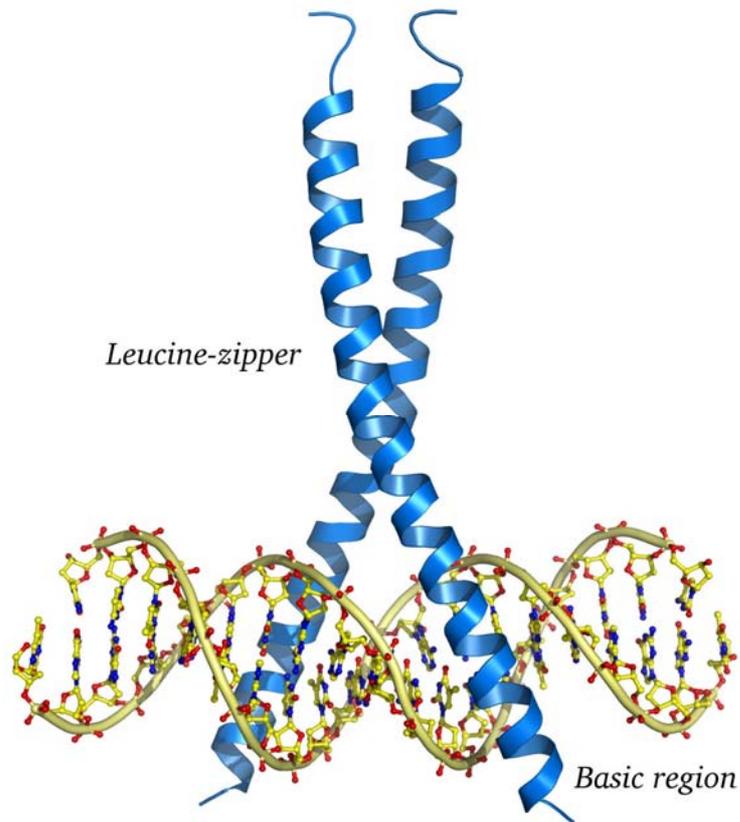


Figura 3- Representação da bZIP GCN4 ligada ao sítio de DNA ATF/CREB (entrada PDB 1DGC).
 Figura preparada com PYMOL (<http://pymol.sourceforge.net>).

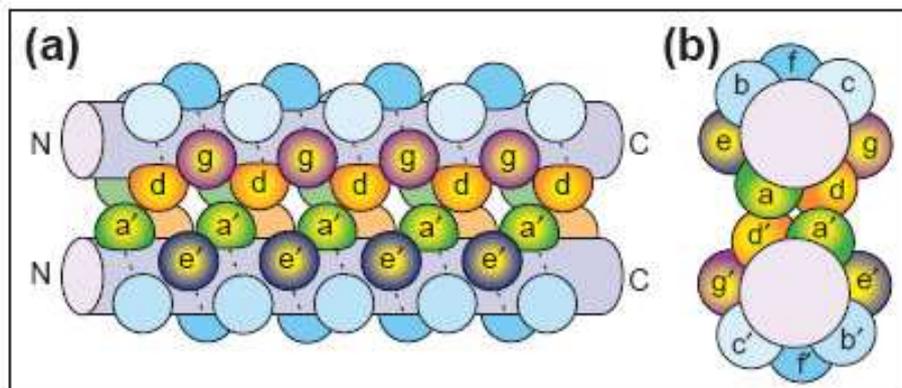


Figura 4- Representação dos heptatos de duas bZIPs interagindo entre si. As posições “a” e “d” são ocupadas por aminoácidos hidrofóbicos (em “d” está situada uma leucina) e nas posições “e” e “g” estão situados aminoácidos carregados. (Fonte: adaptado de Fong *et al.*, 2005).

Adotando a nomenclatura usada na GCN4, um ativador transcricional de levedura, na qual os dois nucleotídeos centrais C e G, do elemento ACGT são designados -0 e +0, respectivamente (Oliphant *et al.*, 1989), temos que o nucleotídeo da posição +2 define o *status* do elemento como *A-box*, *C-box*, *T-box* e *G-box* (Giuliano *et al.*, 1988). Por exemplo, o elemento palindrômico hexamérico **CACGTG** corresponde a uma seqüência *G-box*.

Em Izawa *et al.*, 1993, análises qualitativas e quantitativas da atividade de ligação de bZIPs de plantas ao DNA demonstraram que as proteínas analisadas apresentam elevada afinidade de ligação aos elementos *A-box*, *C-box* e *G-box*, sendo que os nucleotídeos que ocupam as posições ± 3 e ± 4 também interferem na afinidade de ligação. A interação dos fatores de transcrição bZIPs com esses *cis-acting elements* pode proporcionar a máxima ativação de um gene ou o bloqueio da transcrição do mesmo.

A família das bZIPs representa uma das três maiores famílias de fatores de transcrição encontradas na espécie modelo *Arabidopsis thaliana* (Riechmann e Ratcliff, 2000). Essas proteínas participam de um amplo conjunto de processos, como defesa contra patógenos, desenvolvimento de sementes, alongamento celular, desenvolvimento de órgãos, entre outros (Jackoby *et al.*, 2002).

O seqüenciamento do genoma da *Arabidopsis thaliana* permitiu o acesso ao conjunto completo de genes que codificam bZIPs nessa planta, o que possibilitou organizar os fatores bZIP de *Arabidopsis* em 13 sub-famílias de proteínas relacionadas filogeneticamente. Após o projeto de *Data Mining* do SUCEST (*Sugarcane EST Genome Project*, <http://sucest.lad.ic.unicamp.br>) essa classificação para *Arabidopsis* foi usada para identificar e classificar cDNAs que codificam bZIPs em cana-de-açúcar (Vincentz *et al.*, 2001).

Neste trabalho foram utilizadas a seqüência de aminoácidos referentes à região básica e ao zíper de leucinas de duas bZIPs de cana-de-açúcar identificadas no SUCEST: a SCF12 e a SCF5.

A SCF12 (*sugarcane bZIP family 12*) possui 124 aminoácidos. Uma análise comparativa *blastp* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) entre a seqüência primária desta com outras bZIPs mostrou que a SCF12 possui 95% de identidade

com a Atbzip69, uma proteína de *Arabidopsis thaliana* pertencente ao grupo I, segundo Jakoby *et al.*, 2002. Neste grupo estão contidas bZIPs envolvidas na biosíntese de giberelinas e no desenvolvimento vascular da planta. No caso da SCF5 (*sugarcane bZIP family 5*) que possui 104 aminoácidos, a análise *blastp* mostrou que a SCF5 possui 53% de identidade com a Atbzip42, que segundo a classificação pertence ao grupo S, cujos membros estão envolvidos no balanço de consumo e suprimento de carboidratos. Em Schlögl *et al.*, 2004, a SCF5, designada como SCBZIP1, é descrita como uma bZIP com grande afinidade de ligação a elementos *G-box* e *C-box*, e que o fato do mRNA referente a SCF5 ser expresso no início do crescimento da planta sugere que este fator de transcrição pode estar envolvido em diversos processos fisiológicos do desenvolvimento da cana-de-açúcar.

1.5- Técnicas para caracterização estrutural de proteínas

Para a caracterização estrutural das proteínas, foram utilizados o espalhamento dinâmico de luz (DLS, do inglês *Dynamic Light Scattering*), dicroísmo circular (CD, do inglês *Circular Dichroism*) e espalhamento de Raios X a baixos ângulos (SAXS, do inglês *Small-Angle X-ray Scattering*), que são descritos a seguir. Foram também realizados testes de cristalização com vistas à determinação da estrutura atômica das proteínas de interesse.

1.5.1- Dicroísmo circular (CD)

Para compreendermos o fenômeno que está envolvido na técnica de Dicroísmo Circular é necessário antes introduzirmos o conceito de luz natural e luz polarizada. Pela teoria ondulatória, a luz natural, não polarizada, é descrita como uma onda eletromagnética que, a partir de uma fonte, se propaga para todas as direções. A onda eletromagnética possui uma componente de campo magnético (**B**) e outra de campo elétrico (**E**), que são perpendiculares. Quando a onda percorre somente um plano, dizemos que a onda é linearmente polarizada e este

tipo de radiação pode ser obtida pelo uso de lentes e filtros específicos. Na Figura 5 é ilustrada uma onda eletromagnética plana linearmente polarizada, com suas duas componentes, elétrica e magnética (Halliday D. *et al.*, 1995)

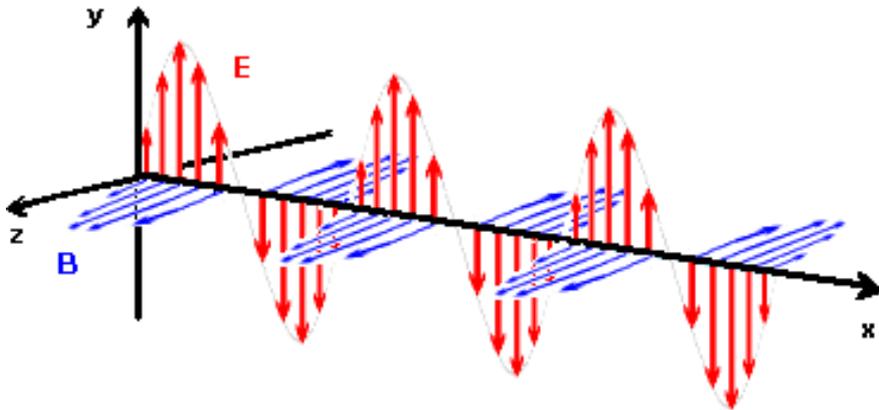


Figura 5- Representação de uma onda plana linearmente polarizada. As componentes de campo elétrico e magnético estão representadas em vermelho e azul, respectivamente.

Na prática podemos representar a luz de forma simplificada, apenas pela sua componente de campo elétrico, como mostrado na Figura 6.

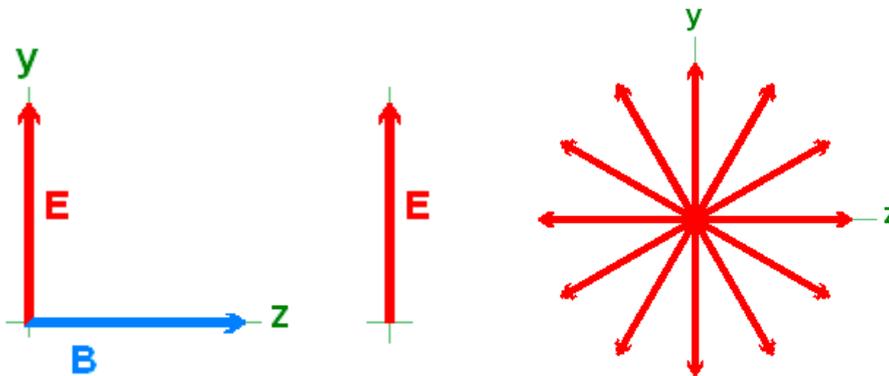


Figura 6- Representação da onda eletromagnética. À esquerda temos a onda eletromagnética linearmente polarizada, observada na direção de propagação e, ao centro, a mesma onda representada de forma simplificada, somente com a componente de campo elétrico. À direita, representação da onda eletromagnética não-polarizada.

Outro tipo de polarização da luz é a polarização circular. A luz circularmente polarizada é produzida quando duas ondas eletromagnéticas linearmente polarizadas de igual amplitude que se propagam em planos perpendiculares com diferença de fase de 90° . Para exemplificar vamos considerar apenas a componente campo elétrico e utilizar o sistema de coordenadas x, y, z da Figura 5, onde existem os planos perpendiculares xy, xz e yz . Se a onda é propagada no plano xy e outra no plano xz , na direção do eixo x e com diferença de fase de 90° , o vetor campo elétrico resultante ao longo do eixo de propagação desenha a forma de uma mola ou espiral de caderno e para um observador que olha na direção da fonte de radiação de frente para o eixo de propagação, o vetor campo elétrico forma um círculo, mostrado na Figura 7. Dependendo se a onda de um plano estiver adiantada ou atrasada em relação a outra, o mesmo observador veria círculo girar no sentido horário ou no sentido anti-horário. Se estiver no sentido horário, a luz é chamada luz circularmente polarizada à direita. No sentido anti-horário, é denominada luz circularmente polarizada à esquerda.

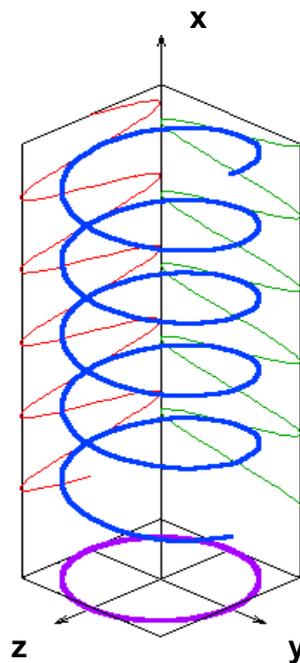


Figura 7- Ilustração esquemática da luz circularmente polarizada. Em azul, propagando-se ao longo do eixo x temos a luz circularmente polarizada, que descreve a forma de uma mola (adaptado de <http://en.wikipedia.org/wiki/Polarization>).

Moléculas assimétricas interagem com luz circularmente polarizada de maneira desigual. Essa diferença na absorção faz com que a amplitude da componente mais absorvida seja menor do que a amplitude da menos absorvida e, como consequência, a projeção da amplitude resultante descreve uma elipse, como mostra a Figura 8. Essa diferença na absorbância das componentes pode ser medida por um espectropolarímetro de CD, conforme a equação 01:

$$\Delta A = A_L - A_R \quad (01),$$

onde A_L corresponde a absorbância da luz circularmente polarizada à esquerda e A_R corresponde a absorbância da luz circularmente polarizada à direita.

Mas geralmente o sinal de absorbância é convertido em elipticidade (θ), expresso em miligraus (Kelly and Price, 1997). Deve ser observado que:

$$(\theta) = \tan^{-1} (b/a) \quad (02),$$

onde b e a são respectivamente o menor e o maior raio que compõem a elipse da Figura 8. Outra unidade muito usada é a elipticidade molar

$$\theta_m = 100 \times \theta / m \times d \quad (03),$$

onde m é a concentração molar da amostra e d é o caminho óptico que a luz percorre pela amostra.

Na faixa do UV distante (170 a 260 nm), no caso de proteínas os cromóforos para a luz circularmente polarizada são as ligações peptídicas. Nesta região ocorre a fraca transição $n \rightarrow \pi^*$ centrada ao redor de 220 nm e a mais intensa transição $\pi \rightarrow \pi^*$ ao redor de 190 nm. Os diferentes tipos de estrutura secundária regular resultam em um espectro de CD característico (Kelly *et al.*, 2005), mostrado na Figura 9.

Existem alguns algoritmos que fornecem uma estimativa da composição de estrutura secundária de uma proteína. A maioria deles utiliza bases de dados que

contêm espectros de CD de proteínas cuja estrutura cristalográfica já está resolvida (Kelly *et al.*, 2005). No caso do UV próximo (260 a 320 nm), os cromóforos preferenciais são as cadeias laterais aromáticas, ligações de dissulfeto e grupos prostéticos (Kelly e Price, 1997). Assim, o UV distante detecta a estrutura secundária e o UV próximo é sensível à estrutura terciária.

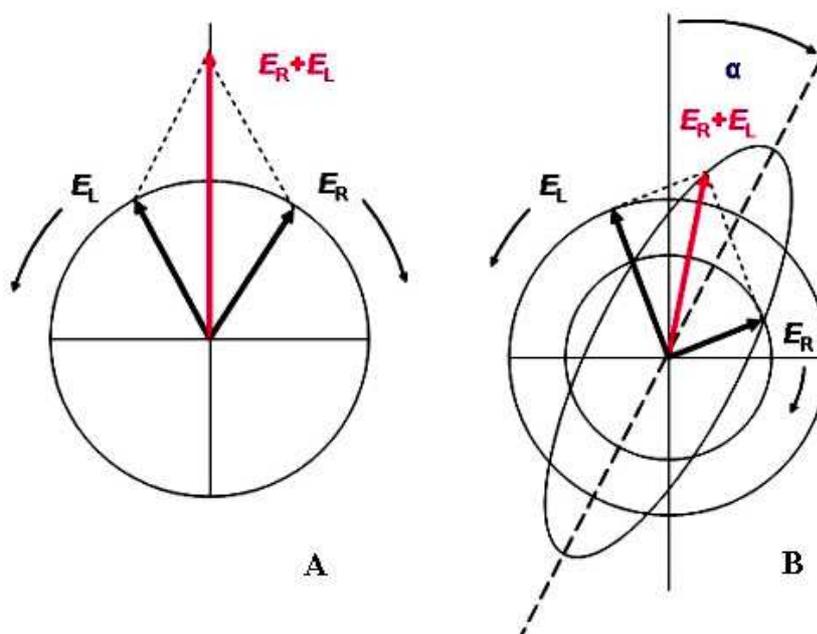


Figura 8- Luz circularmente polarizada e o efeito de CD: A) Superposição das duas componentes circularmente polarizadas de igual amplitude pode representar uma onda eletromagnética linearmente polarizada. B) Absorção diferente das componentes polarizadas direita e esquerda projeta uma elipse.

Apesar de ser possível utilizar dados de CD para quantificar a composição de estruturas secundárias de uma proteína (conteúdo de α -hélices, folhas- β e estruturas randômicas) (Provencher e Glöckner, 1981), esta técnica é principalmente utilizada para estudar mudanças conformacionais e desenvolvimento de proteínas (Fasman, 1996). Uma outra utilização é o

acompanhamento do processo de desnaturação (*unfolding*) e renaturação (*folding*) da proteína, quando esta é submetida a condições desnaturantes (aquecimento, adição de desnaturantes, etc.).

Uma informação útil é a temperatura ou concentração de desnaturante em que 50% das moléculas perdem seu enovelamento. Esses parâmetros podem ser usados para dar medidas comparativas da estabilidade de uma única proteína ou mesmo entre proteínas diferentes (ou mutantes de uma mesma proteína) quando ocorrem alterações no ambiente (Kelly and Price, 1997).

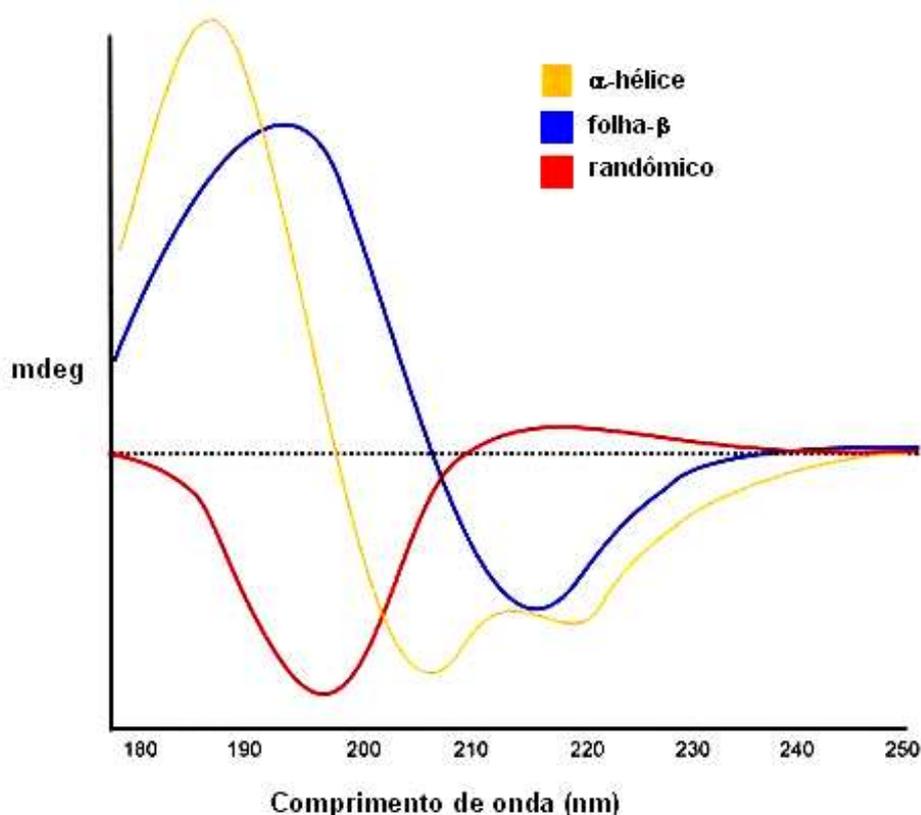


Figura 9- Espectros característicos de CD para os diferentes tipos de estrutura secundária regular.

1.5.2- Espalhamento dinâmico de luz

O espalhamento dinâmico de luz (DLS) é uma técnica pela qual pode-se obter o coeficiente de difusão de uma partícula em solução e o seu raio hidrodinâmico, geralmente calculado com base em modelos esféricos. Esta

técnica pode fornecer informações estruturais da partícula em um curto espaço de tempo, requerendo pequenos volumes de amostra (aproximadamente 60 μL) em concentrações não muito elevadas (de 1 a 10 mg/mL). A técnica é muito sensível a pequenas quantidades de agregados (Papish *et al.*, 2002) e pode-se obter informações sobre tamanho e forma de moléculas em solução entre 1 a 1000 nm de diâmetro (Murphy, 1997). No caso de proteínas, é possível estudar mudanças conformacionais, estado de agregação, estabilidade estrutural, entre outros.

No DLS, um feixe monocromático (~ 800 nm) incide sobre a amostra e a flutuação da intensidade da luz espalhada (numa escala de microsegundos) é registrada. Partículas em solução movem-se ao acaso (movimento Browniano), causando diferenças de concentração em um determinado volume da amostra em função do tempo. Essas flutuações da concentração em partes da amostra determinam flutuações na intensidade de espalhamento medida, ou seja, essas flutuações na intensidade de espalhamento são observadas como variações da concentração no volume de espalhamento no intervalo de tempo medido. Partículas maiores apresentam coeficiente de difusão menor em relação a partículas menores e as flutuações da intensidade de espalhamento ocorrerão mais lentamente (Pecora, 1985). Do coeficiente de difusão se obtém o raio hidrodinâmico da partícula, que relaciona difusão e tamanho da molécula. O raio hidrodinâmico não corresponde efetivamente ao raio da molécula em solução, mas é calculado usando-se uma esfera rígida hipotética cuja difusão seria a mesma da molécula. Na prática, moléculas não são esferas perfeitas, mas o raio calculado a partir das propriedades difusionais é indicativo do tamanho aparente da molécula. Partículas com maior massa teriam maior tamanho e, conseqüentemente, maior raio hidrodinâmico que moléculas de menor massa (Murphy, 1997). Assim, podemos relacionar a massa de uma partícula (que é proporcional ao raio hidrodinâmico) com a intensidade de luz espalhada pela amostra. Como resultado final, obtemos um gráfico de porcentagem de massa das espécies em solução contra o raio hidrodinâmico, a partir do qual é possível averiguar o estado de agregação das diferentes espécies presentes.

1.5.3. Espalhamento de Raios X a Baixos Ângulos (SAXS)

O Espalhamento de Raios X a Baixos Ângulos é uma importante técnica para o estudo de macromoléculas em solução, com massa variando de alguns quilodaltons a vários megadaltons. Excelentes revisões sobre o assunto estão disponíveis, de modo que nos restringiremos a apresentar as principais equações utilizadas neste trabalho (Koch *et al.*, 2003; Svergun and Koch, 2003; Svergun, 2007). Em um experimento de SAXS, a amostra é exposta aos Raios X ($\lambda \sim 1,5 \text{ \AA}$) e a intensidade de espalhamento, $I(q)$, é anotada como proporcional à média da amplitude do espalhamento para todas as orientações, onde \mathbf{q} é o vetor de espalhamento, equação 04.

$$I(q) \propto \langle A^2(\vec{q}) \rangle_{\Omega} \quad (04)$$

O módulo do vetor de espalhamento depende do comprimento de onda da radiação (λ) e do ângulo de espalhamento (2θ), equação 05.

$$|\vec{q}| = \frac{4\pi}{\lambda} \sin\theta \quad (05)$$

O ângulo de espalhamento corresponde ao ângulo formado entre as direções dos feixes incidente (q_0) e espalhado (q_1), Figura 10.

As orientações e posições randômicas das partículas fazem com que o espalhamento seja isotrópico. Com isso, no caso de soluções monodispersas de partículas que não interagem entre si, a intensidade total é proporcional à média temporal e de todas as orientações para uma única partícula (Svergun and Koch, 2002).

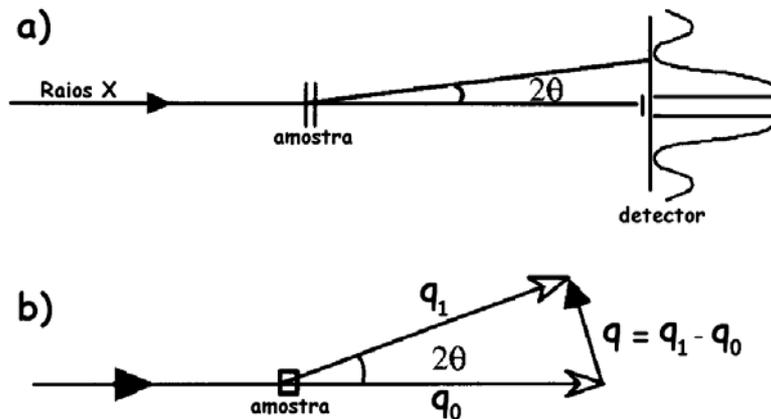


Figura 10- a) representação esquemática do experimento de SAXS. b) representação vetorial do espalhamento de Raios X por uma amostra (adaptado de Vachette and Svergun, 2000).

A técnica de SAXS permite obter informações sobre várias características das macromoléculas, como raio de giro, massa molecular, distância máxima e forma da partícula espalhadora (Glatter and Kratky, 1982).

O raio de giro de uma macromolécula (R_g) é semelhante ao “raio de inércia” da mecânica clássica, sendo definido como a raiz quadrada da distância quadrática média, r , entre os elementos de volume e o centro de massa eletrônico do objeto espalhador, equação 06.

$$R_g^2 = \frac{\int r^2 dv}{v} \quad (06)$$

O raio de giro dá uma medida do grau de distribuição de massa e do tamanho da partícula estudada. O valor de R_g pode, alternativamente, ser obtido a partir da porção da curva de espalhamento para valores de q muito pequenos onde é válida a aproximação de Guinier, equação 07 (Guinier and Fournet, 1955).

$$I(q) \cong I(0) \exp \left(- \frac{1}{3} R_g^2 q^2 \right) \quad (07)$$

Esta equação é válida para sistemas de partículas idênticas e diluídas na região $q \rightarrow 0$ ($1 \leq R_g q \leq 1,3$). Representando a intensidade em escala $\ln I$ versus q^2 , espera-se um comportamento linear com coeficiente angular igual a $-\frac{1}{3}R_g^2$, conforme a equação 8

$$\ln I(q) = \ln I(0) - \frac{1}{3} R_g^2 q^2 \quad (8)$$

Um modo simples de obter a massa molecular de uma macromolécula em solução através da técnica de SAXS é realizando uma extrapolação da intensidade dos Raios X espalhados para o valor de $q = 0$, I_0 . Então, estima-se o valor de I_0 para uma proteína de massa molecular (m) conhecida, normalmente BSA ou lisozima (*padrão*), e com os dados de SAXS para a proteína de interesse (*amostra*) pode-se determinar sua massa molecular pela equação 09.

$$M = \frac{I(0)_{amostra} / C_{amostra}}{I(0)_{padrão} / C_{padrão}} \times m \quad (09),$$

onde $C_{amostra}$ é a concentração da amostra em mg/mL, $C_{padrão}$ é a concentração do padrão (BSA, lisozima) em mg/mL, e m é a massa molecular do padrão em kDa.

À baixos ângulos, os Raios X são insensíveis à estrutura interna da molécula e o espalhamento é dado essencialmente pela forma da partícula. A função $p(r)$, também chamada de “função de distribuição de distâncias”, como o próprio nome sugere, descreve a distribuição de distâncias, r , que podem ser encontradas entre um par de volumes na partícula.

O formato da função de distribuição de distâncias está diretamente relacionado com a forma da partícula, por exemplo, no caso de uma partícula esférica sua $p(r)$ apresenta uma distribuição simétrica ao longo do eixo das distâncias, sendo o diâmetro da esfera igual à distância máxima ($D_{máx}$) entre dois centros espalhadores da partícula (Figura 11).

De posse da curva de espalhamento ou da função $p(r)$ obtida, pode-se recuperar a estrutura tridimensional de baixa resolução da partícula (Svergun e Koch, 2002). De maneira geral, os dados de SAXS coletados são inicialmente tratados utilizando o programa PRIMUS (Konarev *et al.*, 2003). Este programa permite realizar a subtração do espalhamento de Raios X advindo do tampão utilizado e outras facilidades, tais como, média de diferentes medidas, estimativa da intensidade de espalhamento em $q = 0$, I_0 , etc. Uma vez que os dados foram inicialmente tratados é possível avaliar a função $p(r)$ da partícula através do programa GNOM (Svergun, 1992). E, empregando os programas DAMMIM (Svergun, 1999) ou GASBOR (Svergun *et al.*, 2001) pode-se recuperar um modelo de baixa resolução (envelope de SAXS) para o sistema em estudo. Estes programas utilizam a curva de espalhamento para reconstruir a forma da partícula em solução.

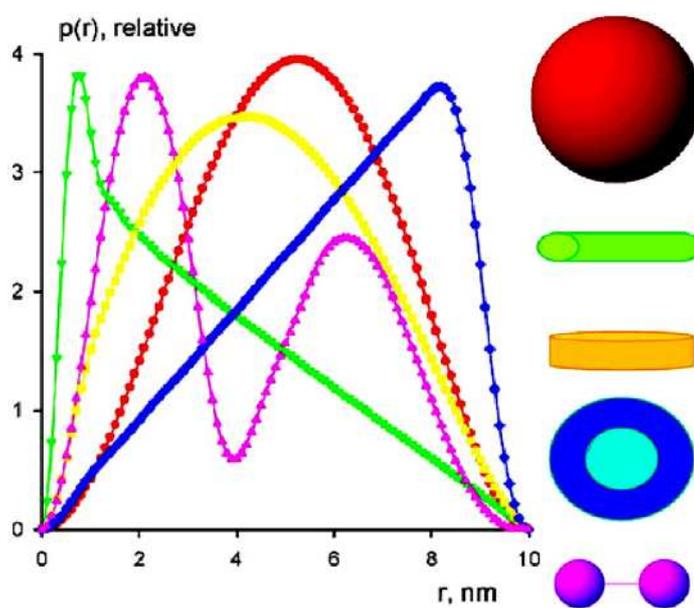


Figura 11- Ilustração da função $p(r)$ para diferentes formas de partículas (adaptado de Svergun & Koch, 2003).

1.5.4- Cristalização de proteínas

Os átomos que compõem uma molécula são separados por distâncias da ordem de 1 Å (ou seja, 0,1 nm). Sendo assim, para se obter informações estruturais desta ordem, é necessária a utilização de ondas eletromagnéticas de comprimento de onda da mesma ordem de grandeza, ou seja, é necessário usar Raios X.

Para os experimentos de cristalografia, tanto de macro como de pequenas moléculas, é necessário antes ter um cristal do composto em estudo. Uma vez que o cristal de proteínas é obtido, este é levado até uma fonte de Raios X. Os raios atingem o cristal e são espalhados pelos elétrons, formando um padrão de difração. Com o uso de computadores, esses dados são processados e são usados para construir um modelo da proteína.

Cristais são arranjos de moléculas ordenados e periódicos, caracterizados por operações de simetria e parâmetros de rede que exatamente definem a disposição e periodicidade das unidades fundamentais (cela unitária) das quais são compostos. A Figura 12 apresenta, com propósitos ilustrativos, o modelo para uma porção de um típico cristal de proteína.

1.5.4.1- Métodos de cristalização

Para obter cristais de um composto, suas moléculas devem ser levadas à supersaturação, estado termodinamicamente instável que pode levar à formação de uma fase cristalina ou amorfa, quando retorna ao equilíbrio. A cristalização de uma macromolécula ocorre pela diminuição lenta de sua solubilidade. Se este processo ocorrer muito rapidamente, ocorre a precipitação (McPherson, 1999).

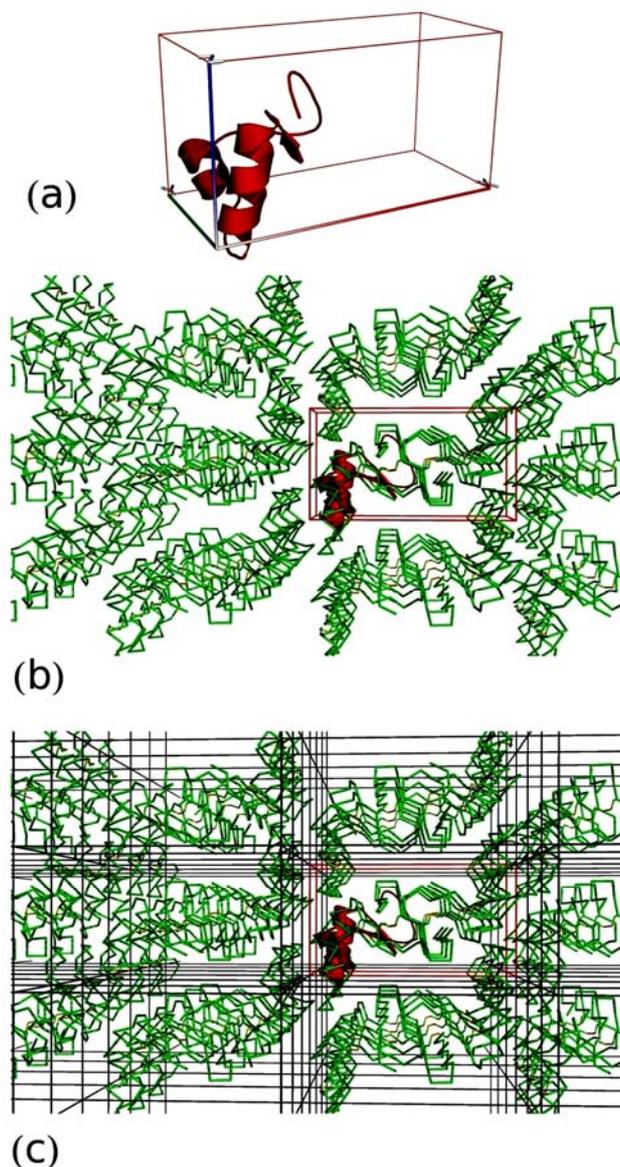


Figura 12- Porção de um típico cristal de proteína (entrada 1EJG do PDB, <http://www.rcsb.org/pdb>, Jelsch *et al.*, 2000), no caso, a pequena proteína crambina de crambe (*crambe abyssinica*; planta com potencial para produção de matéria-prima para biodiesel). **(a)** ilustração do monômero presente em uma das duas unidades assimétricas que compõem a cela unitária (em vermelho) do cristal (grupo de espaço P21, com $a=40.824$ Å, $b=18.498$ Å, $c=22.371$ Å, $\alpha =90.00^\circ$ $\beta =90.47^\circ$ $\gamma =90.00^\circ$); **(b)** visualização do resultado da aplicação das operações de simetria de grupo e operações de translação nos três eixos (moléculas simétricas em verde). Observe-se as regiões de contato intermoleculares e os canais de solvente existentes no cristal; **(c)** representação de algumas celas unitárias do cristal. Figura preparada com PyMOL (<http://pymol.sourceforge.net/>).

O processo de cristalização pode ser dividido em três partes: nucleação, crescimento e cessação do crescimento. A nucleação ocorre quando, numa região de supersaturação, aglomerados são simultaneamente formados e solubilizados até que consigam transpor uma barreira de energia de ativação e começam a crescer, incorporando, de maneira adequada, moléculas de proteína do meio. Assim, o sistema antes supersaturado passa a um estado insaturado (mais estável) e, quando a incorporação de moléculas é igual à saída das moléculas do cristal, o crescimento do mesmo cessa (McPherson, 1999).

Existem diversos métodos para a obtenção de cristais de proteínas, todas eles têm o objetivo de trazer a solução de macromoléculas a um estado de supersaturação. Dentre os métodos, podemos citar: método de cristalização em “*batch*”; o método de cristalização por diálise; e o método por difusão de vapor.

1.5.4.2- Método de difusão de vapor: técnica da gota suspensa

No caso do método de difusão de vapor, atualmente encontram-se materiais para cristalização (placas, soluções de cristalização, reagentes, etc) de diversas marcas comerciais disponíveis no mercado, sendo possível executar e reproduzir experimentos de maneira simples e rápida. Desta forma, o método de difusão de vapor é um dos mais utilizados para a obtenção de cristais de proteína. A Figura 13 ilustra um típico experimento de cristalização pela técnica da gota suspensa. O reservatório contém a solução na condição desejada (pH, concentração de precipitantes, sais, aditivos, etc.). Este é tampado com uma lamínula e o sistema é vedado com graxa de silicone. Porém, antes de tampar o reservatório, uma gota composta pela solução da proteína e pela solução do reservatório (normalmente na proporção de volume 1:1) é colocada sobre a lamínula e esta é utilizada de modo que a gota fique dentro do reservatório, como mostra a Figura 13. A concentração dos componentes da gota (sais, precipitantes, etc) é menor que a concentração no reservatório e o volume da solução (~1000 μL) é muito maior que o volume da gota (~10 μL). Essa diferença faz com que a gota comece a perder solvente até que o sistema atinja equilíbrio, e a proteína

pode atingir um estado de supersaturação que poderá levar a formação do cristal.

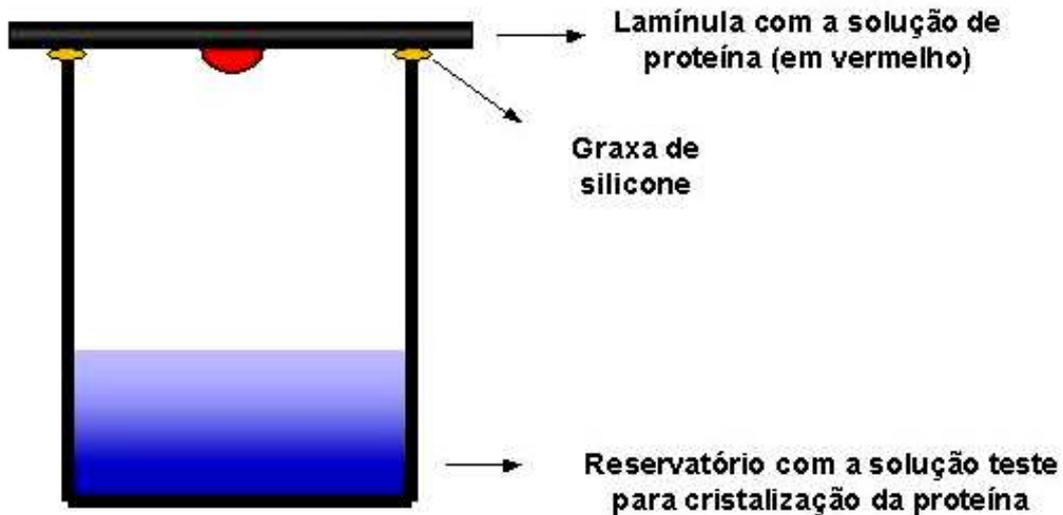


Figura 13- Representação de um experimento de cristalização de proteína pela técnica da gota suspensa.

Geralmente, a cristalização de uma proteína ocorre em duas etapas. O primeiro passo consiste na varredura de uma grande quantidade de condições aleatórias, onde se muda o pH do meio, precipitante, força iônica do meio, aditivos, concentração dos componentes, entre outros. Uma vez que se tem indício de que alguma condição é favorável à formação de cristais da proteína, segue-se ao segundo passo, onde a condição anteriormente determinada é otimizada variando-se os parâmetros sistematicamente, a fim de obter os cristais na qualidade requerida para os experimentos de cristalização. Antes de submeter uma proteína a experimentos de cristalização, é necessário verificar se a proteína está corretamente enovelada (CD) e se está homogênea estruturalmente (DLS, gel-filtração ou gel nativo).

2- Objetivos

Este trabalho teve como objetivo principal obter informações sobre a estrutura de duas proteínas de cana-de-açúcar do tipo bZIP: SCF12 e SCF5. Para isso, alguns objetivos específicos foram traçados:

- expressar as proteínas em bactérias transgênicas (*E. coli*);
- purificar as proteínas por cromatografia líquida;
- realizar estudos por dicroísmo circular de modo a verificar propriedades de estrutura secundária;
- realizar estudos de SAXS e obter um modelo de baixa resolução.

3- Materiais e métodos

A seguir estão descritas separadamente as metodologias e materiais utilizados nos experimentos realizados para as proteínas SCF12 e SCF5. Algumas técnicas complementares, porém, foram realizadas para ambas proteínas:

- Diálise: Para os experimentos de caracterização das bZIPs foi necessário purificá-las a partir de extratos bacterianos. Sendo assim, foi necessário realizar a diálise das amostras para retirar os reagentes advindos das etapas de purificação e adequar o tampão solvente aos experimentos posteriores. Todas as diálises mencionadas neste trabalho foram realizadas utilizando um volume de solução de diálise 100 vezes superior ao volume da amostra, deixadas sob agitação (com barra magnética e agitador magnético) durante uma noite a 4°C.
- Sistema de concentração de proteínas: as amostras de SCF5 e SCF12 foram deixadas na concentração adequada aos experimentos posteriores utilizando o sistema Amicon (Millipore, EUA) com membrana de exclusão de 5000 Da.
- Determinação da concentração de proteína: para a determinação da concentração de proteína de uma amostra foi usado o método de Bradford (Bradford, 1976), utilizando amostras de BSA como padrões para a curva de calibração. As amostras e padrões foram analisadas em espectrofotômetro UV/VIS U-3000 (Hitachi, Japão) com comprimento de onda de 595 nm.
- Determinação da concentração de DNA: a concentração de DNA para os experimentos de ligação com a SCF5 foi determinada por espectrofotometria UV com comprimento de onda a 260 nm.

3.1- Proteína SCF12

O vetor de expressão foi gentilmente cedido pelo Dr. Paulo Sérgio Schlögl. A metodologia para a obtenção do mesmo está resumidamente descrita abaixo.

Do programa SUCEST (<http://sucest.lad.ic.unicamp.br>), foi obtido o cDNA correspondente à região básica e ao zíper de leucinas da proteína SCF12 (clone

RZ2034D03). A seqüência foi amplificada por PCR (*polymerase chain reaction*), clonada no vetor de clonagem pGEM-T Easy (Promega, USA) e subclonada no vetor de expressão pET 28a (Novagen, USA), entre os sítios de restrição *NdeI* e *BamHI*. Foram utilizados os *primers*: direto 5'CAAGAAGCATATGTCCGACGC'3 e reverso 5'GTGGATCCTCATGAGGGCATA'3.

3.1.1- Ensaios de expressão da proteína SCF12 em *E. coli*

3.1.1.1- Influência da cepa de *E. coli* utilizada na produção da SCF12

Para avaliar a expressão da SCF12, foram transformadas bactérias competentes disponíveis em nosso laboratório, do tipo *E. coli* BL21 (DE3) pRil e pLysS (Novagen, EUA) através do método de PEG (Sambrook *et al.*, 1989). As mesmas foram crescidas em placas contendo meio LB sólido (Sambrook *et al.*, 1989) com os agentes de seleção canamicina (50 µg/mL) e cloranfenicol (100 µg/mL) durante 16 h em estufa a 37° C. Uma colônia de cada tipo de célula foi escolhida e adicionada a dois diferentes tubos de cultura contendo 5 mL de meio LB líquido com os mesmos agentes de seleção anteriores. Para a sua multiplicação, as bactérias foram deixadas num agitador a 37° C e 250 rpm, por 16 h. Este meio contendo as bactérias multiplicadas é denominado pré-inóculo. A densidade óptica dos dois pré-inóculos foram determinadas em espectrofotômetro a 600 nm (DO₆₀₀). Em 2 *erlenmeyers* com 20 mL de meio LB e os agentes de seleção, foram adicionados os pré-inóculos até DO₆₀₀ igual a 0,1. Os novos meios de cultura foram deixados a 37° C sob agitação de 250 rpm até DO₆₀₀ 0,6. Após isso, induziu-se com IPTG (Isopropil β-D-1-Tiogalactopiranosídeo) para concentração final de 1 mmol/L e os meios de cultura foram deixados a 37° C e agitação de 250 rpm por 4 horas. De cada *erlenmeyers* foi coletado 1 mL de meio de cultura que, após, foi centrifugado a 15000 g por 1 minuto. Os sobrenadantes foram descartados e os precipitados bacterianos (*pellets*) foram utilizados para análise em gel de poliacrilamida do tipo SDS-PAGE a 15% (Laemmli, 1970).

3.1.1.2- Influência da temperatura na indução

Para avaliar a influência da temperatura na indução da SCF12, as bactérias e condições descritas no item 3.1.1.1, foram induzidas a 28 e 37 °C. Os *pellets* bacterianos foram analisados em SDS-PAGE 15%, conforme o item 3.1.2, descrito abaixo.

3.1.2- Análise das amostras em gel de poliacrilamida (SDS-PAGE)

O gel de poliacrilamida é composto por duas partes: o gel de separação (*resolving gel*), que fica na parte inferior e o gel de empacotamento (*stacking gel*), situado na parte superior. O gel de separação é composto por Tris-HCl 375 mmol/L, pH 8,8; acrilamida/bis-acrilamida (29:1) 15% (v/v); SDS (*sodium dodecyl sulfate*) 0,1% (m/v); APS (persulfato de amônio) 0,1% (m/v) e Temed 0,08% (v/v). O gel de empacotamento é composto por Tris-HCl 126 mmol/L, pH 6,8; acrilamida/bis-acrilamida (29:1) 5% (v/v); SDS 0,1% (m/v); APS 0,1% (m/v) e Temed 0,05% (v/v). Para cada gel, são necessários aproximadamente 3,7 mL de *resolving gel* e 1 mL de *stacking gel* sendo utilizado para seu preparo e corrida o sistema Mini-Protean 3, da Bio-Rad. Para a corrida no SDS-PAGE, 2 volumes da amostra foram coletados e aos mesmos, foi adicionado 1 volume de tampão de amostra 3X (Tris-HCl 50 mmol/L, pH 6,8; glicerol 45% e SDS 2%). Nas análises dos *pellets* bacterianos pós-indução, as células foram lisadas em 200 µL de água ultrapura contendo 100 µL de tampão de amostra 3X. Para a análise das amostras obtidas nos passos de purificação por cromatografia líquida, 20 µL das frações eluídas foram coletados e a eles adicionados 10 µL do tampão de amostra 3X. As amostras contendo o tampão de amostra foram então fervidas por 5 minutos e 20 µL das mesmas foram carregados no gel, submetido à diferença de potencial de 120 V por 1 hora. O gel foi então corado em solução de *Coomassie Blue* R-250, Bio-Rad, (*Coomassie Blue* R-250 0,25%; ácido acético 10 %; etanol 50%) por 1 hora e em seguida descorado em solução de descorante (etanol 10% e ácido acético 5% em água).

3.1.3- Expressão da proteína em média escala

Para a produção da proteína SCF12 foram utilizadas bactérias competentes do tipo *E. coli* BL21 (DE3) pRil. A transformação e indução das células foram feitas conforme o item 3.1.1.1, havendo mudanças apenas nas quantidades usadas: foram produzidos 100 mL de pré-inóculo e este foi utilizado na produção de 2 L de cultura bacteriana em meio LB, que estavam separados em 4 *erlenmeyers* contendo 500 mL de meio LB. Após 4 horas de indução com 1 mmol/L de IPTG, o meio de cultura foi centrifugado em garrafas de polietileno a 4000 *g* por 5 minutos, a 4 °C. O sobrenadante foi descartado e o *pellet* guardado a temperatura de -20°C para trabalhos posteriores.

3.1.4- Purificação

A purificação da proteína SCF12 foi realizada por IMAC (*Immobilized Metal Affinity Chromatography*) com resina de afinidade Ni-NTA (Qiagen, Alemanha). O *pellet* bacteriano foi ressuscitado em 40 mL de tampão A de afinidade (tampão fosfato de sódio 0,050 mol/L, pH 7,2; glicerol 5% (v/v); NaCl 0,10 mol/L) contendo 80 µg/mL de lisozima para auxiliar a lise das bactérias. O extrato foi deixado a 4°C por 1 hora e então submetido a 6 ciclos de sonicação (15 W de potência e 22,5 kHz de frequência) de 40s cada e intervalos de descanso de 1 minuto em gelo. O extrato foi centrifugado a 16000 *g* a 4°C por 20 min. O sobrenadante (fração solúvel) e o precipitado (fração insolúvel) foram analisados por eletroforese em gel SDS-PAGE 15%. A fração solúvel foi incubada com 0,5 mL da resina de afinidade por 1 hora a 4° C. A resina foi depositada em coluna de vidro (capacidade 10 mL) e a proteína foi eluída com o tampão A contendo concentrações crescentes de imidazol (5, 10, 20, 30, 50, 100 e 200 mmol/L). As frações eluídas foram coletadas e analisadas em SDS-PAGE 15% e as que continham a proteína SCF12 pura foram agrupadas e dialisadas em tampão fosfato de sódio 0,050 mol/L, pH 7,2; glicerol 2% (v/v) e Na₂SO₄ 0,050 mol/L durante uma noite. A amostra dialisada foi concentrada por sistema Amicon

(Millipore, EUA) e a concentração foi determinada pelo método de Bradford (Bradford, 1976).

3.1.5- Dicroísmo circular

Para as medidas de dicroísmo circular, a amostra foi previamente dialisada em tampão fosfato de sódio 0,050 mol/L, pH 7,2; glicerol 2% (v/v) e Na₂SO₄ 0,050 mol/L foi diluída para 5,6 µmol/L e transferida para uma cubeta de quartzo de 1 mm de caminho óptico. As medidas abrangeram um comprimento de onda de 190 a 240 nm e foram realizadas num espectropolarímetro JASCO J750 (Jasco, Japão), com largura de banda de 1 nm e tempo de resposta de 1s, no IQ/Unicamp. As medidas foram feitas a temperatura ambiente, com uma velocidade de varredura de 100 nm/min. Foram acumuladas 16 varreduras por amostra e os espectros foram corrigidos pela subtração do branco (tampão). Neste trabalho foi utilizado apenas a região de comprimento de onda onde as medidas são confiáveis, de acordo com o sinal de HT (voltagem que chega ao detector) fornecido pelo equipamento. Para calcular a composição de estruturas secundárias os dados coletados foram analisados pelo programa CONTIN (Provencher and Glöckner, 1981; Van Stokkum *et al.*, 1990) disponível no *Dichroweb Internet Service* (www.cryst.bbk.ac.uk/cdweb/html/home.html) (Lobley and Wallace, 2001; Whitmore and Wallace, 2004) utilizando-se dois bancos de dados disponíveis e adequados ao intervalo de comprimento de onda no qual foram feitas as medidas.

3.2- Proteína SCF5

Os experimentos para clonagem e expressão da proteína SCF5 foram estabelecidos pelo Dr. Paulo S. Schlögl no CBMEG/Unicamp (Centro de Biologia Molecular e Engenharia Genética), durante seus trabalhos de doutorado (Schlögl *et al.*, 2004).

3.2.1- Purificação por troca catiônica

Para a purificação da proteína SCF5, o *pellet* previamente induzido foi ressuspendido em 40 mL de tampão A de troca catiônica (Tris-HCl 0,020 mol/L, pH 7,5; EDTA 5 mmol/L; β -mercaptoetanol 7 mmol/L; NaCl 0,020 mol/L) e submetido a 6 ciclos de sonicação (15 W de potência e 22,5 kHz de frequência) de 40s cada e intervalos de descanso de 1 minuto no gelo. A amostra foi centrifugada a 16000 g por 15 min a 4° C. O sobrenadante foi aquecido a 80° C por 3 min e centrifugada também a 16000 g por 15 min a 4° C. Após, o sobrenadante foi injetado numa coluna sp-sepharose FF (Amersham, USA) no sistema Äkta FPLC (Amersham, USA). A proteína foi eluída utilizando-se tampão B de troca catiônica (Tris-HCl 0,020 mol/L, pH 7,5; EDTA 5 mmol/L; β -mercaptoetanol 7 mmol/L; NaCl 1 mol/L). As frações eluídas foram coletadas e analisadas em SDS-Page 15%. As que continham a proteínas SCF5 pura foram reunidas e dialisadas no tampão adequado e concentradas com o sistema Amicon (Millipore, EUA) para os experimentos posteriores.

3.2.2- Reação de ligação da proteína SCF5 com a fita de DNA

As seqüências de DNA ao qual a proteína SCF5 se liga foi determinada através de técnicas de mobilidade eletroforética (Schlogl *et al.*, 2004). Para a reação de ligação da proteína com o DNA, foram obtidos comercialmente os oligonucleotídeos de seqüência 5'-AAGCTTAGCC**ACGTGG**CACTCGAG-3' e seu complementar. Em negrito está demonstrada a região reconhecida pela bZIP. Para a obtenção do DNA dupla fita, quantidades equimolares dos dois óligos foram reunidas e aquecidas a 94° C e em seguida resfriadas vagarosamente até a temperatura ambiente. Em seguida, as soluções de DNA e proteína (previamente dialisada em tampão Tris-HCl 50 mmol/L, pH 7,5; glicerol 2% (v/v); NaCl 250 mmol/L foram misturadas nas proporções desejadas e deixadas a 37°C por 30 minutos. A concentração de DNA foi determinada por espectrofotometria UV com comprimento de onda a 260 nm.

3.2.3- Espalhamento dinâmico de luz (DLS)

Para o experimento de DLS foi utilizado o equipamento DynaPro situado no Laboratório Nacional de Luz Síncrotron (LNLS).

As amostras para o DLS foram dialisadas previamente em Tris-HCl 50 mmol/L, pH 7,5; glicerol 2% (v/v); NaCl 250 mmol/L preparadas na concentração de 2,4 mg/mL de proteína e foram a elas adicionadas o DNA dupla fita na proporção molar de 1:1 e 2:1 (proteína/DNA), conforme seção 3.2.2.

3.2.4- Medidas de SAXS (*Small-Angle X-ray Scattering*)

Dados de SAXS foram coletados para amostras preparadas da mesma maneira que as utilizadas para o experimento de espalhamento de luz, no tampão Tris-HCl 50 mmol/L, pH 7,5; glicerol 2% (v/v); NaCl 250 mmol/L, alterando a concentração de glicerol e a proporção proteína/DNA.

As medidas foram realizadas no Laboratório Nacional de Luz Síncrotron, na linha D11A - SAXS1, com detector unidimensional, a uma distância amostra-detector de 617,9 mm e comprimento de onda da radiação usada igual a 1,488 Å, abrangendo uma região de valores de q entre: $q_{\min} = 0,01219 \text{ \AA}^{-1}$ - $q_{\max} = 0,5237 \text{ \AA}^{-1}$. Foram coletadas diversas varreduras (*frames*) de 600 s para cada amostra.

As correções necessárias nas intensidades e os seus respectivos erros foram obtidos através do programa TRAT1D (Oliveira C.L.P., 2003). Os processamentos posteriores foram realizados com os programas PRIMUS (Konarev *et al.*, 2003), CRY SOL (Svergun *et al.*, 1995), GNOM (Svergun, 1992), DAMMIN (Svergun, 1999), GASBOR (Svergun *et al.*, 2001), DAMAVER (Volkov & Svergun, 2003) e SUPCOMB (Kozin & Svergun, 2001) do pacote ATSAS (Konarev *et al.*, 2006).

3.2.5- Estudo da estabilidade térmica da proteína SCF5 por Dicroísmo Circular

Para o estudo da estabilidade térmica da SCF5 foram realizados dois experimentos. No experimento 1, foram feitas medidas de CD na região de 240 a 190 nm para uma amostra da proteína na concentração de 16 $\mu\text{mol/L}$ (em tampão fosfato de sódio 0,050 mol/L, pH 7,2; glicerol 2% (v/v); Na_2SO_4 0,20 mol/L). A amostra foi aquecida de 0 a 90°C variando a temperatura em passos de 5° C (1°/min) e em seguida resfriada para 20°C, variando a temperatura em passos de 15°C (1°/min). Para cada temperatura eram aguardados 10 minutos para a estabilização da condição e 3 espectros foram registrados.

No experimento 2, foram coletados dados de CD em diferentes temperaturas (aquecimento e resfriamento) em um comprimento de onda fixo de 222 nm. A concentração da amostra foi de 60 $\mu\text{mol/L}$, em tampão Tris-HCl 0,050 mol/L, pH 7,5; NaCl 0,20 mol/L; glicerol 2% (v/v). Inicialmente a amostra foi aquecida de 0 para 80°C e em seguida resfriada a 0°C novamente. A taxa de aquecimento e resfriamento da amostra foi de 1° C por minuto, tendo sido realizadas medidas de CD a cada a 0,1° C.

3.2.6- Ensaio de cristalização

Para os ensaios de cristalização, a proteína SCF5 foi purificada conforme descrito no item 3.2.1 e então ligada à dupla fita de DNA (seção 3.2.2) na proporção molar de 1:1. Para os ensaios, inicialmente foram utilizados os kits *Crystallization Basic Kit for Proteins* e *Crystallization Extension Kit for Proteins* (Sigma). A concentração de SCF5 empregada foi 2,5 mg/mL e o método empregado foi o da “gota suspensa”. Em cada lamínula foi adicionado 1,5 μL da solução de proteína e 1,5 μL da respectiva solução do kit. Os experimentos foram analisados em microscópio óptico semanalmente. Os mesmos testes foram repetidos com gotas de proporção 2:1 da solução de proteína / solução do kit (3 μL e 1,5 μL , respectivamente), totalizando 8 placas de 24 poços. Na Figura 14 tem-se

a fotografia de uma placa usada nos experimentos de cristalização. As placas foram deixadas a temperatura de 18°C. As soluções presentes nos kits citados acima estão descritas nas Tabelas 1 e 2.



Figura 14- Fotografia de uma placa de cristalização usada nos experimentos para cristalização da proteína SCF5/DNA. Existem 24 reservatórios dispostos no sistema linha x coluna, sendo que há 4 linhas (A, B, C e D) e 6 colunas (1 a 6).

Refinamento de condições

Com base nos resultados obtidos nos testes de cristalização dos kits comerciais citados anteriormente, preparou-se duas novas placas (placas 1 e 2) para abertura da condição 22 do *Crystallization Basic Kit for Proteins* (acetato de sódio trihidratado 0,2M; tris HCl 0,1M, pH 8,5; Polietileno glicol 4000 30% m/v). Sinais de cristais foram obtidos nas condições C1 (tris-HCl 50mM, pH 8,5; PEG 4000 25%; acetato de sódio 0,2M e glicerol 2,5%) e C2 (tris-HCl 50mM, pH 8,5; PEG 4000 30%; acetato de sódio 0,2M e glicerol 2,5%) da placa 1. Na Figura 15 estão descritas as soluções empregadas nas placas 1 e 2. Foram colocadas duas gotas por lamínula, na proporção 1:1 e 1:2 (v/v) de solução de proteína / solução teste (2 μ L + 2 μ L e 3 μ L + 6 μ L, respectivamente). Em todos os poços foram adicionados 500 μ L da respectiva solução teste.

Tabela 1- Descrição das soluções contidas no *Crystallization Basic Kit for Proteins* (Sigma)

#	sal	tampão	precipitante
1	Calcium chloride dihydrate 0,02 M	Sodium acetate trihydrate pH 4,6; 0,1 M	(+/-)-2-Methyl-2,4-pentanediol 30% v/v
2	---	----	Potassium sodium tartrate tetrahydrate 0,4 M
3	---	----	Ammonium phosphate monobasic 0,4 M
4	---	Tris hydrochloride pH 8,5; 0,1 M	Ammonium sulfate 2 M
5	Sodium citrate tribasic dihydrate 0,2 M	HEPES sodium pH 7,5; 0,1 M	(+/-)-2-Methyl-2,4-pentanediol 30% v/v
6	Magnesium chloride hexahydrate 0,2 M	Tris hydrochloride pH 8,5; 0,1 M	Polyethylene glycol 4,000 30% w/v
7	---	Sodium cacodylate trihydrate pH 6,5; 0,1 M	Sodium acetate trihydrate 1,4 M
8	Sodium citrate tribasic dihydrate 0,2 M	Sodium cacodylate trihydrate pH 6,5; 0,1 M	2-Propanol 30% v/v
9	Ammonium acetate 0,2 M	Sodium citrate tribasic dihydrate pH 5,6; 0,1 M	Polyethylene glycol 4,000 30% w/v
10	Ammonium acetate 0,2 M	Sodium acetate trihydrate pH 4,6; 0,1 M	Polyethylene glycol 4,000 30% w/v
11	---	Sodium citrate tribasic dihydrate pH 5,6; 0,1 M	Ammonium phosphate monobasic 1 M
12	Magnesium chloride hexahydrate 0,2 M	HEPES sodium pH 7,5; 0,1 M	2-Propanol 30% v/v
13	Sodium citrate tribasic dihydrate 0,2 M	Tris hydrochloride pH 8,5; 0,1 M	Polyethylene glycol 400 30% v/v
14	Calcium chloride dihydrate 0,2 M	HEPES sodium pH 7,5; 0,1 M	Polyethylene glycol 400 28% v/v
15	Ammonium sulfate 0,2 M	Sodium cacodylate trihydrate pH 6,5; 0,1 M	Polyethylene glycol 8,000 30% w/v
16	---	HEPES sodium pH 7,5; 0,1	Lithium sulfate monohydrate 1,5 M
17	Lithium sulfate monohydrate 0,2 M	Tris hydrochloride pH 8,5; 0,1 M	Polyethylene glycol 4,000 30% w/v
18	Magnesium acetate tetrahydrate 0,2 M	Sodium cacodylate trihydrate pH 6,5; 0,1 M	Polyethylene glycol 8,000 20% w/v
19	Ammonium acetate 0,2 M	Tris hydrochloride pH 8,5; 0,1 M	2-Propanol 30% v/v
20	Ammonium sulfate 0,2 M	Sodium acetate trihydrate pH 4,6; 0,1 M	Polyethylene glycol 4,000 25% w/v
21	Magnesium acetate tetrahydrate 0,2 M	Sodium cacodylate trihydrate pH 6,5; 0,1 M	(+/-)-2-Methyl-2,4-pentanediol 30% v/v
22	Sodium acetate trihydrate 0,2 M	Tris hydrochloride pH 8,5; 0,1 M	Polyethylene glycol 4,000 30% w/v
23	Magnesium chloride hexahydrate 0,2 M	HEPES sodium pH 7,5; 0,1 M	Polyethylene glycol 400 30% v/v
24	Calcium chloride dihydrate 0,2 M	Sodium acetate trihydrate pH 4,6; 0,1 M	2-Propanol 20% v/v
25	---	Imidazole pH 6,5; 0,1 M	Sodium acetate trihydrate 1 M
26	Ammonium acetate 0,2 M	Sodium citrate tribasic dihydrate pH 5,6; 0,1 M	(+/-)-2-Methyl-2,4-pentanediol 30% v/v
27	Sodium citrate tribasic dihydrate 0,2 M	HEPES sodium pH 7,5; 0,1 M	2-Propanol 20% v/v
28	Sodium acetate trihydrate 0,2 M	Sodium cacodylate trihydrate pH 6,5; 0,1 M	Polyethylene glycol 8,000 30% w/v
29	---	HEPES sodium pH 7,5; 0,1 M	Potassium sodium tartrate tetrahydrate 0,8 M
30	Ammonium sulfate 0,2 M	----	Polyethylene glycol 8,000 30% w/v
31	Ammonium sulfate 0,2 M	----	Polyethylene glycol 4,000 30% w/v
32	---	----	Ammonium sulfate 2 M
33	---	----	Sodium formate 4 M
34	---	Sodium acetate trihydrate pH 4,6; 0,1 M	Sodium formate 2 M
35	---	HEPES sodium pH 7,5; 0,1 M	Sodium phosphate monobasic monohydrate 0,8 M e Potassium phosphate monobasic 0,8 M
36	---	Tris hydrochloride pH 8,5; 0,1 M	Polyethylene glycol 8,000 8% w/v
37	---	Sodium acetate trihydrate pH 4,6; 0,1 M	Polyethylene glycol 4,000 8% w/v
38	---	HEPES sodium pH 7,5; 0,1 M	Sodium citrate tribasic dihydrate 1,4 M
39	---	HEPES sodium pH 7,5; 0,1 M	Polyethylene glycol 400 2% v/v e Ammonium
40	---	Sodium citrate tribasic dihydrate pH 5,6; 0,1 M	2-Propanol 2% v/v e Polyethylene glycol 4,000 20% w/v
41	---	HEPES sodium pH 7,5; 0,1 M	2-Propanol 10% v/v e Polyethylene glycol 4,000 20% w/v
42	Potassium phosphate monobasic 0,05 M	----	Polyethylene glycol 8,000 20% w/v
43	---	----	polyethylene glycol 1500 30% w/v
44	---	----	magnesium formate dihydrate 0,2 M
45	Zinc acetate dihydrate 0,2 M	Sodium cacodylate trihydrate pH 6,5; 0,1 M	Polyethylene glycol 8,000 18% w/v
46	Calcium acetate hydrate 0,2 M	Sodium cacodylate trihydrate pH 6,5; 0,1 M	Polyethylene glycol 8,000 18% w/v
47	---	Sodium acetate trihydrate pH 4,6; 0,1M	Ammonium sulfate 2 M
48	---	Tris hydrochloride pH 8,5; 0,1 M	Ammonium phosphate monobasic 2 M
49	Lithium sulfate monohydrate 1 M	----	Polyethylene glycol 8,000 2% w/v
50	Lithium sulfate monohydrate 0,5 M	----	Polyethylene glycol 8,000 15% w/v

Tabela 2- Descrição das soluções contidas no *Crystallization Extension Kit for Proteins* (Sigma)

#	sal	tampão	precipitante
1	Sodium chloride 2,0 M	----	Polyethylene glycol 6,000 10% w/v
2	Sodium chloride 0,5M; Magnesium chloride hexahydrated 0,01 M	----	Hexadecyl tri methyl ammonium bromide 0,01 M
3	---	----	Ethylene glycol 25% v/v
4	---	----	1,4-Dioxane 35% v/v
5	Ammonium sulfate 2,0 M	----	2-Propanol 5% v/v
6	---	----	Imidazole pH 7,0 1 M
7	Sodium chloride 1,5 M	----	Polyethylene glycol 1,000 10% w/v e Polyethylene glycol 8,000 10% w/v
8	---	----	Ethanol 10% v/v
9	Sodium chloride 0,2 M	Sodium acetate trihydrate 0,1 M; pH 4,6	Sodium chloride 2 M
10	Cobalt (II) chloride hexahydrate 0,01M	Sodium acetate trihydrate 0,1 M; pH 4,6	(+/-)-2-Methyl-2,4-pentanediol 30 % v/v
11	Cadmium chloride hydrate 0,1 M	Sodium acetate trihydrate 0,1 M; pH 4,6	1,6-Hexanediol 1 M
12	Ammonium sulfate 0,2 M	Sodium acetate trihydrate 0,1 M; pH 4,6	Polyethylene glycol 400 30% v/v
13	Potassium sodium tartrate tetrahydrate 0,2 M	Sodium acetate trihydrate 0,1 M; pH 4,6	Polyethylene glycol monomethyl ether 2,000 30% w/v
14	Ammonium sulfate 0,5 M	Sodium citrate tribasic dihydrate 0,1 M; pH 5,6	Ammonium sulfate 2 M
15	Sodium chloride 0,5 M	Sodium citrate tribasic dihydrate 0,1 M; pH 5,6	Lithium sulfate monohydrate 1 M
16	---	Sodium citrate tribasic dihydrate 0,1 M; pH 5,6	Ethylene imine Polymer 2% v/v
17	Iron (III) chloride hexahydrate 0,01 M	Sodium citrate tribasic dihydrate 0,1 M; pH 5,6	tert-Butanol 35% v/v
18	---	Sodium citrate tribasic dihydrate 0,1 M; pH 5,6	Jeffamine M-600 10% v/v
19	---	Sodium citrate tribasic dihydrate 0,1 M; pH 5,6	1,6-Hexanediol 2,5 M
20	---	MES monohydrate 0,1 M; pH 6,5	Magnesium sulfate heptahydrate 1,6 M
21	Sodium phosphate monobasic monohydrate 0,1M; Potassium phosphate monobasic 0,1M	MES monohydrate 0,1 M; pH 6,5	Sodium chloride 2 M
22	---	MES monohydrate 0,1 M; pH 6,5	Polyethylene glycol 20,000 12% w/v
23	Ammonium sulfate 1,6 M	MES monohydrate 0,1 M; pH 6,5	1,4-Dioxane 10% v/v
24	Cesium chloride 0,05 M	MES monohydrate 0,1 M; pH 6,5	Jeffamine M-600 30% v/v
25	Cobalt (II) chloride hexahydrate 0,01 M	MES monohydrate 0,1 M; pH 6,5	Ammonium sulfate 1,8 M
26	Ammonium sulfate 0,2 M	MES monohydrate 0,1 M; pH 6,5	Polyethylene glycol monomethyl ether 5,000 30% w/v
27	Zinc sulfate heptahydrate 0,01 M	MES monohydrate 0,1 M; pH 6,5	Polyethylene glycol monomethyl ether 550 25% v/v
28	---	----	Sodium citrate tribasic dihydrate pH 6.5 1,6 M
29	Ammonium sulfate 0,5 M	HEPES 0,1 M; pH 7,5	(+/-)-2-Methyl-2,4-pentanediol 30 % v/v
30	---	HEPES 0,1 M; pH 7,5	Polyethylene glycol 6,000 10% w/v
31	---	HEPES 0,1 M; pH 7,5	Jeffamine M-600 20% v/v
32	Sodium chloride 0,1 M	HEPES 0,1 M; pH 7,5	Ammonium sulfate 1,6 M
33	---	HEPES 0,1 M; pH 7,5	Ammonium formate 2 M
34	Cadmium sulfate hydrate 0,05 M	HEPES 0,1 M; pH 7,5	Sodium acetate trihydrate 1 M
35	---	HEPES 0,1 M; pH 7,5	(+/-)-2-Methyl-2,4-pentanediol 70% v/v
36	---	HEPES 0,1 M; pH 7,5	Sodium chloride 4,3 M
37	---	HEPES 0,1 M; pH 7,5	Polyethylene glycol 8,000 10% w/v e Ethylene glycol 8% v/v
38	---	HEPES 0,1 M; pH 7,5	Polyethylene glycol 10,000 20% w/v
39	Magnesium chloride hexahydrate 0,2 M	Tris 0,1 M; pH 8,5	1,6-Hexanediol 3,4 M
40	---	Tris 0,1 M; pH 8,5	tert-Butanol 25% v/v
41	Nickel (II) chloride hexahydrate 0,01 M	Tris 0,1 M; pH 8,5	Lithium sulfate monohydrate 1 M
42	Ammonium sulfate 1,5 M	Tris 0,1 M; pH 8,5	Glycerol 12% v/v
43	Ammonium phosphate monobasic 0,2 M	Tris 0,1 M; pH 8,5	(+/-)-2-Methyl-2,4-pentanediol 50% v/v
44	---	Tris 0,1 M; pH 8,5	Ethanol 20% v/v
45	Nickel (II) chloride hexahydrate 0,01 M	Tris 0,1 M; pH 8,5	Polyethylene glycol monomethyl ether 2,000 20% w/v
46	Sodium chloride 0,1 M	BICINE 0,1 M ; pH 9,0	Polyethylene glycol monomethyl ether 550 20% v/v
47	---	BICINE 0,1 M ; pH 9,0	Magnesium chloride hexahydrate 2 M
48	---	BICINE 0,1 M; pH 9,0	1,4-Dioxane 2% v/v e Polyethylene glycol 20,000 10% w/v
49	Magnesium chloride 0,1 M	Tris 0,1 M; pH 8,5	PEG 20000 15% w/v
50	---	----	PEG 20000 20 % w/v

Duas novas placas (placas 3 e 4) foram então montadas para averiguar a condição C1 da placa 1. Foram utilizadas gotas na proporção 1:1 (v/v) de solução da proteína / solução teste (2 μ L de amostra e 2 μ L da solução teste). Uma quinta placa foi também montada para avaliar o efeito de aditivos diversos (placa 5). Na Figura 16 estão descritas as condições utilizadas.

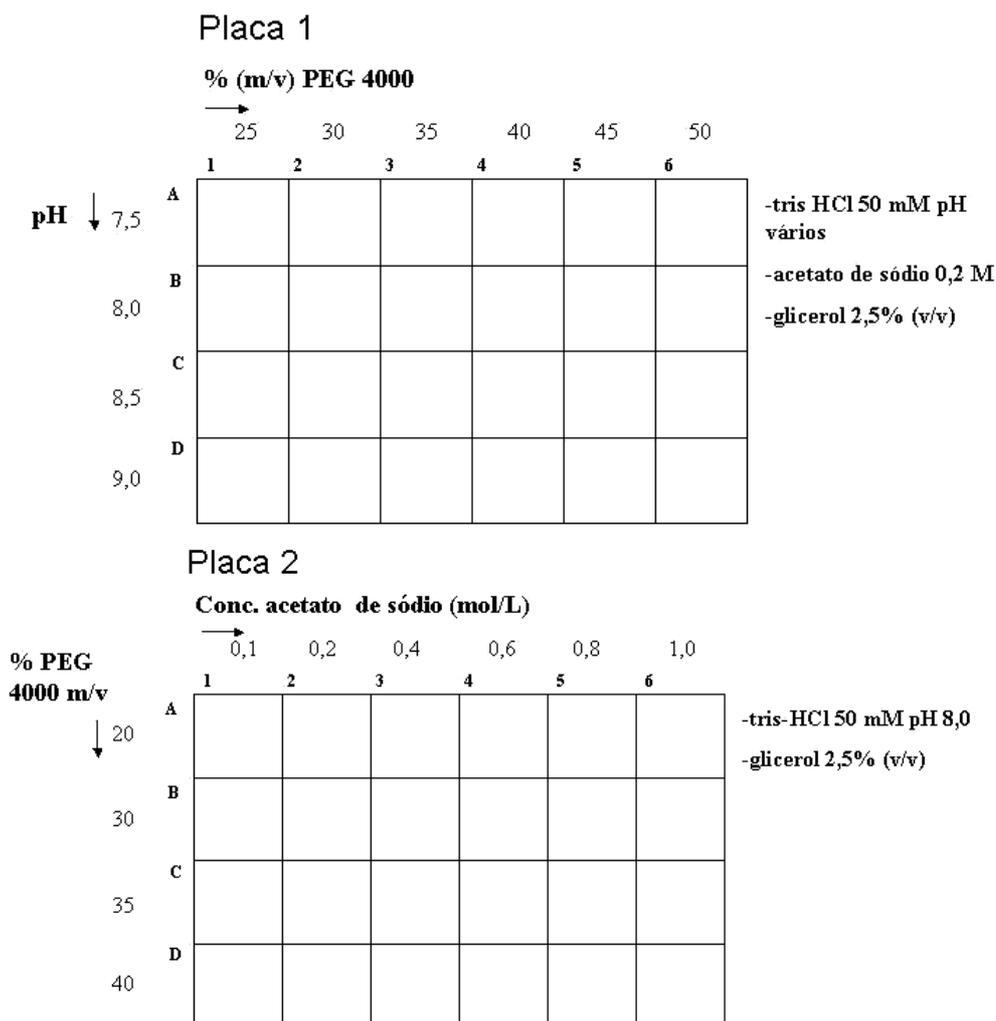


Figura 15- Esquema das placas de cristalização montadas após resultado dos testes iniciais de varredura de condições. Na placa 1 é estudado o efeito do pH e da concentração de PEG 4000 na cristalização da SCF5/DNA, mantendo-se constante as concentrações de acetato de sódio e glicerol. Na placa 2 é variada a concentração de PEG 4000 e a concentração de acetato e sódio, num pH constante 8,0 e concentração de glicerol também constante.

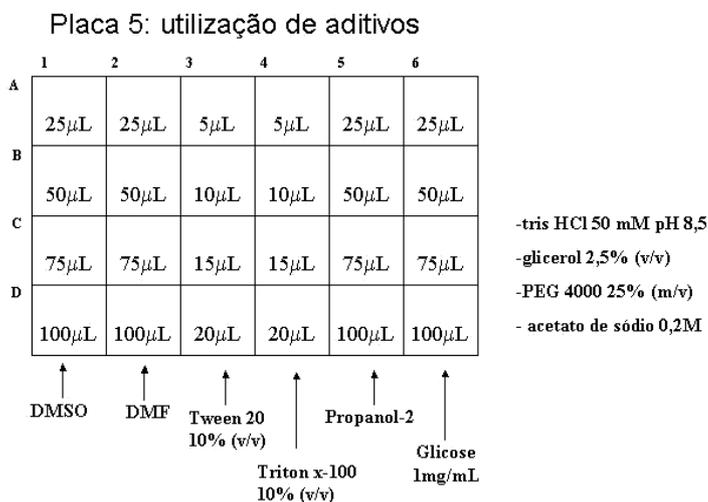
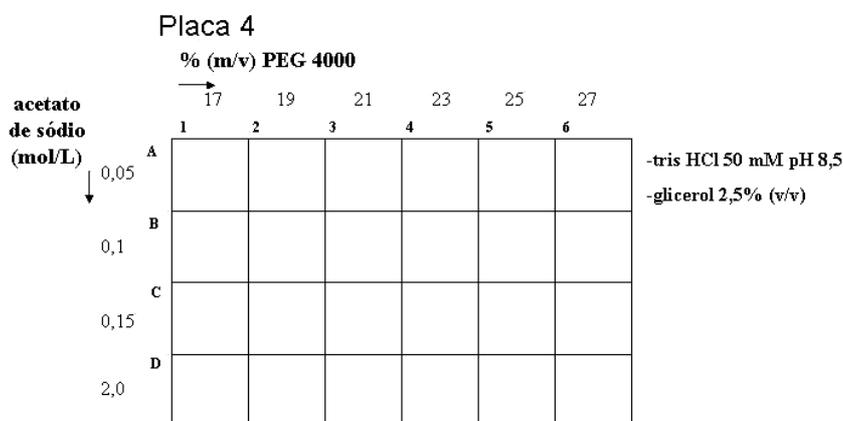
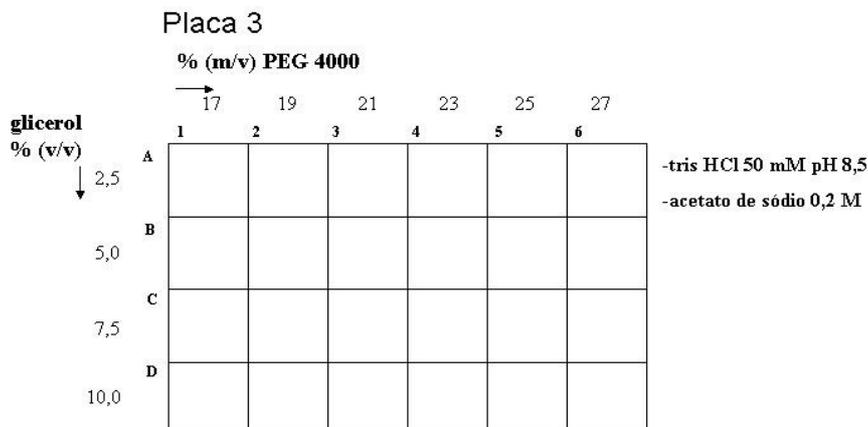


Figura 16- Nova abertura de condições: placa 3 é testado o efeito da utilização diferentes concentrações de PEG 4000 e glicerol e na placa 4, o efeito de diferentes concentrações de PEG 4000 e acetato de sódio. Na placa 5 são testados alguns aditivos, variando-se a quantidade adicionada; coluna 1: DMSO (dimetil sulfóxido); 2: DMF (dimetil-formamida); 3: Tween-20 10% (v/v); 4: Triton X-100 10% (v/v); 5: propanol-2; 6: glicose 1mg/mL. Volume total no poço: 0,5mL

4- Resultados

4.1- Proteína SCF12

4.1.1 Clonagem e expressão

O gene da proteína SCF12 clonado em pET 28a corresponde a região básica e o zíper de leucina de uma proteína do tipo bZIP da família 12 em cana-de-açúcar. A proteína codificada possui uma massa molecular teórica de 14,3 kDa e pI teórico de 9,98 (<http://ca.expasy.org/>). Na Figura 17 está mostrada a seqüência de aminoácidos correspondente a SCF12.

Na Figura 18 são mostrados os resultados de expressão da SCF12 para as *E. coli* das cepas pRil e pLysS, além da influência da temperatura na indução das mesmas. Pode-se verificar que a banda esperada, em torno de 14 kDa, é mais proeminente quando a indução é feita a 28°C, utilizando as bactérias pRil, mostrado na canaleta 4. Nas canaletas 2 e 5, onde foram carregados os extratos bacterianos não-induzidos, na região esperada não é observada nenhuma banda de intensidade equivalente as bandas encontradas nos extratos induzidos.

MSDAKLAELALVDPKRAKRILANRQSAARSKERKMRYIAE LERKVQTLQSEATTL
SAQLAMLQRDTSGLTSENSDLKIRVQTMEEQQVRLQDALNDRLRDEIQLKVATGQ
VNANIGKMGNFGMPS

Figura 17- Seqüência de aminoácidos da proteína SCF12 correspondente a tradução teórica do inserto de DNA clonado no pET. Em verde, está a região básica e, em letras vermelhas, o zíper de leucinas, com o início de cada heptato sombreado em cinza.

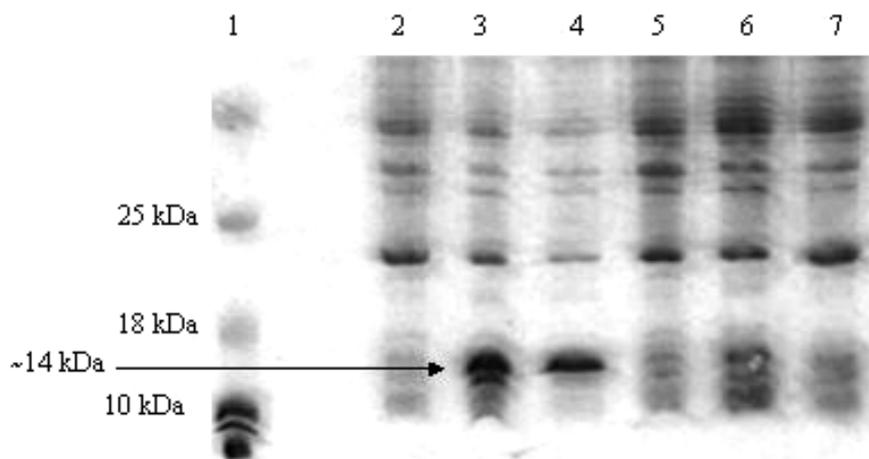


Figura 18- Análise da expressão da proteína de 14,3 kDa SCF12, produzida em *E. coli* BL21 (DE3) pRil e pLysS. 1: marcador de massa molecular Broad Range 2 - 212 kDa (Biolabs, EUA); 2: extrato bacteriano pRil antes da indução; 3: indução pRil a 37° C; 4: indução pRil a 28° C; 5: extrato bacteriano pLysS antes da indução; 6: indução pLysS a 37° C; indução pLysS a 28° C. Pode-se verificar que a indução feita a 28°C com as células pRil parece ser mais eficiente, devido a banda mais proeminente na região em torno de 14 kDa.

4.1.2- Purificação

A purificação da proteína SCF12 foi feita através de cromatografia líquida de afinidade (IMAC). O rendimento obtido de proteína pura foi de aproximadamente 0,5 mg/L de meio de cultura. A Figura 19 mostra o resultado da purificação. Pode-se observar que a partir da canaleta 9, que corresponde a concentração de 100 mmol/L de imidazol, a proteína SCF12 encontra-se praticamente pura, pois é observada a presença de apenas uma banda de proteína, situada na região de ~14 kDa. Para os experimentos posteriores foram aproveitadas as frações de 100 a 400 mmol/L de imidazol (canaletas 9 a 14).

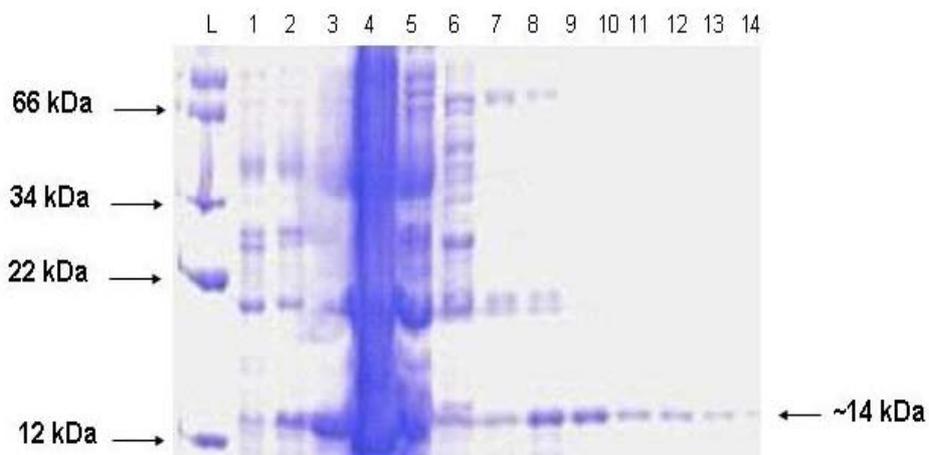


Figura 19- Expressão e purificação da proteína SCF12. L: Marcador de massa molecular; 1: célula não induzida; 2: célula induzida; 3: resina; 4: não interagido; 5 lavagem da resina com tampão A; 6, 7, 8, 9, 10, 11, 12, 13, 14 – eluição com concentrações crescentes de imidazol (5 a 400 mmol/L). As células foram induzidas em LB contendo IPTG 1 mmol/L por 6 horas a 37°C.

4.1.3- Análise de dicroísmo circular (CD)

O espectro obtido pelo experimento de dicroísmo circular é apresentado na Figura 20. Pode-se observar que a proteína encontra-se enovelada a temperatura ambiente e nas condições de concentração (5,6 $\mu\text{mol/L}$) e do tampão utilizado (tampão fosfato de sódio 0,050 mol/L, pH 7,2; glicerol 2% (v/v) e Na_2SO_4 0,050 mol/L). Os pontos de mínimo em torno de 208 e 220 nm são característicos de proteínas onde predominam hélices- α . Uma análise com o programa CONTIN (Provencher, 1982) utilizando dois bancos de dados de referência (bancos de dados 4 e 7) disponíveis no *Dichroweb internet service* (Whitmore and Wallace, 2004) indicou que a SCF12 é composta majoritariamente por hélices- α , concordando com as informações descritas na literatura (Tabela 3, nesta seção). Foram empregados os bancos de dados de referência 4 e 7 porque estes dois eram adequados ao intervalo de comprimento de onda utilizado no experimento.

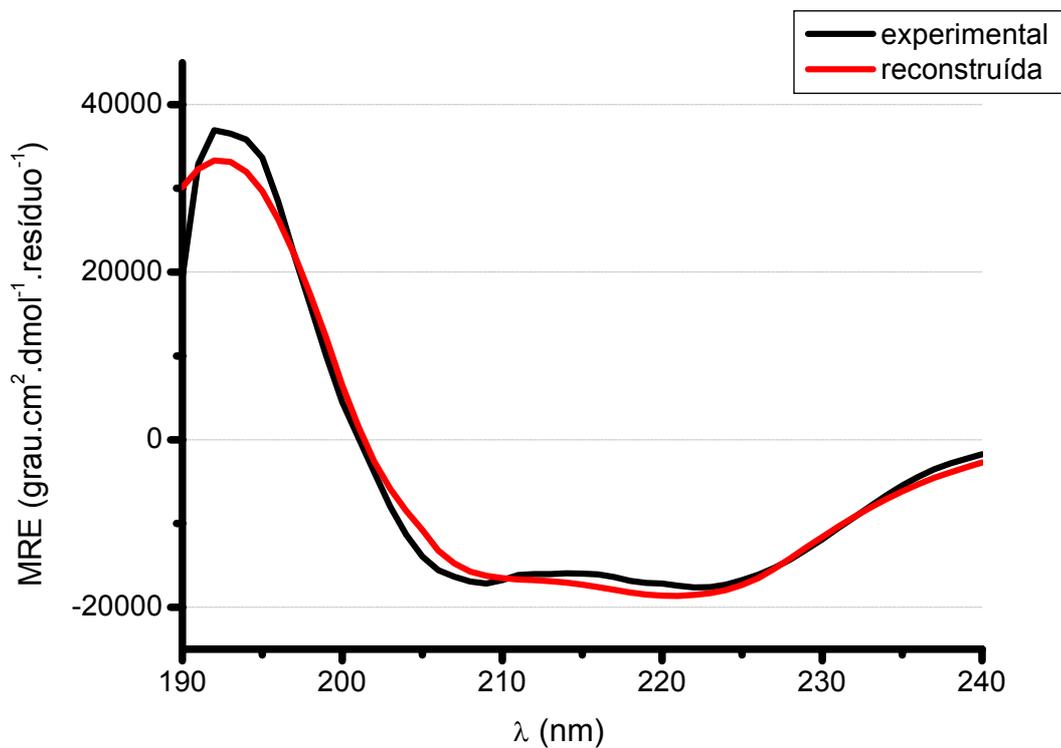


Figura 20 - Espectros de dicroísmo circular da proteína SCF12. A linha preta (experimental) corresponde ao espectro obtido para a proteína SCF12. Em vermelho (reconstruída) está apresentada a curva calculada com base no banco de dados 4 disponível no *Dichroweb*.

Tabela 3- Predição da estrutura secundária da proteína SCF12 usando o programa CONTIN (Provencher and Glöckner, 1981; Van Stokkum *et al.*, 1990)

Banco de dados referência	Composição da estrutura secundária		
	hélices- α (%)	folhas- β (%)	outros (%)
Nº 4	93,4	6,5	0,1
Nº 7	87,3	5,8	6,9

4.2- Proteína SCF5

A SCF5 é uma proteína de 12,5 kDa de massa molecular, sendo composta por 104 aminoácidos (Schlögl *et al.*, 2004). Na Figura 21 é mostrada a seqüência primária da SCF5.

SEESDDYQRSLA **EERRKRRMISNRESARRSRMRK** QKQL
SE **L**WAQVVH **L**RSTNRQ **L**LDQLNH **V**IRDCDR **V**LHENSQ **L**
RDEQTK **L**QQQLEK **L**HVETTESGVMSPDS

Figura 21- Seqüência de aminoácidos da proteína SCF5 correspondente a tradução teórica do inserto de DNA clonado no pET. Em verde, está a região básica e, em letras vermelhas, o zipper de leucinas, com o início de cada heptato sombreado em cinza.

4.2.1 Purificação por troca catiônica

A purificação da proteína SCF5 foi realizada em um sistema de cromatografia líquida Äkta FPLC e o perfil de eluição é mostrado no cromatograma que compõe a Figura 22. A proteína SCF5 é eluída em aproximadamente 6 frações de 2 mL, com concentração da solução B de troca catiônica em torno de 80% (NaCl ~800 mmol/L), sendo o rendimento de aproximadamente 1,5 mg/L de proteína pura por litro de cultura.

As frações eluídas foram coletadas e analisadas em gel SDS-PAGE 15% (Figura 23).

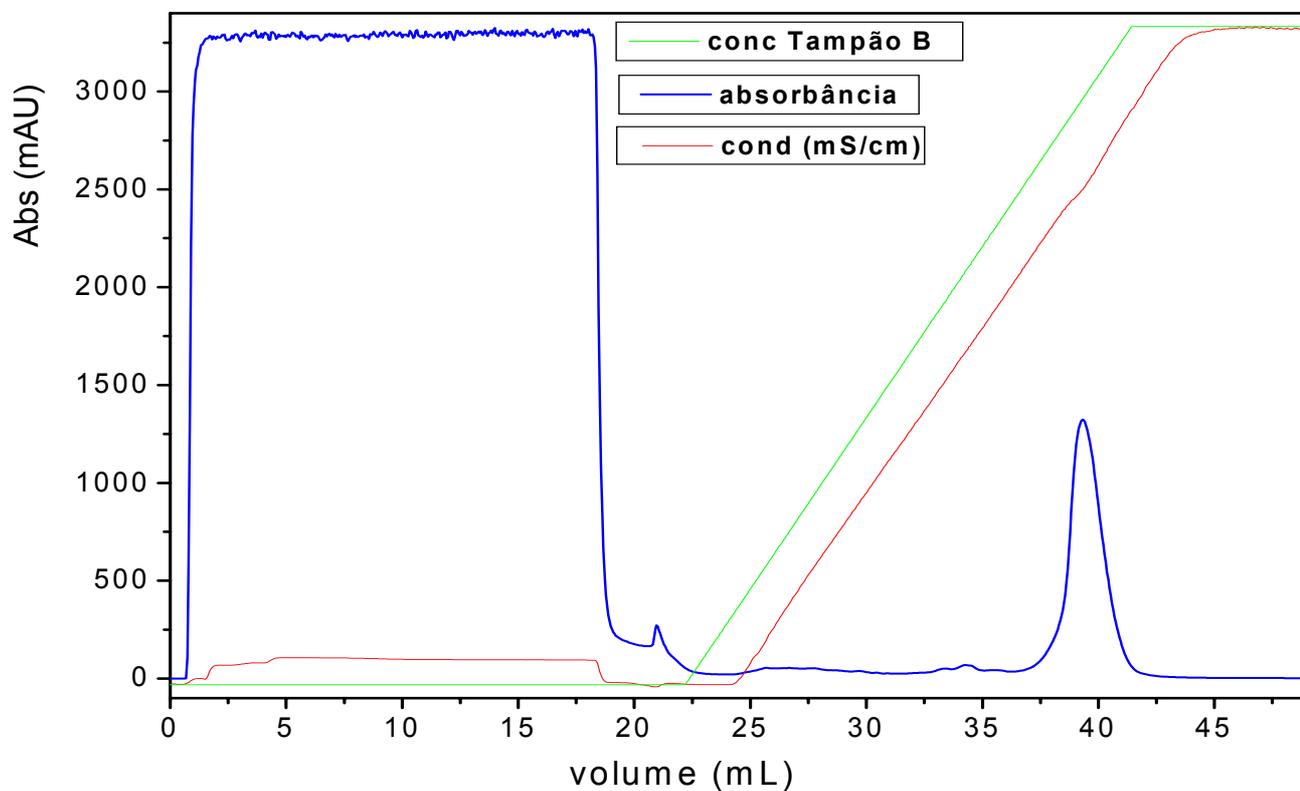


Figura 22- Perfil cromatográfico durante a purificação da SCF5 por troca catiônica em coluna SP-sepharose FF de 1 mL. Em azul está mostrado a absorvância (em mAU) da amostra injetada em função ao volume passado pela coluna. Em verde está a concentração do tampão B de troca catiônica (em %) e em vermelho, a condutividade do meio (em mS/cm). O Volume de amostra injetado foi de aproximadamente 18 mL e o fluxo de 1 ml/minuto. Faixa do gradiente: 0 a 100% do tampão B em 15 volumes de coluna (CV). Os tampões utilizados foram: Tampão A (Tris-HCl 0,020 mol/L, pH 7,5; EDTA 5 mmol/L; β -mercaptoetanol 7 mmol/L; NaCl 0,020 mol/L) e o tampão B (Tris-HCl 0,020 mol/L, pH 7,5; EDTA 5 mmol/L; β -mercaptoetanol 7 mmol/L; NaCl 1 mol/L).

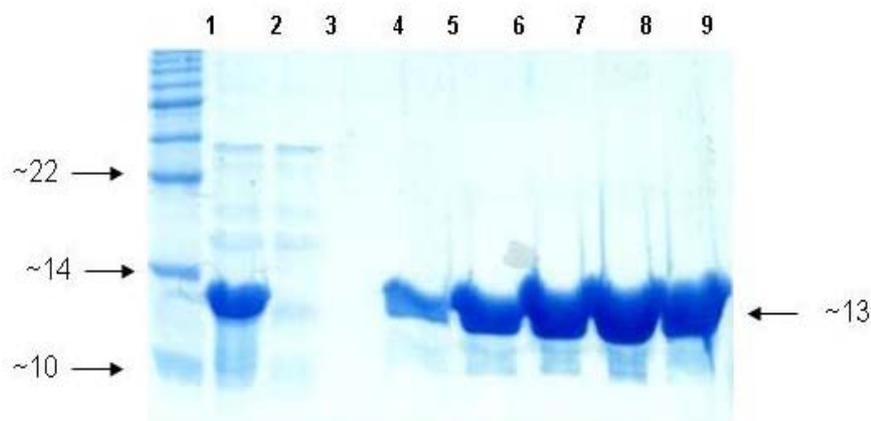


Figura 23- Análise por SDS-PAGE 15% das frações da proteína SCF5 obtidas da cromatografia de troca catiônica: 1- marcador de massa molecular Broad Range (2 - 212 kDa, Biolabs, EUA); 2- amostra antes da coluna; 3, 4- não ligado; 5, 6, 7, 8 e 9- frações eluídas contendo a proteína SCF5.

4.2.2- Estudo da estabilidade térmica por CD

Na seção 3.2.1, menciona-se aquecimento de 80°C em um dos passos para purificação da SCF5. Para observar o comportamento da SCF5 em relação à temperatura, foram coletados dados de CD para diferentes temperaturas entre 0 e 90°C, em passos de 5°C com aquecimento de 1°C/min. De posse dos dados de CD, foram construídos gráficos de elipticidade molar média por resíduo MRE (*Mean Residue Ellipticity*) vs λ . Para diminuir a absorção do tampão e assim obter dados em comprimentos de ondas menores (monitorados pelo sinal de HT), utilizou-se Na₂SO₄ e fosfato de sódio tamponado no lugar de NaCl e tris, respectivamente. Para melhor visualização, na Figura 24 são mostrados apenas os espectros nas temperaturas de 0, 10, 20, 30, 40, 50, 60, 70, 80, e 90°C. Pode-se observar nos espectros que a proteína vai gradativamente perdendo o enovelamento, pois o padrão característico de hélices- α (mínimos em $\lambda = 208$ e 220 nm) vai desaparecendo. Porém, quando a mesma amostra de proteína é então resfriada, com passos de 15°C a 1°C/min, os sinais típicos de hélices- α se restabelecem, resultado é mostrado na Figura 25, que contém os espectros de CD para as temperaturas de 80, 65, 50, 35 e 20°C.

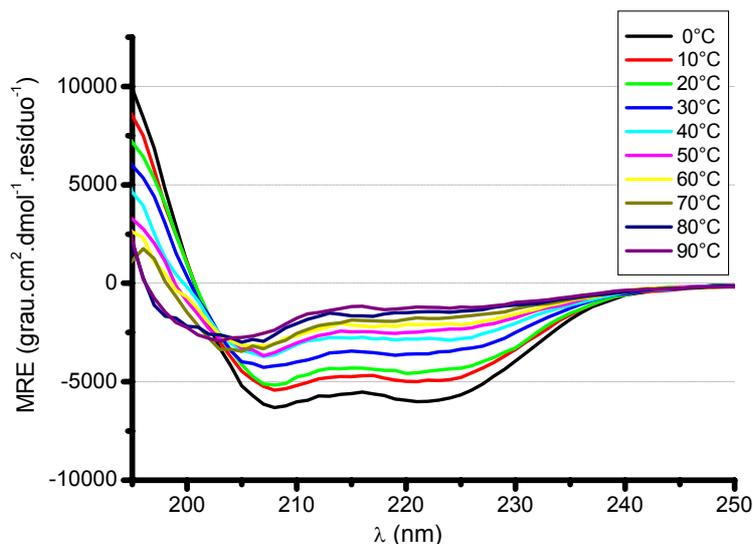


Figura 24- Espectros de CD (MRE vs λ) para a SCF5 em diferentes temperaturas durante o processo de aquecimento de uma amostra de SCF5 de 0 a 90°C, com passos de 5°C e aquecimento a 1°C/min. A proteína foi preparada em tampão fosfato de sódio 0,050 mol/L, pH 7,2; glicerol 2% (v/v); Na₂SO₄ 0,20 mol/L, na concentração de 16 μ mol/L. Medidas realizadas no equipamento Jasco J750 (Jasco, Japão).

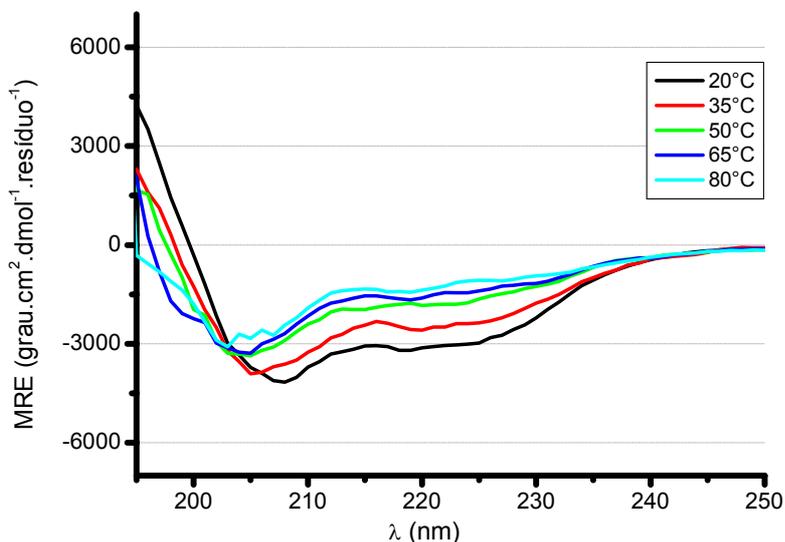


Figura 25- Espectros de CD (MRE vs λ) em diferentes temperaturas durante o processo de resfriamento da SCF5 de 80 a 20°C. Após o aquecimento, a mesma amostra foi submetida ao resfriamento até 20°C, com passos de 15°C a 1°C/min. A proteína foi preparada em tampão fosfato de sódio 0,050 mol/L, pH 7,2; glicerol 2% (v/v); Na₂SO₄ 0,20 mol/L, na concentração de 16 μ mol/L. Medidas realizadas no equipamento Jasco J750 (Jasco, Japão).

Conforme descrito na seção 3.2.5, um experimento de CD adicional foi feito mantendo-se o comprimento de onda constante, igual a 222 nm. Neste caso pôde-se trabalhar com uma concentração de proteína SCF5 maior (65 mmol/L, que gera um sinal mais acentuado) que no experimento de varredura de λ , além de ser possível utilizar o mesmo tampão empregado nos experimentos de DLS e SAXS (Tris-HCl 0,050 mol/L, pH 7,5; NaCl 0,20 mol/L; glicerol 2% (v/v), já que neste λ a interferência causada pelo tipo de íon e concentração da amostra no espectro de CD não é tão acentuada quanto no caso de comprimentos de onda inferiores a 200nm. Neste experimento a amostra foi aquecida de 0 a 80 °C e depois resfriada novamente para 0 °C (Figura 26). Observa-se que o sinal de elipticidade molar começa a diminuir, o que pode significar que a proteína está perdendo o enovelamento, mas com o resfriamento, o sinal volta a ser em torno do inicial, antes do aquecimento. Este resultado concorda com o que foi observado nos dois experimentos anteriores.

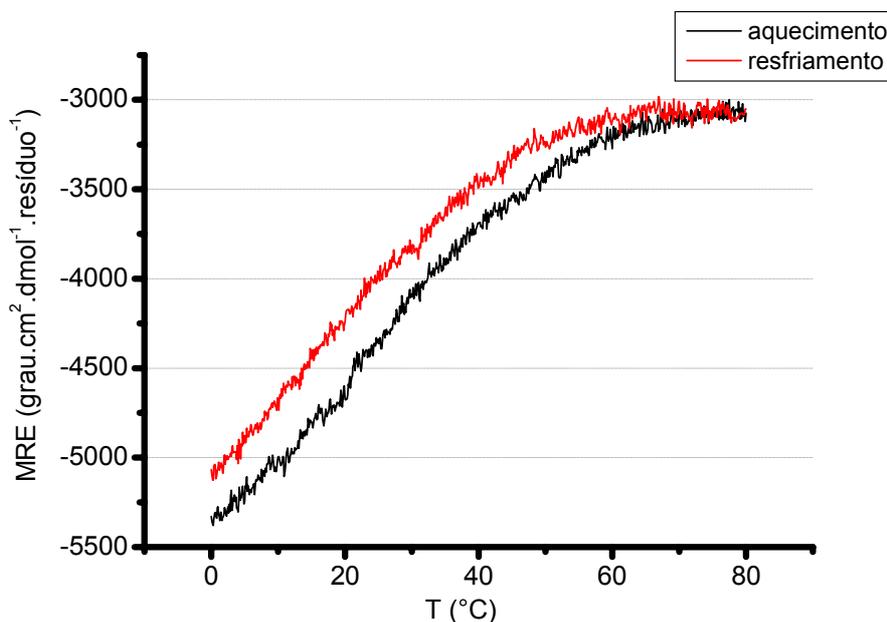


Figura 26- Estabilidade térmica da proteína CSF5: MRE vs temperatura em λ fixo de 222 nm. A proteína foi preparada em tampão Tris-HCl 0,050 mol/L, pH 7,5; NaCl 0,20 mol/L; glicerol 2% (v/v), na concentração de 65 μ mol/L. Medidas realizadas no equipamento Jasco J750 (Jasco, Japão), com aquecimento de 1°C/min, sendo registrados os dados de CD a cada 0,1°C.

4.2.3- Espalhamento Dinâmico de Luz (DLS)

Devido a problemas de precipitação da proteína SCF5, não foi possível obter concentrações superiores a 2,7 mg/mL. Para verificar o estado de dispersão da amostra obtida, para posterior utilização em ensaios estruturais, foi empregada a técnica de espalhamento dinâmico de luz.

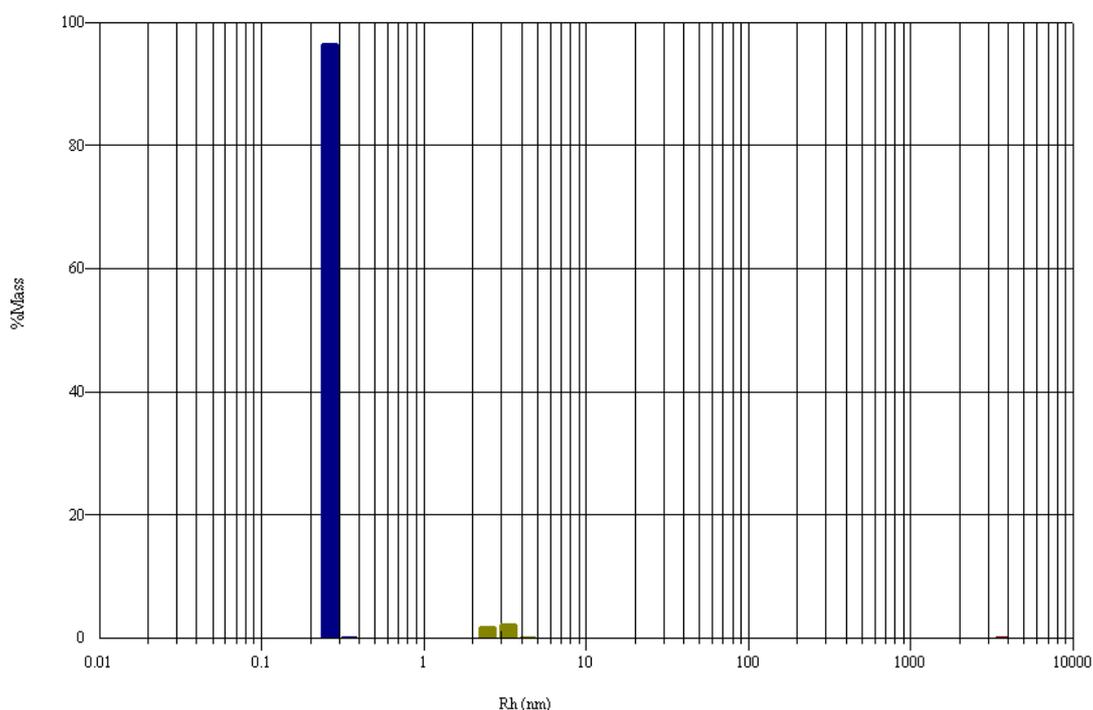


Figura 27- DLS: gráfico referente à distribuição das espécies presentes em solução (SCF5/DNA) em tampão Tris-HCl 50 mmol/L, pH 7,5; glicerol 2% (v/v); NaCl 250 mmol/L, a 2,4mg/mL de proteína e proporção molar bZIP:DNA de 1:1. O gráfico (% massa da amostra total vs raio hidrodinâmico) mostra que 95% da massa da amostra encontra-se no mesmo estado de agregação.

O experimento de DLS determinou a presença de agregados na amostra de proteína sem a presença de DNA. Porém, a amostra que continha o DNA (SCF5/DNA) apresentou-se monodispersa, condição necessária para os ensaios estruturais subsequentes. As barras nas cores azul e verde, presentes na Figura 27, mostram a distribuição de tamanho das partículas presentes no sistema

estudado. A barra em azul indica que aproximadamente 95% das massas presentes no sistema possuem o mesmo raio hidrodinâmico, indicando que o sistema apresenta mono-dispersão suficiente para os estudos estruturais planejados.

4.2.4- Experimento de SAXS (*Small-Angle X-ray Scattering*)

Os procedimentos envolvidos no experimento de SAXS, que abrangem a parte da coleta dos dados de espalhamento e a análise dos mesmos, foram realizados com a participação do aluno de doutorado Marcelo Leite dos Santos, pertencente ao nosso grupo de pesquisa.

Inicialmente os dados de SAXS coletados foram corrigidos, normalizados e os seus respectivos erros foram calculados no programa TRAT1D. Utilizando o programa PRIMUS, os dados de espalhamento dos diferentes *frames*, correspondentes a uma mesma amostra, foram promediados e, em seguida, subtraídos dos dados de espalhamento dos respectivos tampões. Este procedimento permitiu obter uma curva de espalhamento da amostra sem a contribuição do espalhamento de Raios X advindo do tampão. Os tratamentos iniciais dos dados indicaram que a coleta foi bem sucedida e não houve formação de agregados em solução. Na Tabela 4 estão descritas as amostras nas condições utilizadas para a coleta de dados.

Como não observamos diferenças significativas após o processamento dos dados de SAXS nas diferentes condições das amostras, apresentaremos a seguir somente os resultados obtidos para a amostra F5ta11.

Utilizando o programa CRY SOL, calculamos a curva teórica de SAXS para um modelo cristalográfico de complexo bZIP/DNA que foi computacionalmente preparado utilizando uma bZIP de 64 aminoácidos (entrada PDB 1GD2, cadeias E e F) e uma dupla fita de DNA de 24 pares de base (entrada PDB 1L1M, cadeias C e D). Para simplificar, denominaremos este sistema como 1GD2-1L1M (Figura 28a). O ajuste entre a curva de espalhamento experimental da amostra F5ta11 e a

calculada pelo programa CRY SOL para o modelo cristalográfico é apresentado na Figura 28b.

Tabela 4: Amostras submetidas à análise por SAXS.

amostra	proporção molar proteína:DNA	concentração de proteína (mg/mL)	tampão	Número de <i>frames</i> coletados
F5ta11	1:1	2,2	Tris-HCl 50 mmol/L, pH 7,5; glicerol 2% (v/v); NaCl 250 mmol/L	20
F5tb11	1:1	2,7	Tris-HCl 50 mmol/L, pH 7,5; glicerol 10% (v/v); NaCl 250 mmol/L	10
F5ta21	2:1	2,2	Tris-HCl 50 mmol/L, pH 7,5; glicerol 2% (v/v); NaCl 250 mmol/L	10

Observa-se na Figura 28 um bom ajuste entre a curva de espalhamento experimental obtida para o complexo e a curva calculada para o modelo cristalográfico. Este é um indicativo de que, a baixa resolução, o complexo bZIP SCF5/DNA apresenta uma forma semelhante ao modelo. Esta verificação nos encorajou a utilizá-lo para interpretação dos envelopes de SAXS da SCF5.

De posse da curva de espalhamento experimental da amostra, a função de distribuição de distâncias, $\rho(r)$, e o raio de giro, R_g , foram avaliados no programa GNOM. O valor de R_g também foi estimado através da aproximação de Guinier. A curva de espalhamento da amostra F5ta11 e a curva ajustada pelo programa GNOM são apresentadas na Figura 29a. Também são mostradas a função $\rho(r)$ (Figura 29b) e a região de Guinier (Figura 29c), que apresentou o comportamento linear esperado para um sistema monodisperso, considerando a região de $q_{\text{máx}} < 1,3/R_g$ onde é válida esta aproximação (Svergun & Koch, 2003).

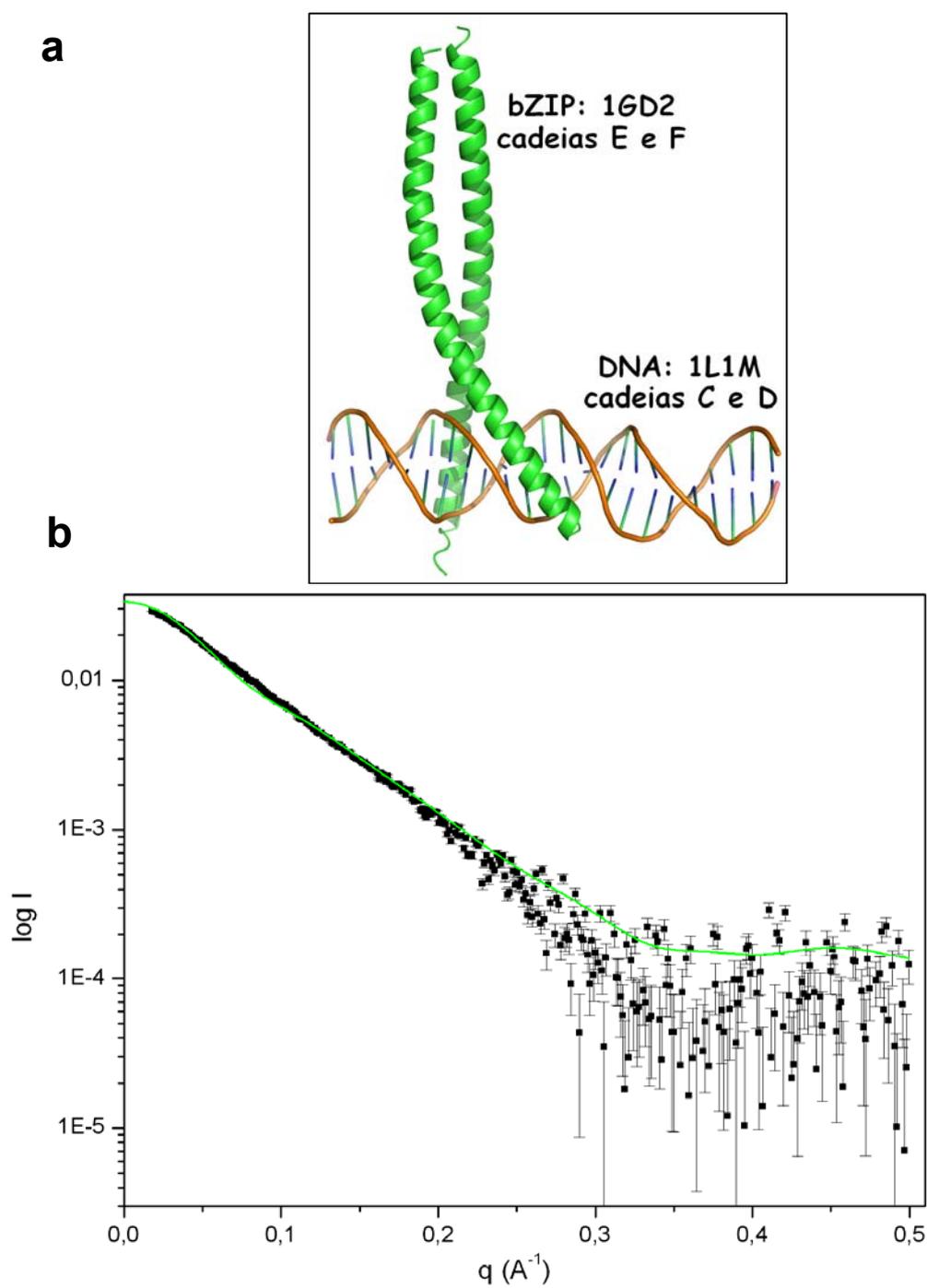


Figura 28- a) Representação da estrutura do complexo 1GD2-1L1M computacionalmente obtida. **b)** Comparação entre as curvas de SAXS experimental da bZIP SCF5/DNA (quadrados com barras de erro) e calculada com o programa CRY SOL para o modelo cristalográfico preparado (linha verde).

O formato da curva $p(r)$ está relacionado com as características de forma e tamanho do complexo em estudo. Modelos cristalográficos disponíveis mostram que complexos bZIP/DNA apresentam a forma alongada (prolata) com duas porções bem definidas: o dímero de proteína e a dupla fita de DNA. A $p(r)$ para esse tipo de molécula corresponde a uma curva assimétrica e inclinada à esquerda, como pode ser verificado pela comparação dos nossos resultados com aqueles correspondentes à partícula cilíndrica em verde apresentada na Figura 11 (seção 1.5.3., pág. 20). A Figura 29a mostra que a curva obtida para o sistema SCF5/DNA é condizente com o esperado e que o complexo apresenta uma dimensão máxima ($D_{máx}$) em torno de 120 Å. Observa-se que alguns pontos na região de mais altos ângulos apresentam erros pequenos incompatíveis com o esperado, de modo que consideramos esses pontos como sendo *outliers*.

Abaixo, na Tabela 5, são apresentados os valores de raios de giro obtidos para o sistema SCF5/DNA através da $p(r)$ e da aproximação de Guinier. Na mesma tabela, também são apresentados os valores de R_g , calculados pelo programa CRY SOL, para os modelos cristalográficos de bZIP/DNA com códigos PDB 1GD2, 1DGC e para o modelo computacionalmente criado 1GD2-1L1M.

Tabela 5- Raios de giro calculados para os modelos ligados ao DNA. Os R_g dos modelos PDBs 1GD2, 1DGC e modelo criado computacionalmente 1GD2-1L1M, foram calculados pelo programa CRY SOL. O R_g do complexo SCF5/DNA foi calculado pelo GNOM e pela aproximação de Guinier.

Modelo	Raio de Giro (Å)
SCF5/DNA - $p(r)$	$30,2 \pm 0,2$
SCF5/DNA - Guinier	27 ± 3
PDB 1GD2	28,3
PDB 1DGC	25,4
1GD2-1L1M	31,3

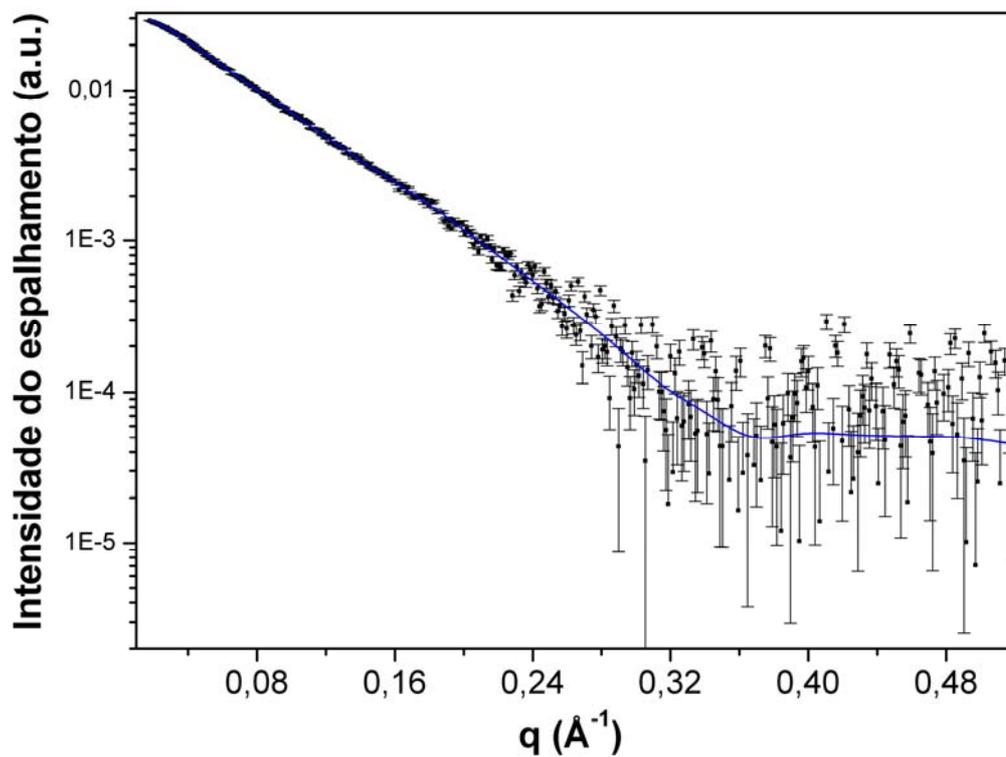
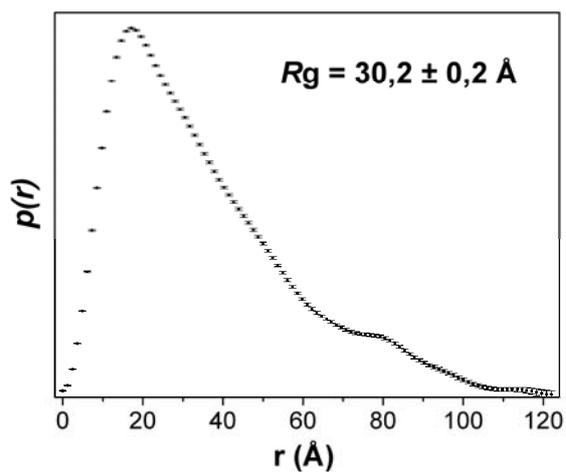
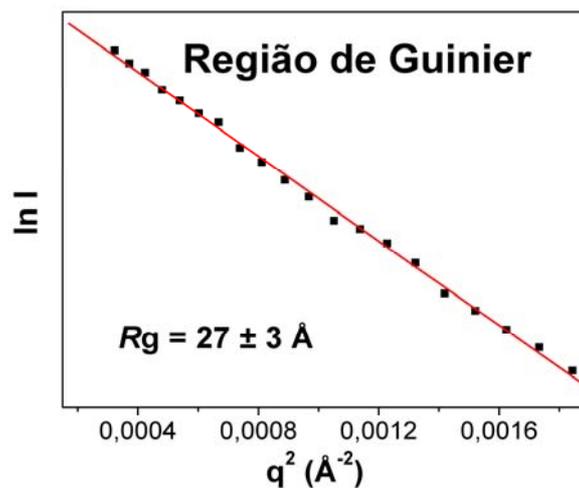
a**b****c**

Figura 29- a) Curva de espalhamento da proteína SCF5 ligada ao DNA (quadrados com barras de erro) e curva ajustada pelo programa GNOM (linha azul). Também são apresentados: b) a função de distribuição de distâncias calculada pelo GNOM e c) o gráfico de Guinier.

Através da comparação dos valores apresentados na Tabela 5 podemos verificar que o valor obtido para o raio de giro do complexo SCF5/DNA através da função $p(r)$ é mais parecido com o do modelo computacionalmente criado (1GD2-1L1M). Este procedimento é mais confiável do que o cálculo de R_g pela região de Guinier, pois para o cálculo da função $p(r)$ são utilizados todos os valores da curva de espalhamento e através da região de Guinier é feita uma análise por regressão linear que utiliza apenas os primeiros pontos da curva de espalhamento. Sendo assim, embora o valor calculado através da região de Guinier ser mais próximo do raio de giro dos modelos do PDB, pode-se sugerir que o complexo em estudo (SCF5/DNA) apresenta dimensões mais próximas do modelo (1GD2-1L1M) do que dos modelos estruturais de alta resolução das bZIPs/DNA (1DGC e 1GD2). Esta é uma verificação interessante uma vez que os modelos cristalográficos de bZIPs disponíveis no PDB apresentam uma quantidade de aminoácidos e de pares de bases menor que o complexo SCF5/DNA.

Os modelos de baixa resolução foram recuperados a partir da curva de espalhamento, utilizando os programas DAMMIN e GASBOR. Os melhores envelopes foram obtidos utilizando o programa GASBOR. Um total de 10 modelos foi construído e um modelo de baixa resolução médio foi obtido com o uso do programa DAMAVER. Empregando-se o programa SUPCOMB, o modelo 1GD2-1L1M foi sobreposto ao envelope de SAXS do complexo SCF5/DNA, Figura 30.

Os espaços não preenchidos observados no modelo de baixa resolução devem-se, provavelmente, ao tamanho real da bZIP que estamos estudando, maior que aquela utilizada como modelo para interpretação do envelope de SAXS. A bZIP SCF5 possui 104 aminoácidos em sua estrutura primária enquanto o modelo cristalográfico (1GD2) possui apenas 64 aminoácidos. A dupla fita de DNA possui o mesmo número de pares de base que a utilizada em nossos experimentos, 24 pb. Os modelos de baixa resolução obtidos sugerem que o fator de transcrição forma um homodímero ligado de forma assimétrica ao DNA.

4.2.5- Ensaio de cristalização

Os testes de cristalização do complexo SCF5/DNA realizados inicialmente com os kits *Crystallization Basic Kit for Proteins* e *Crystallization Extension Kit for Proteins* (Sigma, EUA) mostraram algumas condições que pareceram ser favoráveis à formação de cristais de proteína. Verificou-se que em algumas delas estava presente PEG 4000, em concentração por volta de 30% (m/v). Na Figura 31 são mostradas duas condições (soluções 22 e 31 do *Crystallization Basic Kit for Proteins*, pág. 35) nas quais ocorre uma separação de fases, um indício de que poderia haver formação de cristais, na presença de PEG 4000.

Com base nos resultados obtidos nos testes iniciais de cristalização, foi realizado um refinamento da condição 22 (acetato de sódio trihidratado 0,2M; tris HCl 0,1M, pH 8,5; Polietileno glicol 4000 30% m/v) do *Crystallization Basic Kit for Proteins*. Foram montadas duas novas placas (1 e 2, Figura 15, pág. 37) sendo que, em uma delas, variou-se a concentração de PEG 4000 e o pH (placa 1) e, na outra, a concentração de acetato de sódio e PEG 4000 (placa 2). Alguns pequenos cristais foram obtidos nas condições com PEG 4000 até 30% (m/v) (Figura 32), porém com qualidade insuficiente para a realização de experimentos de difração de Raios X. A melhor condição pareceu ser a C1 da placa 1. Então, na tentativa de se obter cristais de melhor qualidade foram montadas duas novas placas (placas 3 e 4) para o refinamento da condição C1 placa 1 e uma outra placa (placa 5) para verificar o efeito da presença de alguns aditivos (Figura 16, pág. 38). Tais aditivos são substâncias que podem alterar alguns parâmetros na solução de cristalização, tais como força iônica e solubilidade, que podem favorecer o processo de formação e crescimento do cristal de proteína.

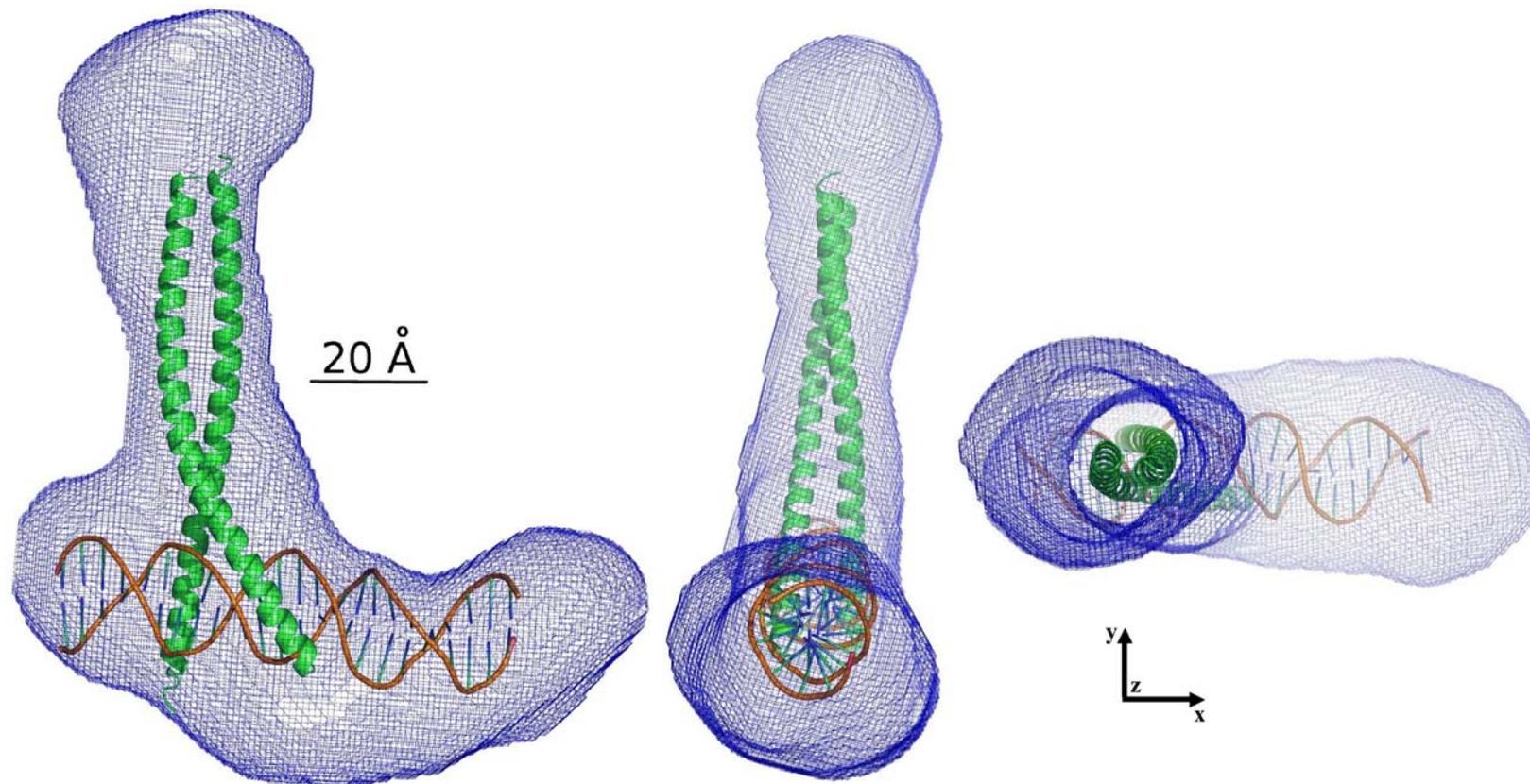


Figura 30- Modelo de baixa resolução (mapa em azul) para o complexo SCF5/DNA obtido através do programa GASBOR a partir da curva experimental, superposto ao modelo 1GD2-1L1M da Figura 28(a). A Figura da esquerda representa o envelope visto no plano xy; as Figuras do centro e da direita representam a perspectiva vista após rotação de 90° no sentido horário em torno do eixo y e anti-horário em torno do eixo x, respectivamente. A barra de escala permite relacionar as distâncias intermoleculares à função $p(r)$ da Figura 29(b). O máximo da $p(r)$ em torno de 20 Å corresponde, aproximadamente, ao diâmetro de seções circulares perpendiculares à direção das hélices que compõem a proteína e à direção das fitas de DNA. Também é interessante notar que o máximo local em torno de 80 Å provavelmente corresponda à distância entre átomos localizados nos glóbulos presentes nas extremidades do modelo. A máxima distância obtida a partir do modelo está em torno de 120 Å, em acordo com a função $p(r)$.

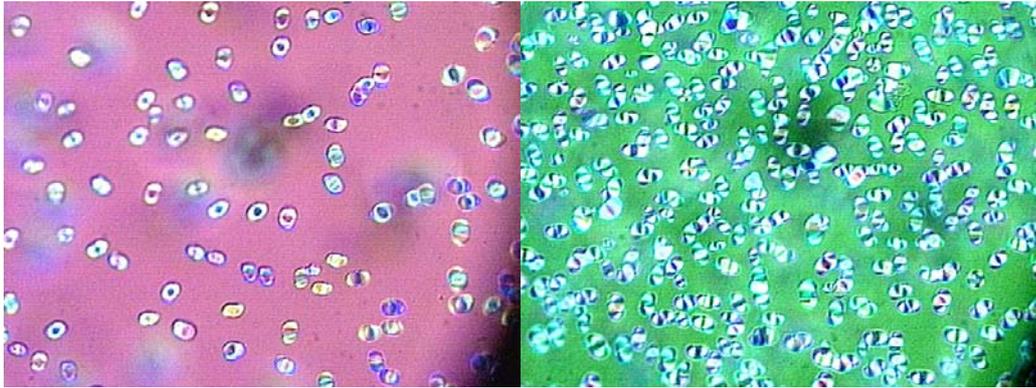


Figura 31- Testes iniciais de cristalização ilustrando separação de fases. A foto da esquerda corresponde à solução 22 do *Crystallization Basic Kit for Proteins* (acetato de sódio trihidratado 0,2M; tris HCl 0,1M, pH 8,5; Polietileno glicol 4000 30% m/v). A foto da direita corresponde à solução 31 do mesmo *kit* (Sulfato de amônio 0,2M; Polietileno glicol 4000, 30% m/v)

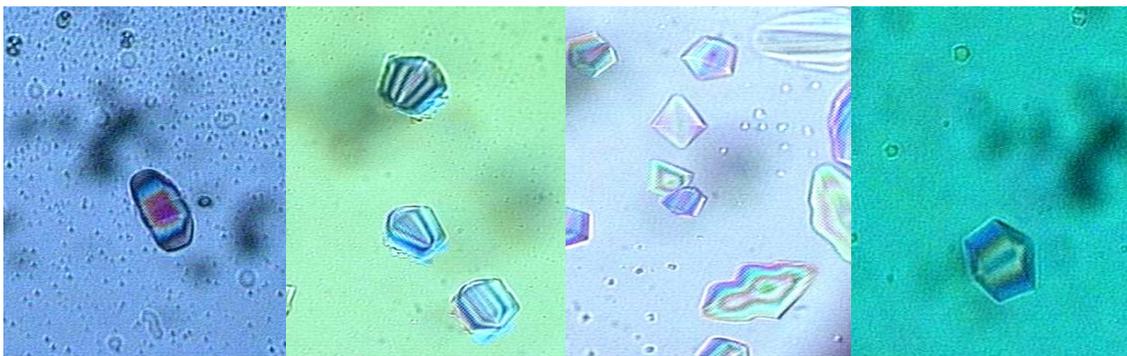


Figura 32- Fotos dos cristais obtidos durante o refinamento da condição 22 do *Crystallization Basic Kit for Proteins*. A terceira foto, da esquerda para a direita, corresponde à condição C1 da placa 1. A maior dimensão destes cristais é da ordem de 20 μm .

Foram obtidos alguns cristais na presença dos aditivos DMF e Tween 20 (placa 5, Figura 16, pág. 38), porém novamente com qualidade inferior à necessária para a coleta de dados de difração. É importante dizer que os cristais eram extremamente frágeis ao toque mecânico, um forte indicativo que realmente eram cristais de proteínas. Na Figura 33 estão mostrados as fotos de alguns

cristais obtidos nas condições utilizadas na placa 5.

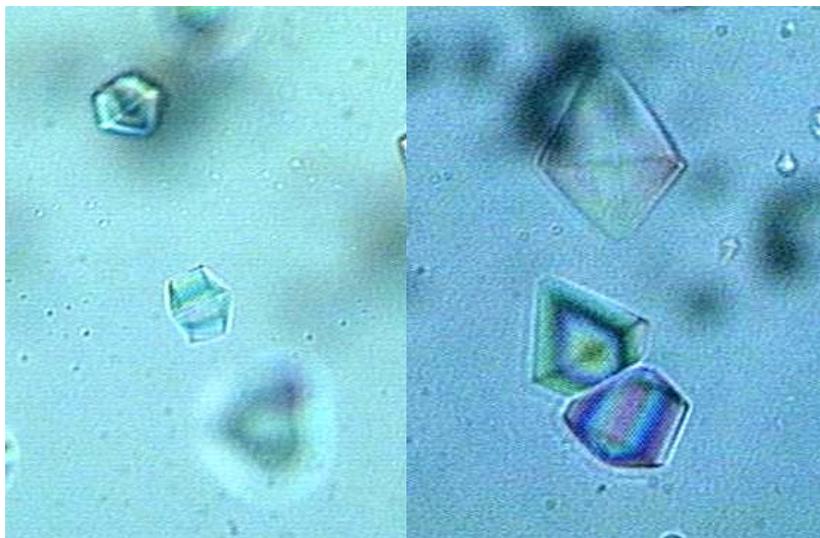


Figura 33- Fotos dos cristais obtidos na placa 5, conforme condições aplicadas: a Figura da esquerda representa a condição A2 e a da direita a condição B3. A maior dimensão destes cristais é da ordem de 20 μm .

5- Discussão

5.1-SCF12

A clonagem do gene da SCF12 no pET28a permitiu a produção da proteína recombinante SCF12. A expressão da proteína a 37°C nas cepas de *E. coli* BL21 (DE3) pRil foi mais eficiente que na cepa pLysS.

O vetor pET28a confere à proteína produzida a cauda de 6 histidinas na posição N-terminal, o que possibilitou sua purificação pelo método de cromatografia líquida de afinidade com resina de metal imobilizado (IMAC - *Immobilized Metal Affinity Chromatography*). As amostras eluídas com concentrações crescentes de imidazol e analisadas em SDS-PAGE mostraram que proteína pura é obtida depois da adição de 100 mmol/L de imidazol. O processo de purificação foi bem sucedido, porém o rendimento foi relativamente baixo, pois parte da proteína ficava retida na resina, provavelmente devido a algum tipo de interação da proteína com a própria matriz da resina. A utilização de concentrações maiores de imidazol na eluição pode ser uma alternativa para contornar este obstáculo. Tentativas com outras resinas foram feitas, mas não foi observada uma melhora significativa no rendimento do processo. Por outro lado, o rendimento de 0,5mg/L de cultura foi suficiente para os estudos propostos.

A proteína SCF12 foi então utilizada num experimento de dicroísmo circular para caracterizar e quantificar a sua estrutura secundária. Inicialmente foi utilizado o tampão Tris-HCl 0,020 mol/L, pH 7,5; EDTA 5 mmol/L; β-mercaptoetanol 7 mmol/L; NaCl 0,100 mol/L, sendo que Tris e NaCl possuem forte absorvância na região do UV distante. Assim, não foi possível obter dados para comprimentos de onda inferiores a 200 nm, o que impossibilitava a análise dos dados nos programas disponíveis no *Dichroweb Internet Service*. Desta forma, foi pesquisado na literatura um novo tampão mais compatível com o experimento de CD, sem comprometer a estabilidade da proteína. Em Barteri *et al.*, (1996) foi descrito que em análises de CD de outra classe de proteína, contendo glicerol no tampão, apresentaram maior quantidade de hélices-α em comparação às amostras sem glicerol. Foi utilizado então um novo tampão, de composição fosfato de sódio

0,050 mol/L, pH 7,2; glicerol 2% (v/v); Na₂SO₄ 0,050 mol/L, possibilitando leituras de absorvância em comprimentos de onda até 190 nm. Este fato possibilitou a utilização do programa CONTIN para analisar os novos dados de CD coletados. Foi escolhido este programa porque era o mais adequado à faixa de λ em que foi possível realizar o experimento (240 a 190nm). A análise mostrou que a SCF12 encontra-se enovelada a temperatura de 20°C, no tampão mencionado, sendo composta majoritariamente de hélices- α , o que é coerente com o que é relatado na literatura (Hurst, 1995).

Devido ao fato de a primeira amostra de SCF12 não possuir dados coletados para λ inferiores a 200 nm, não foi possível a análise pelo CONTIN. Porém, para uma rápida comparação, foram analisados os dados coletados da primeira amostra (sem glicerol) e da segunda amostra (com glicerol) no próprio programa na JASCO (que utiliza um banco de dados do próprio programa que vem junto do equipamento). A nova amostra, que continha glicerol, apresentou menor porcentagem de composição randômica na proteína, em comparação à amostra inicialmente medida sem glicerol. Este fato foi tomado como um indicativo de que o glicerol pode ajudar na estabilidade estrutural da proteína SCF12. Não foram realizados testes de estabilidade térmica para a SCF12 porque, diferentemente da SCF5, não é necessário o passo de aquecimento a 80°C durante o processo de purificação.

Para a determinação de uma seqüência de DNA ao qual a proteína SCF12 se ligaria, foram realizados dois experimentos de mobilidade eletroforética diferencial (*Electrophoretic mobility shift assay - EMSA*), onde foram testadas 8 seqüências diferentes de DNA, porém, sem sucesso. Até o momento, não existe nenhum modelo cristalográfico de bZIP disponível sem que a mesma esteja ligada a uma fita de DNA. O fato de não se ter uma seqüência de DNA ligada à bZIP compromete a obtenção de um sistema estável em solução, condição esta necessária tanto para a cristalização de proteínas, como para os experimentos de SAXS. Desta forma, não foi possível a realização de experimentos de SAXS para a SCF12.

5.2- SCF5

Os dados de clonagem e a metodologia para expressão da SCF5 foram obtidos pelo Dr. Paulo S. Schlögl no CBMEG/Unicamp (Centro de Biologia Molecular e Engenharia Genética), durante seus trabalhos de doutorado. As metodologias encontram-se na literatura (Schlögl *et al.*, 2004). A parte do gene que codifica a região básica e o zíper de leucinas da proteína SCF5 foi inserido no vetor de expressão pET3c (Novagen, EUA) e para a expressão foram utilizadas células de *E. coli* do tipo BL21 (DE3) pLysS. A proteína obtida possui 104 aminoácidos, tendo uma massa molecular de 12,5 kDa e pI teórico de 8,97 (<http://ca.expasy.org>).

A purificação da proteína SCF5 também está descrita na literatura, bem com algumas características funcionais. Porém, com o intuito de minimizar o tempo gasto no processo de purificação da proteína, foi estabelecido um novo protocolo no qual o sobrenadante obtido, após aquecimento a 80°C por 3 minutos e centrifugação da fração solúvel do extrato bacteriano (descrito em Schlögl *et al.*, 2004), é passado em seguida em coluna de troca catiônica ao invés da cromatografia de fase reversa. Essa medida permitiu a dispensa de pré-tratamentos de precipitação com sulfato de amônio e polietilenoimina (PEI) sem comprometer o rendimento e a pureza da proteína recuperada. Apesar de ser comumente usado pH em torno de 5 em cromatografia de troca catiônica, optou-se inicialmente por testar um pH mais próximo de condições fisiológicas. Mesmo na condição de pH 7,5 utilizada nos tampões de troca catiônica, a SCF5 interage fortemente com a resina da coluna e começa a ser eluída quando o gradiente está em torno de 80% de tampão B (concentração de aproximadamente 800 mmol/L de NaCl), resultando em uma purificação de proteína muito satisfatória.

Como a purificação da proteína SCF5 tem um passo de aquecimento a 80°C, foi realizado um experimento para avaliar a estabilidade térmica da proteína por CD. Os experimentos de aquecimento mostraram que a proteína vai perdendo o enovelamento gradativamente até que, a 80°C ela se encontra desnaturada. Porém, com o resfriamento da amostra, a SCF5 vai novamente recuperando sua

estruturção e o fato de ela ligar-se ao DNA após o processo de purificação, descrito nos experimentos de DLS e SAXS, mostra que a proteína se re-enovela corretamente. Neste sentido, Schlögl *et al.* também confirmaram que a proteína mantinha atividade.

Para os experimentos de SAXS, foi sintetizada uma fita de DNA ao qual a proteína SCF5 se liga. Foram realizadas duas coletas de SAXS. Na primeira, trabalhou-se com a proteína SCF5 dissolvida em Tris-HCl 0,020 mol/L, pH 7,5; EDTA 5 mmol/L; β -mercaptoetanol 7 mmol/L; NaCl 0,020 mol/L com e sem a presença de DNA. Houve a formação de agregados em ambas as amostras, impossibilitando a conclusão do experimento. Para a segunda coleta, mudou-se o tampão no qual estava dissolvida a proteína. O resultado obtido no experimento de CD para a proteína SCF12, que mostrou que a amostra contendo glicerol apresentava menor quantidade de estrutura randômica, foi tomado como um indicativo de que o glicerol poderia estabilizar a estrutura da SCF5, impedindo a formação dos agregados.

Inicialmente, seriam coletados dados de SAXS para a proteína SCF5 sem DNA e para a SCF5/DNA. Porém, antes da coleta na linha de SAXS as amostras da proteína foram analisadas por DLS. O resultado foi satisfatório para a amostra SCF5/DNA, mas no caso da proteína sem DNA houve novamente a formação de agregados. Desta forma, foram coletados dados de SAXS apenas para o complexo SCF5/DNA. Foram utilizadas 3 amostras conforme a Tabela 4 (pág. 50). Porém, o resultado do processamento dos dados foi muito semelhante e o modelo mostrado na Figura 30 corresponde à amostra F5ta11.

A utilização do complexo bZIP/DNA gerado computacionalmente (modelo 1GD2-1L1M) possibilitou a comparação, pelo programa CRY SOL (Svergun *et al.*, 1995), entre a curva de espalhamento experimental e a curva calculada. A curva de espalhamento obtida para o complexo SCF5/DNA apresenta um comportamento muito semelhante a do complexo 1GD2-1L1M, sendo um indicativo de que os dados experimentais de espalhamento estão de acordo com o esperado. Outro aspecto importante é que o raio de giro calculado, tanto para os modelos cristalográficos quanto para a SCF5/DNA apresentam valores próximos.

O modelo obtido é formado por um homodímero da SCF5 ligado a uma molécula do DNA e é alongado, como esperado para complexos bZIP/DNA. O modelo bZIP/DNA inserido no envelope da SCF5/DNA, construído a partir das estruturas disponíveis no PDB, possui um número muito menor de aminoácidos, explicando os espaços não preenchidos no modelo da Figura 30. Outro fator que é possível observar é que a interação proteína/DNA ocorre de forma não simétrica. Isso pode ocorrer devido à maneira com que os oligonucleotídeos foram sintetizados, onde o sítio que a SCF5 reconhece não está no centro da molécula de DNA, conferindo essa assimetria ao modelo.

Apesar de que o modelo obtido por SAXS seja de baixa resolução, este é o primeiro modelo de uma bZIP de planta, já obtido experimentalmente. O modelo permite distinguir entre algumas possibilidades de ligação, pois existem bZIPs que interagem de maneira diferente entre si e com a fita de DNA (Nair *et al.*, 2003). Além disso, os experimentos de SAXS levaram à confirmação de que a proteína encontra-se monodispersa em solução, na forma dimérica e ligada a uma fita de DNA.

Quanto aos resultados de cristalização, apesar das condições testadas até o presente momento, não foram obtidos cristais apropriados para os experimentos com os Raios X. Um fator limitante enfrentado é que não foi possível obter concentrações da SCF5 superiores a 2,8 mg/mL, pois ocorria precipitação acima desta concentração. Ainda assim, uma outra bZIP, GCN4, de *Saccharomyces cerevisiae*, foi cristalizada na concentração em torno de 3 mg/mL (König e Richmond, 1993). Tal fato mostra que, em relação à concentração de proteína, a cristalização da SCF5 é viável, requerendo a realização de novos ensaios a fim de se obter cristais adequados para os experimentos de difração.

6- Trabalhos adicionais

Os objetivos propostos inicialmente para este projeto visavam o estudo estrutural de duas proteínas bZIP de cana-de-açúcar, SCF5 e SCF12. Porém, paralelamente às atividades do projeto inicial, foram realizados outros trabalhos com outras duas proteínas, a UGPase de cana-de-açúcar, e a Aldose Redutase de milho.

6.1- UGPase

A UGPase (UTP:a-D-glucose-1-phosphate uridilil-transferase, EC 2.7.7.9, MW: 38,1 KDa) é uma proteína de cana-de-açúcar que está envolvida na biosíntese de sacarose. Do programa SUCEST/FAPESP foi obtido o clone, no vetor pSPORT, com o cDNA correspondente à proteína UGPase (SUCEST *cluster consensus*: SCJLRT2051F02.g). Os vetores de clonagem e expressão foram obtidos pelo doutorando Marcelo Leite dos Santos e pela aluna de iniciação científica Aline da Costa Lima (pertencentes ao grupo de pesquisa coordenado pelo Prof. Dr. Ricardo Aparicio), supervisionados pela Dra. Juliana Felix, no Laboratório Genoma Funcional - CBMEG, coordenado pelo Prof. Dr. Marcelo Menossi Teixeira. Participamos diretamente no estabelecimento dos protocolos de expressão e purificação desta proteína e em coletas de dados de SAXS. Os resultados de SAXS são promissores e a proteína encontra-se em cristalização.

6.2- Aldose Redutase (AR)

A caracterização bioquímica da AR fez parte dos trabalhos da então aluna de doutorado Sylvia Morais de Sousa (CBMEG), orientada pelo Prof. Dr. José Andrés Yunes. Para a caracterização estrutural da AR foi estabelecida uma parceria com grupo de pesquisa do Prof. Dr. Ricardo Aparicio, orientador do presente trabalho. A AR (EC 1.1.1.21) é uma enzima membro da superfamília de aldo-ceto redutases, que é formada por proteínas monoméricas (~35 kDa) do tipo

(α/β) 8-*barrel*, que usam NADP/NADPH para catalisar a redução de aldeídos e cetonas, monossacarídeos, cetoesteróides, prostaglandinas e outros substratos específicos (Jez & Penning, 2001). Os experimentos de cristalização e coleta de dados de cristalografia foram bem sucedidos e foram publicados no artigo: Kiyota E., de Sousa S. M. , dos Santos M. L. , da Costa Lima A. , Menossi M. , Yunes J. A. and Aparicio R., (2007) Crystallization and preliminary X-ray diffraction analysis of maize aldose reductase. *Acta Cryst.*, **F63**, 990-992, incluído no anexo 1 desta dissertação.

7- Conclusão e Perspectivas

- Foi estabelecido um novo protocolo para purificação das proteínas SCF5 e SCF12;
- Os experimentos de dicroísmo circular mostraram que as proteínas apresentavam-se enoveladas nas condições estudadas e, no caso da SCF12, foi possível estimar elementos da sua estrutura secundária;
- Através dos experimentos de espalhamento de Raios x à baixos ângulos foi obtido um modelo para a proteína SCF5 ligada a DNA.
- Não foi possível, pela técnica de *gel shift*, determinar uma seqüência específica de DNA ao qual a proteína SCF12 se liga. Novos experimentos poderão ser realizados, testando novas seqüências;
- Novos testes de cristalização, utilizando uma fita de DNA menor à empregada nos experimentos iniciais, poderão ser implementados;
- Coleta de dados de SAXS poderá ser realizada para a proteína SCF12 ligada a DNA;
- A Aldose Redutase foi cristalizada e já foi obtido um modelo cristalográfico que atualmente encontra-se em refinamento;
- Um modelo de SAXS foi obtido para a UGPase e experimentos de cristalização estão em andamento.

Anexo 1

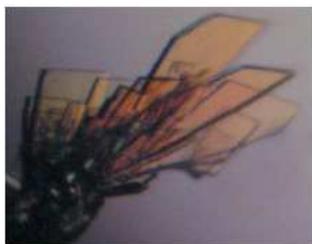
Eduardo Kiyota,^{a,b} Sylvia Morais de Sousa,^{b,†} Marcelo Leite dos Santos,^a Aline da Costa Lima,^a Marcelo Menossi,^c José Andrés Yunes^d and Ricardo Aparicio^{a*}

^aLaboratório de Biologia Estrutural, Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas-SP, Brazil, ^bCentro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas-SP, Brazil, ^cDepartamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, Campinas-SP, Brazil, and ^dLaboratório de Biologia Molecular, Centro Infantil Boldrini, Campinas-SP, Brazil

† Current address: Horticultural Sciences Department, University of Florida, Gainesville, Florida, USA.

Correspondence e-mail:
aparicio@iqm.unicamp.br

Received 24 September 2007
Accepted 23 October 2007



© 2007 International Union of Crystallography
All rights reserved

Crystallization and preliminary X-ray diffraction analysis of maize aldose reductase

Maize aldose reductase (AR) is a member of the aldo-keto reductase superfamily. In contrast to human AR, maize AR seems to prefer the conversion of sorbitol into glucose. The apoenzyme was crystallized in space group $P2_12_12_1$, with unit-cell parameters $a = 47.2$, $b = 54.5$, $c = 100.6$ Å and one molecule in the asymmetric unit. Synchrotron X-ray diffraction data were collected and a final resolution limit of 2.0 Å was obtained after data reduction. Phasing was carried out by an automated molecular-replacement procedure and structural refinement is currently in progress. The refined structure is expected to shed light on the functional/enzymatic mechanism and the unusual activities of maize AR.

1. Introduction

Proteins of the aldo-keto reductase (AKR) superfamily are monomeric (α/β)₈-barrel proteins of approximately 35 kDa in molecular weight which use NAD(P)(H) to catalyze the reduction of aldehydes and ketones, monosaccharides, ketosteroids, prostaglandins and other specific substrates (Jez & Penning, 2001). AKRs occur in prokaryotes and eukaryotes, including yeast, plant, amphibia and mammals (Jez *et al.*, 1997). Aldose reductase (AR; EC 1.1.1.21) is a member of the AKR superfamily (Bohren *et al.*, 1989) that has a monomeric NADPH-dependent oxidoreductase activity with broad substrate specificity. In some species, AR is able to catalyze the reduction of glucose to sorbitol in the polyol pathway (Jez *et al.*, 1997; Ueda *et al.*, 2004). This pathway is especially relevant to diabetic patients. Hyperglycaemia causes an increased flux of glucose through the polyol pathway, which elicits various metabolic imbalances and early damage to tissues that undergo insulin-independent uptake of glucose, such as the lens, retina, peripheral nerve and renal glomerulus. For this reason, AR inhibition has been considered to be an attractive approach to halting long-term diabetic complications (Srivastava *et al.*, 2005). Although the structure of AR has been extensively studied in humans (Rondeau *et al.*, 1992; Wilson *et al.*, 1992; Biadene *et al.*, 2007), little is known about ARs in plants. Some evidence suggests that plant ARs are involved in abiotic stress resistance. For example, freezing treatment induces AR expression in brome grass (Lee & Chen, 1993) and in *Xerophyta viscosa* AR is expressed under dehydration conditions (Mundree *et al.*, 2000).

Barley AR is expressed in embryos with a pattern that correlates with the seed-maturation phase (Bartels *et al.*, 1991; Roncarati *et al.*, 1995). Likewise, rice AR is expressed at high levels when the seeds reach maturity (Sree *et al.*, 2000). In maize, on the other hand, AR is broadly expressed but is apparently more abundant in embryos, especially in the earlier phases of development. Maize AR not only displays the typical activity of AKR-family members, but is also able to convert sorbitol into glucose, an activity which seems especially adapted to maize seeds and allows developing maize embryos to metabolize substantial amounts of sorbitol from the endosperm (de Sousa *et al.*, work to be published). We have initiated a study that aims to understand the apparent preference of maize AR for sorbitol over glucose as a substrate, a feature that is the opposite of that observed in humans and fungi. Here, we report the crystallization of

crystallization communications

apo maize aldose reductase and its preliminary crystallographic analysis.

2. Material and methods

2.1. Cloning, expression and purification

Detailed methodology will be published elsewhere (de Sousa *et al.*, work to be published). The gene encoding maize AR (DQ517521) was amplified and cloned into the vector pET28a (Novagen). Competent *Escherichia coli* strain BL21(DE3) pRil cells were transformed with positive recombinant plasmid and cultured in LB broth with 50 $\mu\text{g ml}^{-1}$ kanamycin and 100 $\mu\text{g ml}^{-1}$ chloramphenicol. Lactose (100 mM) was added to induce protein expression and incubation proceeded for an additional 6 h. The cells were harvested and pellets were resuspended in affinity buffer (50 mM sodium phosphate buffer pH 7.2) containing 100 mM NaCl and 5% glycerol. Lysozyme was added to a final concentration of 1.0 mg ml⁻¹. The suspension was incubated for 1 h at 277 K and sonicated. Insoluble debris was removed by centrifugation and the clarified supernatant was used for protein purification by IMAC (immobilized metal-affinity chromatography). The eluted maize AR was dialysed in anion-exchange buffer (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 7 mM β -mercaptoethanol and 20 mM NaCl) and then purified on a Q-Sepharose FF anion-exchange chromatography column (1 ml; Amersham Biosciences, USA) using an ÄKTA FPLC system (Amersham Biosciences, USA). Bound proteins were eluted using a NaCl gradient. Protein concentration and purity was analyzed by SDS-PAGE. More accurate estimations for purified maize AR were made based on the absorbance at 280 nm, using a calculated extinction coefficient of 1.824 g l⁻¹ cm⁻¹ (Pace & Schmid, 1997).

2.2. Crystallization

Initial crystallization conditions were screened in Tissue Culture Test Plates 24 (TPP) by the hanging-drop method at 293 K, using the sparse-matrix method (Jancarik & Kim, 1991) implemented in the Crystallization Basic and Extension Kits for Proteins (Sigma). Imperfect crystals were obtained in various conditions and were used as a guide for further optimization. Good diffracting crystals were obtained in a condition similar to condition 22 of the Crystallization Basic Kit. The optimum reservoir solution, consisting of 26% PEG 4000 (Sigma/Fluka), 0.2 M sodium acetate (Vetec) and 0.1 M Tris-HCl pH 6.5 (Vetec), was mixed with protein solution (10 mg ml⁻¹ in water) in equal amounts and equilibrated against reservoir solution.

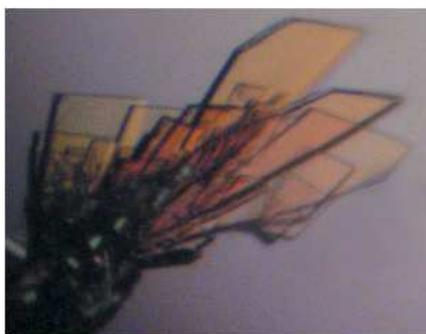


Figure 1
Representative crystals of maize AR grown as clusters of plates with maximum dimensions of approximately 400 \times 200 \times 50 μm .

Table 1
Crystal parameters and data-collection statistics

Values in parentheses are for the last resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (\AA)	
<i>a</i>	47.2
<i>b</i>	54.5
<i>c</i>	100.6
Solvent content (%)	34.4
Molecules in ASU	1
Resolution range (\AA)	50.32–2.00 (2.11–2.00)
No. of images	129
No. of observed reflections	77147 (11331)
No. of rejected reflections	2592
No. of unique reflections	18151 (2593)
Multiplicity	4.3 (4.4)
Completeness (%)	99.8 (99.8)
R_{sym} (%)	11.8 (38.5)
$\langle I/\sigma(I) \rangle$	9.7 (2.8)
Wilson plot B factor (\AA^2)	243

Crystals were obtained as clusters of plates and grew to full size in two weeks at 293 K (Fig. 1). Attempts to obtain single crystals by the use of additives, seeding and other strategies were not successful. However, single plates manually separated from the initial clusters exhibited good morphology and size and proved to be of sufficient quality for data collection.

2.3. Data collection and processing

Cryocrystallographic techniques (Garman & Schneider, 1997) were employed to prevent radiation damage. Crystals were briefly soaked in a cryoprotectant solution containing 15% (v/v) ethylene glycol and were rapidly frozen in a gaseous nitrogen stream at 100 K (Oxford Cryosystems). Data were collected by the oscillation method on beamline D03B-MX1 at the Laboratório Nacional de Luz Síncrotron (LNLS, Campinas-SP, Brazil; Polikarpov *et al.*, 1998) using a MAR CCD 165 detector. The X-ray wavelength was 1.425 \AA with a crystal-to-detector distance of 78.2 mm, giving an outer-edge resolution of 1.8 \AA . The oscillation range was 1 $^\circ$ and the exposure time per image was 240 s. Data reduction was performed with the packages *MOSFLM* and *SCALA* (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 1997). A summary of crystal parameters and data-collection statistics is presented in Table 1.

3. Results and discussion

Initial attempts were made to solve the crystal structure of maize AR using homologous protein structures available in the Protein Data Bank. The program *MATTHEWS_COEF* (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 1997) was used to calculate the Matthews coefficient (Matthews, 1968) and the Matthews probability (Kantardjiev & Rupp, 2003). Across all resolution ranges, including the high-resolution limit, the probability of $V_M = 1.9 \text{\AA}^3 \text{Da}^{-1}$ and a solvent content of 34.4% was estimated to be 1.0, unequivocally indicating the presence of one molecule in the asymmetric unit. Primary sequence searches and sequence alignments were made with *ENTREZ* and *BLAST* (Altschul *et al.*, 1997). Molecular-replacement trials were performed manually using the program *AMoRe* (Navaza, 1994; Winn *et al.*, 1997) and although there was an indication of possible solutions, the low contrast obtained suggested the application of automated methods.

Accordingly, the automated procedure for molecular replacement implemented in the program *MrBUMP* (Keegan & Winn, 2007) was adopted. *MrBUMP* employed the programs *FASTA* (Pearson &

Lipman, 1988) and *CLUSTALW* (Chenna *et al.*, 2003) for the template search, *MOLREP* (Vagin & Teplyakov, 1997; Winn *et al.*, 1997) for molecular replacement and *REFMAC* (Murshudov *et al.*, 1997; Winn *et al.*, 1997) for initial evaluation of the possible solutions. The most successful model was prepared with *CHAINSAW* (Schwarzenbacher *et al.*, 2004; Winn *et al.*, 1997) from an homologous structure of pig aldose reductase complexed with sorbinil (PDB code 1ah0; Urzhumtsev *et al.*, 1997), the 11th hit in the primary *BLAST* search (48% sequence identity, corresponding to 138 of 285 amino-acid residues). After 30 cycles of automated restrained refinement, the *R* factor and *R*_{free} (5% of the total reflections) were 37.9% and 43.3%, respectively.

The low contrast in the solutions obtained from the manual molecular-replacement trials and the current relatively high value of *R*_{free} seem to be an indication that there are important differences in the three-dimensional structures of maize AR and other members of the AKR family, which could be related to its diverse properties as mentioned above. Refinement is in progress and the structural details will be reported in due course.

This work was supported in part by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via grant 01/07546-1 (to MM). SMS was supported by a postdoctoral fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). MLS received a PhD fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). ACL was funded by a studentship from Serviço de Apoio ao Estudante (SAE)/Unicamp. We are also grateful to the LNLS D03B-MX-1 beamline staff and Professors Anita J. Marsaioli and Inês Joeques (IQ/Unicamp) for providing part of the laboratory facilities.

References

- Altschul, S. F., Madden, T., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1990). *Nucleic Acids Res.* **25**, 3389–3402.
- Bartels, D., Engelhardt, K., Roncarati, R., Schneider, K., Rotter, M. & Salamini, F. (1991). *EMBO J.* **10**, 1037–1043.
- Biadene, M., Hazemann, I., Cousido, A., Ginell, S., Joachimiak, A., Sheldrick, G. M., Podjarny, A. & Schneider, T. R. (2007). *Acta Cryst. D* **63**, 665–672.
- Bohren, K. M., Bullock, B., Wermuth, B. & Gabbay, K. H. (1989). *J. Biol. Chem.* **264**, 9547–9551.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). *Nucleic Acids Res.* **31**, 3497–3500.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D* **50**, 760–763.
- Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Jez, J. M., Flynn, T. G. & Penning, T. M. (1997). *Biochem Pharmacol.* **54**, 639–647.
- Jez, J. M. & Penning, T. M. (2001). *Chem. Biol. Interact.* **130–132**, 499–525.
- Kantardjiev, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Keegan, R. M. & Winn, M. D. (2007). *Acta Cryst. D* **63**, 447–457.
- Lee, S. P. & Chen, T. H. (1993). *Plant Physiol.* **101**, 1089–1096.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Mundree, S. G., Whittaker, A., Thomson, J. A. & Farrant, J. M. (2000). *Planta*, **211**, 693–700.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst. D* **53**, 240–255.
- Navaza, J. (1994). *Acta Cryst. A* **50**, 157–163.
- Pace, C. N. & Schmid, F. X. (1997). *Protein Structure: A Practical Approach*, 2nd ed, edited by T. E. Creighton, pp. 253–259. New York: IRL Press.
- Pearson, W. R. & Lipman, D. J. (1988). *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
- Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R. C., Arruda, P., Leite, A. & Craievich, A. (1998). *Nucl. Instrum. Methods A*, **405**, 159–164.
- Roncarati, R., Salamini, F. & Bartels, D. (1995). *Plant J.* **7**, 809–822.
- Rondeau, J. M., Tête-Favier, F., Podjarny, A., Reymann, J. M., Barth, P., Biellmann, J. F. & Moras, D. (1992). *Nature (London)*, **355**, 469–472.
- Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst. D* **60**, 1229–1236.
- Sree, B. K., Rajendrakumar, C. S. V. & Reddy, A. R. (2000). *Plant Sci.* **160**, 149–157.
- Srivastava, S. K., Ramana, K. V. & Bhatnagar, A. (2005). *Endocr. Rev.* **26**, 380–392.
- Ueda, H., Kawanishi, K. & Moriyasu, M. (2004). *Biol. Pharm. Bull.* **27**, 1584–1587.
- Urzhumtsev, A., Tête-Favier, F., Mitschler, A., Barbanton, J., Barth, P., Urzhumtseva, L., Biellmann, J. F., Podjarny, A. & Moras, D. (1997). *Structure*, **5**, 601–612.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Wilson, D. K., Bohren, K. M., Gabbay, K. H. & Quijcho, F. A. (1992). *Science*, **257**, 81–84.
- Winn, M., Dodson, E. J. & Ralph, A. (1997). *Methods Enzymol.* **277**, 620–633.

8- Referências

Barteri M., Gaudio M. C. and Santucci R., (1996). Influence of glycerol on the structure and stability of ferric horse heart myoglobin: a SAXS and circular dichroism study. *Biochim. Biophys. Acta*, **129**, 551-58.

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

Chiu R., Angel P. and Karin M. (1989) Jun-B differs in its biological properties from, and is a negative regulator of c-jun. *Cell*, **59**, 979-986.

Fasman G. D., (1996) Circular dichroism and the conformational analysis of biomolecules. New York: Plenum Press.

Foster R., Izawa T. and Chua N. H., (1994) Plant bZIP proteins gather at ACGT elements. *FASEB J.*, **8**, 192-200.

Giuliano G., Pichersky E., Malik V. S., Timko M. P., Scolnik P. A. and Cashmore A. R., (1988) An evolutionarily conserved protein binding sequence of a plant light-regulated gene. *Proc. Natl. Acad. Sci. USA* , **85**, 7089-7093

Glatter O. and Kratky O., (1982). Small Angle Scattering. London: Academic Press.

Guinier, A. and Fournet G., (1955). Small-angle scattering of X-rays, *John Wiley & Sons*, New York).

Holstege F. C. P. and Yg R. A., (1999) Transcriptional regulation: contending with complexity. *Proc. Natl. Sci. U. S. A.*, **96**, 2-4.

Halliday D., Resnick R. and Walker J., (1995) Fundamentos de Física 4. 4ª Ed LTC Editora – Rio de Janeiro.

Hurst H. C., (1995) Transcription Factors: bZIP. *Protein Profile* vol. 1 (issue 1), Academic Press, London, UK.

Izawa T., Foster R. and Chua N. H., (1993) Plant bZIP protein DNA binding specificity. *J. Mol. Biol.*, **230**, 1131-1144

Jakoby M., Weisshaar B., Dröger-Laser W., Vicente-Carbajosa J., Tiedemann J., Kroj T. and Parcy F., (2002) BZIP transcription factors in *Arabidopsis*, *Trends Plant Sci.*, **7**, 106–111.

Jelsch C., Teeter M. M., Lamzin V., Pichon-Pesme V., Blessing R. H. and Lecomte C., (2000) Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc.Natl.Acad.Sci.USA*, **97**, 3171-3176.

Jez J. M. and Penning T. M., (2001) The aldo-keto reductase (AKR) superfamily: an update. *Chem. Biol. Interact.*, **130-132**, 499-525.

Keller W., König P. and Richmond T. J., (1995) Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *J. Mol. Biol.*, **254**, 657-667.

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

Kelly S. M., Jess T. J. and Price N. C., (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta*, **1751**, 119-139.

Kiyota E., de Sousa S. M. , dos Santos M. L. , da Costa Lima A. , Menossi M. , Yunes J. A. and Aparicio R., (2007) Crystallization and preliminary X-ray diffraction analysis of maize aldose reductase. *Acta Cryst.*, **F63**, 990-992.

Koch M. H. J., Vachette P. and Svergun D. I., (2003) "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution". *Q. rev. biophys.*, **36**, 147-227.

Konarev P. V., Volkov V. V., Sokolova A. V., Koch M. H. J. and Svergun D. I., (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.*, **36**, 1277-1282.

Konarev P. V., Petoukhov M. V., Volkov V. V. and Svergun D. I., (2006) ATSAS 2.1, a program package for small-angle scattering data analysis *J. Appl. Cryst.*, **39**, 277-286

König P. and Richmond T. J., (1993). The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.*, **233**, 139-154.

- Kornberg R. D., (1999) Eukaryotic transcriptional control. *Trends in Cell Biol.*, **9**, M46-M49.
- Kouzarides T. and Ziff E., (1989). Leucine zipper of *fos*, *jun* and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature*, **340**, 568-571.
- Kozin M. B. and Svergun D. I., (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.*, **34**, 33-41.
- Laemmli U. K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature*, **227**, 680-685.
- Lee S., Reth A., Maletzus D., Sevilla M. and Kennedy C., (2000) Characterization of a major SAS of *nif*, *fix* and associated genes in a sugarcane endophyte, *Guconacetobacter diazotrophicus*. *J. Bacteriol.*, **182**, 7088-91.
- Lehninger L. A., Nelson D. L. and Cox M. M., (2000) Lehninger Principles of Biochemistry. 3^a ed. New York: Worth Publishers.
- Lehninger L. A., Nelson D.L. and Cox M. M., (2005). Principles of Biochemistry. Cap. 3, 75-115. New York, Freeman and Company.
- Lobley A. and Wallace B.A. (2001) DICHROWEB: A Website for the analysis of protein secondary structure from circular dichroism spectra. *Biophysical J.*, **80**, 373a.
- Lupas A., (1996) Coiled-coils: new structure and new functions. *Trends Biochem. Sci.*, **21**, 375-82.
- MacPherson A., (1999) Crystallization of Biological Molecules. Cold Spring Harbor Laboratory Academic Press.
- Marris E., (2006) Drink the best and drive the rest. *Nature - Business Features*, **444**, 670-671
- Murphy R. M., (1997) Static and dynamic light scattering of biological molecules: what can we learn? *Cur. Opin. biotech.*, **8**, 25-30.
- Nair S. K. and Burley S. K., (2003) X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA: Molecular Bases of Regulation by Proto-Oncogenic Transcription Factors. *Cell*, **112**, 193-205

Oliphant A. R., Brandl C. J. and Struhl K., (1989) Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mot. Cell. Biol.*, **9**, 2944-2949.

Oliveira C. L. P., (2003). TRAT1D - Computer Program for SAXS Data Treatment, *LNLS Technical Manual* MT012003.

O' Shea E. K., Klemm J. D., Kim P. S. and Albert T., (1991). X-ray Structure of the GCN4 leucine-zipper, a two-stranded, parallel coiled-coil. *Science*, **254**, 539-544.

Papish A. L., Tari L. W. and Vogel H. J., (2002) Dynamic light scattering studies of Colmoduli-target peptide complexes. *Biophys. J.*, **83**, 1455-1464.

Pecora R., (1985) Dynamic light scattering – Application of photon correlation spectroscopy. New York: Plenum Press.

Pessoa A., Roberto I. C., Menossi M., Dos Santos R. R., Ortega S. and Penna T. C. V., (2005) Perspectives on bioenergy and biotechnology in Brazil. *Appl. Biochem. Biotechnol.*, **121**, 59-70.

Provencher S. W., (1982) CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comp. Phys. Commun.*, **27**, 229-242.

Provencher S. W. and Glöckner J., (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*, **20**, 33-37.

Riechmann J. L. and Ratcliffe O. J., (2002) A genome perspective on plant transcription factors. *Cur. Opin. Plant Biol.*, **3**, 423-434.

Sambrook, J., Fritsch, E. F. and Maniats, T., (1989) Molecular Cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Schlögl P. S., Kobarg J., Moreau V. H., Leite A., Sabino A. A., Eberlin M. N. and Menossi M., (2004) Expression, purification and characterization of a novel bZIP protein from sugarcane, *Plant Sci.*, **167**, 583-595.

Smeal T., Angel P. and Karin M., (1989) Different requirements for formation of Jun: Jun and Jun: Fos complexes. *Genes Dev.*, **12B**, 2091-2100.

Stryer L., (1995) Biochemistry. 4th Edition. *W. H. Freeman and Company. New York.*

Svergun D. I., (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.*, **25**, 495-503.

Svergun D., Barberato C. and Koch M. H. J., (1995). CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.*, **28**, 768–773.

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

Svergun D. I., Petoukhov M. V. and Koch M. H. J., (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946-2953.

Svergun D. I. and Koch M. H. J., (2002) Advances in structure analysis using small-angle scattering in solution. *Current Opinion in Structural Biology*, **12**, 654–660.

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

Svergun D. I., (2007) Small-angle scattering studies of macromolecular solutions. *J. Appl. Cryst.*, **40**, 10-17.

Vachette P. and Svergun D. I., (2000) Small-angle X-ray scattering by solutions of biological macromolecules. *Structure and Dynamics of Biomolecules*, 199-237, New York: Oxford University Press.

Van Stokkum I. H. M., Spoelder H. J. W., Bloemendal M., Van Grondelle R. and Groen F. C. A., (1990) Estimation of protein secondary structure and error analysis from CD spectra. *Anal. Biochem.* **191**, 110-118

Vettore A. L., Yunes J. A., Cord Neto G., da Silva M. J., Arruda P. and Leite A., (1998) The molecular and functional characterization of an *Opaque-2* homologue protein from *Coix* and new classification of plant bZIP proteins. *Plant Mol. Biol.*, **36**, 249–263.

Vincenz M., P. S., Corrêa L.G., F Kuhne. and Leite A., (2001) Phylogenetic relationships between *Arabidopsis* and sugarcane bZIP transcriptional regulatory factors. *Genet. Mol. Biol.*, **24**, 55-60.

Voet D., Voet J. G. e Pratt C. W., (2000). Fundmentos de Bioquímica. Artes Médicas Sul. Porto Alegre - Brasil.

Volkov V. V. and Svergun D. I., (2003). Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.*, **36**, 860-864.

Whitmore L. and Wallace B. A., (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucl. Acids Res.*, **32**, 668-673.