UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



Thiago Carlos Cagliari

"Análise da expressão de chaperonas moleculares em plantas e

clonagem, purificação e caracterização inicial das proteínas

Hsp100 e Hsp90 de cana-de-açúcar."

Este exemplar corresponde à redação finada tese defendida pelo(a) candidato (a) †th'AGo CARLOS CAGUAN

e aprovada pela Comissão Julgadora.

Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Biologia Funcional e Molecular, na área de Bioquímica.

Orientador: Prof. Dr. Carlos Henrique Inácio Ramos

Campinas, 2009

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

C117e	Cagliari, Thiago Carlos Análise da expressão de chaperonas moleculares em plantas e clonagem, purificação e caracterização inicial das proteínas Hsp100 e Hsp90 de cana-de-açúcar / Thiago Carlos Cagliari. – Campinas, SP: [s.n.], 2009.
	Orientador: Carlos Henrique Inácio Ramos. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.
	 Chaperones moleculares. Proteínas de choque- térmico HSP100. Proteínas de choque-térmico HSP90. Ramos, Carlos Henrique Inácio. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.
	(rcdt/ib)

Titulo em inglês: Expression analysis of plant molecular chaperones and cloning, purification and primary characterization of the proteins Hsp100 and Hsp90 from sugarcane. **Palavras-chave em inglês**: Molecular chaperones; HSP100 heat shock proteins; HSP90 heat shock proteins.

Área de concentração: Bioquímica.

Titulação: Doutor em Biologia Funcional e Molecular.

Banca examinadora: Carlos Henrique Inácio Ramos, Gonçalo Amarante Guimarães Pereira, Leonardo Fernandes Franceto, Fernanda Canduri, Flávio Henrique da Silva.

Data da defesa: 08/05/2009.

Programa de Pós-Graduação: Biologia Funcional e Molecular.

Campinas, 08 de Maio de 2009

Profa. Dra. Fernanda Canduri

Prof. Dr. Leonardo Fernandes Franceto

Prof. Dr. Gonçalo Amarante Guimarães Pereira

BANCA EXAMINADORA

Prof. Dr. Carlos Henrique Inácio Ramos (Orientador)

lo Mano Assinatura

uauda Assinatura

nard Emands Assinatura

Assinatura

Assinatura

Assinatura

Assinatura

Assinatura

Prof. Dr. Flávio Henrique da Silva

Profa. Dra. Leila Maria Beltramini

Prof. Dr. Cláudio Chrysostomo Werneck

Profa. Dra. Eneida de Paula

Agradecimentos

Agradeço ao Laboratório Nacional de Luz Síncrotron (LNLS) e Instituto de Química da UNICAMP pelas instalações. Agradeço ao Conselho Nacional para o Desenvolvimento da Pesquisa (CNPq), ao Fogarty International Research Collaboration Award (FIRCA do NIH) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento ao laboratório do Prof. Dr. Carlos H.I. Ramos. Agradeço à FAPESP pelo apoio financeiro na forma de bolsa de doutorado direto que permitiu a realização desta tese de doutorado. Sou grato aos professores do curso de Biologia Funcional e Molecular, em especial a Profa. Dra. Helena Coutinho e a secretária Andréia A. Vigilato, por toda compreensão e ajuda nas questões burocráticas.

Devo um grande agradecimento ao Prof. Dr. Carlos Ramos por ter dado a oportunidade de realizar este projeto, e também por toda a orientação e ajuda durante estes anos de trabalho.

Agradeço ao Prof. Dr. Júlio César Borges, por sua constante colaboração e incentivo durante este projeto.

Obrigado aos técnicos e amigos do LNLS e IQ pela ajuda e conselhos úteis a este projeto de pesquisa, e também por tornarem mais leves os momentos de dificuldades e frustrações.

Finalmente gostaria de agradecer a toda minha família pelo apoio e incentivo ao longo dos anos, desde a graduação e iniciação científica até este momento. Agradeço à minha esposa Erika pela compreensão e paciência durante as noites, madrugadas e finais de semana que dediquei a este trabalho. Agradeço aos meus pais pelo grande incentivo e em especial à minha mãe, pessoa que resume as palavras amor, abnegação, perseverança e cuidado. Aos meus irmãos Pedro e Izabella, por sempre terem acreditado em mim mesmo nos momentos mais difíceis.

Muito obrigado a todos, Thiago Carlos Cagliari

Índice

Agradecimentos	4
Índice	5
Lista de abreviações	8
Resumo	12
Abstract	13
1 Introdução	14
1.1 Enovelamento protéico e as chaperonas moleculares	14
1.2 Hsp90	17
1.2.1 Dados gerais	17
1.2.2 Estrutura e atividade	18
1.3 Hsp100	23
1.3.1 Dados gerais	23
1.3.2 Desagregação mediada por chaperonas	27
1.3.3 Modelos de desagregação protéica dependente do sistema Hsp100/Hsp70	29
1.4 Co-chaperonas	33
1.5 Co-chaperona p23	34
1.5.1 Dados gerais	34
1.5.2 Estrutura da p23 humana	35
2 Objetivos	38
3 Materiais e métodos	39
3.1 Data mining e bioinformática	39
3.2 Clone	39
3.3 Oligonucleotídeos para PCR	39
3.4 PCR	40
3.5 Gel de agarose e determinação da concentração de DNA	40
3.6 Vetores	41
3.6.1 Vetor de clonagem/transferência	41
3.6.2 Vetor de expressão	41
3.7 Linhagens de Escherichia coli	43
3.8 Meios de cultura	43
3.9 Clonagem	44
3.9.1 Clonagem em pGEM-T	44
3.9.2 Subclonagem em pET28a	45
3.10 Precipitação de produto de ligação	45
3.11 Obtenção de bactérias eletro-competentes	45
3.12 Obtenção de bactérias competentes para transformação por choque térmico.	46
3.13 Transformação de bactérias competentes	46
3.14 Extração de DNA plasmidial	47
3.15 Seleção dos clones transformados	47
3.16 Seqüenciamento de DNA	47

3.17 Expressão de proteínas	47
3.18 Lise de bactérias	48
3.19 Tampões	48
3.20 Purificação das proteínas recombinantes	49
3.20.1 Cromatografia de afinidade	49
3.20.2 Cromatografia de exclusão molecular	49
3.20.3 Cromatografia de troca iônica	49
3.21 Eletroforese em gel de poliacrilamida (SDS-PAGE)	50
3.22 Determinação da concentração de proteínas	50
3.23 Espectros de dicroísmo circular (CD)	51
3.24 Emissão de fluorescência	52
3.25 Determinação da concentração de nucleotídeos	52
3.26 Espalhamento dinâmico de luz	53
3.27 Ensaio de ligação baseado em resina de purificação (tipo "pull-down")	54
3.28 Modelagem molecular por homologia	54
3.29 Ultracentrifugação analítica	55
3.29.1 Velocidade de sedimentação	55
3.29.2 Sedimentação em equilíbrio	56
4 Hsp100 de cana-de-açúcar	57
4.1 Resultados Hsp100 de cana-de-açúcar	57
4.1.1 Clonagem, expressão e purificação	57
4.1.2 Caracterização por dicroísmo circular da proteína Hsp100	64
4.1.3 Caracterização por fluorescência da proteína Hsp100	68
4.1.4 Estabilidade térmica da Hsp100	70
4.1.5 Espalhamento dinâmico de luz da proteína Hsp100	72
4.1.6 Ultracentrifugação analítica: sedimentação em equilíbrio	73
4.1.7 Ultracentrifugação analítica: velocidade de sedimentação	75
4.1.8 Modelagem molecular por homologia	77
4.1.9 Interação com agregados protéicos	81
4.2 Discussão Hsp100 de cana-de-açúcar	83
4.2.1 Expressão e purificação	83
4.2.2 Dicroísmo circular	83
4.2.3 Fluorescência	84
4.2.4 Estabilidade medida por CD	84
4.2.5 Espalhamento dinâmico de luz	85
4.2.6 Ultracentrifugação analítica	85
4.2.7 Modelagem molecular por homologia	87
4.2.8 Interação com agregados	87
4.3 Conclusões Hsp100 de cana-de-açúcar	89
5 Hsp82 de cana-de-açúcar	91
5.1 Resultados Hsp82 de cana-de-açúcar	91
5.1.1 Clonagem do cDNA da Hsp82	91

5.1.2 Expressão e purificação da Hsp82 de cana-de-açúcar	92
5.1.3 Caracterização por CD da proteína Hsp82	98
5.1.4 Caracterização por fluorescência da proteína Hsp82	103
5.1.5 Análise por espalhamento dinâmico de luz	108
5.1.6 Ultracentrifugação analítica – velocidade de sedimentação	109
5.1.7 Modelagem molecular por homologia	111
5.2 Discussão Hsp82 de cana-de-açúcar	114
5.2.1 Expressão e purificação	114
5.2.2 Dicroísmo circular	115
5.2.3 Fluorescência	115
5.2.4 Estudos de estabilidade empregando uréia	115
5.2.4 Ultracentrifugação analítica e espalhamento dinâmico de luz	116
5.2.5 Modelagem molecular por homologia	118
5.3 Conclusões Hsp82 de cana-de-açúcar	120
6 Co-chaperona humana p23	121
6.1 Resultados p23 humana	121
6.1.1 Clonagem e expressão da co-chaperona p23	121
6.1.2 Purificação da co-chaperona p23 humana	124
6.1.3 Caracterização da proteína p23 humana por dicroísmo circular	127
6.1.4 Caracterização da proteína p23 humana por fluorescência	136
6.1.5 Desenovelamento químico da proteína p23 monitorado por CD e fluorescência	137
6.1.6 Análise hidrodinâmica da proteína p23	145
6.2 Discussão p23 humana	151
6.2.1 Dicroísmo circular	151
6.2.2 Fluorescência	151
6.2.3 Estabilidade medida por CD e fluorescência	152
6.2.4 Ultracentrifugação analítica e espalhamento dinâmico de luz	153
6.3 Conclusões p23 humana	154
7 Referências	155
8 Anexo: artigos	170
8.1 Artigo: Genetics and Molecular Biology, 28, 3 (suppl), 520-528 (2005)	170
8.2 Artigo: Journal of Plant Physiology 164: 505-513 (2007)	196
9 Curriculum Vitae	215

Lista de abreviações

- Σ: somatório;
- ε: coeficiente de absortividade molar;
- η: viscosidade;
- η_{tex} : viscosidade do solvente na temperatura experimental;
- η_{t20} : viscosidade do solvente a 20 °C;
- η_w: viscosidade da água;
- λ : comprimento de onda;
- Θ: elipticidade;
- ρ: densidade;
- ρ_f: densidade do fluído;
- $\rho_{20,W}$: densidade do solvente em água e na temperatura de 20 °C;
- ρ_{ex}: densidade do solvente nas condições experimentais;
- ω : velocidade angular;
- Å: angström (1x10⁻¹⁰ metros);
- AMP: adenosina mono-fosfato;
- ADP: adenosina difosfato;
- ATP: adenosina trifosfato;
- BLAST: ferramenta de procura (Basic Local Alignment Search Tool);
- ^oC: unidade de temperatura graus Celsius;
- C: concentração;
- CEM: cromatografia de exclusão molecular;
- CD: dicroísmo circular;
- cDNA: DNA complementar;
- D: coeficiente de difusão;
- D_{20,w}: coeficiente de difusão em água e na temperatura de 20 °C;
- $D^{0}_{20,w}$: coeficiente de difusão em água, na temperatura de 20 °C e concentração zero de

proteínas;

Deg: grau (degree);

EDL: espalhamento dinâmico de luz;

EDTA: ácido etilenodiaminotetracético;

EMR: elipticidade molar residual;

EST: etiqueta de seqüência expressa (Expressed Sequence Tag);

Exp.: Dados experimentais;

f: coeficiente friccional;

f/f0: razão friccional ou fator de Perrin;

F_c: Força centrípeta;

F_{Emp}: Força de empuxo;

F_f: Força friccional;

P_i: fosfato inorgânico;

FPLC: rápida cromatografia líquida de proteínas (Fast Protein Liquide

Chromatography);

g: aceleração da gravidade;

Gdn-Cl: cloridrato de guanidina;

Hop: proteína organizadora das Hsp70-Hsp90 (Hsp70-Hsp90 organizing protein);

Hsp: proteína de choque térmico (Heat shock protein);

Hsp82: proteína de choque térmico de 82 kDa;

Hsp100: proteína de choque térmico de 100 kDa;

Hsp90: proteína de choque térmico de 90 kDa;

Hsp70: proteína de choque térmico de 70 kDa;

Hsp40: proteína de choque térmico de 40 kDa;

 IF_{λ} : intensidade de fluorescência em um determinado comprimento de onda;

IMAGE Consortium: Integrated Molecular Analysis of Genomes and their Expression

Consortium;

IPTG: Isopropiltiol-β-D-galactosídeo;

 k_d : constante de Boltzmann (1,381 x 10⁻²³ J K⁻¹ (SI) ou 3,298 x 10⁻²⁴ cal K⁻¹); K: unidade de temperatura – Kelvin; kDa: kiloDalton; LNLS: Laboratório Nacional de Luz Síncrotron; l: comprimento do caminho ótico; LB: Luria-Bertani; m: massa; MM: massa molecular; n: estequiometria da ligação; NCBI: National Center of Biotechnology Information; ND: não determinado; nm: nanômetro (10⁻⁹ metros); pb: par de bases; P: Fator de Perrin ou razão friccional; PCR: reação em cadeia da DNA polimerase (DNA Polimerase Chain Reaction); PDB: banco de dados de estrutura em alta resolução de proteínas (Protein Data Bank); r: raio; r_{esf}: raio da esfera; R: constate dos gases (8,314 J mol⁻¹ K⁻¹ (SI) ou 1,987 cal mol⁻¹ K⁻¹); R_s: raio de Stokes; s: coeficiente de sedimentação; s*: coeficiente de sedimentação aparente; s_{20,w}: coeficiente de sedimentação em água e na temperatura de 20 °C; s⁰_{20,w}: coeficiente de sedimentação em água, na temperatura de 20 °C e corrigido para eliminar o efeito de não-idealidade da concentração de proteínas ; S: Svedberg (10⁻¹³ segundos); SDS-PAGE: eletroforese em gel de poliacrilamida em presença de duodecil-sulfato de sódio

(sodium dodecyl-sulphate polyacrylamide gel electrophoresis);

SMC: sítio múltiplo de clonagem;

- T: temperatura absoluta;
- T_m: temperatura média;
- UA: unidade arbitrária;
- UCA: ultracentrifugação analítica;
- v: velocidade;
- V_{bar}: volume parcial específico;
- [O]: elipticidade molar residual média
- <λ>: centro de massa espectral de fluorescência.

Resumo

As proteínas são macromoléculas que possuem importância vital para o funcionamento celular, participando da maioria das reações biológicas e também como componentes estruturais. Para que uma proteína possa exercer sua função, precisa atingir sua estrutura nativa através de um processo denominado enovelamento protéico. Neste contexto, as chaperonas moleculares são proteínas capazes de auxiliar no enovelamento de outras proteínas, atuando na prevenção de agregados, desagregação, translocação, ativação, entre outros. Dentre os muitos tipos de chaperonas existentes, neste trabalho foram abordadas as chaperonas das famílias Hsp100 e Hsp90, as quais estão relacionadas aos processos de desagregação e auxílio do enovelamento de proteínas-substrato, respectivamente. O presente trabalho pretendeu produzir as proteínas recombinantes Hsp100 e Hsp82 de cana-de-açúcar para a caracterização de suas respectivas relações estrutura-função. Para isto foram empregadas técnicas como: dicroísmo circular, fluorescência, espalhamento dinâmico de luz e ultracentrifugação analítica. Assim, foi observado que a força iônica do meio é capaz de influenciar a estrutura quaternária da proteína Hsp100, a qual se apresenta hexamérica em menores concentrações de sal. Além disto, é capaz de reconhecer agregados protéicos formados pelas proteínas luciferase e citrato sintase em ensaios in vitro. Já a proteína Hsp82 apresentou uma estrutura dimérica, a qual não é influenciada pela presença de nucleotídeos e apresenta grande estabilidade térmica. Finalmente, a proteína p23 humana, a qual é responsável por auxiliar a proteína Hsp90 no enovelamento de muitas proteínas/complexos protéicos, também foi caracterizada. Foram observados indícios de que a região C-terminal, rica em resíduos de aminoácidos carregados, pode possuir algum grau de estruturação, apesar de alguns estudos na literatura indicarem o contrário. O estudo das chaperonas de cana-de-açúcar foi direcionado por um trabalho prévio de anotação de seqüências relacionadas às chaperonas moleculares no banco de dados do projeto SUCEST (Sugarcane EST Genome Project), o qual foi realizado por nosso grupo de pesquisa. Além disto, são apresentados os resultados da anotação das seqüências relacionadas às chaperonas de eucalipto no banco de dados FORESTs (Eucalyptus Genome Sequencing Project Consortium), possibilitando futuros estudos com estas proteínas.

Abstract

Proteins are macromolecules that are vital to the functioning cell, participating in most of the biological reactions as well as structural components. To perform its function, a protein need to achieve its native structure through a process called protein folding. In this context, the molecular chaperone proteins are able to assist in the folding of other proteins, acting in the prevention of aggregation, disaggregation, translocation, activation, among others. From all types of existing chaperones, here were highlight the Hsp100 and Hsp90 families, which are related to processes of disaggregation and assistance of substrateprotein folding, respectively. This study sought to produce the recombinant proteins Hsp100 and Hsp82 from sugar cane for the characterization of their structure-function relationships. In order to do this, some techniques were employed such as: circular dichroism, fluorescence, dynamic light scattering and analytical ultracentrifugation. As a result, it was observed that the ionic strength of the solvent is capable of influencing the quaternary structure of protein Hsp100, which presents as a hexamer in lower salt concentrations. Furthermore, it is capable of recognizing protein aggregates formed by luciferase protein and citrate synthase in in vitro essays. The Hsp82 protein showed a dimeric structure, which was not influenced by the presence of nucleotides and presented a great thermal stability. Finally, the human protein p23, which is responsible for assisting in the Hsp90 protein folding of many proteins/protein complexes, was also characterized. In spite of some studies indicating the contrary, we observed evidence that the C-terminal region, which is rich in charged amino acid residues, can possible have some structure. The sugarcane chaperones study was guided by a previous chaperone sequence annotation work in the SUCEST (Sugarcane EST Genome Project) databank performed by our research group. In addition, results regarding chaperone sequences annotation in the eucalyptus databank (FORESTs - Eucalyptus Genome Sequencing Project Consortium) were presented here as well, which can also lead to future chaperone proteins function and structure studies.

1 Introdução

1.1 Enovelamento protéico e as chaperonas moleculares

Em geral, as proteínas só podem desempenhar suas respectivas atividades biológicas após atingirem suas estruturas nativas, as quais são obtidas através de um processo conhecido como enovelamento protéico. Este processo é basicamente determinado pela própria següência de aminoácidos da proteína, ou seja, toda a informação necessária para o enovelamento e formação da estrutura nativa está contida em sua estrutura primária (Anfinsen, 1973). O processo de enovelamento envolve, principalmente, um evento denominado colapso hidrofóbico. Este se baseia na observação de que o estado nativo das proteínas globulares apresenta um núcleo hidrofóbico formado pelas cadeias laterais de aminoácidos apolares, enquanto que a maioria das porções polares das cadeias laterais de resíduos polares ou carregados está exposta ao solvente na superfície da proteína. Assim, a estabilização energética conferida à proteína através do seqüestro das cadeias laterais hidrofóbicas do ambiente aquoso circundante geraria a estabilização dos intermediários da via de enovelamento. Logo, o colapso hidrofóbico é um evento relativamente precoce na via de enovelamento, ocorrendo antes da formação de muitas estruturas secundárias e contatos presentes apenas na estrutura terciária totalmente enovelada. Em última análise, o colapso hidrofóbico estabiliza a estrutura nativa da proteína compensando termodinamicamente a entropia desfavorável provocada pela formação desta estrutura organizada. Os aspectos abordados acima podem ser encontrados nas seguintes revisões sobre enovelamento protéico: Baldwin (1989), Kuwajima (1989), Kim e Baldwin (1990), Ptitsyn (1991), Fersht (1994), Privalov (1996), Jaenicke (1999), Price (2000), Ramos e Ferreira (2005).

O processo de enovelamento é influenciado por várias forças, como a energia livre (potencial termodinâmico que mede o trabalho obtido em um sistema isotérmico e isobárico), a concentração protéica do meio, a formação de ligações dissulfeto e alterações pós-traducionais, tais como glicosilação, adenilação, etc. (Ruddon *et al.*, 1997). Em um meio com alta saturação protéica, podem ocorrer interações das regiões hidrofóbicas de proteínas não-enoveladas ou parcialmente enoveladas, formando agregados protéicos que precipitam nas células e tecidos (Jaenicke *et al.*, 1987; Beissinger e Buchner,

1998; Fink, 1999; Ramos e Ferreira, 2005). As falhas no processo de enovelamento, com formação de agregados protéicos, podem estar relacionadas com algumas doenças humanas degenerativas e amilóides como: Alzheimer (Agorogiannis *et al.*, 2004), câncer (Thor *et al.*, 1991; Tetu *et al.*, 1992) e outras (Beissinger e Buchner, 1998; Chai *et al.*, 1999; Ruddon *et al.*, 1996; Thomas *et al.*, 1995).

Neste contexto, um importante mecanismo de auxílio ao processo de enovelamento *in vivo* de muitas proteínas é baseado na participação de proteínas conhecidas como chaperonas moleculares ou proteínas de choque térmico (*heat shock proteins* ou *Hsp*) (Borges e Ramos, 2005; Ramos, 2007). As chaperonas moleculares são proteínas capazes de assistir ao enovelamento de outras proteínas, deste modo prevenindo interações errôneas e conseqüente formação de agregados protéicos dentro das células (figura 1). No entanto, ao realizarem tal tarefa, não interferem na estrutura tridimensional final ou nativa das proteínas auxiliadas, respeitando o conceito formulado por Anfinsen (Anfinsen, 1973), o qual afirma ser a seqüência primária que determina a estrutura nativa. Segundo Hendrick e Hartl (1995), uma chaperona molecular pode ser definida como: *"Uma proteína que se liga e estabiliza uma conformação de outra forma instável de outra proteína, facilitando seu destino correto in vivo: seja no enovelamento, formação oligomérica, transporte para um compartimento sub-celular específico ou alternância controlada entre conformações ativas e inativas"* (tradução).

As chaperonas moleculares foram descritas inicialmente como proteínas de choque-térmico ou "heat-shock proteins" (Hsp), porque muitas apresentam sua síntese induzida em células submetidas a condições de estresse por calor. As chaperonas apresentam, de modo geral, afinidade por seqüências inespecíficas de resíduos de aminoácidos hidrofóbicos na superfície dos polipeptídios não-enovelados. Além de auxiliar o enovelamento protéico, as chaperonas também participam de outras atividades, tais como: 1) transporte, através de membranas, de polipeptídios recém sintetizados para as organelas; 2) prevenção da agregação protéica e interações não desejadas/produtivas durante o enovelamento; 3) resolubilização/desagregação de agregados protéicos, bem como o encaminhamento dos mesmos para vias de degradação (Borges e Ramos, 2005). Também estão relacionadas ao controle intracelular hormonal corticóide e a outros fatores de transcrição, assim como na montagem de complexos protéicos a partir de suas subunidades (Fink, 1999; Hendrick *et al.*, 1993; Ruddon *et al.*, 1997).



Figura 1: Esquema generalizado de funcionamento das chaperonas moleculares. A maioria das chaperonas utiliza a energia de hidrólise do ATP no enovelamento de proteínas substrato desenoveladas ou parcialmente enoveladas, porém sem interferir na conformação final da mesma (modificado de "Principles of Biochemestry", editora W. H. Freeman, terceira edição).

1.2 Hsp90

1.2.1 Dados gerais

A família Hsp90 está presente em todos os filos estudados, com exceção de *archaea*, no qual parece estar ausente. Nos eucariotos, a presença de pelo menos uma isoforma citoplasmática da Hsp90 é essencial para a sobrevivência sob qualquer condição ambiental (Borkovich *et al.*, 1989), enquanto que a homóloga em bactéria, HtpG, é dispensável sob condições normais de crescimento (Spence *et al.*, 1990). A Hsp90 citoplasmática é uma das proteínas mais abundantes, representando aproximadamente 1% do total das proteínas mesmo na ausência de choque-térmico (Lai *et al.*, 1984), o qual pouco afeta sua expressão. A concentração intracelular da Hsp90 está localizada predominantemente no citoplasma e em menor quantidade no núcleo (Schlatter *et al* 2002). Através de experimentos de fusão da Hsp90 com a proteína GFP (Green Fluorescent Protein) observou-se um aumento da coloração periplasmática e intranuclear após um choque-térmico, o que indica que a proteína Hsp90, fusionada ao GFP, migra para estas regiões celulares (Langer et el., 2002). Além disso, já foram encontradas evidências de que a Hsp90 estaria também na superfície celular (Ferrarini *et al.*, 1992).

Em seres humanos estão presentes duas isoformas citoplasmáticas da Hsp90, a alfa e a beta; em levedura também ocorrem duas isoformas da Hsp90, a Hsp82 e a Hsc82; já *Drosophila melanogaster* apresenta apenas um gene, Hsp83. A expressão das isoformas humanas é diferencialmente regulada, sendo que a isoforma alfa é regulada por calor, já a expressão da isoforma beta é constitutiva e pode ser induzida por fatores de crescimento. A família das Hsp90 é formada por proteínas que possuem grande identidade na seqüência de resíduos de aminoácidos, revelando indício de alto grau de conservação evolutiva. Por exemplo, a identidade entre a Hsp90 humana e a Hsp90 de levedura é de aproximadamente 60%, já entre a humana e a de bactéria (*E. coli* – HtpG) é de 40%.

A Hsp90 também atua como um capacitor para a evolução em *Drosophila melanogaster* (Rutherford e Lindquist, 1998) e *Arabdopsis thaliana* (Queitsch *et al.*, 2002) regulando manifestações fenotípicas causadas por mutações crípticas, ou seja, mutações que não se manifestam espontaneamente no fenótipo destes organismos. Acredita-se que a Hsp90 de cloroplasto pode ser derivada da forma encontrada no retículo endoplasmático e não de cianobactérias, enquanto que a Hsp90 mitocondrial seria derivada de eubactérias e não de alfa-proteobactérias como a *E.coli* (Emelyanov, 2002).

1.2.2 Estrutura e atividade

A proteína Hsp90 foi identificada, inicialmente, como uma das inúmeras proteínas de choque térmico conservadas. Assim como outras grandes classes de chaperonas moleculares, a Hsp90 também apresenta uma atividade de proteção, prevenindo a agregação inespecífica de proteínas parcial ou totalmente desenoveladas (Wiech et al., 1992). No entanto, a Hsp90 parece ser menos promíscua em relação às outras chaperonas, pois interage com um subconjunto específico do proteoma (Picard, 2002). Outra característica específica da Hsp90 é seu papel regulador na indução de alterações conformacionais em proteínas substrato enoveladas/nativas que levam à ativação ou estabilização das mesmas (Jakob et al., 1995). Recentemente, as estruturas tridimensionais da Hsp90 completa de Escherichia coli, levedura e do retículo endoplasmático (Grp94) de cão foram obtidas através de cristalografia de raios-X (Shiau et al., 2006; Ali et al., 2006; Dollins et al., 2007; Pearl e Prodromou, 2006). Juntamente com os dados de seqüência, os dados estruturais revelaram que enquanto a Hsp90 manteve sua estrutura geral de domínios das bactérias ao homem, mudanças específicas parecem ter ocorrido no sentido de adaptar a Hsp90 ao ambiente protéico mais complexo da célula eucariótica. Desta forma, a Hsp90 eucariótica apresenta uma "seqüência carregada" que conecta os domínios N-terminal e intermediário (Middledomain ou M-domain), possui também uma extensão do domínio C-terminal, a qual inclui a motivo conservado "MEEVD". Esta região representa o principal sítio de interação para uma série de cochaperonas (Richter et al., 2005), as quais ajudam a Hsp90 no enovelamento e ativação de suas proteínas-substrato em eucariotos.

A Hsp90 forma um dímero flexível (Radanyi *et al.*, 1989), e cada monômero consiste de três domínios: o Domínio-N conectado por uma longa seqüência carregada (eucariotos) ao Domínio-M, o qual é seguido pelo domínio de dimerização C-terminal. O Domínio-N possui um sítio de ligação ao ATP posicionado profundamente em seu interior (Prodromou *et al.*, 1997), onde o ATP se liga de uma forma

torcida e incomum. A hidrólise de ATP pela Hsp90 é um tanto lenta: a Hsp90 de levedura hidrolisa uma molécula de ATP a cada um ou dois minutos (Panaretou et al., 1998; Scheibel et al., 1998), enquanto que a Hsp90 humana um ATP a cada vinte minutos (0,04 min⁻¹) (McLaughlin et al., 2002). No caso da homóloga em levedura, foi comprovado que esta atividade ATPásica é essencial para sua função (Panaretou et al., 1998; Obermann et al., 1998). A lenta hidrólise do ATP sugere que rearranjos conformacionais complexos da Hsp90 estão acoplados a esta reação e que estas alterações representam o passo limitante da reação. Os primeiros passos destas mudanças conformacionais foram elucidados recentemente em detalhe: após a ligação do ATP, um curto segmento do Domínio-N chamado de "tampa do ATP" (ATP-lid) muda sua posição e se move cobrindo o sítio de ligação (Richter et al., 2006). Isto libera um curto segmento N-terminal de sua posição original (Richter et al., 2002). Em uma reação subseqüente, este segmento se liga ao respectivo Domínio-N da outra subunidade do dímero, provocando a dimerização transiente entre os dois domínios (Ali et al., 2006; Richter et al., 2006). Estes rearranjos N-terminais resultam em mudanças conformacionais adicionais através de todo o dímero da Hsp90, levando à formação de um dímero torcido e compacto, no qual os domínios N e M se associam e a distância entre os domínios M é encurtada em 40 Å (Ali *et al.*, 2006). A associação dos domínios N e M completa, em cada uma das subunidades, o sítio ativo desta "ATPase dividida" (figura 2). Recentemente, uma progressão similar de passos foi demonstrada para a homóloga endoplasmática Grp94 (Richter et al., 2002), TRAP1 mitocondrial (Dollins et al., 2007; Leskovar et al., 2008) e Hsp90 humana (Richter et al., 2008). Portanto, o cenário delineado acima parece ser um mecanismo ATPásico ubiguamente conservado para a Hsp90. Interessantemente, o modo incomum pelo qual o ATP se liga à Hsp90 é perfeitamente mimetizado por alguns compostos naturais como a geldanamicina e o radicicol. Estes são inibidores altamente específicos e potentes da atividade ATPásica da Hsp90 (Roe et al., 2009), bloqueando a maturação de proteínas-substrato e eventualmente resultando na degradação das mesmas (Whitesell et al., 1994). Como muitas proteínas-substrato da Hsp90 são quinases que podem ser desreguladas no desenvolvimento do câncer, derivados de inibidores da Hsp90 estão sendo atualmente investigados como terapêuticas anti-câncer em estágio de testes clínicos (Sharp e Workman, 2006).

Modelos atuais assumem que as alterações conformacionais associadas com a hidrólise de ATP são necessárias para atingir ou manter um estado ativado das proteínas-substrato. Os receptores de

hormônio esteróide (RHEs) são um exemplo bem conhecido de como muitos co-fatores interagem com a Hsp90 de uma maneira seqüencial para montar uma maquinaria chaperona funcional (Smith, 1993; Smith et al., 1993). Esta sucessão ordenada de diferentes associações entre as proteínas do sistema Hsp90 pode agora ser analisada, já que muitos co-fatores da Hsp90 apresentam uma forte preferência pela ligação a conformações específicas da Hsp90. A ligação de um RHE à Hsp90 requer a cooperação desta com a chaperona Hsp70 e seu co-fator Hsp40 (Smith et al., 1993). Além disto, ambas chaperonas se tornam fisicamente ligadas através de uma proteína adaptadora chamada Hop/Sti1. Esta cochaperona se liga através de domínios formados por pequenas repetições de tetratricopeptídeos (TPR) à extremidade C-terminal da Hsp70 e Hsp90 (Scheufler et al., 2000). Aparentemente a Hsp70 estabiliza o RHE em uma conformação capaz de ser reconhecida e de se ligar ao dímero da Hsp90, porém esta hipótese ainda necessita de mais evidências experimentais. A transferência de substrato entre os complexos da Hsp70 e Hsp90 ainda não foi esclarecida. Porém as proteínas Hop/Sti1, que funcionam como uma ponte entre os dois sistemas, poderiam selecionar moléculas de Hsp90 em uma conformação competente para a ligação ao substrato, somando-se a um aumento da concentração local das Hsp70 -Hsp90. Para a progressão do ciclo chaperona, Hsp70 e Hop/Sti1 precisam se dissociar do complexo, enquanto outras co-chaperonas específicas como peptidil prolil cis/trans isomerases (PPIases) e a p23/Sba1 se integram ao mesmo (Smith, 1993). Estas PPIases, assim como a Hop/Sti1, também possuem um domínio TPR que reconhece o motivo MEEVD e promove a ligação ao C-terminal da Hsp90. O segundo co-fator, p23/Sba1, se associa com a conformação N-terminal dimerizada da Hsp90 (Prodromou et al., 2000; Richter et al., 2004), fazendo parecer que o dramático rearranjo conformacional do estado aberto para o fechado da Hsp90 ocorre neste estágio do ciclo chaperona da Hsp90. Esta conformação fechada é metaestável e retorna para o estado aberto após a hidrólise do ATP (Richter et al., 2004). Por sua vez a proteína-substrato se dissocia da Hsp90, permitindo uma nova revolução do ciclo. Os primeiros passos do ciclo para maturação de quinases sinalizadoras são variações do tema descrito acima. Porém neste caso, a co-chaperona quinase-específica Cdc37 parece se associar primeiro com as proteínassubstrato em suas formas inativas, e então este complexo pode se ligar à Hsp90 (Kimura et al., 1997). Ainda não é possível ter certeza sobre a necessidade das chaperonas e co-chaperonas Hsp70, Hsp40 e Hop/Sti1 na ativação destas quinases (Caplan et al., 2007). Desde que a Cdc37 inibe parcialmente a

atividade ATPásica da Hsp90 (Siligardi *et al.*, 2002), é plausível especular que um bloqueio momentâneo da atividade ATPásica facilitaria a transferência de substrato à Hsp90 em geral. Até o momento mais de uma dúzia de co-fatores distintos foram identificados (Richter *et al.*, 2005). Seu grande número não encontra paralelo com os outros sistemas chaperonas, sendo que a maioria destes co-fatores se liga à Hsp90 com afinidade sub-micromolar.



Figura 2: Estrutura cristalográfica da proteína Hsp90 de levedura ligada à co-chaperona p23 (Sba1) e detalhe do sítio de ligação de ATP. A) Os monômeros da Hsp90 estão representados em azul e laranja, enquanto que os monômeros da co-chaperona p23 estão representados em vermelho e verde. "N" e "C" indicam as extremidades N e C-terminal da Hsp90, respectivamente. Os aminoácidos correspondentes à região carregada da Hsp90 estão pontilhados. O modelo superior esquerdo apresenta o nucleotídeo AMP-PNP ligado ao N-terminal da Hsp90. A seqüencia EEVD (C-terminal), está apenas indicada por não

ter estrutura tridimensional resolvida (Modificada de Ali *et al.*, 2006). B) Detalhe do sítio de ligação ao ATP da proteína Hsp90 situado na região N-terminal. O ATP é representado como esferas e bastões (átomos de carbono: cinza; nitrogênio: azul; oxigênio: vermelho, fósforo: laranja) e a Hsp90 é representada como uma superfície sólida (carregada negativamente: vermelho; carregada positivamente: azul e eletrostaticamente neutra: cinza). Modificado de Prodromou *et al.*, 1997.

1.3 Hsp100

1.3.1 Dados gerais

As Hsp100, relacionadas à família das Clps (*caseinolytic protease*), têm atividade ATPásica e estão envolvidas nas mais diversas atividades biológicas tais como tolerância ao estresse, proteólise, transposição de DNA e regulação gênica (Parsell *et al.*, 1994; Schirmer *et al.*, 1996; Schirmer *et al.*, 1998). A função mais importante das Hsp100 seria a de auxiliar na solubilização de proteínas que estariam formando agregados (Schirmer *et al.*, 1996) e/ou atuando na desorganização de agregados protéicos em seus componentes (Levtchenco *et al.*, 1997). Esta família de chaperonas moleculares pode ser subdividida em duas classes. A classe I, composta pelas subfamílias ClpA, B, C e D descritas em *E. coli* e sua homóloga descrita em levedura (*S. cerevisiae*), a Hsp104. A classe II é composta pelas subfamílias ClpM, N, X e Y (Schirmer *et al.*, 1996).

As proteínas da família Hsp100, que pertencem à classe I, possuem dois domínios altamente conservados (Parsell *et al.*, 1994; Schirmer *et al.* 1996; Schirmer *et al.*, 1998). Em cada um destes domínios existe um sítio de ligação ao ATP específico para a classe de proteínas Hsp100, cuja seqüência consenso é [GX₂GXGKT] (Schirmer *et al.*, 1998). Em geral, adotam uma forma hexamérica na presença de ATP, formando uma estrutura semelhante a um anel (Parsell *et al.*, 1994). Dois exemplos da classe I das Hsp100, as chaperonas ClpA e ClpB de *E. coli*, apresentam atividade ATPase semelhante, diferindo em características físico-químicas e funcionais (Schirmer *et al.*, 1998). Apenas a ClpB proveniente da bactéria *Thermus thermophilus* possui a estrutura de todos os seus domínios resolvida (Figura II), sendo que o cruzamento dos dados de cristalografia com dados de microscopia crio-eletrônica possibilitaram demonstrar que esta ClpB forma hexâmeros (Lee *et al.*, 2003). As proteínas da classe II são menores e possuem apenas um dos dois sítios de ligação a nucleotídeos adenosina (Schirmer *et al.*, 1996). As proteínas ClpA e ClpX de *E. coli* apresentam atividade chaperona intrínseca e também agem como parte de complexos proteolíticos, diferindo da Hsp104 (ClpB) que funciona como uma chaperona independente de componentes proteolíticos (Grimaud *et al.*, 1998; Schirmer *et al.*, 1996). A ClpA e ClpX

são capazes de formar complexos com uma serino-proteinase denominada Clp Protease, na proporção de 2:1 (ClpAP, ClpXP e ClpAXP) (Grimaud *et al.*, 1998).

A proteína ClpB e suas ortólogas envolvidas na desagregação protéica são compostas, como mencionado acima, por um domínio N-terminal e dois domínios de ligação à nucleotídeos (domínios AAA), os quais são fundamentais para a formação do hexâmero e para a função chaperona (Schirmer *et al*, 1996; Zolkiewski, 2006). Apesar das muitas funções diferentes, todos os membros da super-família AAA possuem características importantes, tais como: estrutura oligomérica, domínios de ligação aos nucleotídeos tipo "Walker" homólogos e a capacidade de utilizar a energia derivada da hidrólise do ATP para remodelar seus substratos (Vale, 2000; Ammelburg *et al*, 2006). A estrutura por cristalografia de raios-X do monômero da ClpB de *Thermus thermophilus* juntamente com a reconstituição por microscopia crio-eletrônica de seu oligômero definiu a estrutura e orientação de seus domínios (figura 3, Lee *et al.*, 2003). Esta ClpB forma um hexâmero no qual se distinguem duas camadas, formadas pelos domínios AAA-1 (anel superior) e AAA-2 (anel inferior), as quais circundam um poro axial com diâmetro de aproximadamente 16 Å (Lee *et al*, 2003; Diemand and Lupas, 2006).



Figura 3: Estrutura cristalográfica da proteína ClpB de Thermus thermophilus.* A) Estrutura do monômero indicando os principais domínios: N-terminal, AAA-1, Central e AAA-2. B) Estrutura do hexâmero obtida através da combinação dos dados de cristalografia de raios-x com microscopia crioeletrônica. O modelo está mostrando o potencial eletrostático nas superfícies superior e inferior, respectivamente (azul: potencial positivo; vermelho: potencial negativo). * Modificado de Lee *et al.*, 2003.

Um estudo envolvendo a mudança estrutural sofrida pela ClpB de *T. thermophilus* mostrou que o anel superior desenvolve uma mudança conformacional dirigida pela ligação ao ATP, enquanto que o anel inferior permanece inalterado (Lee *et al*, 2007). O domínio AAA-1 contém ainda uma região definida como uma estrutura super espiralada (*coiled coil*) em forma de hélice que transpassa os limites do hexâmero em direção ao meio externo, chamada de domínio central (*middle domain*), e que é constituída de quatros hélices alfa (Schirmer *et al*, 1996; Lee *et al*, 2003). As análises genéticas e bioquímicas atribuem o correto funcionamento destas chaperonas desagregadoras à mobilidade desta estrutura/domínio, pois quando sua capacidade de movimentação foi reduzida, a chaperona perdeu parcial ou completamente sua atividade biológica (Lee *et al*, 2003). Finalmente, este domínio central é o principal diferencial entre os membros das sub-famílias, como as proteínas ClpA e ClpX, que possuem uma função dupla porque atuam como subunidades reguladoras de proteases (Wojtkowiak *et al*, 1993) e também como chaperonas moleculares (Wickner *et al*, 1994; Wawrzynow *et al*, 1995). Já quanto ao domínio N-terminal da ClpB, que se encontra ligado ao domínio AAA-1 e portanto situado no topo do anel hexamérico, acredita-se que auxiliaria a interação entre os agregados ou substratos parcialmente desenovelados com a chaperona.

1.3.2 Desagregação mediada por chaperonas

Alterações nas condições ambientais, como as variações de temperatura, estão constantemente desafiando o estado nativo das proteínas. Sob condições desnaturantes, a ocorrência de conformações não nativas é favorecida, o que pode levar à agregação protéica intracelular. No entanto, a sobrevivência celular nestas condições é tremendamente favorecida por um pré-condicionamento em altas temperaturas, denominado termotolerância adquirida, a qual ocorre devido ao aumento da síntese de chaperonas no período de condicionamento (Li e Werb, 1982). Sanchez e Lindquist (1990) demonstraram que células de levedura expressando a chaperona Hsp104 são aproximadamente 10.000 vezes mais resistentes as altas temperaturas do que as células sem esta chaperona. Assim, as linhagens de levedura expressando a proteína Hsp104 são muito mais resistentes a altas temperaturas do que células em que esta chaperona está ausente. Resultados similares foram obtidos para as proteínas ClpB e Hsp101, proteínas homólogas em células bacterianas e de plantas, respectivamente (Kitagawa et al, 1991; Squires et al, 1991; Queitsch et al, 2000). Já em estudos in vivo foram relacionadas à remoção de agregados intracelulares induzidos por aumento de temperatura às chaperonas responsáveis pela termotolerância, especificamente as proteínas Hsp104 de Saccharomyces cerevisiae (Parsell et al, 1994) e ClpB de Escherichia coli (Mogk et al, 1999), ambas proteínas da grande família AAA⁺ (ATPases Associadas com diversas Atividades celulares). Porém, muitos estudos in vitro demonstraram que as chaperonas ClpB e Hsp104 não são capazes de desagregar substratos de forma efetiva sem a ajuda de outras chaperonas. Neste caso, as chaperonas da família Hsp100 interagem com o sistema Hsp70, que em procariotos é composto pelas proteínas DnaK e as co-chaperonas DnaJ e GrpE, enquanto que em S. cerevisiae a interação ocorre com a chaperona Ssa e a co-chaperona Ydj (DnaK/Ssa e DnaJ/Ydj são homólogas à Hsp70 e Hsp40 humanas, respectivamente) (Glover and Lindquist, 1998; Goloubinoff et al, 1999; Motohashi et al, 1999; Zolkiewski, 1999). Fenótipos relacionados à agregação protéica são observados tanto em mutantes da proteína ClpB como também das proteínas DnaK, DnaJ e GrpE, sendo que nestes últimos o fenótipo representa o resultado de múltiplos fatores. Isto porque as proteínas que compõe o sistema Hsp70 atuam não somente no processo de desagregação, mas também em muitos outros processos celulares vitais, o que inclui o enovelamento de proteínas nascentes (Deuerling et al,

1999; Teter *et al*, 1999), o remodelamento de complexos protéicos (Liberek *et al*, 1988; Zzaman *et al*, 2004; Goldfless *et al*, 2006) e a regulação da resposta ao choque térmico (Gamer *et al*, 1992; Liberek and Georgopoulos, 1993), dificultando a análise dos efeitos referentes somente à agregação protéica. A eliminação dos agregados intracelulares pode ocorrer por desagregação mediada pelo sistema bichaperona Hsp70-Hsp100 e subseqüente reenovelamento, ou simplesmente através de proteólise. Estudos analisando a remoção de agregados em linhagens de *E. coli* na qual os genes que codificam algumas proteases ou chaperonas envolvidas com a desagregação foram deletados, revelaram que a ausência da ClpB é mais comprometedora para esta função específica (Laskowska *et al*, 1996a; Mogk *et al*, 1999).

A importância do processo de desagregação mediado pela chaperona ClpB foi evidenciada por estudos mutagênicos nos quais a ClpB foi modificada para atuar na via de degradação ao invés da via de desagregação. O experimento mostrou que a linhagem mutante não apresenta termotolerância mesmo quando a eliminação dos agregados é mantida, destacando a importância da recuperação das proteínas anteriormente agregadas e não somente a desagregação em si (Weibezahn et al, 2004). Conclusões semelhantes sobre a importância dos processos de desagregação e reenovelamento foram também observadas em estudo envolvendo os eventos de replicação e tradução mitocondriais em S. cerevisiae. O uso da mitocôndria neste tipo de estudo se torna interessante devido ao fato de que a maioria de suas proteínas são codificadas no núcleo, portanto dependentes de importação. Assim, através do bloqueio da tradução de genes nucleares temos a eliminação de proteínas recém sintetizadas, e a recuperação das funções mitocondriais se torna totalmente dependente das proteínas pré-existentes. Estes estudos revelaram que a recuperação da replicação do DNA mitocondrial (Germaniuk et al, 2002), tradução (Schmitt et al, 1996) e também morfologia mitocondrial normal (Lewandowska et al, 2006) seguida de severo estresse por calor se baseiam mais na reativação de proteínas termicamente inativadas residindo no interior da mitocôndria, do que na importação de proteínas recém-sintetizadas. Os dados experimentais demonstram que a desagregação e a reativação são favorecidas sobre a proteólise, o que é compreensível quando comparados os gastos de energia envolvidos em cada processo, até a obtenção de uma nova proteína nativa. Também é preciso levar em conta a questão do tempo, uma vez que ao final do processo de desagregação e reenovelamento temos uma proteína pronta para desempenhar sua

atividade, ao passo que no final da proteólise a célula dispõe apenas de uma mistura de aminoácidos "reciclados".

Finalmente, vários estudos buscaram investigar se as proteínas Hsp70 e Hsp100 teriam outras funções, além da desagregação, no controle da qualidade protéica quando em condições de estresse. Pelham (1986) propôs que uma das funções do sistema Hsp70 seria a de prevenir a agregação protéica durante o estresse, o que já pôde ser demonstrado *in vitro* para todas as chaperonas Hsp70 estudadas, servindo até como um padrão de atividade para esta família chaperona. O problema de analisar esta questão in vivo é que mutações na Hsp70 podem influenciar tanto a prevenção da agregação como também a desagregação. Além disso, ambos os processos podem resultar da interação das chaperonas Hsp70 com os agregados e, portanto, a simples associação da Hsp70 com agregados não é capaz de discriminar entre as funções de desagregação e de prevenção. Já a natureza da atividade da Hsp70, a qual é baseada na retenção e liberação de substratos mediada por suas co-chaperonas, parece desafiar a função de prevenir a agregação, uma vez que tal processo requer interações mais estáveis com as proteínas-substrato. Alguns trabalhos levantaram a possibilidade que a co-chaperona GrpE (fator de liberação de nucleotídeos ADP da Hsp70/DnaK) poderia trabalhar como um sensor térmico e que, em altas temperaturas, poderia mudar a conformação da homóloga procariótica da Hsp70, DnaK, tornandoa apta para a ligação ao substrato (Groemping and Reinstein, 2001; Grimshaw et al, 2003). Alguns trabalhos relacionam a atividade chaperona da Hsp70 com a proteólise, sugerindo que o sistema Hsp70 seria capaz de interagir com proteases de maneira similar às chaperonas Hsp100 (Park et al, 2007). Mutações em genes que codificam chaperonas Hsp100 resultam em alta e estável agregação protéica intracelular e, como estas chaperonas não são capazes de prevenir a agregação protéica em estudos in vitro, é possível assumir que o principal papel da família Hsp100 seria o de desagregar proteínas e não de prevenir sua agregação. Assim, a principal atividade do sistema bi-chaperona Hsp70-100 seria a desagregação e não a prevenção da agregação, atividade esta que é, por sua vez, eficientemente desempenhada pelas proteínas que compõe a família das proteínas de choque térmico de baixa massa molecular ou, simplesmente, "small heat shock proteins".

1.3.3 Modelos de desagregação protéica dependente do sistema Hsp100/Hsp70

A descrição exata do mecanismo pelo qual as proteínas são desagregadas ainda não é conhecida, porém alguns modelos já foram propostos. A princípio, imaginou-se que as chaperonas ClpB interagissem com os agregados no primeiro passo do processo de desagregação, alterando a estrutura dos mesmos para que se tornem reconhecíveis à chaperona Hsp70 (Ben-Zvi and Goloubinoff, 2001). Outro modelo proposto para o processo de desagregação seria a quebra dos agregados maiores em agregados menores pela ClpB (Lee et al, 2003) que seriam em seguida processados pela Hsp70. Recentemente, foi proposto que o sistema Hsp70 é o verdadeiro responsável pela interação com os agregados protéicos durante o passo inicial da desagregação (Weibezahn et al, 2004; Zietkiewicz et al, 2004). Bukau e colaboradores adicionaram o motivo "IGL" ao domínio AAA-2 da ClpB, , de forma que a proteína mutante fosse capaz de interagir com a subunidade proteolítica ClpP (Weibezahn et al, 2004). Na presença deste mutante, proteínas agregadas eram eficientemente degradadas somente na presença das proteínas do sistema Hsp70, enfatizando a importância dos processos de translocação pelo poro central da ClpB bem como a interação dos substratos com o sistema Hsp70. Uma conclusão similar sobre a ordem em que os eventos ocorrem durante a desagregação foi dirimida através de um estudo envolvendo cinética, no qual foi demonstrado que a pré-incubação dos agregados com o sistema Hsp70, mas não com a ClpB, alterou as características físicas dos agregados e acelerou o início da reação de desagregação (Zietkiewicz et al, 2004). Além disto, também foi observado que o sistema Hsp70 é capaz de extrair polipeptídeos dos agregados na ausência da ClpB (Zietkiewicz et al, 2006), porém os polipeptídeos não são capazes de readquirir o enovelamento nativo, o que sugere que a ClpB/Hsp100 estaria mais envolvida com o processo de reenovelamento do que com a retirada dos polipeptídeos dos agregados, ou seja, a desagregação propriamente dita.

Um trabalho recente envolvendo mutagênese demonstrou que o chamado domínio central da ClpB conecta a atividade de translocação com a atividade chaperona Hsp70, sugerindo que esta introduz polipeptídeos na entrada do canal/poro da ClpB/Hsp100 através da interação com este domínio (Haslberger *et al*, 2007). Esta translocação dos polipeptídeos pelo poro central da ClpB/Hsp100 é um passo importante nos processos de desagregação e reativação das proteínas afetadas, como mostrado

por Bukau e colaboradores (Weibezahn *et al*, 2004). A importância da integridade do canal para a atividade das chaperonas Hsp100 foi demonstrada por estudos de mutagênese (Lum *et al*, 2004; Schlieker *et al*, 2004). No entanto, apesar de muitos trabalhos mostrarem a cooperação entre o sistema Hsp70 e a Hsp100, o exato modo como isto acontece ainda é incerto, porém já se sabe que esta interação é altamente específica, pois tanto eventos de desagregação *in vitro* como *in vivo* só ocorrem quando ambas as proteínas são originárias de um mesmo organismo (Glover and Lindquist, 1998; Krzewska *et al*, 2001).

Mesmo não se conhecendo o mecanismo exato da translocação através das chaperonas desagregadoras, estudo realizado com a proteína ClpA, a qual atua como chaperona e como subunidade reguladora de protease, mostrou que a ligação dos substratos a três "loops" expostos na superfície interna do poro central é fundamental para o desenovelamento dos polipeptídeos. Este trabalho propõe também que a força mecânica exercida sobre o substrato se origina de um destes "loops", o qual por sua vez é movido pela hidrólise de ATP, resultando na translocação do substrato pelo poro e seu conseqüente desenovelamento (Hinnerwisch et al, 2005). Mais recentemente foi relatado que mesmo na ausência do sistema Hsp70, as chaperonas desagregadoras ClpB (E. coli) e Hsp104 (S. cerevisiae) são capazes de realizar a reação de desagregação, porém com baixa eficiência e na presença do nucleotídeo modificado e não-fisiológico ATPyS (Schaupp et al, 2007; Doyle et al, 2007a, b). Esta observação levanta a possibilidade de que, durante a reação de desagregação, o sistema chaperona Hsp70 não apenas desempenha o papel de entregar o polipeptídeo retirado dos agregados no poro central da Hsp100, mas também influencia o ciclo ATPase da Hsp100 de uma maneira que é mimetizada pelas condições artificiais obtidas na presença da mistura ATPyS e ATP. Estudos de mutagênese recentes revelaram que a ligação do ATP ao domínio AAA-1 regularia a atividade chaperona da Hsp104, uma vez que é capaz de disparar a ligação de polipeptídeos e estimula fortemente a hidrólise de ATP no AAA-2, o que implica que a hidrólise de ATP neste último domínio seja importante para a reação desagregação (Schaupp et al, 2007). Após a ação da ClpB/Hsp100, a cadeia polipeptídica recém liberada é então capaz de readquirir sua estrutura nativa. Por analogia com o enovelamento co-traducional da cadeia polipeptídica que emerge do ribossomo, o mecanismo de translocação poderia permitir o enovelamento seqüencial da cadeia polipeptídica emergindo do poro central da ClpB. Portanto, é muito provável que o papel das chaperonas Hsp70 não seja restrito apenas a desemaranhar polipeptídeos dos agregados e posicionar o polipeptídio na ClpB antes da translocação, mas também promova o enovelamento do polipeptídio que emerge do canal da Hsp100, como estabelecido para o enovelamento pós-traducional. Isto explica porque a reação completa de desagregação *in vitro* não pode ser dividida em etapas seqüenciais dependentes da Hsp70 ou da Hsp100 (Goloubinoff *et al*, 1999; Zietkiewicz *et al*, 2004).

Finalmente, a eficiência da desagregação é quase completamente dependente das propriedades físicas dos agregados protéicos, como demonstrado por estudos pioneiros em que a ação do sistema chaperona Hsp70 sozinho foi suficiente para a desagregação de certos substratos (Skowyra *et al*, 1990; Ziemienowicz *et al*, 1993). Outros substratos, no entanto, requerem a cooperação de ambas chaperonas para serem desagregados (Glover and Lindquist, 1998; Goloubinoff *et al*, 1999; Motohashi *et al*, 1999; Zolkiewski, 1999). Além da influência específica do substrato, as condições em que a agregação foi provocada também determinam a necessidade do envolvimento da ClpB/Hsp100 na desagregação (Ben-Zvi and Goloubinoff, 2002; Lewandowska *et al*, 2007). O sistema Hsp70 é, aparentemente, suficiente para o enovelamento correto quando os polipeptídeos agregados apresentam uma conformação quase nativa e/ou não estão "aprisionados" em estruturas não-nativas estáveis. Nas situações em que as proteínas formadoras do agregado se encontram mais distantes do estado nativo, incluindo o aumento de estruturas tipo folhas-beta, necessitariam da atividade translocadora da Hsp100 para sua correta reativação (Lewandowska *et al*, 2007).

1.4 Co-chaperonas

Interações das chaperonas moleculares Hsp70 e Hsp90 com proteínas clientes conhecidas envolvem, comumente, muitas proteínas chamadas de co-chaperonas. Dentre as atividades atribuídas co-chaperonas estão a regulação da atividade ATPásica da chaperona ou troca de nucleotídeos no sítio catalítico desta (Prodromou e Pearl, 2003; Mayer et al., 2001), o recrutamento de chaperonas e cochaperonas adicionais (Chen e Smith, 1998), direcionamento da proteína cliente complexada para a degradação (Hohfeld et al., 2001) e localização subcelular, ou seja, direcionamento para diferentes organelas/compartimentos celulares (Pratt et al., 1999). Um dos sistemas chaperona/proteínas clientes mais estudados é a via de montagem para complexos receptores esteróides, nos quais múltiplas chaperonas e co-chaperonas participam em um processo dinâmico que envolve múltiplos passos (Pratt e Toft, 1997; 2003). Neste processo, a co-chaperona Hop (Hsp70-90 organizing protein), a qual pode se ligar a ambas Hsp70 e Hsp90, realiza um papel crucial no recrutamento da Hsp90 ao complexo Hsp70receptor pré-existente (Chen e Smith, 1998; Dittmar et al., 1996; Kosano et al., 1998), promovendo a progressão da montagem e maturação final do complexo. Já a co-chaperona p23 é a menor parceira da maquinaria da Hsp90, possuindo massa molecular em torno de 23 kDa e estrutura relativamente simples (Weikl et al 1999; Weaver et al 2000). Assim como a Hop, a co-chaperona p23 também faz parte do complexo de montagem e maturação de receptores esteróides (Felts e Toft, 2003), além de ser encontrada em complexo com outras proteínas cliente da Hsp90, tais como a telomerase (Holt et al., 1999) e o fator de transcrição Hsf1 (Nair et al., 1996). Finalmente, as proteínas da família PDI (protein disulfide isomerase) são enzimas que possuem um sítio catalítico com a seqüência CXXC (X: qualquer aminoácido, C: Cisteína), a qual é capaz de catalisar a troca da ligação dissulfeto com o substrato ou entre substratos. A função principal das PDIs é promover o enovelamento oxidativo de proteínas presentes no retículo endoplasmático, muitas vezes atuando sobre as mesmas proteínas clientes da Grp94 (representante da família Hsp90) (Appenzeller-Herzog e Ellgaard, 2008).

1.5 Co-chaperona p23

1.5.1 Dados gerais

A proteína p23 é encontrada de leveduras a humanos, sendo também encontradas com prováveis representantes em plantas (Garcia-Ranea *et al* 2002). Foi encontrada inicialmente como integrante do complexo da Hsp90 com o receptor de progesterona (Johnson *et al.*, 1994), no entanto também pode ser encontrada em complexos de outras proteínas clientes da Hsp90, incluindo a FES tirosina quinase (Nair *et al* 1996), polimerases (Holt *et al* 1999), transcriptase reversa (Hu *et al* 2002) e outras. A p23 se liga *in vitro* a uma conformação específica da Hsp90, promovendo a estabilização da ligação da proteína Hsp90 ao ATP e estimulando a hidrólise deste (ver adiante). Além disto, ainda atua como chaperona passiva, ou seja, possui atividade de prevenção da agregação de proteínas desnaturadas (Bose *et al* 1996; Freeman *et al* 1996). Alguns estudos mostram que a p23 ainda pode continuar atuando sobre proteínas clientes já liberadas pela Hsp90 (Freeman *et al* 2000; Freeman and Yamamoto 2002).

Como já mencionado, a proteína p23 faz parte do complexo de receptor esteróide com a Hsp90. Além disto, pode ser encontrada ligada à Hsp90, sem a presença de proteínas clientes, em lisados celulares (Johnson and Toft 1994). A presença da p23 em complexos isolados de lisados e também formados por proteínas purificadas (*in vitro*) depende da presença de adenosina trifosfato (ATP) (Johnson and Toft 1994; Sullivan *et al.*, 1997). Isto indicou que os nucleotídeos desempenham um papel importante na regulação da estrutura e função da Hsp90, acabando por desencadear uma série de trabalhos sobre a ligação e hidrólise do ATP pela Hsp90 (Young *et al.*, 2001). A ligação da Hsp90 ao ATP provoca uma alteração conformacional em que são formados contatos adicionais na região N-terminal (Prodromou *et al* 2000). Neste estágio, o ATP fica aprisionado pela Hsp90 e já se encontra comprometido com o processo de hidrólise catalisado pela chaperona (Weikl *et al* 2000; Sullivan *et al* 2002). A proteína p23 se liga à Hsp90 neste estágio e estabiliza o estado conformacional da Hsp90 ligado ao ATP. Para que a ligação da p23 com a Hsp90 corra, são necessários a presença dos resíduos que compõe o domínio N-terminal da Hsp90 e também 200 resíduos do domínio central desta chaperona (exceto resíduos da região carregada). Além disto, estes resíduos do domínio central precisam estar em uma configuração dimérica (Chadli *et al.*, 2000). Em uma série de experimentos, Sullivan e outros (2002) mostraram que a

remoção do ATP provoca a lenta transformação do estado conformacional "ATP ligado" da Hsp90 para o estado "não ligado". A adição do p23 desacelera esta transformação, aumentando a estabilização da conformação "ATP ligada" da Hsp90, apesar da relativa baixa afinidade da p23 pela Hsp90. Panaretou e colaboradores (2002) demonstraram que a p23 de levedura, denominada de Sba1, pode inibir fracamente a atividade da Hsp90 deste organismo. A Hsp90 humana e de levedura podem ter sua atividade ATPásicas estimuladas pela proteína ativadora Aha1 (*activator of Hsp90 ATPase*), sendo este estímulo inibido de forma eficiente pela presença da p23 de levedura. De forma semelhante, foi demonstrado que a p23 humana pode inibir a atividade ATPásica da proteína Hsp90 humana mesmo quando esta se encontra estimulada pela presença de uma proteína cliente (Mc-Laughlin *et al.*, 2002).

1.5.2 Estrutura da p23 humana

A estrutura cristalográfica da proteína p23 humana foi obtida na ausência de 35 resíduos de aminoácidos do seu C-terminal (Weaver *et al.*, 2000). Apesar da p23 ser um monômero em solução, um dímero ligado por ponte dissulfeto foi caracterizado, no qual cada subunidade é composta de 8 folhas beta que formam um sanduíche compacto antiparalelo, compreendendo os resíduos 1-110 da proteína (figura 4). Cadeias laterais hidrofóbicas alternadas das folhas opostas formam um núcleo hidrofóbico centrado no "empilhamento" dos anéis dos resíduos Phe19 e Phe38 (interação aromática). A região C-terminal ao resíduo 110 parece ser desestruturada (resíduos 110-160), fato consistente com o estudo de Weikl e colaboradores (1999). Baseado na sensibilidade a proteases e análises de espectros de dicroísmo circular, este estudo determinou que a proteína p23 é formada por uma estrutura em folha beta muito estável com uma cauda C-terminal flexível.

Quando a proteína p23 foi descoberta, esta parecia ser um membro da família das pequenas proteínas de choque térmico, as quais são chaperonas de tamanho similar. No entanto, a seqüência primária da p23 mostrou baixa similaridade com as destas pequenas chaperonas. Surpreendentemente, uma comparação das estruturas cristalográficas revelou estruturas em beta-sanduíche altamente homólogas para a proteína p23 e para as chaperonas MjHsp16.5 de *Methanococcus jannischii* e Hsp16.9B de trigo (Weaver *et al* 2000; van Montfort *et al* 2001). Através de alinhamentos de múltiplas

seqüencias, foi sugerido que estas proteínas possuem os resíduos conservados mais antigos do que pode ser uma grande família de proteínas que variam em tamanho e função mas que contem este motivo estrutural como parte de suas estruturas (Garcia-Ranea et al 2002). O "loop" altamente conservado da p23 (entre as fitas 7 e 8) se alinha bem ao "loop" presente na fita 6 das pequenas proteínas de choque térmico o qual, juntamente com o C-terminal flexível, forma "braços" capazes de interagir com o dímero vizinho formando a estrutura dodecamérica da proteína MiHsp16.5. Têm sido formulado que o choque térmico causa a separação do dodecâmero em dímeros, liberando portanto estes "braços" para as funções chaperona. Assim como a p23, o C-terminal flexível das pequenas proteínas de choque térmico contém muitos resíduos de aminoácidos carregados, os quais parecem ser importantes para a manutenção da solubilidade quando esta se encontra complexada com proteínas desnaturadas (Pasta et al 2002). O caráter iônico da região C-terminal ainda pode ser aumentado através da fosforilação de três resíduos de serina (W. Sullivan e D. Toft, comunicação pessoal). A remoção da cauda C-terminal da proteína p23 impede tanto sua atividade chaperona passiva sobre proteína desnaturadas como sua atividade co-chaperona sobre o vetor de progesterona (Weikl et al 1999; Weaver et al 2000). No entanto, a cauda C-terminal é dispensável na ligação à Hsp90 (Weikl et al 1999; Weaver et al 2000). Portanto, um modelo semelhante às proteínas similares a Hsp20 pode ser proposto para a p23, no qual a face contendo o "loop" b7-b8 pode cooperar com o C-terminal móvel para se ligar e trabalhar com a Hsp90 para promover a atividade chaperona sobre a proteína cliente (Felt e Toft, 2003).


Figura 4: Estrutura cristalográfica do dímero da p23 humana.* Dímero da proteína p23-C35 (aminoácidos correspondentes ao C-terminal deletado) resolvido por difração de raios-x. Ponte dissulfeto entre os resíduos Cys58 de cada monômero indicada em verde. *Modificado de Weaver *et al.*, 2000.

2 Objetivos

O objetivo geral deste trabalho foi obter as proteínas Hsp100 e Hsp82 de cana-de-açúcar e a cochaperona p23 do sistema Hsp90 humano e estudar as propriedades estruturais destas proteínas para compreender mais sobre a relação entre a estrutura e função. Para isso, foram utilizadas técnicas de biofísica molecular, tais como: dicroísmo circular, fluorescência, espalhamento dinâmico de luz e ultracentrifugação analítica.

Os objetivos específicos deste trabalho foram:

1. Produzir e caracterizar a proteína Hsp100 de cana-de-açúcar, determinando sua estrutura quaternária e os possíveis efeitos estruturais e funcionais causados pela presença de nucleotídeos e força iônica do meio.

2. Produzir e caracterizar a proteína Hsp82 de cana-de-açúcar, buscando compreender o efeito dos nucleotídeos sobre a dimerização do seu domínio N-terminal e sua relação com a atividade.

3. Produzir a proteína p23 humana para caracterizar a estruturação da sua região C-terminal, permitindo observar suas implicações na relação estrutura-função desta co-chaperona.

3 Materiais e métodos

3.1 Data mining e bioinformática

O cDNA que codifica as proteínas de interesse foram inicialmente identificados através de pesquisa no banco de dados de EST (*Expressed Sequence Tag*), utilizando seqüências de chaperonas e cochaperonas previamente identificadas no banco de dados *GenBank* do NCBI (National Center of Biotechnology Information), através da utilização da ferramenta de bioinformática *BLAST* (*Basic Local Alignment Search Tool*). As seqüências de aminoácidos das proteínas Hsp100 e Hsp82 de *Saccharum officinarum* (cana-de-açúcar), foram obtidas através de buscas no banco de dados do projeto SUCEST (Sugar Cane EST Genome Project – FAPESP) (Borges *et al.*, 2001).

3.2 Clone

Os clones que codificam as proteínas Hsp100 (SCQSST3115C07) e Hsp82 (SCQGLR1062C04) de cana-de-açúcar foram obtidos comercialmente junto à FUNEP (Fundação de Estudos e Pesquisas em Agronomia, Medicina Veterinária e Zootecnia), que comercializa cDNAs obtidos durante o projeto SUCEST – FAPESP.

3.3 Construção dos oligonucleotídeos iniciadores para PCR

Os oligonucleotídeos para clonagem e sequenciamento foram adquiridos comercialmente e estão descritos na tabela I.

Oligonucleotídeos	Finalidade	Sítio de	Seqüência 5' – 3'
		Restrição	
TC82CAN-S (5´)	Clonagem Hsp82	Ndel	GTCTCGCCACATATGGCCTCG
TC82CAN-AS (3′)	Clonagem Hsp82	<i>Eco</i> RI	CTCGGGGAATTCTTAGTCGAC
H82CAN1-S (5′)	Sequenciamento	-	TCAAGGATGATCAGTTG
	Hsp82		
H82CAN2-S (5´)	Sequenciamento	-	AGAACAAGATCCTCAAG
	Hsp82		
H82CAN3-S (5′)	Sequenciamento	-	AAGGTGATCAAGGAGGT
	Hsp82		
TCP23-S (5')	Clonagem p23	BamHI	TAGGATCCATGCAGCCTGCTTCTGCAAAG
TCP23-AS (3')	Clonagem p23	EcoRI	ATGAATTCTATTCCTTACTCCAGATCTGG

Tabela I: Oligonucleotídeos utilizados no projeto

TCPDI-S (5')	Clonagem PDI	BamHI	TAGGATCCATGTCGCACCCGTCCCCCA
TCPDI-AS (3')	Clonagem PDI	<i>Eco</i> RI	ATGAATTCTTTCTAAGCAAACATTTT
TCHop-S (5')	Clonagem Hop	<i>Eco</i> RI	ATGAATTCATGGAGCAGGTCAATGAGCT
TCPHop-AS (3')	Clonagem Hop	HindIII	ATAAGCTTTCACCGAATTGCAATCAGACC
SEQ100A (5´)	Sequenciamento	-	ACGGCGCTGATCAAGGCCAT
	Hsp100		
SEQ100B (5')	Sequenciamento	-	CGCGGAGCCGAGCGTGCCCG
	Hsp100		
SEQ100B (5')	Sequenciamento	-	CGACCGGTGTGGGCAAGACC
	Hsp100		

3.4 PCR

Nos experimentos de PCR (*polymerase chain reaction*) foi utilizado o kit PCR SuperMix[®] (Life Technologies), de acordo com o protocolo do fabricante. A reação foi realizada no termociclador *Mastercycler Personal* (Eppendorf), com temperatura de 95 °C para o desenovelamento do cDNA molde. Para a reação de amplificação foi escolhida a menor temperatura de anelamento entre os dois oligonucleotídeos utilizados. A etapa de extensão foi realizada a 72 °C, e o tempo calculado em função da extensão do cDNA-alvo. Assim, foi utilizado aproximadamente 1 minuto de extensão para cada 1000 pares de base do DNA a ser amplificado.

3.5 Gel de agarose

Os DNAs foram analisados através de eletroforese (sistema Horizon[®] 58 – Life Technologies) em gel de agarose contendo 10 mg/mL de agarose em tampão TBE (Tris-Borato 45 mM – pH 8,0, EDTA 1 mM). Após a aplicação do DNA no gel, este foi submetido a uma corrente elétrica de 75 Volts, por aproximadamente uma hora e trinta minutos. Em seguida o gel foi incubado em tampão TBE, contendo 0,5 µg/mL de brometo de etídeo, por 30 minutos, e visualizado no transluminador de UV T1202 (Sigma[™]). Utilizamos a câmera digital DC120 (Kodak[™]) acoplada a um computador contendo o programa EDAS120 (*Electrophoresis Documentation and Analysis System* 120 – Kodak) para o registro das imagens dos géis.

3.6 Vetores

3.6.1 Vetor de clonagem/transferência

Para a clonagem dos produtos de PCR produzidos a partir de oligonucleotídeos flanqueados para os sítios de restrição adequados, foi utilizado o vetor pGEM-T easy[®] (Promega – figura 5). A utilização do vetor de transferência facilita o processo de clonagem uma vez que não é necessário digerir o produto de PCR antes da ligação. Isto evita a digestão de fragmentos de DNA que exibem sítios de restrição em suas extremidades, fato este que diminui consideravelmente a eficiência de clivagem da maioria das enzimas de restrição.



Figura 5: Vetor de clonagem pGEM[®]-T Easy.

3.6.2 Vetor de expressão

Para a expressão das proteínas de interesse, seus respectivos ORFs (*open reading frame*) foram subclonados no vetor de expressão pET28a (Novagen). Este vetor está sob o controle do sistema T7, no qual o cDNA alvo é clonado sob o controle do promotor do gene 10 do bacteriófago T7 (Studier & Moffatt, 1986). A expressão do cDNA alvo ocorre somente na presença da RNA polimerase desse vírus, devido à especificidade do promotor (Kochetkov *et al.*,1998). As cepas BL21(DE3) possuem, integrado ao seu genoma, o prófago do bacteriófago T7, que é lisogênico. Neste prófago, a expressão da T7 RNA

polimerase está sob o controle do promotor lacUV5 do operon *lac*, que é mantido reprimido na presença de glicose e na ausência de seu indutor natural a 1,6 aldolactose, um derivado da galactose. O produto do gene *lac I* representa o repressor que, quando ligado ao operador presente no prófago do T7, impede a expressão da T7 RNA polimerase e, conseqüentemente, a transcrição do cDNA alvo (Dubendorff & Studier, 1991). Porém, na presença de IPTG (isopropil-β-D tiogalactopiranosídeo), um análogo do indutor natural (1,6-aldolactose), o repressor se desligará do operador liberando a expressão da T7 RNA polimerase do bacteriófago T7. Esta, por sua vez, irá atuar na transcrição do cDNA alvo clonado no vetor do sistema pET (Studier *et al.*,1990). Como a T7 RNA polimerase é cinco vezes mais eficiente que a RNA polimerase da *E. coli* e está sendo produzida em grandes quantidades, em pouco tempo quase todo o produto protéico sintetizado pela bactéria será, praticamente, apenas o codificado pelo cDNA alvo. Outra vantagem desse promotor é sua especificidade, ou seja, a RNA polimerase da *E. coli* não transcreve o cDNA alvo, o que permite controlar o momento da expressão da proteína de interesse (Studier & Moffatt, 1986).

Em cada extremidade do sítio de clonagem múltipla do vetor pET28a (figura 6), existe uma seqüência de nucleotídeos que codifica seis histidinas que poderão estar adicionadas à extremidade amino-terminal e/ou carboxi-terminal da proteína produzida, dependendo da estratégia de clonagem. A adição desta seqüência de histidinas permite a purificação do polipeptídio através de cromatografia de afinidade, utilizando para isto uma resina contendo íons de níquel imobilizados (Westra *et al.*, 2001).



Figura 6: Vetor de expressão pET-28a.

3.7 Linhagens de Escherichia coli

A linhagem DH5α foi utilizada para a produção e a manutenção de todos os plasmídios. Dentre os motivos para sua utilização, destacamos: 1) não é capaz de produzir a enzima endonuclease I, permitindo preparações plasmidiais melhores devido à ausência de digestão inespecífica deste tipo de DNA; 2) Não carrega o plasmídeo F+ para a conjugação, garantindo sua biossegurança; 3) presença dos operons deoR e nupG, que permitem a captura de grandes plasmídeos; 4) porção alfa do gene da beta-galactosidase deletada (lacZdeltaM15), o que permite busca de colônias em ensaios de "azul e branco"; 5) presença do gene hsdR17, o qual permite a clonagem de DNAs não metilados, como os provenientes de PCR.

A linhagem BL21(DE3) possui baixos níveis de produção de proteases, que poderiam degradar a proteína de interesse, e possui ainda o gene (DE3), responsável pela síntese da enzima T7 RNA polimerase controlado pelo promotor *lac*UV5 e ativado pela adição de IPTG.

A linhagem BL21(DE3)pLysS, além das características acima descritas, apresenta expressão aumentada de lisozima, que inibe a T7 RNA polimerase, regulando a expressão da proteína de interesse clonada no vetor do sistema pET. Esta linhagem é utilizada quando a proteína de interesse se apresenta tóxica para a bactéria, mesmo quando em baixas taxas de indução, ou problemas na etapa de lise das bactérias induzidas.

3.8 Meios de cultura

O meio de cultura LB (Luria-Bertani) foi utilizado para o crescimento de bactérias, preparação de células competentes, expressão de proteínas e manutenção de bactérias transformadas com o vetor de interesse. Este meio é composto por: triptona 10 g/L, extrato de levedura 5 g/L, NaCl 10 g/L. Para o meio sólido foi acrescentado 15 g/L de ágar.

O meio de cultura SOC foi utilizado para crescimento de células transformantes por choque térmico e eletroporação. Composição: triptona 20 g/L, extrato de levedura 5 g/L, NaCl 10 mM, KCL 2,5 mM, MgCl₂ 10 mM, Mg₂SO₄ 10 mM e Glicose 20 mM. A glicose foi acrescentada, após a esterilização em autoclave do restante dos reagentes, a partir de uma solução estoque 1,0 M filtrada com filtro de 0,22 μm.

O meio de cultura PSI foi utilizado na preparação de bactérias competentes por choque térmico. Composição: peptona 16 g/L, extrato de levedura 10 g/L, NaCl 5 g/L, KH_2HPO_4 0,4 g/L e Na_2HPO_4 0,5 g/L, pH 7,6.

3.9 Clonagem

3.9.1 Clonagem em pGEM-T

As clonagens no vetor pGEM-T foram realizadas com o kit de clonagem pGEM-T easy[®] (Promega) de acordo com o protocolo do fabricante.

3.9.2 Subclonagem em pET28a

Os vetores de expressão e transferência contendo o cDNA de interesse foram submetidos a digestão exaustiva com as respectivas enzimas – conforme o protocolo de cada fabricante (concentração de DNA, unidades de enzima, temperatura e tempo de incubação). Após a digestão, as amostras foram submetidas à eletroforese em gel de agarose 1% e coradas com brometo de etídeo. As bandas, correspondentes ao fragmento de interesse e ao vetor de expressão digerido, foram extraídas do gel e purificadas usando o kit Gel Clean up (Eppendorff) – conforme o protocolo do fabricante. Após a purificação, o inserto e o vetor digerido foram incubados com a enzima T4 DNA ligase (USB), por 14 horas a 14 °C.

3.10 Precipitação de produto de ligação

O DNA contido na reação de ligação foi precipitado com acetato de sódio e etanol absoluto, evitando a interferência do sal presente no tampão de ligação na eficiência de transformação por eletroporação em DH5α. Acrescentou-se 1/10 do volume de ligação de acetato de sódio 3 M pH 5,2 e 2,5 vezes o volume de ligação de etanol absoluto. Após agitação do material, este foi armazenado a -20 °C por pelo menos 1 hora, centrifugado a 14.000 rpm (rotor F45-30-11 - Eppendorf) por 15 minutos a 4 °C e o sobrenadante descartado. Em seguida, foram adicionados 2,5 vezes o volume de ligação de etanol 70% e repetiu-se a centrifugação nas mesmas condições, o sobrenadante foi então descartado e o precipitado de DNA seco a 37 °C.

A obtenção dos clones para expressão das proteínas de interesse foi feita por eletrotransformação (ver item 3.11) de bactérias *E. coli* BL21(DE3), ou BL21(DE3)pLysS. O DNA plasmidial para esta transformação foi purificado pelo kit Miniprep Eppendorf conforme protocolo do fabricante, não necessitando ser precipitado.

3.11 Obtenção de bactérias eletro-competentes

A cepa de interesse foi pré-inoculada em 100 mL de LB e crescida durante aproximadamente 16 horas. O pré-inóculo (5 mL) foi diluído em 500 mL de meio LB e as células cresceram a 37 °C sob agitação

de 250 rpm até a $DO_{600 nm}$ (densidade ótica) atingir aproximadamente 0,7. As células foram armazenadas em banho de gelo por 20 minutos e centrifugadas a 4 °C por 15 minutos a 15.000xg. O sobrenadante foi descartado e as células ressuspendidas em 500 mL de glicerol 10% gelado. Esta operação foi repetida com três volumes diferentes: 250 mL, 20 mL e finalmente com 2 mL. As células foram armazenadas em alíquotas de 100 µL a -80 °C.

3.12 Obtenção de bactérias competentes para transformação por choque térmico.

A cepa de interesse foi pré-inoculada em 5 mL de meio PSI e crescida por 16 horas a 37 °C, sob agitação de 200 rpm. Em seguida adicionamos o pré-inóculo em 200 mL de meio PSI e incubamos novamente a 37 °C e 200 rpm até a cultura atingir $DO_{600 nm}$ 0,7. A cultura foi então centrifugada a 4000 rpm (rotor F45-30-11 – Eppendorf), por 10 minutos, a 4 °C. O sobrenadante foi descartado e o pellet ressuspendido em 25 mL de Tampão de transformação I (KOH 30 mM, ac. acético 30 mM, MgCl₂ 50 mM, RbCl₂ 100 mM, CaCl₂ 13 mM, 13,5 % Glicerol (v/v), pH 5,8) e deixado em gelo por quinze minutos. As bactérias foram então centrifugadas a 4000 rpm, 4 °C, por 15 minutos; o sobrenadante foi descartado e o pellet ressuspendido em 8 mL de Tampão de transformação II (MOPS (Ácido 3-[N-morfolino]propanosulfônico) 0,084g/L, CaCl₂ 100 mM, RbCl₂ 10 mM, 13,5 % Glicerol (v/v), pH 7,0). As células foram divididas em alíquotas de 150 ou 75 μ L em tubos de microcentrífuga de 1,5 mL, congeladas em nitrogênio líquido e armazenadas a -80 °C.

3.13 Transformação de bactérias competentes

As transformações de bactérias pelo método de eletroporação, utilizando o eletroporador MicroPulserTM Electroporation Apparatus (Bio-Rad), foram realizadas conforme protocolo do fabricante. Nas transformações de bactérias competentes por choque térmico, cerca de 400 ng de DNA foram incubados em um tubo de microcentrífuga, contendo 75 μ L de bactérias competentes, por trinta minutos em gelo. O tubo foi então transferido para um banho-maria a 42°C por 90 segundos e imediatamente recolocado no gelo por mais 90 segundos. Adicionamos 800 μ L de meio SOC à transformação e incubamos a 37°C e 200 rpm por uma hora. Finalmente, cerca de 200 μL da transformação foram adicionados a placa contendo meio LB e o antibiótico seletivo.

3.14 Extração de DNA plasmidial

A extração de DNA plasmidial foi realizada através do método de lise alcalina em pequena escala como descrito em Sambrook *et al* (1989). Para a obtenção de DNA plasmidial com um grau de pureza maior, principalmente para as reações de PCR e seqüenciamento de DNA, utilizou-se o mesmo método de lise alcalina, porém com o auxílio do kit de extração de DNA plasmidial, kit Miniprep Eppendorf, de acordo com o protocolo do fabricante.

3.15 Seleção dos clones transformados

A seleção dos clones transformados foi feita através da análise de restrição do DNA plasmidial com as enzimas adequadas, utilizando as especificações do fabricante de cada enzima.

3.16 Seqüenciamento de DNA.

A reação de PCR, para seqüenciamento de clones, foi feita utilizando-se o kit Big Dye Terminator (Perkin Elmer) e foi conduzida em termociclador Mastercycler Personal (Eppendorf). O seqüenciamento foi feito em seqüenciador automático 377 ABI Prism (Perkin Elmer).

3.17 Expressão de proteínas

O sistema de expressão utilizado para a produção das proteínas recombinantes é o sistema pET descrito por Studier *et al.* (1990). Bactérias *E. coli* da linhagem BL21(DE3), transformadas com pET28a contendo o cDNA que codifica a proteína de interesse, foram crescidas em meio de cultura LB contendo 30 μ g/mL de Kanamicina, sob agitação constante de 200 rpm a 37 °C (incubadora refrigerada com agitação, TECNAL TE-421). Este crescimento foi monitorado até que a DO_{600 nm} atingisse 0,7 - 0,8. Atingida esta fase de crescimento, as culturas de células BL21(DE3) eram induzidas à expressão com a adição de IPTG (isopropiltiol- β -D-galactosídeo) na concentração final de 0,5 mM. Para determinar o melhor tempo de indução para cada proteína recombinante, 50 mL de cultura de células foram induzidos

nas condições descritas acima e alíquotas de 1 mL foram retiradas a cada hora. Estas alíquotas foram preparadas e avaliadas por SDS-PAGE (Laemmli, 1970), como descrito no item 3.20. As culturas foram centrifugadas a 2.590 x g por 10 min e a 4 °C, e o sedimento foi armazenado à -20 °C.

3.18 Lise de bactérias

As bactérias que expressaram as proteínas recombinantes foram lisadas por pulsos com auxílio do sonicador Sonifier 450 (Branson) em tampão de lise contendo Tris-HCl 50 mM pH 8,0, KCl 500 mM e EDTA 1 mM (15 a 20 mL/L de cultura) após 30 minutos de incubação em gelo com 5U de DNase (GIBCO BRL) e 30 µg/mL de lisozima (Sigma), a lise foi centrifugada a aproximadamente 18.000 x g, por 30 min, a 4 °C. O sobrenadante foi filtrado em filtro Millipore 0,45 µm.

3.19 Tampões

Os tampões empregados nas diversas purificações e experimentos receberam as seguintes siglas:

1) TNB500 (Tampão da Hsp100): Tris-HCl 25 mM pH 7,5, NaCl 500 mM, β-mercaptoetanol 1mM.

2) TNB250 (Tampão da Hsp100): Tris-HCl 25 mM pH 7,5, NaCl 250 mM, β-mercaptoetanol 1mM.

3) TN150 (Tampão da Hsp82): Tris-HCl 50 mM pH 7,5, NaCl 150 mM.

4) HEP500 (Tampão da p23): Hepes 40 mM pH 7,5, NaCl 500 mM, β-mercaptoetanol 5 mM.

5) HEP100 (Tampão da p23): Hepes 40 mM pH 7,5, NaCl 100 mM, β-mercaptoetanol 5 mM.

6) TNB100 (Tampão da p23): Tris-HCl 25 mM pH 7,5, NaCl 100 mM, β-mercaptoetanol 5 mM.

3.20 Purificação das proteínas recombinantes

3.20.1 Cromatografia de afinidade

Para a purificação das proteínas recombinantes fusionadas a uma cauda de poli-histidina, foi utilizada a coluna de afinidade *HiTrap Chelanting* 5 mL (Ammershan Pharmacia Biotech) previamente carregada com níquel – de acordo com o protocolo do fabricante. O processo de purificação foi feito com a coluna de afinidade acoplada ao sistema de FPLC ('Fast Protein Liquid Chromatography') *ÄKTA FPLC* (Amershan Pharmacia Biotech). O protocolo básico de purificação foi feito da seguinte forma: equilíbrio da coluna com 3 VC (volume de coluna) de tampão – fosfato de sódio 25 mM, NaCl 500 mM, imidazol 20 mM e glicerol 1% pH 7,4 (tampão SB – *start buffer*) – seguida pela injeção da amostra e eluição da fração não ligante com 5 VC. Após estes passos, a amostra foi eluída em um gradiente de 20 VC entre o tampão SB e o tampão fosfato de sódio 25 mM, NaCl 500 mM, imidazol 500 mM e glicerol 1% pH 7,4 (tampão com softa de sódio 25 mM, NaCl 500 mM, imidazol 500 mM e glicerol 1% pH 7,4 (tampão a softa de sódio 25 mM, NaCl 500 mM e glicerol 1% pH 7,4 (tampão a softa de sódio 25 mM, NaCl 500 mM e glicerol 1% pH 7,4 (tampão a softa de sódio 25 mM, NaCl 500 mM e glicerol 1% pH 7,4 (tampão EB – *elution buffer*). Todo o processo foi realizado com monitoramento constante da absorbância a 280 nm, da temperatura e da condutância.

3.20.2 Cromatografia de exclusão molecular

Foi utilizada uma coluna HiLoad Superdex 200 pg 16/60 (Amershan Pharmacia Biotech), acoplada ao sistema de FPLC ('Fast Protein Liquid Chromatography') *ÄKTA FPLC* (Amershan Pharmacia Biotech). Todos os tampões utilizados possuíam pelo menos 150 mM de NaCl ou outro sal para prevenir interações inespecíficas com a resina, o que afetaria o volume/tempo de eluição das mesmas. Detalhes específicos quanto à concentração de sal utilizada em cada purificação se encontram nas legendas das figuras.

3.20.3 Cromatografia de troca iônica

A purificação da proteína Hsp82 envolveu uma etapa de cromatografia de troca iônica em uma coluna DEAE de 50 mL (Amershan Pharmacia Biotech), acoplada ao sistema *ÄKTA FPLC* (Amershan). A proteína, purificada por afinidade, foi extensivamente dialisada no tampão Tris-Cl 50 mM pH 7,5 e posteriormente submetida à purificação em um gradiente entre o tampão da diálise e o tampão Tris-Cl

50 mM pH 7,5, NaCl 1 M. O conteúdo das frações eluídas no gradiente foi visualizado por SDS-PAGE (ver item 3.20). A eluição foi realizada em um fluxo médio de 3 mL/minuto e frações entre 8 e 10 mL.

3.21 Eletroforese em gel de poliacrilamida (SDS-PAGE)

As amostras coletadas em cada passo do processo de indução e de purificação foram diluídas em tampão de amostra para eletroforese (Tris-HCl 50 mM pH 6,8; DTT 100 mM; SDS 2%; azul de bromofenol 0,1% e glicerol 10%), fervidas durante 5 min e centrifugadas por 2 min em microcentrífuga Eppendorf à temperatura ambiente. Estas amostras foram então submetidas, juntamente com um padrão de massa molecular, à eletroforese em SDS-PAGE (Laemmli, 1970). A eletroforese foi conduzida com a utilização do sistema *Mini-Protean II Dual Slab Cell* (Bio-Rad), com gel de empacotamento 5%, gel de separação 10% - 15% (dependendo da proteína), e voltagem constante de 200 V por aproximadamente 1 hora. O gel foi corado por 30 min em etanol:ácido acético:água 5:1:15 (v:v:v) e Coomassie Brilliant Blue R (Bio-Rad) 0,25% e descorado em ácido acético:etanol:água 3:2:35 (v:v:v).

3.22 Determinação da concentração de proteínas

A concentração de proteínas foi determinada espectroscopicamente através do método descrito primeiramente por Edelhock (1967), utilizando os seguintes comprimentos de onda: 282, 280, 279, 278 e 276 nm (valores de absorbância). As amostras foram diluídas em tampão (pH 6,5) contendo fosfato de sódio na concentração final de 2 mM e guanidina-HCl 6 M.

Os coeficientes de absorbância molar foram calculados utilizando a seguinte fórmula:

$$\epsilon(\lambda)(M^{-1}cm^{-1}) = n_{Trp} x \epsilon_{Trp} + n_{Tyr} x \epsilon_{Tyr} + n_{Cys} x \epsilon_{Cys}$$

Na qual, $\varepsilon(\lambda)$ é o coeficiente de absorbância molar (M⁻¹cm⁻¹) da proteína de interesse em um determinado comprimento de onda; (λ); n_{Trp} – o número de triptofanos da proteína; n_{Tyr} – o número de tirosinas e n_{Cys} o número de cisteínas; ε_{Trp} , ε_{Tyr} e ε_{Cys} são os coeficientes de absorbância molar do triptofano, da tirosina e da cisteína, respectivamente. Os valores dos coeficientes de absorbância molar destes aminoácidos variam de acordo com o comprimento de onda (λ) e foram calculados com o auxílio

do programa ProtParam Tool (http://ca.expasy.org/tools/protparam.html) de acordo com Gill *et al.* (1989). Os valores de $\varepsilon(\lambda)$ foram determinados a partir da seguinte equação (Beer–Lambert):

A=ɛ./.C

Na qual, A é a absorbância; *I*, comprimento do caminho ótico em cm e C, concentração molar (M). A concentração molar determinada corresponde à média das concentrações encontradas para cada comprimento de onda entre 276 e 282 nm. As leituras de absorbância, para cada comprimento de onda relacionado, foram realizadas em triplicada em um espectrofotômetro Jasco V530 UV/VIS.

3.23 Espectros de dicroísmo circular (CD)

Os espectros de CD foram feitos utilizando-se um espectropolarímetro J-810 (JASCO), cubetas de quartzo com caminho ótico entre 1 cm e 0,1 cm e proteínas diluídas de forma a fornecer um sinal entre - 20 e - 40 miligraus a 222 nm. Os programas utilizados para o registro e o tratamento dos dados foram Spectra manager[®] (JASCO) e Origin[®] versões 6.1 e 7.5 (OriginLAB Corporation). A temperatura da cubeta foi mantida constante a 20°C através de um sistema interno de controle (Peltier Type Control System PFD 425S- Jasco) e controlador externo de temperatura NESLAB RTE Series (NESLAB). Os valores obtidos na leitura de CD (miligraus) foram convertidos para elipcidade molar residual (EMR) que é definida pela seguinte equação:

$EMR = (\theta \times MM)/(c \times I \times n)$

Na qual "θ" é a elipcidade em miligraus, "MM" é a massa molecular da proteína (g/mol), "c" é a concentração da proteína (mg/mL), "l" é o comprimento do caminho ótico (cm) e "n" é o número de resíduos de aminoácidos da respectiva proteína. A porcentagem de estrutura secundária tipo hélice alfa, folha beta-pregueada e estrutura randômica foram calculadas com o auxílio do programa CDNN Deconvolution (G. Böhm *et al.*, 1992).

No caso de desenovelamento, os sinais de CD e fluorescência também podem ser convertidos para valores de variação de sinal, através da equação:

$$\Delta \text{ sinal} = (S_{\text{Obs}} - S_{\text{D}})/(S_{\text{N}} - S_{\text{D}})$$

51

Na qual S_{Obs} é o sinal observado em uma dada concentração de agente desnaturante, S_D o sinal correspondente à proteína desenovelada, assumido como o de maior concentração de agente desnaturante. S_N é o sinal correspondente à proteína nativa, assumido como a condição sem agente desnaturante.

3.24 Emissão de fluorescência

Os experimentos de emissão de fluorescência foram realizados utilizando-se um fluorímetro Aminco Bowman® Series 2 (SLM-AMINCO), cubetas de quartzo com caminho ótico entre 1 cm e 0,1 cm e proteínas em concentrações entre 2 e 8 μ M. Os programas utilizados para registro dos dados foram SLMAB2 (SLM-AMINCO) e Origin[®] versões 6.1 e 7.5 (OriginLAB Corporation). A temperatura da cubeta foi mantida constante à 20°C através de um controlador de temperatura digital Polyscience, quando assim for indicado. Os espectros de emissão fluorescência foram monitorados utilizando-se comprimento de onda (λ) de excitação a 295 nm, com a abertura para passagem de luz de 2 nm e de emissão a 330 nm (abertura 4 nm). Os pontos para a composição do espectro de emissão foram coletados a cada 1 nm dentro do intervalo de 305 a 420 nm.

Os dados de intensidade de fluorescência em função da concentração de agente desnaturante foram convertidos para valores de centro de massa espectral (λ) segundo a equação:

$$\langle \lambda \rangle = \sum \lambda i F i / \sum F i$$

Onde λi representa cada comprimento de onda utilizado e *Fi* a intensidade de fluorescência em λi .

3.25 Determinação da concentração de nucleotídeos

A concentração de nucleotídeos adenosina foi determinada, assim como a concentração das proteínas, por técnicas espectrométricas utilizando a equação de Beer-Lambert:

52

Os nucleotídeos ADP e ATP, ambos sais de sódio, foram de proveniência Sigma. Os nucleotídeos foram diluídos até uma leitura entre 0,5 e 1,0 em λ de 259 nm. Para isto, foi utilizado o ε (M^{-1 ·} cm⁻¹) de 15.400 para nucleotídeos adenosina, obedecendo a correlação absorbância 280/260 nm de 0,16 (Sambrook *et al.*, 1989), após subtração do respectivo branco. A equação de Beer-Lambert foi resolvida e corrigida pelo fator de diluição para a determinação da concentração de nucleotídeos.

3.26 Espalhamento dinâmico de luz

Os dados referentes ao coeficiente de difusão (valor D) foram estimados por experimentos de espalhamento de luz dinâmico (EDL), os quais foram realizados em um equipamento *DynaPro-MS/X* (Protein Solutions) em temperatura constante de 20 °C. As amostras foram previamente centrifugadas a 14000 x g por 10 minutos a 20 °C. A função de auto-correlação foi medida em um ângulo de espalhamento de 90 graus, com aprox. 300 - 400 acumulações de 10 segundos cada e analisados pelo programa fornecido pelo fabricante. O EDL também foi utilizado como forma de controlar a monodispersidade das amostras utilizadas em praticamente todos os experimentos realizados. Os dados brutos do coeficiente de difusão foram convertidos para valores de D_{20,W} através da seguinte equação:

$$D_{20,W} = [D_{obs} \cdot (\eta_A / \eta_{H20})]/10$$

Na qual D_{obs} é o dado observado, η_A a viscosidade da amostra e η_{H20} a viscosidade da água a 20 °C. A unidade de viscosidade utilizada foi *Pa*·*s* (Pascal·segundo). A partir dos dados de $D_{20,W}$ em função da concentração de proteína, foi possível determinar o valor de $D_{20,W}^0$ utilizando a equação:

$$D_{20,W}^{0} = D_{20,W} - K_{ap} \cdot c$$

Na qual K_{ap} é o valor da constante de equilíbrio aparente e c a concentração de proteína em mg/ml.

3.27 Ensaio de ligação baseado em resina de purificação (tipo "pull-down")

A proteína Luciferase a 2 μM foi desnaturada termicamente pela incubação a 42 °C por 10 minutos na ausência e na presença da proteína Hsp100 e resfriada a temperatura ambiente. A solução foi incubada então com 50 μL de resina Talon Metal Chelate[®] (Clontech) em 50% de tampão Tris-Cl 25 mM pH 7,5, NaCl 250 mM, β-mercaptoetanol 1 mM por 1 hora na temperatura ambiente. A mistura foi centrifugada por 1 minuto a 10.000 x g (4 °C), e o precipitado contendo a resina foi lavado duas vezes no tampão Tris-Cl 25 mM pH 7,5, NaCl 250 mM, β-mercaptoetanol 1 mM contendo 15 mM de imidazol. As proteínas ligadas à resina foram eluídas na presença no tampão Tris-Cl 25 mM pH 7,5, NaCl 250 mM, β-mercaptoetanol 1 mM contendo 15 mM de imidazol. As proteínas ligadas à resina foram eluídas na presença no tampão Tris-Cl 25 mM pH 7,5, NaCl 250 mM, β-mercaptoetanol 1 mM contendo a penas as proteínas eluídas foi concentrada através de precipitação com acetona 80% (Merck) e visualizados por SDS-PAGE. O protocolo de ligação de complexos protéicos à resina Talon[®] foi adaptado de Borges *et al.* (2005).

3.28 Modelagem molecular por homologia

A modelagem por homologia das proteínas de interesse foi realizada através do programa Modeller (versão 7v7), desenvolvido pelo grupo do professor Andrej Sali (*Department of Biopharmaceutical Sciences – UCLA*). Este software utiliza a modelagem comparativa por homologia de estruturas tridimensionais protéicas, através da satisfação de restrições espaciais. Para as modelagens foram utilizados alinhamentos entre as proteínas-alvo e seqüências de proteínas homólogas com estrutura em alta resolução conhecidas. Os alinhamentos gerados pelo programa MULTALIN *multiple alignment* (<u>http://npsa-pbil.ibcp.fr/cgi-bin/align multalin.pl</u>), foram corrigidos manualmente para eliminar os "gaps" entre as seqüências utilizadas, e também formatados para que fossem reconhecidos pelo programa Modeller. Os modelos gerados foram avaliados pelos programas PROCHECK (<u>http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html</u>) e WHATIF (<u>http://swift.cmbi.kun.nl</u> /WIWWWI/), e visualizados através dos programas Sybyl (Tripos), Weblab e Rasmol.

3.29 Ultracentrifugação analítica (AUC)

Os experimentos de AUC foram realizados em uma ultracentrífuga *Beckman Optima XL-A* (Beckman Coulter), em temperatura constante de 20 °C, utilizando o rotor AN-60Ti. A varredura da absorbância em função do raio foi realizada em comprimentos de onda entre 230 e 280 nm dependendo da concentração de proteína e da composição do tampão.

3.29.1 Velocidade de sedimentação

Os experimentos de velocidade de sedimentação foram realizados em velocidades apropriadas para a sedimentação de cada proteína e estão descritos em detalhe no item "Resultados". As interpretações dos dados de absorbância versus o raio da célula foram realizadas através de ajustes de regressão não-linear pelo programa Origin[®] (Microcal). Os métodos de análise de dados Second Moment (Goldberg, 1953) e Sedimentation Time Derivative (Stafford, 1994) foram utilizadas para analisar os dados de velocidade de sedimentação. O cálculo do valor do s é importante para caracterizar mudanças no tamanho e na forma das macromoléculas causadas por mudanças nas condições do meio. Porém, o valor obtido pelos programas, o s*, não leva em consideração as contribuições do solvente e da temperatura nos cálculos executados pelos programas descritos acima. Assim, o valor de s* calculado deve ser sempre extrapolado das condições observadas para as condições padrões (água e temperatura de 20°C), ou S_{20,w} (Laue, 2001). Este valor pode ser calculado pela seguinte relação:

 $s_{0,W} = s_{,B} \cdot (\eta_{T,B}/\eta_{20}) \cdot [(1 - \overline{\upsilon} \cdot \rho) 20/(1 - \overline{\upsilon} \cdot \rho)_{T,B}]$

Na qual, $S_{T,B}$ é o valor obtido experimentalmente (*T* para *temperature* e *B* para *buffer*), $\eta_{T,B}$ a viscosidade do tampão da amostra, η_{20} a viscosidade da água a 20 °C, \overline{v} o volume parcial específico (ml/g), ρ_{20} a densidade da água a 20 °C e $\rho_{T,B}$ a densidade do tampão de amostra também nesta temperatura. Esta extrapolação elimina as contribuições do solvente e temperatura no valor de s* que não foram considerados pelos cálculos iniciais determinados nos métodos de análise, como discutido acima. O programa Sednterp (www.jphilo.mailway.com /download.htm) foi utilizado para realizar a operação descrita pela equação acima e também para estimar os seguintes valores: volume parcial específico a 20°C (V_{bar}), calculado a partir da seqüência primária de aminoácidos da proteína, viscosidade e

densidade do tampão. O valor de S_{20,w} pode ser medido em várias concentrações e a extrapolação da curva para o valor zero permite a obtenção de S⁰_{20,w}, o qual representa uma propriedade única da macromolécula sem os efeitos dependentes da concentração da proteína na solução (Laue, 2001). Assim, mudanças observadas no valor de S⁰_{20,w} ocasionadas por diferentes condições experimentais tais como temperatura, pH, adição de ligantes e outros, etc podem ocorrer somente em decorrência de variações na massa ou de mudanças conformacionais da macromolécula (Laue, 2001).

3.29.2 Sedimentação em equilíbrio

Os experimentos de sedimentação em equilíbrio também foram realizados em pelo menos três velocidades diferentes. As amostras foram aceleradas a uma velocidade inicial constante por aproximadamente 16 horas e, através de leituras sucessivas, foi verificado se o estado de equilíbrio fora alcançado (detalhes indicados nas legendas). Após a obtenção do equilíbrio, foram coletados aproximadamente 25 cumulativas por amostra. O método de *Self-Association* foi utilizado para analisar os dados de sedimentação em equilíbrio utilizando diferentes modelos de associação para cada proteína. A distribuição de proteína pela célula foi interpretada pela seguinte equação (Johnson, 1981):

$$C = C_0 e \left[\frac{MM \left(1 - V_{bar} \rho \right) \omega^2 \left(r^2 - r_0 \right)}{2RT} \right]$$

Na qual *C* é a concentração de proteína em uma determinada posição radial r; C_0 é a concentração de proteína na posição radial r_0 ; ω é a velocidade angular do rotor e *e* representa uma função exponencial. As análises foram feitas com um conjunto mínimo de 3 a 9 curvas oriundas de experimentos me pelo menos 3 velocidades diferentes em uma mesma concentração de proteína. Isto foi feito utilizando três concentrações distintas de cada proteína e o resultado final dado pela média dos resultados. O programa Sednterp foi utilizado para estimar os valores de V_{bar} e densidade para a realização dos cálculos.

4 Hsp100 de cana-de-açúcar

4.1 Resultados Hsp100 de cana-de-açúcar

4.1.1 Clonagem, expressão e purificação

O seqüenciamento completo deste cDNA foi através da utilização de oligonucleotídeos universais, que reconhecem seqüencias do vetor, e também de oligonucleotídeos específicos, desenhados para reconhecer regiões internas da seqüência. Na figura 7 temos a seqüência de nucleotídeos que codifica esta proteína e também sua seqüência de aminoácidos propriamente dita.

GAGTACCGCAAGTACGTGGAGAAGGACGCGGCGGTTCGAGCGGCGGTTCCAGCAGGTGTTCGTCGCGGAGCCGAGCGTGCCCGACACGATCAGCATCTGAGGGGGGCTCAAGGAGAAGTA AGCGCCGCGAGGAGCTGCAGGTTCACCCTGCAGGAGGCCGAGCGCCGGATGGACCTGGCCCGTGTGGCCCGATCTCAAGTACGGCGCCCCTCCAGGAGATCGACGCCGCAATCGCCAAGCTGG AGAGCGAAACAGGGGAGAACCTGATGCTCACTGAAACCGTCGGCCCTGAACAAATCGCAGAGGTGGTGAGCCGGTGGACAGGTATTCCGGTGACCCGGCTTGGCCAGAACGACAAGGAGA GGCTGGTCGGCCTGGCTGACAGGCTTCACCAGAGGGTGGTCGGCCAAACGGAGGCGGTGAACGCCGTCGCAGAGGCGGTGCTGAGGTCAAAGGCCGGTCTTGGCAGGCCACAGCAGTCC ATGAGATCGTGATCTTCGATCCTCTGTCCCACGAGCAACTGAGGAAGGTGGCTCGCCTTCAGATGAAGGACGTCGCCGTCCGCCGTCCGCGAGAGGGGGCATCGCCTCTGGCCGTGACCGACGCC MGSSHHHHHHSSGLVPRGSHMNPDNFTHKTNEALVAAHEAASEAGHAQLTPLHLAAALVADKGGILRQAITGASGGDGAAGDSFERVLSKALKKLPSQSPPPDSVPA STALIKAIRRAQSAQKKRGDSHLAVDQLLLGLLEDSQISDCLKEAGVSAARVRAELEKLRGGGGRRVESASGDTNFQALKTYGRDLVEQAGKLDPVIGRDEEIRRVVRILS <u>RRTKNNPVLIGEPGVGKTAVVEGLAQRIVRGDVPSNLLDVRLIALDMGALVAGAKYRGEFEERLKSVLKEVEEAEGKVILFIDEIHLVLGAGRTEGSMDAANLFKPMLAR</u> **GQLRCIGATTLEEYRKYVEKDAAFERRFQQVFVAEPSVPDTISILRGLKEKYEGHHGVRIQDRALVVAAQLSARYIMGRHLPDKAIDLVDEACANVRVQLDSQPEEIDNLE** RKRIOLEVELHALEKEKDKASKARLVEVKKELDDLRDKLOPLTMKYRKEKERIDEIRKLKORREELOFTLOEAERRMDLARVADLKYGALOEIDAAIAKLESETGENLMLTE TVGPEQIAEVVSRWTGIPVTRLGQNDKERLVGLADRLHQRV<u>VGQTEAVNAVAEAVLRSKAGLGRPQQSTGSFLFLGPTGVGKTELAKAFAEQLFDDENLLVRIDMSEY</u> <u>QEVRRHFRPELLNRLDEIVIFDP</u>LSHEQLRKVARLQMKDVAVRLAERGIALAVTDAALDIILSLSYDPVYGARPIRRWIEKRVVTQLSKMLIQEEIDENCTVYIDAAPAKDEL AYRVDRSGGLVNAETGLKSDILIQVPNDAVRSDAAQAVKKMRIMEEDEDGMDEE

Figura 7: Dados referentes às seqüências da Hsp100 – Histag. A seqüência de cDNA da proteína Hsp100 é composta de 2739 nucleotídeos, que codificam 912 aminoácidos. A seqüência destacada em azul corresponde à cauda de poli-Histidinas e outros aminoácidos codificados pelo sítio de clonagem múltipla do vetor pET-28a. A proteína Hsp100 com a cauda poli-His possui uma massa molecular de 103,1 kDa. Os

dois resíduos de triptofano estão destacados em vermelho e os dois domínios de ligação aos nucleotídeos estão sublinhados.

A figura 8 mostra o alinhamento da seqüência da proteína Hsp100 de cana-de-açúcar e as seguintes proteínas: Hsp101 de *Arabdopisis thaliana* (gi:18410584), ClpB de *Thermus thermophilus* (gi:38492937) e Hsp104 de *Saccharomyces cerevisiae* (gi:6323002).

SG_Hsp100	1	MNPO-NFTHKTNERALVAAHEAASEAGHAQLTPLHLAAALVADKGGILRQAITGASGGDGAAGDSFERVLSKALKKLPSQSPPPDSVPASTALIKAIRRAQ	100
AT_Hsp101	1	MNPE-KFTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDFGIFPQAISSAGGENAAQSAERVINQALKKLPSQSPPPDDIPASSALIKAIRRAQ	100
TT_C1pB	1	MNLE-RHTQAAREALAQAQVLAQRKHQAIDLPHLWAVLLKDERSLAWRLLEKAGADPKALKELQERELARLPKVEGAEVGQYLTSRLSGALNRAE	100
SC_Hsp104	1	MNDQTQFTERALTILTLAQKLASDHQHPQLQPIHILAAFIETPEDGSVPYLQNLIEKGRYDYDLFKKVVNRNLVRIPQQOPAPAEITPSYALGKVLQDAA	100
SG_Hsp100	101	SAQKKRGDSHLAVDQLLLGLLEDSQISDCLKEAGVSAARVRAELEKLRGGGGRRVESASGDINFQALKTYGRDLVEQAGKLDPVIGRDEEIRRVVR	200
AT_Hsp101	101	AAQKSRGDTHLAVDQLIMGLLEDSQIRDLLNEVGVATARVKSEVEKLRGKEGKKVESASGDINFQALKTYGRDLVEQAGKLDPVIGRDEEIRRVVR	200
TT_ClpB	101	GLMEELKDRVVAVDILVLALAEATFGLFGLEALKGALKELRGGKTVQIEHAESIYNALEQYGIDLIRLAAEGKLDPVIGRDEEIRRVIQ	200
SC_Hsp104	101	KIQKQQKDSFIAQDHILFALFNDSSIQQIFKEAQVDIEAIKQQALELRGNTRIDSRGADTNTPLEYLSKYAIDMTEQARQGKLDPVIGREEEIRSIIR	200
SG_Hsp100	201	ILSRRTKNNPVLIGEPGVGKTAVVEGLAQRIVRGDVPSNLLDVRLIALDMGALVAGAKYRGEFEERLKSVLKEVEEAEGKVILFIDEIHLVLGAGRTEGS	300
AT_Hsp101	201	ILSRRTKNNPVLIGEPGVGKTAVVEGLAQRIVKGDVPNSLTDVRLISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVILFIDEIHLVLGAGKTEGS	300
TT_ClpB	201	ILLRRTKNNPVLIGEPGVGKTAIVEGLAQRIVKGDVPEGLKGKRIVSLQMGSLLAGAKYRGEFEERLKAVIQEVVQSQGEVILFIDELHTVVGAGKAEGA	300
SC_Hsp104	201	VLARRIKSNPCLIGEPGIGKTAIIEGVAQRIIDDDVPTILQGAKLFSLDLAALTAGAKYKGDFEERFKGVLKEIEESKTLIVLFIDEIHMLMGNGKDD	300
SG_Hsp100	301	MDAANLFKPMLARGQLRCIGATTLEEYRKYVEKDAAFERRFQQVFVAEPSVPDTISILRGLKEKYEGHHGVRIQDRALVVAAQLSARYIMGRHLPDKAID	400
AT_Hsp101	301	MDAANLFKPMLARGQLRCIGATTLEEYRKYVEKDAAFERRFQQVYVAEPSVPDTISILRGLKEKYEGHHGVRIQDRALINAAQLSARYITGRHLPDKAID	400
TT_C1pB	301	VDACNMLKPALARGELRLIGATTLDEYR-EIEKOPALERRFQEVYVDEPTVEETISILRGLKEKYEVHHGVRISDSAIIAAATLSHRYITERRLEDKAID	400
SC_Hsp104	301	AANILKPALSRGQLKVIGATTNNEYRSIVEKDGAFERRFQKIEVAEPSVRQTVAILRGLQPKYEIHHGVRIDSALVTAAQLAKRYLPYRRLPDSALD	400
SG_Hsp100	401	LVDEACANVRVQLDSQPEEIDNLERKRIQLEVELHALEKEKDKASKARLVEVKKELDDLRDKLQPLTMKYRKEKERIDEIRKLKQRREELQFTLQEAE	500
AT_Hsp101	401	LVDEACANVRVQLDSQPEEIDNLERKRMQLEIELHALEREKDKASKARLIEVRKELDDLRDKLQPLTMKYRKEKERIDEIRRLKQKREELMFSLQEAE	500
TT_ClpB	401	LIDEAAARLRMALESAPEEIDALERKKLQLEIEREALKKEKDPDSQERLKAIEAEIAKLIEEIAKLREWEREREILKKLREAQHRLDEVRREIELAE	500
SC_Hsp104	401	LVDISCAGVAVARDSKPEELDSKERQLQLIQVEIKALERDEDADSTIKDRLKLARQKEASLQEELEPLRQRYNEEKHGHEELTQAKKKLDELENKALDAE	500
SG_Hsp100	501	RRMDLARVADLKYGALQEIDAAIAKLESETGENLMLTETVGPEQIAEVVSRWTGIPVTRLGQNDKERLVGLADRLHQRVVGQTEAVNAVAEAV	600
AT_Hsp101	501	RRYDLARAADLRYGAIQEVESAIAQLEGTSSEENVMLTENVGPEHIAEVVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVGQNQAVNAVSEAI	600
TT_ClpB	501	RQYDLNRAAELRYGELPKLEAEVEALSEKLRGARFVRLEVTEEDIAEIVSRWTGIPVSKLLEGEREKLLRLEEELHKRVVGQDEAIRAVADAI	600
SC_Hsp104	501	RRYDTATAADLRYFAIPDIKKQIEKLEDQVAEEERRAGANSMIQNVVDSDTISETAARLTGIPVKKLSESENEKLIHMERDLSSEVVGQMDAIKAVSNAV	600
SG_Hsp100	601	LRSKAGLGRPQQSTGSFLFLGPTGVGKTELAKAFAEQLFDDENLLVRIDMSEYMEQHSVARLIGAPPGYVGHEEGGQLTEQVRRPYSVILFDEVEKAHV	700
AT_Hsp101	601	LRSRAGLGRPQQPTGSFLFLGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVSRLIGAPPGYVGHEEGGQLTEAVRRPYSVILFDEVEKAHV	700
TT_ClpB	601	RRARAGLKDPNRFIGSFLFLGPTGVGKTELAKTLAATLFDTEEAMIRIDMTEYMEKHAVSRLIGAPPGYVGYEEGGQLTEAVRRPYSVILFDEIEKAHP	700
SC_Hsp104	601	RLSRSGLANPRQP-ASFLFLGLSGSGKTELAKKVAGFLFNDEDMMIRVDCSELSEKYAVSKLLGTTAGYVGYDEGGFLTNQLQYKPYSVLLFDEVEKAHP	700
SG_Hsp100	701	AVFNTLLQVLDDGRLTDGQGRTVDFRNTVIIMTSNLGAEHLLAGMVGKNSMKVARDLVMQEVRRHFRPELLNRLDEIVIFDPLSHEQLRKVARLQMKDVA	800
AT_Hsp101	701	AVFNTLLQVLDDGRLTDGQGRTVDFRNSVIIMTSNLGAEHLLAGLTGKVTMEVARDCVMREVRKHFRPELLNRLDEIVVFDPLSHDQLRKVARLQMKDVA	800
TT_ClpB	701	DVFNILLQILDDGRLTDSHGRTVDFRNTVIILTSNLGAEHLLAGLTGKVTMEVARDEVFKVLQQHFRPEFLNRLDEIVVFRPLTKEQIRQIVEIQLSVLR	800
SC_Hsp104	701	DVLTVMLQMLDDGRITSGQGKTIDCSNCIVIMTSNLGAEFINS-QQGSKIQESTKNLVMGAVRQHFRPEFLNRLSSIVIFNKLSRKAIHKIVDIRLKEIE	800
SG_Hsp100	801	VRLAERGIALAVTDAALDIILSLSYDPVYGARPIRRWIEKRVVTQLSKMLIQEEIDENCTVYIDAAPAKDELAYRVDRSGGLVNAETGLKSDILIQVP	900
AT_Hsp101	801	VRLAERGVALAVTDAALDYILAESYDPVYGARPIRRWMEKKVVTELSKMVVREEIDENSTVYIDAGAGDLVYRVE-SGGLVDASTGKKSDVLIHIA	900
TT_ClpB	801	ARLAEKRISLELTEAAKDFLAERGYDPVFGARPLRRVIQRELETPLAQKILAGEVKEGDRVQVDVGPAGLVFAVP	900
SC_Hsp104	801	ERFEQNDKHYKLNLTQEAKDFLAKYGYSDDMGARPLNRLIQNEILNKLALRILKNEIKDKETVNVVLK-KGKSRDENVPEAAECLEVLP	900
SG_Hsp100 AT_Hsp101 TT_ClpB SC_Hsp104	901 901 901 901	NDAVRSDAAQAVKKMRIMEEDEDGMDEE 931 NGPKRSDAAQAVKKMRIEEIEDDDNEEMIED 931 ARVEA	

Figura 8: Alinhamento entre a Hsp100 de cana-de-açúcar e proteínas homólogas. Entre os resíduos, 31 % são idênticos (vermelho), 22 % são muito¹ similares (verde), 8 % são pouco² similares (azul) e 39 % são diferentes (preto). Entre as seqüências alinhadas temos: *"SG_Hsp100"*: Hsp100 de cana-de-açúcar (912 resíduos); *"AT_Hsp101"*: Hsp101 de *A. thaliana* (911 resíduos); *"TT_ClpB"*: ClpB de *T. thermophilus* (854 resíduos) e *"SC_Hsp104"*: Hsp104 de *S. cerevisiae* (908 resíduos). Nota-se que as regiões mais

conservadas ficam restritas aos dois domínios de ligação a nucleotídeos (*NBDs – nucleotide binding domains: resíduos Q164 a E413 e Q555 a E843*), fundamentais para a oligomerização e atividade ATPásica destas proteínas. (¹por exemplo: dois aminoácidos carregados positivamente; ²por exemplo: dois aminoácidos simplesmente polares)

A expressão da proteína Hsp100 foi realizada utilizando *Escherichia coli,* BL21 (DE3), por um período de 4-5 horas. A figura 9 mostra um gel de poliacrilamida (10 %), contendo as alíquotas retiradas a cada hora após a indução, após a adição do indutor IPTG, e também a fração antes da indução. Como é possível observar no gel se destaca, após algumas horas, uma banda de indução com massa molecular entre 97 e 116 kDa, que corresponde à expressão da proteína Hsp100 de cana-de-açúcar.



Figura 9: Análise da expressão em pequena escala do clone pET28a-Hsp100. Bactérias BL21(DE3), transformadas com o vetor pET28a-Hsp100, foram cultivadas em 50 mL de LB contendo 30 µg/mL de canamicina até uma DO_{600nm} igual a 0,6, sendo acrescido IPTG para uma concentração final de 0,5 mM. Alíquotas de 0,5 mL de cultura foram colhidas no tempo 0 de indução (não-induzido - NI) e a cada hora seguinte até 5 horas. Após centrifugação, o sedimento das alíquotas foi ressuspendido em 0,1 mL de tampão de amostra de proteína, sendo aplicados cerca de 10 µL no gel de poliacrilamida a 10%. A seta

maior indica a banda correspondente à indução da Hsp100. Já as setas menores indicam as proteínas do padrão de massa molecular que possuem 97 e 116 kDa, respectivamente.

Após a indução, procedemos com a lise das bactérias segundo descrito no item "materiais e métodos". Depois da realização da lise, constatamos que a proteína Hsp100 de cana estava presente tanto na fração precipitada quanto na sobrenadante (aproximadamente 50 %), sendo que esta última foi submetida à purificação por cromatografia líquida. Após o equilíbrio da coluna *Hitrap*[®] *Chelating* (Amershan) com o tampão SB e aplicação da amostra, realizamos um passo de lavagem com 15 % do tampão EB, o que equivale a aproximadamente 75 mM de imidazol, por 3 volumes de coluna (VC). Em seguida, realizamos um gradiente de concentração com 20 VC, o qual teve início em 15% e término em 100% do tampão EB, com o intuito de eluir a Hsp100 com cauda de Histidina da coluna. O gel de poliacrilamida a 10 % das frações eluídas na purificação por afinidade está mostrado na Figura 10.



Figura 10: Análise da purificação por afinidade da proteína Hsp100 de cana-de-açúcar por gel de poliacrilamida (10%). A purificação foi realizada através de cromatografia de afinidade por níquel em coluna *HiTrap Chelanting* acoplada ao sistema *ÄKTA FPLC* como descrito acima. P: padrão de massa molecular (BIORAD); fração 1: fração protéica que não se ligou à coluna; fração 2: proteínas eluídas na primeira lavagem com 15 % do tampão EB (75 mM de imidazol); frações 3 a 8: proteínas eluídas durante

o gradiente de eluição após a lavagem inicial. A proteína Hsp100 começou a ser eluída em uma concentração de aproximadamente 107 mM de imidazol, o que representa 21 – 22 % do tampão EB, se encontrando principalmente nas frações 4, 5, e 6. Cada fração corresponde a um volume de 5 mL. A seta indica a banda de 97 kDa do padrão de massa molecular.

O cromatograma da purificação por afinidade da proteína Hsp100 de cana-de-açúcar está mostrado na figura 11. A fração "A" corresponde às proteínas do sobrenadante do lisado que não interagiram com a resina e que, portanto, foram eluídas durante o processo de injeção sem se ligar à coluna. A fração "B" corresponde às proteínas eluídas durante a lavagem com 15 % do tampão EB e a fração "C" as proteínas eluídas no gradiente entre os tampões SB e EB. Como afirmado anteriormente, a proteína Hsp100 foi eluída, predominantemente, na fração C do cromatograma.



Figura 11: Cromatograma de purificação da Hsp100 de cana por cromatografia de afinidade. A) proteínas que não se ligaram à coluna *Hitrap chelating global 5 mL* (Amershan Pharmacia Biotech); B) proteínas que foram eluídas com aproximadamente 75 mM de imidazol; C) proteínas eluídas durante o gradiente de concentração de imidazol*, nas quais está incluída a proteína Hsp100 de cana-de-açúcar

com cauda de Histidinas. (*Cem por cento de imidazol corresponde, neste caso, a uma concentração de 500 mM)

As frações 4, 5 e 6 (Figura 10), que apresentaram a proteína Hsp100, foram incubadas com 5 unidades de fosfatase alcalina por cerca de 16 horas a 4 °C (vide materiais e métodos). Finalizada a etapa de purificação por afinidade, o próximo passo de purificação da Hsp100 de cana-de-açúcar foi a eliminação dos agregados e contaminantes da amostra através de uma cromatografia de exclusão molecular (CEM). A Figura 12 mostra o gel de poliacrilamida (10 %) das amostras purificadas por CEM, representando as frações purificadas. A proteína foi eluída em torno de 72 minutos ou 180 mL após a aplicação da amostra na coluna HiLoad Superdex 200 pg 26/60, correspondendo a fração 2 (F2) do cromatograma (Figura 13).



Figura 12: Análise da purificação por cromatografia de exclusão molecular (CEM) da proteína Hsp100 de cana-de-açúcar por gel de poliacrilamida 10%. A amostra de proteína Hsp100 de cana proveniente da cromatografia de afinidade foi submetida à purificação através de cromatografia de exclusão molecular em coluna *HiLoad Superdex 200 pg 26/60* acoplada ao sistema *ÄKTA FPLC*. P: padrão de massa molecular (BIORAD); AF: amostra proveniente da afinidade aplicada à coluna de exclusão molecular; F1: proteínas eluídas em 95 mL após iniciada a purificação, ou seja, proteínas e/ou agregados grandes que não interagiram com os poros da resina; F2: fração eluída em 140 mL de purificação, da qual a Hsp100 é

majoritária; F3: proteínas eluídas em 170 mL após iniciada a CEM, correspondem a contaminantes de massa molecular intermediária; F4: proteínas eluídas em 185 mL da purificação, correspondem a contaminantes de menor massa molecular (volumes de eluição aproximados). A seta indica a banda de 97 kDa do padrão de massa molecular.



Figura 13: Cromatograma de purificação da Hsp100 de cana por cromatografia de exclusão molecular (CEM). Frações F1, F2, F3 e F4 correspondem às proteínas eluídas em 95, 140, 170 e 185 mL, respectivamente, após o início da purificação por CEM. A proteína Hsp100 foi eluída, em sua maior parte, na fração F2. Também é possível observar, além da eliminação de agregados, a eliminação de contaminantes de menor massa molecular (frações 3 e 4). A purificação por CEM da Hsp100 foi realizada no tampão TNB 500, em um fluxo constante de 2 mL/min.

Após este procedimento de purificação, a amostra da Hsp100 foi concentrada em um concentrador Centricon[®] Plus-20 e sua concentração determinada segundo o método de Edelhock (vide materiais e métodos).

4.1.2 Caracterização por dicroísmo circular da proteína Hsp100

A proteína Hsp100 recombinante de cana-de-açúcar foi analisada através de duas técnicas distintas, dicroísmo circular (CD) e fluorescência. A estrutura secundária da proteína Hsp100 foi observada através de CD, o que permitiu, através da deconvolução do espectro, estimar a quantidade de cada elemento da estrutura secundária da mesma. A curva de elipcidade molar residual por comprimento de onda no UV distante indicou que a proteína se encontra estruturada (figura 14), possuindo um perfil de proteínas ricas em estruturas do tipo hélice-alfa, como predito pelo programa de deconvolução CDNN (Böhm *et al.*, 1992). Ainda segundo esta predição, as estruturas em hélice-alfa compreenderam aproximadamente 54 % da estrutura secundária total da proteína, seguida pelas estruturas randômicas com cerca de 21 %. Os efeitos da concentração de NaCl sobre a estrutura secundária da proteína foram analisados através da medida de duas amostras independentes contendo 250 e 500 mM de NaCl, respectivamente (figura 14). Na tabela I está apresentada a quantidade calculada de cada elemento de estrutura secundária em função da concentração de NaCl.





e temperatura constante de 20°C. As leituras foram feitas utilizando-se cubetas de quartzo de 1 mm de caminho ótico e espectropolarímetro de CD J-810.

Tabela I: Predição da quantidade de estrutura secundária da proteína Hsp100 de cana-de-açúcar (er	ro
de aproximadamente 10 %).	

	250 m	nM NaCl	500 mM NaCl		
Estrutura Secundária	205-260 nm	n 210-260 nm	205-260 nm	210-260 nm	
Hélice-alfa	56 %	53 %	55 %	52 %	
Folha beta antiparalela	4 %	5%	5 %	5 %	
Folha beta paralela	5 %	5 %	5 %	5 %	
Volta beta	14 %	14 %	14 %	14 %	
Randômica	20 %	22 %	21 %	23 %	
Total	99 %	99 %	100 %	99 %	

Os efeitos da adição de nucleotídeos sobre a estrutura secundária da proteína também foram estudados por dicroísmo circular, já que a proteína possui dois sítios de ligação a nucleotídeos por monômero. Utilizamos uma concentração de proteína de 4 µM em tampão TNB500, contendo ADP ou ATP 200 µM (figura 15). Não foi possível, através desta sonda, a detecção de alterações na estrutura secundária da Hsp100 de cana-de-açúcar após a adição dos nucleotídeos.



Figura 15: **Caracterização de estrutura secundária monitorada por dicroísmo circular na região do UVdistante Hsp100.** Foram utilizadas as seguintes condições experimentais: 4,0 μM proteína no tampão TNB500, suprido com ATP ou ADP para uma concentração final de 200 μM, e temperatura constante de 20 °C. As leituras foram realizadas em cubetas de quartzo com 1 mm de caminho ótico e espectropolarímetro de CD J-810.

A Tabela II mostra a deconvolução dos dados de CD da proteína Hsp100 na presença dos nucleotídeos ATP e ADP. A adição destes nucleotídeos praticamente não afetou a distribuição dos elementos de estrutura secundária da Hsp100.

	AI	DP	AT	Р	Sem A	ditivos
Estrutura Secundária	205-260 nm	210-260 nm	205-260 nm	210-260 nm	205-260 nm	210-260 nm
Hélice-alfa	58 %	55 %	54 %	51 %	55 %	52 %
Folha beta antiparalela	4 %	4 %	5 %	5%	5 %	5 %
Folha beta paralela	4 %	5 %	5 %	6 %	5 %	5 %
Volta beta	13 %	13 %	14 %	14 %	14 %	14 %
Randômica	19 %	21 %	21 %	23 %	21 %	23 %
Total	98 %	98 %	99 %	99 %	100 %	99 %

Tabela II: Predição da quantidade de estrutura secundária da proteína Hsp100 de cana-de-açúcar na presença e ausência dos nucleotídeos ATP e ADP (erro de aproximadamente 10 %).

4.1.3 Caracterização por fluorescência da proteína Hsp100

Como mencionado anteriormente (figura 7), a proteína Hsp100 de cana apresenta dois triptofanos, os quais foram utilizados como sonda local da estrutura terciária. Os dois triptofanos estão localizados no segundo domínio de ligação a nucleotídeos, portanto servem apenas para monitorar o enovelamento desta região da Hsp100. A figura 16 apresenta as medidas de emissão de fluorescência para a Hsp100 em quatro condições diferentes: sem aditivos, na presença de 6 M Guanidina-Cl, 1 mM de ATP e 1 mM de ADP. Tanto a condição sem aditivos como a condição desnaturante (6 M Guanidina), apresentaram um máximo de emissão em 352 ± 1 nm, enquanto que na condição com ADP o máximo foi de 350 ± 1 nm e com ATP de 349 ± 1 nm. A tabela III mostra os valores de máximo de emissão, centro de massa espectral e integral para cada uma das condições analisadas.



Figura 16: Espectro de emissão de fluorescência da proteína Hsp100. A proteína Hsp100 a 2 μ M, no tampão TNB500, apresentou um máximo de emissão de 352 nm com e sem o agente desnaturante cloreto de guanidina a 6 M. A adição de 1 mM do nucleotídeo ADP alterou o máximo de emissão para 350 nm e do nucleotídeo ATP para 349 ± 1 nm. As amostras, média de três experimentos, foram excitadas em 295 nm e os dados coletados a cada 1 nm no intervalo de 305 a 420 nm.

Condição	Centro de massa espectral	Integral (320 - 380 nm)	Máximo de emissão (nm)
s/ aditivos	357	424,5	352 ± 1
c/ guanidina	357	182,2	352 ± 1
АТР	357	404,2	349 ± 1
ADP	356	407,2	350 ± 1

Tabela III: Parâmetros de emissão de fluorescência da proteína Hsp100 de cana-de-açúcar (erro de aproximadamente 10 %).

4.1.4 Estabilidade térmica da Hsp100

Os experimentos de estabilidade térmica da Hsp100, monitorados por dicroísmo circular, foram realizados na ausência e presença dos nucleotídeos ATP, ADP e AMP. Estes desenovelamentos visaram observar alguma estabilização da proteína quando da ligação destes nucleotídeos aos respectivos sítios presentes na estrutura da Hsp100. Como pode ser observado na figura 17, o perfil de desenovelamento da proteína nativa foi levemente alterado pela adição dos nucleotídeos. A presença de ATP e AMP parece ter provocado uma pequena estabilização de algum domínio em relação à proteína nativa (s/ aditivos), na faixa de temperatura entre 60 e 70 °C. Já a presença de ADP, nesta mesma faixa de temperatura, parece não ter provocado o mesmo efeito, chegando a apresentar uma quantidade de estrutura significativamente menor mesmo em relação à condição nativa. Porém, em todas as condições testadas a temperatura média de desenovelamento não foi alterada, ficando entre 58 ± 1 °C e 60 ± 1 °C. Em todas as condições analisadas o desenovelamento da proteína Hsp100 se mostrou irreversível (dados não mostrados).



Figura 17: Desenovelamento térmico da proteína Hsp100 acompanhado por dicroísmo circular na presença e na ausência de nucleotídeos. A proteína Hsp100 purificada foi submetida a um desenovelamento térmico na presença dos nucleotídeos ATP, ADP e AMP. Condições: 1 μ M de proteína no tampão TNB500 suprido com 200 μ M de ATP, ADP ou AMP. As leituras foram feitas utilizando-se cubetas de quartzo com 1 mm de caminho ótico e uma velocidade de aquecimento de 1 °C por minuto. Os dados, média de três experimentos independentes, foram coletados em um espectropolarímetro de CD J-810 e tratados com o programa Origin[®].

4.1.5 Espalhamento dinâmico de luz da proteína Hsp100

Utilizamos o espalhamento dinâmico de luz para obter o coeficiente de difusão da Hsp100 em diferentes condições experimentais. O valor *D* está relacionado ao formato e tamanho da molécula, já que moléculas menores se difundem mais rapidamente que as maiores e vice-versa. A figura 18 mostra que a proteína Hsp100 de cana-de-açúcar apresentou um valor de D⁰_{20,W} igual a 2,38 x 10⁻⁷ cm²/s no tampão TNB250 suprido com 1 mM de ATP. O parâmetro D⁰_{20,W} representa o coeficiente de difusão normalizado para a temperatura de 20 °C, água como solvente e concentração protéica extrapolada para "0 mg/mL", evitando distorções causadas pela viscosidade do meio, por exemplo. A adição de outros nucleotídeos não promoveu nenhuma alteração importante no coeficiente de difusão (dados não mostrados).



Figura 18: Caracterização do coeficiente de difusão em função de ligantes e concentração de NaCl por EDL. A proteína Hsp100 em tampão TNB250 suprido com 1mM ATP, foi analisada por espalhamento de luz dinâmico. Os valores aqui mostrados foram coletados pelo programa Dynamics V6[™].
4.1.6 Ultracentrifugação Analítica: Sedimentação em equilíbrio

Os experimentos de sedimentação em equilíbrio (SE) foram realizados em diversas concentrações e velocidades, visando determinar a massa molecular do oligômero da Hsp100. A figura 19 contém os dados do experimento de SE da proteína Hsp100, em três concentrações diferentes, no tampão TNB250 contendo 200 µM de ATP. O ajuste em conjunto destas três condições permitiu inferir a massa molecular da proteína Hsp100 em aproximadamente 616 ± 21 kDa. Estes dados foram analisados com uma versão específica para ultracentrifugação analítica do programa Origin[®], fornecido juntamente com a ultracentrífuga.





C)

Figura 19: Experimentos de SE da proteína Hsp100. A) SE da proteína a 0,5 mg/mL no tampão TNB250 contendo 200µM do nucleotídeo ATP e velocidade de 3000 rpm. B) SE da proteína a 0,75 mg/mL no tampão TNB250 contendo 200µM de ATP e velocidade de 3000 rpm. C) SE da proteína a 1,0 mg/mL no tampão TNB250 contendo 200µM de ATP e velocidade de 3000 rpm. A massa calculada para o oligômero, utilizando estes três conjuntos de dados, foi de 616 ± 21 kDa , o que indica que nas condições analisadas a proteína Hsp100 se encontra na forma hexamérica. As amostras foram mantidas por pelo menos 8 horas na mesma velocidade para garantir que o equilíbrio fosse atingido. Os experimentos foram realizados em uma ultracentrífuga modelo X-LA e em diversas velocidades e concentrações protéicas, sempre a 20 °C.

4.1.7 Ultracentrifugação Analítica: Velocidade de Sedimentação

A proteína Hsp100 foi submetida a experimentos de velocidade de sedimentação visando determinar, principalmente, seu coeficiente de sedimentação. A proteína foi submetida a experimentos de VS, no tampão TNB250 contendo 200 μ M de ATP γ S, em diversas concentrações. O cálculo do valor de S_{20,W} (coeficiente de sedimentação à 20°C e em água) a partir do S experimental em cada concentração de proteína permitiu determinar, através de uma equação linear, o valor de S⁰_{20,W}, ou seja, o valor de S_{20,W} normalizado para uma concentração de proteína igual a zero. Este parâmetro é mais confiável na medida em que os possíveis efeitos de concentração são minimizados. A figura 20 apresenta os experimentos de VS para a proteína Hsp100 e também a análise em função da concentração de proteína contendo a regressão linear utilizada. Os dados de VS foram tratados inicialmente com o programa SEDFIT (http://www.analyticalultracentrifugation.com) e em seguida com o programa Origin[®].



Figura 20: Experimentos de velocidade de sedimentação (VS) da proteína Hsp100. No gráfico inserido, temos a distribuição do coeficiente de sedimentação da proteína Hsp100 a 1 mg/mL no tampão TNB250

suprido com 200 μ M de ATP γ S. Já no gráfico maior temos os valores de S_{20,W} calculados para cinco concentrações distintas, bem como a reta que ajusta os mesmos. O valor de S⁰_{20,W} para a proteína Hsp100 nestas condições foi de 16,2 ± 0,4 S. Todos os experimentos foram realizados em uma ultracentrífuga modelo X-LA, a uma velocidade de 16000 rpm e 20°C. As curvas de c(s*) x s* foram ajustadas com função Gaussiana, enquanto que os dados de S_{20,W} x concentração protéica foram ajustados por uma regressão linear.

4.1.8 Modelagem molecular por homologia

Para a modelagem por homologia da proteína Hsp100 de cana-de-açúcar foi utilizada a estrutura cristalográfica da proteína ClpB de *Thermus thermophilus* (código PDB: 1QVR), resolvida a uma resolução de 3,0 Å na presença do ligante AMPPNP. Os modelos foram gerados pelo programa Modeller e analisados pelo programa PROCHECK. O cDNA que codifica a proteína Hsp100 foi totalmente seqüenciado e a seqüência de resíduos de aminoácidos da proteína utilizada na modelagem (vide fig. 1). A seqüência da Hsp100 de cana-de-açúcar é composta de 912 aminoácidos enquanto que a proteína ClpB de *T. thermophilus* apresenta 854 aminoácidos. Apesar desta diferença na estrutura primária, as duas cadeias possuem aproximadamente 50 % de identidade entre si (figura 21). A seqüência que codifica a proteína Hsp100 de cana-de-açúcar será depositada no banco de dados do NCBI em ocasião oportuna. A figura 22 apresenta o modelo por homologia da proteína Hsp100 de cana-de-açúcar com seus domínios individualizados e também o gráfico de Ramachandram, mostrando os ângulos torsionais *Phi e Psi* da estrutura (obtido através do programa PROCHECK). Para obter este modelo final foram gerados e testados 20 modelos distintos.

	10	20	30	40	50	60	70	80
							1	
Hsp100_cana	MNPDNFTHKTNEA	LVAAHEAASEA	AG <mark>H</mark> AQLTPL <mark>H</mark> I	LAAALVADKGG	ILRQAITG <mark>A</mark> S	GGD <mark>GA</mark> AG <mark>D</mark> SE	'E <mark>R</mark> VL <mark>S</mark> KA <mark>L</mark> KK	(LPSQSP
ClpB_thermus	MNLERWTQAAREA	LAQAQVLAQR	4K <mark>H</mark> QAIDLP <mark>H</mark>	LWAVLLKDERS	LAWRLLEK <mark>A</mark> -	<mark>GA</mark> DP <mark>K</mark> AI	KELQERELAF	LPKVEG
	** :.:** :**	*. *: *	* : *:	* * . * : * : .	: : : *	** .::	:.: .: * :	**
Consenso	MN2222T2222EA	L22A222A222	22H222222H1	L2A2L22D222	222222222AS	GGDGA22222	22222222L22	LP2222
	90	100	110	120	130	140	150	160
		1	1	1	1	1	1	1
Hsp100 cana	PPDSVPASTALIK	AIRRAOSAOKI	KRGDSHLAVD) LLLGLLEDSO	ISDCLKEAGV	SAARVRAELE	KLRGGGGRRV	ESASGD
ClpB thermus	AEVGOYLTSRLSG	ALNRAEGLME	ELKDRYVAVD		PGLPG	-LEALKGALF	ELRGGRTV	OTEHAE
1 _ 1	*	* * *	* * ***	* * * * * :		::. *:	**** * *	
Consenso	22222222222L22	A22RA222222	222D222AVD2	2L2L2L2E22Q	ISDCL22222	S222222212	2LRGGGGR2V	7222222
	170	180	190	200	210	220	230	240
		1		1	Í		1	i
Hsp100 cana	TNFOALKTYGRDL	VEOAGKLDI	PVIGRDEETRI	RVVRTLSRRTK	NNPVLIGEPG	VGKTAVVEGI	AORTVRGDVF	SNLLDV
ClpB_thermus	STYNALEQYGIDL	TRLAAEGKLDI	PVIGRDEEIR	RVIQILLRRTK	NNPVLIGEPG	VGKTAIVEGI	AQRIVKGDVF	EGLKGK
	:.::**: ** **	* ****	*******	**::** ****	*******	*****:****	*****:***	• • •
Consenso	2222AL22YG2DL	222AAEGKLDI	PVIGRDEEIR	RV22IL2RRTK	NNPVLIGEPG	VGKTA2VEGI	AQRIV2GDVF	22L222
	250	260	270	280	290	300	310	320
		1	1	1	1	1	1	1
Hsp100 cana	RLIALDMGALVAG	AKYRGEFEERI	KSVLKEVEEA	AEGKVILFIDE	IHLVLGAGRT	EGSMDAANLE	KPMLARGOLF	CIGATT
ClpB thermus	RIVSLOMGSLLAG	AKYRGEFEERI	KAVIOEVVO	SOGEVILFIDE	LHTVVGAGKA	EGAVDAGNMI	KPALARGELF	LIGATT
	* • • • * • * • * • * • * *	******	** ** ** *	* * * * * * * * *	* * * * * * * • •	**::**.*:	** **** **	: *****
Consenso	R222L2MG2L2AG	AKYRGEFEERI	LK2V22EV222	22G2VILFIDE	2H2V2GAG22	EG22DA2N22	KP2LARG2LF	2IGATT
	330	340	350	360	370	380	390	400

		I	I	I	I	I.	I	1
Hsp100_cana ClpB_thermus	LEEYRKYVEKDAAF LDEYR-EIEKDPAL	ERRFQQVFV ERRFQPVYV	AEPSVPDTIS DEPTVEETIS	ILRGLKEKYE(ILRGLKEKYE	GHHGVRIQDRA VHHGVRISDSA	ALVVAAQLSAF AIIAAATLSHF	RYIMGRHLPDK RYITERRLPDK	AIDLVD
Consenso	L2EYRK22EKD2A2	ERRFQ2V2V	2EP2V22TIS	ILRGLKEKYE:	2HHGVRI2D2	A222AA2LS2F	RYI22R2LPDK	AIDL2D
	410	420	430	440	450	460	470	480
Hsp100_cana ClpB_thermus	EACANVRVQLDSQF EAAARLRMALESAF	'EEIDNLERK 'EEIDALERK	RIQLEVELHA KLQLEIEREA	LEKEKDKASKI LKKEKDPDSQI	ARLVEVKKELI ERLKAIEAEI) D l rdklqpl1 A <mark>klteeiakl</mark> f	'MKY <mark>RKE</mark> KERI RAE <u>W</u> EREREIL	DEI <mark>R</mark> KL RKL <mark>R</mark> EA
Consenso	**.*.:*: *:* * EA2A22R22L2S2F	**** **** EEID2LERK	::***:* .* 22QLE2E22A	*:**** *: L2KEKD22S2:	** :: *: 2RL22222E22	.* ::: * 22L222222L2	::.:*:* : 222222E2E22	::*: 222R22
	490	500	510	520	530	540	550	560
Hsp100_cana ClpB_thermus	KQRREELQFTLQEA QHRLDEVRREIELA	ERRMDLARV ERQYDLNRA	ADLK <mark>YGAL</mark> QE AELRYGELPK	ID <mark>A</mark> AIAK <mark>LES</mark> I LE <mark>A</mark> EVEA <mark>LSE</mark> I	ETGENLMLTE: KLRGARFVRLI	IVGPEQIAEVN EVTEEDIAEIN	/SR <u>W</u> TGIPVTR /SR <u>W</u> TGIPVSK	LGQNDK
Consenso	::* :*:: :: * 22R22E2222222A	**: ** *. ER22DL2R2	*:*:** * : A2L2YG2L22	::* : * 22A2222L22:	: : : 2222222222222	* *:***:* 2V22E2IAE2\	********* /SRWTGIPV22	* :.:: L22222
	570 I	580 I	590 I	600 I	610	620	630	640
Hsp100_cana ClpB_thermus	ERLVGLADRLHQRV EKLLRLEEELHKRV	VGQTEAVNA VGQDEAIRA	VAEAVLRSKA VADAIRRARA	GLGRPQQSTG	SFLFLGPTGV(SFLFLGPTGV(GKTELAKAFAE GKTELAKTLAZ	EQLFDDENLLV ATLFDTEEAMI	RIDMSE RIDMTE
Consenso	E2L22L222LH2RV	VGQ2EA22A	VA2A22R22A	GL22P2222G	SFLFLGPTGV	GKTELAK22A2	22LFD2E2222	RIDM2E
	650 I	660 I	670 I	680 I	690 I	700 	710 	720
Hsp100_cana ClpB_thermus	YMEQHSVARLIGAF YMEKHAVSRLIGAF	PGYVGHEEG PGYVGYEEG	GQLTEQVRRR	PYSVILFDE <mark>V</mark> PYSVILFDE I I	EKAHVAVFNTI EKAHPDVFNII	LLQVLDDGRL1 LLQILDDGRL1	ID <mark>GQ</mark> GRTVDFR	NTVIIM
Consenso	YME2H2V2RLIGAF	PGYVG2EEG	GQLTE2VRRR	PYSVILFDE21	EKAH22VFN21	LLQ2LDDGRL1	D22GRTVDFR	NTVII2
	730	740	750	760	770	780	790	800
Hsp100_cana ClpB_thermus	TSNLGAEHLLAGMV TSNLGSPLILEGLQ	GKNSMKVA <mark>R</mark> KG <u>W</u> PYERI <mark>R</mark>	DLVMQEVRRH DEVFKVLQQH	FRPELLNRLDI FRPEFLNRLDI	EIVIFDPLSHI EIVVFRPLTKI	EQLRKV <mark>AR</mark> LQN EQIRQIVEIQI	IKDVAVRLAER SYLRARLAEK	GIALAV RISLEL
Consenso	*****: :* *: TSNLG2222L2G22	222222222	* *:: :::* D2V222222H	FRPE2LNRLD	***:* **:: EIV2F2PL22I	**:*: EQ2R22222Q2	222222RLAE2	*:* : 2I2L22
	810	820	830	840	850	860	870	880
Hsp100_cana ClpB_thermus	TDAALDIILSLSYE TEAAKDFLAERGYE	PVYGARPIR PVFGARPLR **•****	RWIEKRVVTQ RVIQRELETP	LSKMLIQEEI LAQKILAGEVI	DENCTVYIDA KEGDRVQVDV(* * *	APAKDELAYR GPAG	/DRSGGLVNAE	TGLKSD
Consenso	T2AA2D222222YD	PV2GARP2R	R2I22222T2	L22222222E2:	2E222V22D22	2PA2DELAYR\	/DRSGGLVNAE	TGLKSD
	890 I	900 I	910 					
Hsp100_cana ClpB_thermus	ILIQVPNDAVRSDA LVFAVPARVEA	AQAVKKMRI	MEEDEDGMDE	E 				
Consenso	2222VPNDA2R22A	AOAVKKMRT	MEEDEDGMDE	F.				

Figura 21: Alinhamento das seqüências de aminoácidos da proteína Hsp100 de cana-de-açúcar e ClpB de *T. thermophilus* utilizado na modelagem. As seqüências da Hsp100 e da ClpB possuem 50 % de aminoácidos idênticos (resíduos em vermelho), 20 % de aminoácidos muito similares (resíduos em verde), 6 % de aminoácidos pouco similares (resíduos em azul) e 25 % de aminoácidos diferentes (resíduos em preto). O alinhamento foi realizado pelo programa *CLUSTALW* (http://npsa-pbil.ibcp.fr/cgi-

bin/align_clustalw.pl). A proteína ClpB de *T. thermophilus* possuí 6 triptofanos, enquanto que a Hsp100 de cana-de-açúcar possui apenas 2 (aminoácidos grifados).



Figura 22: Modelo por homologia da proteína Hsp100 e o respectivo gráfico de Ramachandran. A) Representações do modelo da Hsp100, destacando os dois Triptofanos (vermelho) presentes na estrutura primária. Na representação da direita, os dois domínios de ligação a nucleotídeos estão destacados pelos círculos vermelho e azul, domínios 1 e 2, respectivamente. A região de "junção" entre os dois subdomínios que compõe o DLN1 (amarelo), exclusiva das chaperones da classe ClpB, faz parte

do chamado "pequeno domínio D1" e está relacionada a atividade de desagregação da Hsp100. O domínio N-terminal está representado em roxo. B) O gráfico de Ramachandran, o qual mostra os valores em graus dos ângulos de torsão Phi e Psi para cada resíduo de aminoácido da seqüência protéica, indicou a presença de 90,1 % dos resíduos nas regiões mais favorecidas (vermelha), 8,7 % nas regiões adicionalmente permitidas (amarela), 0,9 % nas regiões generosamente permitidas (bege) e 0,3 % nas regiões não permitidas. Os triângulos representam os ângulos Phi e Psi dos resíduos de glicina da cadeia. A proteína ClpB de *T. thermophilus*, utilizada como modelo, apresentou 82,5 % dos resíduos nas regiões mais favorecidas do gráfico de Ramachandran e não apresentou nenhum resíduo nas regiões não permitidas.

4.1.9 Interação com agregados protéicos

A proteína Hsp100 foi submetida a experimentos de interação com agregados protéicos visando compreender melhor sua modulação pelos nucleotídeos, bem como a comprovação de que estamos caracterizando uma proteína biologicamente ativa. Assim, dois tipos de substratos foram empregados: as proteínas luciferase (Promega) e citrato sintase (Sigma), na proporção estequiométrica de 1:1 em relação à proteína Hsp100. A figura 23 apresenta o gel de poliacrilamida a 10 % do experimento de interação, o qual foi realizado na presença e ausência do nucleotídeo ATPγS (análogo de ATP não-hidrolizável) na concentração de 5 mM.



Figura 23: Ensaio de interação entre a proteína Hsp100 e dois substratos diferentes. O experimento com cada substrato foi desenvolvido independentemente, sendo posteriormente analisado no mesmo gel de poliacrilamida (10 %). Após a agregação dos substratos via aquecimento, estes foram incubados com a proteína Hsp100 por uma hora a 30 °C, ou apenas tampão, no caso dos controles negativos. Após a eluição das proteínas da resina *Talon[®] Metal Chelate*, estas foram precipitadas em acetona 80% e ressuspendidas em tampão de amostra de proteína. As setas indicam as bandas do padrão de massa molecular. Em cada pista está indicada a presença (+) ou ausência (-) das respectivas proteínas, na seguinte ordem: luciferase livre (sem sofrer agregação prévia), luciferase agregada a 42 °C, Hsp100, luciferase agregada incubada com a Hsp100 na presença de

ATPγS, citrato sintase livre (cs – não agregada), cs a 42 °C, Hsp100, cs agregada incubada com Hsp100, cs agregada incubada com a Hsp100 na presença de ATPγS.

4.2 Discussão Hsp100 de cana-de-açúcar

4.2.1 Expressão e purificação

O clone pET28a-Hsp100 apresentou uma boa expressão, permitindo o prosseguimento para a etapa de purificação. Inicialmente, a amostra purificada por cromatografia de afinidade apresentou excesso de proteínas contaminantes, o que foi melhorado após o desenvolvimento do protocolo de purificação descrito no texto. Após a purificação por afinidade a Hsp100-His foi incubada com a enzima fosfatase alcalina, visando eliminar os vestígios de nucleotídeos que poderiam interferir nos resultados futuros com a Hsp100 (Borges e Ramos, 2006).

4.2.2 Dicroísmo circular

Os dados de dicroísmo circular revelaram que a Hsp100 purificada apresenta-se estruturada e que é constituída, em sua maioria, de estruturas do tipo hélice alfa. Esta característica também é observada em outras proteínas da família Hsp100, como na ClpB de *Escherichia coli* (Barnett et al., 2000). A partir de medidas em diferentes concentrações de NaCl (100 mM, 250 mM e 500 mM) foi possível observar que a proteína Hsp100 de cana-de-açúcar é muito sensível à força iônica do meio, uma vez que precipitou na concentração de 100 mM de NaCl. A diferença de sinal observada entre esta condição e as demais se deve a depleção devido à precipitação, e não a uma eventual perda de estrutura por parte da proteína. Por outro lado, o aumento da concentração de NaCl de 250 para 500 mM não provocou um efeito perceptível no espectro de dicroísmo circular da Hsp100, ou seja, parece não exercer influência sobre sua estrutura secundária. Estas informações permitiram determinar os limites seguros de concentração de NaCl a serem utilizados nos experimentos com esta proteína.

A adição dos nucleotídeos (ATP e ADP) não afetou, de forma significativa, a estrutura da Hsp100 vista por dicroísmo circular. Isto porque não foram observadas alterações nas quantidades de elementos de estrutura secundária (hélice-alfa, folha beta, etc) quando da deconvolução dos respectivos espectros. Estes resultados eram esperados, uma vez que os nucleotídeos parecem induzir movimentações entre os domínios da proteína (Lee et al., 2003), ou seja, em sua estrutura terciária, e não promovem mudanças na estrutura secundária dos mesmos.

4.2.3 Fluorescência

O espectro de emissão de fluorescência demonstrou que os dois triptofanos presentes na seqüência primária de aminoácidos da Hsp100 estão em um ambiente bastante exposto ao solvente. A característica do espectro que leva a esta conclusão é a de que o máximo de emissão de fluorescência, observado para a Hsp100 de cana-de-açúcar (350 nm), está próximo ao máximo de emissão do triptofano livre em solução (355 nm). É importante salientar que este máximo de emissão de fluorescência é similar ao encontrado para a ClpB de *E. Coli* (350 nm), a qual também possui dois resíduos de triptofano (Barnett et al., 2000).

A proteína Hsp100 foi submetida a experimentos de fluorescência na presença de nucleotídeos e do agente desnaturante cloreto de guanidina, em concentrações de 1 mM e de 6 M, respectivamente. Os resíduos de triptofano da proteína sem aditivos e na presença de guanidina apresentaram um máximo de emissão muito próximo ao do triptofano livre em solução, demonstrando que se encontram expostos ao solvente em ambas as condições. A intensidade de fluorescência em todas as condições foi praticamente o dobro da intensidade com cloreto de guanidina, o que pode ser um indício de hiper-fluorescência intrínseca aos resíduos de triptofano da Hsp100. Já a adição dos nucleotídeos não alterou de forma significativa nem a intensidade de fluorescência e nem o máximo de emissão, em relação às outras condições testadas. Os valores muito próximos de centro de massa espectral de fluorescência em todas as condições utilizadas são um forte indício de que as alterações provocadas pelos ligantes e pela guanidina são mínimas. Assim, qualquer mudança estrutural causada pelos mesmos parece não afetar os ambientes em que estão inseridos os triptofanos.

4.2.4 Estabilidade medida por CD

O desenovelamento térmico da Hsp100 revelou-se irreversível, ou seja, a proteína não retornou ao enovelamento nativo após desnaturação por calor. A temperatura de transição do estado nativo para o desenovelado foi de aproximadamente 58 °C, portanto muito próxima da temperatura de transição (56 °C) observada para a ClpB de *E. Coli* (Barnett et al., 2000). O desenovelamento térmico da proteína Hsp100 na presença do nucleotídeo ADP revelou uma pequena mudança na estabilidade em relação à proteína sem aditivos e na presença de ATP e AMP, porém esta não se mostrou significativa. Houve

também, na faixa entre 60 e 70 °C, uma aparente estabilização de algum domínio da proteína com os nucleotídeos ATP e AMP em relação a condição nativa e, principalmente, na condição com ADP. Porém, todas estas alterações são muito sutis, e não representam diferenças significativas que justifiquem alguma especulação em relação à alteração da estabilidade térmica da Hsp100 na presença de nucleotídeos.

4.2.5 Espalhamento dinâmico de luz

Os experimentos de EDL se mostraram muito úteis não só para controlar a monodispersidade das amostras da Hsp100, mas também para corroborar os efeitos causados pela concentração de sal e presença de nucleotídeos na oligomerização da proteína. Através do EDL foi possível notar um aumento do coeficiente de difusão (*D*) quando diminuímos a concentração de NaCl da solução protéica. Como o valor do *D* reflete o tamanho da proteína e é inversamente proporcional a este, o oligômero formado pela Hsp100 em 500 mM de NaCl é aparentemente menor do que o oligômero formado na condição com 250 mM. Assim, enquanto que em 250 mM de NaCl obtivemos um D de aproximadamente 2,38 x 10^{-7} cm²/s, em 500 mM obtivemos um D de aproximadamente 2,9 x 10^{-7} cm²/s (dados em 500 mM NaCl não mostrados). Estes dados em relação ao efeito da força iônica sobre a Hsp100 também puderam ser observados através de ultracentrifugação analítica, como explicado no próximo tópico.

4.2.6 Ultracentrifugação analítica

Em geral, os dados de massa molecular do oligômero da proteína Hsp100 de cana observados por sedimentação em equilíbrio (SE) corroboraram os dados para outras proteínas da família. A proteína apresentou na SE uma massa molecular de aproximadamente 616 kDa, o que corresponde à formação de um hexâmero. A formação de hexâmeros também é observada para as proteínas Hsp104 de levedura (Bosl et al., 2005), ClpB de *Escherichia coli* (Akoev et al., 2004) e Hsp101 de tabaco (Gallie et al., 2002). Os conjuntos de dados que permitiram observar com clareza a formação do hexâmero foram aqueles de menor concentração de sal, nos quais o ATP ou o MgATP foram adicionados. Como visto por espalhamento dinâmico de luz, a presença de 0,5 M NaCl é suficiente para deslocar a reação de oligomerização para a formação de complexos com menor massa molecular. O deslocamento da reação

no sentido de formação do hexâmero, tanto para proteína Hsp104 (levedura) como para a proteína ClpB (*E. coli*) é substancialmente favorecido pela presença de ATPγS (análogo do ATP não hidrolisável) ou o próprio ATP (Bosl et al., 2005; Akoev et al., 2004). Da mesma forma, os dados de SE da Hsp100 de cana em que estes nucleotídeos não foram adicionados apresentaram uma ambigüidade que pode estar relacionada a outras espécies oligoméricas discretas nesta condição, impossibilitando a obtenção de um consenso na análise destes dados. Apesar desta ressalva, foi possível observar a formação do hexâmero da proteína Hsp100 através da SE.

Os valores de coeficiente de sedimentação (S_{20,W}) da proteína Hsp100 de cana, obtidos por experimentos de velocidade de sedimentação (VS), também foram similares aos encontrados na literatura. A proteína Hsp100 apresentou um valor de S_{20,w} de aproximadamente 16,7 a 1 mg/mL enquanto que a proteína Hsp104 de levedura apresentou um S de 16,5 (Bosl et al., 2005) e a ClpB de E.coli 16,5 S (Akoev et al., 2004). O coeficiente de sedimentação da proteína em um meio contendo 0,5 M de NaCl sofreu uma alteração para um S de aproximadamente 9,7, indicando que o tamanho do oligômero diminuiu uma vez que este é diretamente proporcional ao valor de S. Novamente, este dado está de acordo com o observado por EDL. Os dados sem a adição de nucleotídeos apresentaram uma distribuição heterogênea de coeficientes de sedimentação, indicando que a amostra possui várias espécies diferentes. Desta forma, foi dificultada a tarefa de determinar valores de referência para esta condição em particular. Este fato corrobora dados da literatura para proteínas homólogas, como Hsp104 de levedura, ClpB de Escherichia coli e Hsp100 de tabaco (Bosl et al., 2005; Akoev et al., 2004; Gallie et al., 2002), que indicam que o hexâmero somente é estabilizado na presença de ATP ou ATPyS. Isto é confirmado pelos dados coletados na presença do nucleotídeo ATPyS, que apresentaram uma espécie majoritária e uma discreta fração com sedimentação equivalente ao dobro desta. Os dados de velocidade de sedimentação na presença de ATPyS apresentaram um valor de 16,19 \pm 0,4 S (S⁰_{20,W}), semelhante ao obtido para as proteínas Hsp104 e ClpB, ambas com 16,5 S (Bosl et al., 2005 e Akoev et al., 2004). A massa do oligômero calculada a partir da relação entre coeficiente de sedimentação e de difusão, resulta em uma massa de aproximadamente 680 kDa. Este massa é próxima ao valor teórico esperado para o hexâmero, de aprox. 620 kDa e também ao valor obtido através de sedimentação em equilíbrio para a condição com ATP, de aprox. 616 kDa.

4.2.7 Modelagem molecular por homologia

O seqüenciamento completo do cDNA que codifica a proteína Hsp100 de cana-de-açúcar possibilitou a construção de um modelo tridimensional da proteína baseado em modelagem por homologia com a proteína homóloga ClpB (*T. thermophilus*). A proteína Hsp100 possui grande identidade seqüencial nos domínios de ligação aos nucleotídeos com a proteína ClpB, no entanto as regiões N e C-terminal apresentam menor identidade. O menor comprimento da cadeia da ClpB obrigounos a omitir grande parte dos aminoácidos do C-terminal da Hsp100 na estrutura final do modelo. A estrutura da ClpB (PDB: 1QVR) possui uma resolução relativamente baixa de 3 Å, o que acabou também limitando a qualidade do modelo, pois este nunca terá uma qualidade superior à estrutura cristalográfica que foi utilizada em sua construção. Mesmo assim, o programa Modelller conseguiu gerar um modelo com uma boa qualidade em termos de ângulos torsionais e distâncias entre átomos.

4.2.7 Interação com agregados

A principal característica das proteínas pertencentes à família Hsp100/ClpB é a capacidade de reconhecer e se ligar à agregados protéicos (Maurizi e Xia, 2004). Os experimentos foram realizados com a proteína Luciferase porque alguns trabalhos mostram que as chaperonas da família Hsp100 são capazes de se ligar a esta proteína no estado agregado (Mosser *et al.*, 2004; Zietkiewicz *et al.*, 2004). Além deste ensaio, tentamos também realizar o mesmo experimento tendo como substrato a proteína Citrato sintase, já utilizada em outros experimentos de agregação realizados em nosso grupo de pesquisa. Para cada substrato foram feitos experimentos na ausência e presença de ATPγS, uma vez que outras proteínas da família Hsp100 necessitam estar no estado-ATP ligado para reconhecer agregados (Maurizi e Xia, 2004; Bosl *et al.* 2006). Desta forma tentamos evitar a hidrólise do nucleotídeo e a modulação negativa da atividade de reconhecimento desta proteína.

Os experimentos foram aperfeiçoados de modo que se tornasse possível detectar a presença da proteína Luciferase ligada à proteína Hsp100, apesar da ocorrência de uma pequena banda de Luciferase no controle negativo (sem a chaperona). Isto se deve à dificuldade de eliminar os agregados durante a lavagem da resina para a remoção de proteínas não-ligadas ou ligadas inespecificamente, já que após cada lavagem a resina é centrifugada, levando consigo agregados insolúveis. Apesar da ligação aos

agregados de Luciferase ter sido comprovada, ainda não está claro o papel do nucleotídeo na modulação desta atividade, uma vez que foi possível ver a ligação da chaperona ao agregado em todas as condições testadas. Isto tanto pode ser uma atividade específica da Hsp100 de cana, como pode ser algum artefato do experimento. Da mesma forma, não é possível distinguir diferenças quanto à ligação da proteína Hsp100 ao agregado de Citrato sintase entre as condições com e sem o nucleotídeo ATPγS, embora também tenha sido possível detectar uma fraca ligação entre chaperona e este substrato.

4.3 Conclusões Hsp100 de cana-de-açúcar

- A proteína Hsp100 de cana-de-açúcar foi clonada e totalmente seqüenciada, apresentando uma identidade de 31 % em relação às proteínas Hsp104 (*Saccharomyces cerevisiae*), Hsp101 (*Arabdopisis thaliana*) e ClpB (*Thermus thermophilus*).
- As análises de CD indicaram que a proteína se encontra estruturada e que é constituída majoritariamente por hélices alfa. Além disto, a presença de nucleotídeos não é capaz de alterar o espectro de CD da proteína Hsp100.
- Espectros de CD obtidos em diferentes concentrações de NaCl permitiram observar que a proteína Hsp100 sofre precipitação em baixas concentrações de sal.
- A avaliação da estabilidade térmica medida por CD revelou que a proteína Hsp100 apresenta uma temperatura de transição, entre os estados enovelado e desenovelado, de aproximadamente 58
 ^oC. O desenovelamento se mostrou irreversível e a adição de nucleotídeos não foi capaz de alterar significativamente a estabilidade térmica vista por CD.
- A análise de emissão de fluorescência permitiu observar que os dois triptofanos presentes na proteína Hsp100 estão expostos ao solvente. As adições dos nucleotídeos ATP e ADP não foram capazes de alterar o perfil observado para a condição sem ligantes.
- Experimentos de espalhamento dinâmico de luz permitiram determinar o coeficiente de difusão para a proteína Hsp100 (2,38 x 10⁻⁷ cm²/s), o qual não foi alterado pela presença dos nucleotídeos.
- Através de experimentos de sedimentação em equilíbrio foi possível determinar a massa molecular do oligômero da proteína Hsp100 na presença de ATP (616 ± 21 kDa).
- Os experimentos de velocidade de sedimentação na presença ATPγS permitiram determinar o coeficiente de sedimentação para a proteína Hsp100, o qual apresentou um valor de 16,2 ± 0,4 S.
- Utilizando os dados de EDL e de velocidade de sedimentação foi possível estimar a massa molecular do oligômero como de aproximadamente 675 kDa.
- O seqüenciamento da Hsp100 de cana-de-açúcar permitiu a construção de um modelo por homologia baseado na identidade com a proteína ClpB de *T. thermophilus*.

 A proteína Hsp100 foi capaz de reconhecer agregados formados pelas proteínas substrato luciferase e citrato sintase através de ensaios de interação utilizando resina de purificação, demonstrando que a proteína obtida apresenta atividade biológica.

5 Hsp82 de cana-de-açúcar

5.1 Resultados Hsp82 de cana-de-açúcar

5.1.1 Clonagem da Hsp82

O cDNA que codifica a proteína Hsp82, foi utilizado como molde para a reação de PCR com oligonucleotídeos específicos, gerando um produto de 2.100 pb. A figura 24 A apresenta o padrão de digestão do vetor pGEMT com as enzimas de restrição *Nde*I e *Eco*RI, mostrando o inserto correspondente ao cDNA da proteína Hsp82. O seqüenciamento parcial das extremidades deste DNA plasmidial confirmou a clonagem do cDNA da proteína Hsp82 em pGEMT. O clone pGEMT-Hsp82 foi utilizado para a obtenção do inserto clivado com as enzimas *Nde*I e *Eco*RI para a subclonagem em pET28a. A figura 24 B apresenta o padrão de digestão do vetor pET28a-Hsp82, com as enzimas *Nde*I e *Eco*RI, mostrando o inserto que corresponde ao cDNA da Hsp82. A clonagem do cDNA que codifica a proteína Hsp82 no vetor pET28a foi confirmada através do seqüenciamento deste clone.

A)



B)



Figura 24: Clonagem do cDNA que codifica a proteína Hsp82 no vetor de transferência pGEM-T easy e no vetor de expressão pET-28a. A) Gel de agarose 1% corado com brometo de etídio demonstrando a digestão do clone pGEMT-Hsp82 com as enzimas *Ndel e Eco*RI. A clonagem do cDNA da Hsp82 em pGEMT foi confirmada através de seqüenciamento. B) Gel de agarose 1% corado com brometo de etídio demonstrando a digestão do clone pET28a-Hsp82 com as enzimas de restrição *Nde*I e *Eco*RI. A clonagem do cDNA completo que codifica a proteína Hsp82 de cana-de-açúcar no vetor pET28a foi confirmada pelo seqüenciamento com os oligonucleotídeos específicos descritos em materiais e métodos. 1)

minipreparação do DNA plasmidial pET28a-Hsp82; 2) clone pET28a-Hsp82 digerido com as enzimas de restrição *Nde*I e *Eco*RI.

5.1.2 Expressão e purificação da Hsp82 de cana-de-açúcar

A Hsp82 de cana-de-açúcar foi expressa em *E.coli* BL21 (DE3)pLysS por um período de cinco horas à 37 °C. A figura 25 mostra alíquotas retiradas a cada hora após indução, sendo possível observar o aumento de uma banda relativa a uma proteína com massa molecular entre 66 e 97 kDa que fica visível a partir de 2 horas de indução. Após a lise da cultura induzida, a proteína Hsp82 foi encontrada tanto na fração precipitada quanto na solúvel, sendo esta submetida a várias etapas de purificação.



Figura 25: Análise da indução do clone pET28a Hsp82 com IPTG 0,5 mM por gel de poliacrilamida (10%). Bactérias *E. coli* BL21 (DE3)pLysS, transformadas com o vetor pET28a-Hsp82 foram crescidas em 50 mL de LB contendo canamicina 30 μg/mL até DO_{600nm} igual a 0,7, no qual foi acrescentado IPTG para uma concentração final de 0,5 mM. Alíquotas de 1 mL de cultura foram colhidas antes da indução e a cada hora após o início da mesma (não-induzido - NI). Após centrifugação, os precipitados das alíquotas foram ressuspendidos em 0,1 mL de tampão de amostra de proteína e 0,01 mL foram aplicados no gel. O

traço indica a proteína de 97 kDa correspondente ao padrão de massa molecular e a seta indica a banda de indução da proteína Hsp82.

Inicialmente, a proteína Hsp82 foi submetida a uma purificação por afinidade na coluna *Hitrap Chelating® Global 5 mL*, uma vez que possui uma cauda de histidina fusionada a sua extremidade Nterminal. A purificação apresentou três frações protéicas, eluídas em condições diferentes de concentração de imidazol, como mostrado na figura 26A. A primeira fração (F₁) corresponde às proteínas que não interagiram com a resina de purificação, sendo eluídas antes do início do gradiente de imidazol. A segunda fração (F₂), foi eluída com aproximadamente 50 mM de imidazol e a última fração (F₃) foi eluída em aprox. 120 mM de imidazol. A proteína Hsp82 de cana-de-açúcar foi encontrada majoritariamente na terceira fração, como pode ser visto na figura 26B.

A)





Figura 26: Purificação da proteína Hsp82 por afinidade. Após a indução da proteína Hsp82 a cultura foi lisada por sonicação no tampão Tris-HCl 50 mM pH 8,0, KCl 500 mM e EDTA 1 mM na presença de aproximadamente 1 μg/mL de lisozima. A fração solúvel, obtida após a centrifugação do lisado por 30 minutos a 18000 x g, foi submetida à purificação por afinidade na coluna *HiTrap Chelating Global*[®] conectada ao cromatógrafo Äkta FPLC. A) Cromatograma* de purificação da proteína Hsp82 mostrando as frações eluídas da coluna (F₁, F₂ e F₃) pelo gradiente de imidazol entre os tampões SB e EB. B) Gel de poliacrilamida (10 %) mostrando as proteínas eluídas em cada uma das três frações presentes no cromatograma. A proteína Hsp82 foi encontrada na fração F₃, como pode ser notado pela banda indicada pela seta vermelha, com massa molecular entre 97 e 66 kDa. *Absorbância a 280 nm: azul; gradiente de imidazol: vermelho. Setas pretas indicam as bandas correspondentes às proteínas do padrão de massa molecular.

A proteína Hsp82 foi submetida em seguida a uma purificação por troca iônica em um gradiente de NaCl. As frações purificadas por afinidade foram extensivamente dialisadas no tampão T50 e então submetidas à purificação. A figura 27A apresenta a purificação por troca iônica da proteína Hsp82, a qual apresentou três frações distintas de eluição: a primeira fração (F₁) foi eluída em 250 mM de NaCl, a segunda fração (F₂) foi eluída em 315 mM de NaCl e a terceira fração (F₃) foi eluída em 490 mM de NaCl (concentrações aproximadas). A figura 27B apresenta o gel de poliacrilamida (10%) mostrando que a Hsp82 de cana-de-açúcar se encontra predominantemente na fração 2 (F2).



Figura 27: Purificação da proteína Hsp82 por troca iônica. Após a purificação por afinidade, a proteína Hsp82 foi submetida a uma purificação em coluna DEAE Sepharose conectada ao cromatógrafo Äkta FPLC. A) Cromatograma* de purificação da proteína Hsp82 mostrando as frações eluídas da coluna (F₁, F₂ e F₃) em um gradiente de NaCl. B) Gel de poliacrilamida (10 %) mostrando as proteínas eluídas em cada uma das três frações presentes no cromatograma. A proteína Hsp82 está presente na fração F₂, como pode ser notado pela banda indicada pela seta vermelha (massa molecular entre 97 e 66 kDa). * Absorbância a 280 nm: azul; gradiente de NaCl: vermelho. Setas pretas indicam as bandas correspondentes às proteínas do padrão de massa molecular.

Após a purificação por troca iônica, a proteína Hsp82 foi submetida a uma terceira e última etapa de purificação, por exclusão molecular. As frações da troca iônica que continham a Hsp82 foram reunidas e concentradas para um volume de aproximadamente 10 mL, sendo aplicadas na coluna Superdex 200 26/60 em seguida. A figura 28A apresenta o cromatograma desta purificação, sendo

possível identificar três frações eluídas durante a purificação. A proteína Hsp82 foi eluída em duas frações distintas, a primeira fração (F₁) em um volume de 143 mL e a segunda fração (F₂) eluída em 165 mL, já a terceira fração (F₃), contendo os contaminantes ainda presentes na amostra, foi eluída em 202 mL (volumes aproximados). A figura 28B mostra o gel de poliacrilamida que correspondente a esta purificação, no qual é possível distinguir as proteínas presentes em cada uma das frações observadas no cromatograma.





Figura 28: Purificação da proteína Hsp82 por exclusão molecular. A proteína Hsp82 foi submetida a uma purificação isocrática no tampão TN150, em uma coluna Superdex 200 26/60 conectada ao equipamento Äkta FPLC. A) Cromatograma de purificação da proteína Hsp82 mostrando as frações eluídas da coluna (F₁, F₂ e F₃). B) Gel de poliacrilamida (10 %) mostrando as proteínas eluídas em cada uma das três frações

presentes no cromatograma. A proteína Hsp82 esta presente nas frações F₁ e F₂, como pode ser notado pela bandas indicadas (setas vermelhas), que possuem massa molecular entre 97 e 66 kDa.

5.1.3 Caracterização por CD da proteína Hsp82

A proteína Hsp82 teve sua estrutura secundária caracterizada por dicroísmo circular, em quatro condições diferentes: sem aditivos, com 200 μ M de ATP, ADP, AMP e ATP γ S (figura 29). Em todas as condições analisadas, com e sem a presença dos ligantes, a proteína Hsp82 apresentou um perfil de polipeptídeo estruturado. Conforme mostrado na tabela IV, não houve diferença significativa entre a condição sem aditivos (tampão TN150) e na presença dos nucleotídeos, como previsto pelo programa de deconvolução CDNN. A proteína Hsp82 apresentou aproximadamente 33 ± 3 % de estruturas hélice alfa, 17 ± 2 % de estruturas folha beta e 33 ± 3 % de estruturas randômicas.



Figura 29: Espectro de dicroísmo circular (CD) da proteína Hsp82. A proteína Hsp82, já purificada, foi submetida a análises de CD a uma concentração de 2 μ M no tampão TN150, em cubetas de 1 mm de caminho ótico. Os experimentos foram realizados em um espectropolarímetro acoplado a um banhomaria externo e a um controlador interno de temperatura tipo Peltier. Os dados foram tratados com o programa Origin[®], sendo que cada espectro representa uma média de dez leituras em triplicata.

Tabela IV: Predição da quantidade de estrutura secundária da proteína Hsp82 poli-His de cana-deaçúcar na presença e ausência de nucleotídeos através do programa CDNN (erro de aprox. 10 %).

Estrutura Secundária	S/ Aditivos	ATP	ADP	AMP	ΑΤΡγS
Hélice alfa	33 %	33 %	33 %	32 %	32 %
Folha beta anti-paralela	8 %	8 %	8 %	8 %	8 %
Folha beta paralela	9 %	9 %	9 %	9 %	9 %
Volta Beta	17 %	17 %	17 %	17 %	17 %
Randômica	33 %	33 %	33 %	34 %	34 %
Total	100 %	100 %	100 %	100 %	100 %

A proteína Hsp82 também foi submetida a um desenovelamento térmico monitorado por CD em um comprimento de onda fixo de 222 nm. Primeiramente, a Hsp82 foi submetida a um desenovelamento de 10 °C a 90 °C na ausência e presença de 200 μ M de ATP, ADP e AMP e 2 μ M proteína. Para todas as condições o aquecimento não provocou o desenovelamento completo da proteína (figura 30), demonstrando uma grande resistência ao desenovelamento térmico da Hsp82 de cana-de-açúcar.



Figura 30: Desenovelamento térmico acompanhado por CD da proteína Hsp82. A proteína Hsp82 a uma concentração de 2 μ M, foi submetida a um desenovelamento térmico acompanhado por CD em comprimento de onda 222 nm. A) Desenovelamento da proteína Hsp82 monitorado por CD, entre 10 °C e 90 °C, em quatro condições: sem aditivos, ATP, ADP e AMP (200 μ M cada). Os dados foram analisados através do programa Origin[®].

A proteína Hsp82 foi submetida a ensaios de desenovelamento químico utilizando uréia como agente desnaturante na presença e ausência de diferentes nucleotídeos. A figura 31 mostra o desenovelamento monitorado a 222 nm da proteína Hsp82 na ausência e presença dos nucleotídeos ATP, ADP e AMP, a uma concentração de 200 µM. Os pontos foram obtidos em intervalos de 0,25 M de uréia, abrangendo uma faixa de concentração entre zero e 8 M deste agente desnaturante.



Figura 31: Desenovelamento por uréia da proteína Hsp82 de cana-de-açúcar monitorada por CD a 222 nm. Desenovelamento químico por uréia da proteína Hsp82 sem aditivo (nativa) e na presença de ATP, ADP e AMP. Os experimentos foram realizados em uma concentração de 0,7 mg/mL de proteína e solução estoque de 9,5 M de uréia, diluída no tampão de diálise (Tris-Cl 50 mM, pH 7,5, NaCl 150 mM).

A figura 32 mostra os mesmos dados de desenovelamento por uréia (figura 31) como delta de sinal, ou seja, os dados foram normalizados assumindo-se que o valor no ponto 0 M de uréia correspondem à proteína totalmente enovelada e no ponto 8 M de uréia totalmente desenovelada. Através da análise do delta de sinal é possível observar mais facilmente as diferenças entre as condições estudadas, uma vez que apenas variações relativas são consideradas.



Figura 32: Delta de sinal do desenovelamento por uréia da proteína Hsp82 de cana-de-açúcar monitorada por CD. Dados normalizados de desenovelamento por uréia da proteína Hsp82 em todas as condições testadas. A adição dos nucleotídeos (ATP, ADP e AMP), não alterou significativamente a estabilidade da proteína.

5.1.4 Caracterização por fluorescência da proteína Hsp82

Foram obtidos dados de fluorescência estática para a proteína Hsp82 nas mesmas condições em que foram realizados os experimentos de dicroísmo circular, porém utilizando a proteína a 0,1 mg/mL. A figura 33 mostra o espectro de emissão da proteína Hsp82 na ausência e presença de 200 µM dos nucleotídeos ATP, ADP e AMP.



Figura 33: Espectro de emissão de fluorescência estática para a proteína Hsp82 de cana-de-açúcar na ausência e presença de nucleotídeos. Os máximos de emissão para as condições analisadas foram: 331 nm (sem aditivos), 330 nm (ATP) e 328 nm (ADP e AMP). Já o centro de massa espectral para cada condição foi : 343 ± 1 nm (s/ aditivos), 343 ± 1 nm (ATP), 337 ± 1 nm (ADP) e 338 ± 1 (nm).

A figura 34 apresenta o monitoramento do desenovelamento químico da proteína Hsp82 por fluorescência, mostrando a variação do centro de massa espectral em função da concentração de uréia para as condições com e sem nucleotídeos. O desenovelamento sem aditivos e com ATP apresentaram perfis semelhantes, enquanto que os dados na presença de ADP e AMP apresentaram um perfil diferente. A figura 35 mostra os dados de desenovelamento monitorados por fluorescência da proteína Hsp82 como delta de sinal, permitindo comparar as variações entre cada uma das condições estudadas.



Figura 34: Desenovelamento por uréia da proteína Hsp82 de cana-de-açúcar monitorada por fluorescência. Variações dos centros de massa espectrais durante o desenovelamento por uréia nas condições sem aditivos e na presença de ATP, ADP e AMP, acompanhados por fluorescência.



Figura 35: Delta de sinal do desenovelamento por uréia da proteína Hsp82 de cana-de-açúcar monitorada por fluorescência. Dados normalizados de desenovelamento por uréia em todas as condições testadas. A adição dos nucleotídeos aumentou a estabilidade da proteína Hsp82 ao desenovelamento químico quando observado por fluorescência.

A figura 36 mostra a diferença de sinal entre fluorescência e dicroísmo circular entre cada uma das condições estudadas (s/ aditivos, ATP, ADP e AMP), permitindo comparar a estabilidade da estrutura secundária com a estrutura terciária próxima aos triptofanos.





Figura 36: Comparação dos valores de delta de sinal do desenovelamento por uréia da proteína Hsp82 monitorada por CD e fluorescência. A) condição sem aditivos; B) condição com ATP (200 μ M); C) condição com ADP (200 μ M); D)) condição com AMP (200 μ M). Os dados foram tratados com o programa Origin[®] utilizando a fórmula: (S_{obs} – S₀)/(S₀-S₈), sendo S_{obs} o sinal observado, S₀ o sinal em zero M de uréia e S₈ o sinal em 8 M de uréia.

5.1.5 Análise por espalhamento dinâmico de luz

A proteína Hsp82 teve suas características hidrodinâmicas estudadas por espalhamento dinâmico de luz. Esta técnica permite observar a monodispersidade da amostra e também permite determinar o coeficiente de difusão de proteínas em solução. Os experimentos foram realizados na ausência e presença dos nucleotídeos ATP e ADP a uma concentração de 1 mM. A figura 37 apresenta os dados de coeficiente de difusão, normalizados para condições padrão de temperatura e solvente (20 °C e água), em função da concentração de Hsp82. Através do ajuste destes valores com uma função linear, foi possível determinar o valor de D⁰_{20,W} para cada uma das condições experimentais. A proteína Hsp82 não apresentou alterações significativas em seu coeficiente de difusão em função da presença dos nucleotídeos.



Figura 37: Espalhamento de luz dinâmico (DLS) da proteína Hsp82 de cana-de-açúcar. A proteína Hsp82 apresentou um valor de $D^{0}_{20,W}$ de aproximadamente 3,2 x 10^{-7} cm²/seg na condição sem aditivos, 3,3 x 10^{-7} cm²/seg na presença de 1mM ATP e 2,9 x 10^{-7} cm²/seg na presença de 1mM ADP. Na figura os símbolos representam os valores obtidos em cada concentração de proteína e as retas representam os ajustes lineares. Os dados de DLS da proteína Hsp82 foram coletados no tampão TN150.
5.1.6 Ultracentrifugação analítica – velocidade de sedimentação

Os experimentos de velocidade de sedimentação da proteína Hsp82 foram conduzidos em nove concentrações protéicas diferentes (entre 0,3 e 1,2 mg/mL). Como pode ser observado na figura 38, foram obtidos dados em quatro condições experimentais diferentes: sem aditivos e na presença de 200 μ M de ATP, ADP e ATP γ S. Os gráficos representam os coeficientes de sedimentação normalizados para condições de solvente e temperatura (água a 20 °C) para cada concentração protéica, valor conhecido como S_{20,W}. O ajuste dos valores de S_{20,W} com uma regressão linear permitiu a obtenção dos valores do S⁰_{20,W}, ou seja, o coeficiente de sedimentação também normalizado quanto à concentração protéica.





Figura 38: Ultracentrifugação analítica da proteína Hsp82 de cana-de-açúcar (velocidade de sedimentação). A) Experimento de velocidade de sedimentação na condição nativa, que apresentou um coeficiente de sedimentação de aprox. 5,3 S. B) Velocidade de sedimentação na condição com ATP, que apresentou um valor de aprox. 6,2 S. C) Velocidade de sedimentação na condição com ATPγS, que apresentou um valor de aprox. 5,4 S. D) Velocidade de sedimentação na condição nativa, que apresentou um valor de aprox. 5,8 S.

D)

C)

5.1.7 Modelagem molecular por homologia

Para a modelagem por homologia da proteína Hsp82 de cana-de-açúcar foram utilizadas as estruturas cristalográficas 1AH6, 2CGE e 2CG9 (figura 39), que correspondem ao domínio N-terminal, ao domínio intermediário mais o domínio C-terminal e finalmente à Hsp90 completa de *Saccharomyces cerevisiae*. Inicialmente foi realizado, para a construção do modelo, um alinhamento das seqüências através do programa CLUSTAW, sendo este depois corrigido manualmente. Foram gerados 40 modelos através do programa MODELLER (versão 7.7), que posteriormente foram analisados e selecionados através do programa PROCHECK. A figura 40 apresenta o melhor modelo por homologia da proteína Hsp82 de cana-de-açúcar e também o gráfico de Ramachandran, mostrando os ângulos torsionais Phi e Psi entre seus resíduos. Este modelo apresentou 88,2 % dos resíduos nas regiões mais favorecidas (548 aminoácidos), 9,3 % dos resíduos nas regiões adicionalmente permitidas (58 aminoácidos), 1,9 % dos resíduos nas regiões generosamente permitidas (12 aminoácidos) e 0,5 % dos resíduos nas regiões proibidas (3 aminoácidos).



Figura 39: Estruturas cristalográficas da Hsp90 de *S. cerevisiae* utilizadas na modelagem da proteína Hsp82. A) N-terminal da proteína Hsp90 de *S. cerevisiae* (PDB: 1AH6); B) Domínios intermediário e C-terminal da proteína Hsp90 de levedura (PDB: 2CGE); C) Estrutura completa da proteína Hsp90 de levedura (PDB: 2CGE); A identidade entre a seqüência da Hsp90 de levedura e a de cana-de-açúcar é de aproximadamente 70 %.



C) N-terminal: MGSSHHHHHHSSGLVPRGSHM C-terminal: IDEDEAPEADTDMPPLEDDAGESKMEEVD

Figura 40: Modelo por homologia da proteína Hsp82 e gráfico de Ramachandran mostrando os ângulos Phi e Psi. A) Modelo do tipo fita da proteína Hsp82 de cana-de-açúcar, com as cadeias laterais dos sete resíduos de triptofano indicadas em azul. B) Gráfico de Ramachandran indicando os ângulos torcionais Phi e Psi de toda a proteína. Os resíduos indicados em vermelho correspondem aos resíduos pertencentes às regiões generosamente permitidas (12 resíduos) e proibidas (3 resíduos). C) Seqüências N e C-terminais que não estão incluídas no modelo construído. A seqüência N-terminal corresponde à cauda adicionada pelo vetor de expressão pET28a e a seqüência C-terminal não possui correspondente estrutural no *Protein Data Bank*. Juntos estes dois fragmentos correspondem a 50 resíduos (aprox. 5,5 kDa), que representam 6,8 % de todos os resíduos da Hsp82.

113

5.2 Discussão Hsp82 de cana-de-açúcar

As proteínas da família da Hsp90 são de grande importância para os eucariotos, uma vez que pelo menos uma isoforma citoplasmática é essencial para a sobrevivência destes organismos em qualquer condição ambiental (Borkovich et al., 1989; Buchner 1999). Além disto, a Hsp90 é uma das proteínas mais abundantes no citoplasma, podendo representar até 1% de todo o conteúdo protéico da célula mesmo na ausência de choque térmico (Lai et al., 1984). Em eucalipto, foram encontrados 10 clusters codificando isoformas citoplasmáticas que, em conjunto com outras següências relacionadas à família Hsp90, representam 18% de todo conteúdo de chaperonas moleculares expressos por esta espécie (Cagliari et al., 2005). Já a cana-de-açúcar apresenta 4 clusters codificando Hsp90, podendo expressar até 9 proteínas diferentes (Borges et al., 2006). Tanto em cana como em eucalipto, foram encontrados ESTs relacionados à família Hsp90 em todos os tecidos estudados, o que destaca a importância destas proteínas para todo o organismo e não apenas a um tecido específico. Recentemente, a Hsp90 foi implicada no estresse biótico por interagir com produtos de genes de resistência em plantas (Hubert et al., 2003; Liu et al., 2004; Lu et al., 2003). Dada a importância da família Hsp90 como um todo e as possibilidades biotecnológicas no desenvolvimento de cultivares mais resistentes, o estudo da Hsp82 de cana-de-açúcar se torna de grande interesse científico e comercial. O presente estudo visa não só estabelecer as condições de expressão e purificação para a proteína Hsp82, mas também determinar suas características bioquímicas e a relação destas com o desempenho de suas funções.

5.2.1 Expressão e purificação

A proteína Hsp82 foi expressa em grandes quantidades e purificada com relativa facilidade. Inicialmente, a purificação apenas por afinidade e exclusão molecular apresentou um contaminante importante com massa molecular um pouco inferior à Hsp82. A inserção de uma etapa de purificação por troca iônica antes da exclusão molecular resolveu o problema, pois este contaminante foi eluído em uma concentração diferente de NaCl em relação à Hsp82. A proteína foi eluída em duas frações distintas na purificação por exclusão molecular, talvez representando o dímero e o monômero, uma vez que as proteínas da família Hsp90 formam homodímeros (Radanyi *et al.*, 1989; Chadli et al.,2000).

5.2.2 Dicroísmo circular

O espectro de CD da Hsp82 de cana-de-açúcar revelou que a proteína se encontra estruturada e que a adição de nucleotídeos não provocou mudanças significativas em sua estrutura. A principal diferença entre a condição sem aditivos e as condições com nucleotídeos foi a diminuição do sinal correspondente às hélices alfa e o aumento do sinal correspondente às folhas beta. Apesar destas alterações, não foi possível notar nenhuma alteração na estabilidade térmica da proteína Hsp82 provocada pela presença dos nucleotídeos. Em todas as condições estudadas, a Hsp82 apresentou uma grande estabilidade térmica, não sendo possível atingir seu desenovelamento mesmo a 90 °C. Devido à importância destas proteínas, necessárias para o crescimento de qualquer organismo eucarioto sob qualquer condição ambiental (Borkovich *et al.*, 1989), e a condição séssil das plantas, é esperado que possuam uma estabilidade térmica como a observada para a Hsp82 de cana-de-açúcar.

5.2.3 Fluorescência

Os dados de fluorescência estática da proteína Hsp82 de cana-de-açúcar foram coletados na ausência e presença dos nucleotídeos ATP, ADP e AMP. Não foi notada nenhuma alteração significativa no centro de massa espectral entre as condições estudadas. No entanto, como a proteína Hsp82 possui 5 resíduos de triptofano, é possível que eventuais mudanças locais ocorridas acabem por se "anular" em termos do perfil espectral resultante. Assim, uma eventual mudança no ambiente químico de um resíduo pode ser compensada pela mudança em outro, fazendo com que estas alterações não sejam notadas.

5.2.4 Estudos de estabilidade empregando uréia

A proteína Hsp82 de cana-de-açúcar foi submetida a experimentos de desenovelamento químico, monitorado pelas sondas de dicroísmo circular e fluorescência, coletados na ausência e na presença dos nucleotídeos ATP, ADP e AMP. Apesar das pequenas diferenças observadas, nenhuma mudança significativa na estabilidade da proteína alvo foi observada quando da adição destes nucleotídeos nas duas sondas utilizadas. No entanto, a estrutura terciária da proteína Hsp82 parece ser afetada antes da secundária quando do incremento da concentração de uréia, como pode ser observado através da comparação entre os valores normalizados (delta de sinal) entre os experimentos de CD e fluorescência. Assim, em uma mesma concentração do agente desnaturante, temos um menor sinal de fluorescência em relação ao sinal de CD na maior parte das concentrações de uréia analisadas. Uma possibilidade é a de que apenas os nucleotídeos não sejam suficientes para promover a dimerização da região N-terminal, pois se o dímero se apresentasse em uma conformação completamente fechada seria mais compacto e mais resistente ao desenovelamento químico do que a condição sem aditivos. Desta forma, apesar de provocar pequenas mudanças em relação à condição sem aditivos, os nucleotídeos não são capazes de promover uma maior estabilização da proteína via dimerização do N-terminal. Entretanto, como vários eventos parecem ocorrer entre o estado nativo e o desenovelado, é possível que o "rompimento" da dimerização do domínio N-terminal esteja ocorrendo no início do processo de desenovelamento. Porém a semelhança entre os experimentos com e sem os nucleotídeos indica que isto provavelmente não está ocorrendo.

5.2.5 Ultracentrifugação analítica e espalhamento dinâmico de luz

A análise por espalhamento dinâmico de luz permitiu observar uma única espécie oligomérica, ou seja, as amostras da proteína Hsp82 de cana-de-açúcar se apresentaram monodispersas. Os dados de EDL para a proteína sem aditivos e na presença de ATP apresentaram valores de coeficiente de difusão ligeiramente maiores do que na condição com ADP. Isto poderia ser um indicativo de que nestas duas condições a proteína se apresenta mais compacta do que na condição com ADP, porém a diferença entre os valores não é significativa para permitir tal diferenciação. Garnier e colaboradores (2002) obtiveram um valor de $D^0_{20,W}$ igual a 3,3 x 10^{-7} cm²/seg para a proteína Hsp90 de porco, valor muito similar ao encontrado para a proteína Hsp82 na ausência de nucleotídeos (3,2 x 10^{-7} cm²/seg) e idêntico ao encontrado na presença de ATP (3,3 x 10^{-7} cm²/seg). Os valores de D para a proteína Hsp82 de cana-de-açúcar nestas duas condições se apresentaram, portanto, compatíveis com o conhecimento estabelecido até o momento para a família Hsp90.

Os dados de velocidade de sedimentação (ultracentrifugação analítica) foram coletados em quatro condições diferentes: sem aditivos e na presença de 200 μ M dos nucleotídeos ATP, ADP e ATP γ S. Os valores de coeficiente de sedimentação obtidos experimentalmente foram de 5,3 ± 0,1 S para condição sem aditivos, 6,2 ± 0,5 S para condição com ATP, 5,8 ± 0,3 para a condição com ADP e 5,4 ± 0,5 S para a

condição com ATPγS. Estes valores obtidos para a proteína Hsp82 são compatíveis com os observados para proteínas Hsp90 de outros organismos, tais como 6,2 S para a Hsp90 de camundongo (Koyasu *et al.*, 1986), 5,6 S para a HtpG de *Escherichia coli* (Spence e Georgopoulos, 1989) e 6,1 S para a Hsp90 citoplasmática de porco (Garnier *et al.*, 2002) (valores aproximadas). É importante salientar que todas estas medidas citadas de coeficiente de sedimentação foram obtidas na ausência de nucleotídeos e que, no caso da Hsp90 de porco, o objetivo era identificar o efeito de cátions divalentes sobre a oligomerização daquela proteína. Portanto, acreditamos que o estudo da proteína Hsp82 de cana-de-açúcar na presença de nucleotídeos por ultracentrifugação analítica, somados aos de EDL e outros, acrescentará importantes informações a respeito das proteínas da família da Hsp90.

Apesar das variações observadas entre as condições com e sem nucleotídeos nos experimentos de velocidade de sedimentação, é possível observar uma tendência de aumento do coeficiente de sedimentação quando da adição de nucleotídeos, principalmente na condição com ATP. O dímero da Hsp90 de levedura (código PDB: 2CG9), o qual foi obtido através de uma co-cristalização com a proteína Sba1 (homóloga da co-cochaperona humana p23) e o nucleotídeo AMPPNP, apresentou uma conformação chamada de "fechada" (Maruf et al., 2006). Esta conformação recebeu este nome devido ao fato de que as proteínas Hsp90 apresentam uma conformação alongada e constantemente dimerizada através do seu domínio C-terminal. Assim, quando ocorre a concomitante dimerização do domínio N- terminal oposto, afirma-se que o dímero adotou um estado conformacional "fechado". A utilização Hydropro (http://leonardo.fcu.um.es/macromol/programs/hydropro/ do programa hydropro.htm), permite estimar o valor do coeficiente de sedimentação de proteínas a partir de seus arquivos de estrutura em formato pdb. Deste modo, após a remoção dos dados relativos às cochaperonas, o arguivo correspondente à proteína Hsp90 de levedura (PDB 2CG9) foi submetido à análise com o programa Hydropro. Curiosamente, o coeficiente de sedimentação teórico da estrutura cristalográfica obtida foi de aproximadamente 7,1 S. Este valor pode ser utilizado como parâmetro de comparação com a proteína de cana-de-açúcar uma vez que ambas possuem praticamente o mesmo tamanho e, provavelmente, a proteína Hsp82 também se encontra dimerizada pela região C-terminal. Assim, um dímero completamente "fechado" da proteína Hsp82 apresentaria um valor acima de 7 S nas condições com nucleotídeo, fato que não foi corroborado por nossos experimentos de

ultracentrifugação analítica. Apesar de termos observado um ligeiro aumento dos valores de coeficiente de sedimentação na condição com ATP (6,2 S) em relação à proteína sem aditivos, este ainda se encontra inferior ao valor para o dímero totalmente fechado (aproximadamente 7 S). Como a estrutura cristalográfica foi obtida na presença de uma proteína homóloga à p23 humana, podemos especular que a ligação a esta proteína, e não somente à ligação aos nucleotídeos, seja necessária para provocar o fechamento completo do dímero da Hsp90.

Considerando-se os dados de velocidade de sedimentação (ultracentrifugação analítica) e espalhamento dinâmico de luz, podemos afirmar que a proteína Hsp82 se apresentou como um dímero em todas as condições estudadas. Este resultado é compatível com o observado para outras proteínas da família Hsp90 (Radanyi *et al.*, 1989; Pearl e Prodromou, 2006).

5.2.6 Modelagem molecular por homologia

Para a construção do modelo por homologia do monômero da proteína Hsp82 foi utilizado o programa MODELLER (http://www.salilab.org/modeller/). Este programa aplica restrições espaciais, obtidas a partir de estruturas em alta resolução fornecidas, para gerar um modelo com base em um alinhamento entre seqüências de polipeptídeos diferentes. As restrições espaciais impostas ao modelo da Hsp82 foram obtidas das estruturas cristalográficas 1AH6, 2CGE e 2CG9, as quais correspondem a porções da Hsp90 levedura. A identidade entre a seqüência da Hsp82 de cana-de-açúcar e a Hsp90 de levedura é de aproximadamente 70 %, o que permitiu construir um bom modelo já que o mínimo recomendável para tais rotinas é algo em torno de 30 % (http://www.salilab.org/modeller/FAQ.html). Além disto, a escolha de mais de uma estrutura melhora a qualidade dos modelos obtidos, pois o programa dispõe de mais restrições para modelar a sequência de interesse. Segundo o programa PROCHECK (http://www.biochem.ucl.ac.uk/ ~roman/procheck/procheck.html), a análise de 118 estruturas resolvidas a pelo menos 2,0 Å e com um *R-factor* (fator de confiança: mede a discrepância entre o modelo cristalográfico e os dados experimentais de difração de raios-X) menor do que 20 % permite esperar que um bom modelo apresente mais de 90 % dos resíduos nas regiões mais favorecidas do gráfico de Ramachandran (indicador dos ângulos torsionais Phi e Psi). No caso do modelo da proteína Hsp82 obtivemos 88,2 % dos resíduos nesta região mais favorável. No entanto, isto se deve ao fato de que apenas a estrutura 1AH6 possui uma resolução inferior a 2,0 Å, enquanto que as estruturas 2CGE e 2CG9 apresentam uma resolução de 3,0 e 3,1 Å, respectivamente. As três estruturas apresentaram a seguinte porcentagem de resíduos nas regiões mais favorecidas do gráfico de Ramachandran: 91,8 % para 1AH6, 83,5 % para 2CGE e 69,5 % para 2CG9. Desta forma o modelo obtido, no que se refere aos ângulos torcionais da cadeia principal, apresenta uma boa qualidade em relação às estruturas utilizadas como modelo.

5.3 Conclusões Hsp82 de cana-de-açúcar

- O clone da Hsp82 de cana-de-açúcar foi totalmente seqüenciado, apresentando uma identidade de 70 % em relação a proteínas Hsp90 de Saccharomyces cerevisiae.
- A caracterização da proteína Hsp82 por dicroísmo circular mostrou que a proteína se encontra estruturada. A adição dos nucleotídeos não provoca alterações significativas nas quantidades dos elementos de estrutura secundária.
- Os cinco resíduos de triptofano da proteína Hsp82 se apresentaram parcialmente protegidos do solvente, segundo os espectros de emissão de fluorescência. A adição dos nucleotídeos ATP, ADP e AMP não foi capaz de alterar o centro de massa espectral em relação à condição sem aditivos.
- O desenovelamento térmico acompanhado por CD mostrou que a proteína Hsp82 apresenta alta estabilidade térmica, pois não sofreu desenovelamento mesmo a 90 °C.
- A proteína Hsp82 não apresentou diferenças de estabilidade entre as condições estudadas por desenovelamento químico na presença de uréia (CD e fluorescência). No entanto, a comparação entre as duas sondas permitiu observar que a estrutura terciária foi afetada antes da secundária quando do incremento da concentração de uréia.
- A proteína Hsp82 se mostrou monodispersa nos experimentos de espalhamento dinâmico de luz, apresentando um coeficiente de difusão de aproximadamente 3,2 x 10⁻⁷ cm²/seg. Os nucleotídeos ATP e ADP praticamente não alteraram o D em relação à proteína sem ligantes.
- Os experimentos de velocidade de sedimentação permitiram observar um pequeno aumento do coeficiente de sedimentação nas condições com nucleotídeos (ATP, ADP, AMP e ATPγS) em relação à proteína sem aditivos.
- O seqüenciamento da Hsp82 de cana-de-açúcar permitiu a construção de um modelo por homologia baseado na identidade com a proteína Hsp90 de *S. cerevisiae*.

6 Co-chaperona humana p23

6.1 Resultados co-chaperona p23

6.1.1 Clonagem e expressão da co-chaperona p23

O cDNA correspondente a co-chaperona humana p23 foi amplificado por PCR com oligonucleotídeos específicos, e em seguida ligados ao vetor de transferência p-GEM T-easy. A figura 41 mostra um gel de agarose (1%) corado com brometo de etídio mostrando a digestão dos plasmídeos p-GEM T-easy contendo os cDNAs que codificam as co-chaperonas com a enzima de restrição *Eco*RI.



Figura 41: Clonagem do cDNA da co-chaperona p23 no vetor de transferência p-GEM T-easy. O produto de PCR amplificado a através de oligonucleotídeos específicos, foram ligados ao vetor p-GEM T-easy conforme protocolo do fabricante. Gel de agarose 1% corado com brometo de etídio mostrando as seguintes amostras: P) Padrão 500 pb; 1) vetor p-GEM T-easy-p23, 2) plasmídeo digerido com a enzima *Eco*RI.

O cDNA da co-chaperona foi subclonado no vetor de expressão pET-28a através da digestão do respectivo vetor de transferência p-GEM T-easy com as seguintes enzimas: *Bam*HI e *Eco*RI. Após a digestão o cDNA foi ligado ao vetor pET-28a e o vetor resultante foi confirmado através de seqüenciamento. A figura 42 apresenta a seqüência de aminoácidos da co-chaperona p23, bem como os aminoácidos codificados pelo vetor pET-28a.

```
pET28a-p23

1 MGSSHHHHHH SSGLVPRGSH MASMTGGQQM GRGSMQPASA KWYDRRDYVF IEFCVEDSKD 60

61 VNVNFEKSKL TFSCLGGSDN FKHLNEIDLF HCIDPNDSKH KRTDRSILCC LRKGESGQSW 120

121 PRLTKERAKL NWLSVDFNNW KDWEDDSDED MSNFDRFSEM MNNMGGDEDV DLPEVDGADD 180

181 DSQDSDDEKM PDLE 194
```

Figura 42: Seqüência de aminoácido da co-chaperona p23. A seqüência de aminoácido em azul corresponde aos aminoácidos codificados pelo sítio de clonagem múltipla do vetor pET28a, já a seqüência que corresponde a co-chaperona está em preto. O vetor vetor pET28a-p23 codifica um polipeptídio com massa molecular de 22,2 kDa (194 resíduos).

O vetor de expressão obtido foi transformado por choque térmico em células competentes *E. coli* BL21 (DE3), e induzidas a 37 °C por quatro horas na presença do agente indutor IPTG. A figura 43 apresenta um gel de poliacrilamida mostrando a expressão da co-chaperona p23.



Figura 43: Expressão da co-chaperona p23 humana. A cultura contendo as bactérias transformadas com o vetor pET28a-p23 foi induzidas com 0,5 mM de IPTG por quatro horas a 37 °C e sob agitação de 200 rpm. O gel de poliacrilamida (10 %) mostra a expressão (seta vermelha), e o tempo de indução. A massa molecular aparente da co-chaperona p23 corresponde ao esperado (aprox. 22 kDa), como pode ser observado através da comparação com as proteínas do padrão de massa molecular.

6.1.2 Purificação da co-chaperona p23 humana

A proteína p23 foi expressa fusionada a uma cauda de Histidinas (vide figura 42), o que permitiu sua purificação por cromatografia de afinidade. O sobrenadante do lisado celular contendo a proteína p23-His foi purificado na coluna *Hitrap Chelating*[®] *Global 5 mL*. A purificação resultou três frações protéicas distintas, sendo que a proteína p23-His foi eluída majoritariamente na fração três. A figura 44 apresenta o cromatograma de purificação contendo as frações eluídas e o gel de poliacrilamida da fração 3, que contem a proteína p23-His.



Figura 44: Purificação da proteína p23-His humana por afinidade. A) Cromatograma da purificação da proteína p23-His indicando a absorbância a 280 nm (azul) e a concentração de imidazol (vermelho). As proteínas foram eluídas em três frações distintas, e a proteína p23-His se encontra predominantemente na terceira fração. B) Gel de poliacrilamida (12%) contendo as alíquotas que compõe a fração três. A proteína p23-His está indicada pela seta, possuindo uma massa molecular de 23 kDa. Proteínas com massa molecular indicada pertencem ao padrão de massa molecular.

A)

Visando eliminar possíveis agregados da amostra, a fração previamente purificada por afinidade foi submetida a uma purificação por exclusão molecular. Utilizando uma coluna Superdex 200 Hiload 26/60 a proteína p23-His foi purificada, sendo eluída predominantemente na segunda fração da purificação. A figura 45 apresenta o cromatograma e o gel de poliacrilamida da purificação por CEM.



Figura 45: Purificação da proteína p23-His humana por cromatografia exclusão molecular. A) Cromatograma de eluição acompanhado por absorbância a 280 nm. B) gel de poliacrilamida (12%) indicando as alíquotas correspondentes às frações apontadas no cromatograma. Proteínas com massa molecular indicada pertencem ao padrão de massa molecular.

6.1.3 Caracterização da proteína p23 humana por dicroísmo circular

A proteína p23-His foi submetida a ensaios de dicroísmo circular, visando estudar seus elementos de estrutura secundária. A figura 46 apresenta o espectro da proteína p23-His por CD, o qual apresentou uma banda na região de 230 nm que provavelmente corresponde aos 5 triptofanos presentes na seqüência de aminoácidos. O espectro também apresentou uma banda em aproximadamente 206 nm, a qual possivelmente está relacionada à estrutura em folha beta pregueada presente nesta proteína.



Figura 46: Espectro de dicroísmo circular da proteína p23-His humana. Dados foram coletados no tampão HEP500 em uma concentração de proteína de aproximadamente 6mg/mL e cubeta circular com caminho óptico de 0,1 mm.

Os resultados das análises dos espectros pelo programa "CDNN deconvolution" estão mostrados na tabela V. A proteína p23 apresenta, predominantemente, estruturas do tipo folha beta antiparalela, e pequeno percentual de estruturas do tipo alfa-hélice.

Tabela V – Deconvolução do espectro de CD da proteína p23-His utilizando o programa CDNN (erro de aproximadamente 10 %).

	180-260	185-260	190-260	195-260	200-260	205-260	210-260
Alfa-hélice	13%	8%	7%	8%	7%	4%	4%
Folha antiparalela	50%	54%	50%	42%	42%	44%	45%
Folha paralela	2%	4%	4%	5%	6%	5%	5%
Volta beta	19%	19%	17%	19%	20%	19%	19%
Randômica	26%	27%	29%	33%	37%	35%	35%
Total	110%	112%	107%	107%	112%	107%	108%

O perfil de desenovelamento térmico da proteína p23-His foi monitorado entre 20°C e 90°C e apresentou um comportamento não reversível. O desenovelamento foi acompanhado no comprimento de onda de 222 nm, exibindo um pequeno ganho de sinal em temperaturas superiores a 50°C (figura 47).



Figura 47: Desenovelamento térmico da proteína p23-His monitorado por CD em 222 nm. A proteína p23-His no tampão HEP500 apresentou um pequeno ganho de sinal em temperaturas superiores a 50 °C, em uma velocidade de aquecimento de 60°C/hora. Quadrados: dado experimental; linha vermelha: ajuste dos dados.

O monitoramento do espectro em diferentes temperaturas permitiu detectar mudanças importantes no mesmo. A banda em 230 nm, correspondente aos triptofanos, deixou de existir em temperaturas superiores a 50°C, enquanto que a banda em 222 nm apresentou um aumento significante em temperaturas mais elevadas, como pode ser observado na figura 48.



Figura 48: Desenovelamento térmico monitorado por CD da proteína p23-His. Com o aumento da temperatura a banda em 230 nm perdeu gradativamente a intensidade, enquanto que a banda em 222 nm apresentou um aumento nas temperaturas mais elevadas.

A proteína p23-His foi digerida com a protease trombina visando eliminar uma possível influência da cauda de histidinas na caracterização por dicroísmo circular, tendo em vista a pequena massa molecular da p23. A figura 49 apresenta o espectro da p23 clivada em três concentrações diferentes.



Figura 49: Espectros da proteína p23 clivada com trombina. A co-chaperona p23 digerida com trombina foi submetida à análise por CD em três concentrações diferentes: 1, 2 e 3 mg/mL.

Tabela VI - Predição do conteúdo de estrutura secundária da p23 clivada, na concentração de 1 mg/mL, no tampão TNB100 (erro de aproximadamente 10 %).

	190-260	195-260	200-260	205-260	210-260
Alfa-hélice	7%	8%	7%	6%	7%
Folha antiparalela	29%	30%	32%	33%	35%
Folha paralela	3%	4%	5%	5%	5%
Volta beta	24%	24%	25%	23%	20%
Randômica	37%	36%	38%	35%	36%
Total	100%	102%	107%	102%	103%

Para suprimir a influência dos triptofanos no espectro da proteína p23 foi adicionado o reagente N-bromosuccinamida (NBS), capaz de se ligar à cadeia lateral dos triptofanos e "apagar" o sinal correspondente a estes resíduos no dicroísmo circular. A figura 50 mostra os espectros de CD da proteína p23 clivada com trombina nos quais concentrações crescentes de NBS foram adicionadas. A partir de NBS a 5 mM a banda em 230 nm começa a ser suprimida e com 10 mM NBS o efeito dos triptofanos sobre o espectro é totalmente apagado, passando a exibir um aumento da banda na região de 220 nm e, principalmente, na região de 210 nm.



Figura 50: Espectros de CD da proteína p23 clivada incubada com reagente NBS. A proteína p23 no tampão TNB100 e concentração de 1 mg/mL foi incubada com 1, 5 e 10 mM de NBS. Na concentração de 10 mM NBS o sinal correspondente aos triptofanos, em 230 nm, foi totalmente suprimida.

	195-260	200-260	205-260	210-260
Alfa-hélice	8%	7%	7%	7%
Folha antiparalela	29%	31%	31%	34%
Folha paralela	4%	5%	5%	5%
Volta beta	25%	26%	24%	19%
Randômica	37%	39%	35%	35%
Total	103%	108%	102%	100%

Tabela VII - Predição do conteúdo de estrutura secundária da p23 clivada, na concentração de 1 mg/mL, no tampão TNB100 contendo 1mM NBS (erro de aproximadamente 10 %).

Tabela VIII - Predição do conteúdo de estrutura secundária da p23 clivada, na concentração de 1 mg/mL, no tampão TNB100 contendo 5mM NBS (erro de aproximadamente 10 %).

	195-260	200-260	205-260	210-260
Alfa-hélice	9%	9%	8%	9%
Folha antiparalela	28%	29%	29%	32%
Folha paralela	5%	5%	5%	6%
Volta beta	24%	25%	23%	19%
Randômica	36%	38%	35%	35%
Total	102%	106%	100%	101%

Tabela IX - Predição do conteúdo de estrutura secundária da p23 clivada, na concentração de 1 mg/mL, no tampão TNB100 contendo 10mM NBS (erro de aproximadamente 10 %).

	200-260	205-260	210-260
Alfa-hélice	23%	28%	30%
Folha antiparalela	13%	13%	14%
Folha paralela	5%	6%	6%
Volta beta	22%	21%	18%
Randômica	33%	32%	32%
Total	96%	100%	100%

O perfil do desenovelamento térmico da proteína p23 clivada com trombina apresentou uma perda de sinal em 230 nm na medida em que a temperatura é elevada, como pode ser observado na figura 51.



Figura 51: Desenovelamento térmico monitorado por CD da proteína p23 clivada com trombina. O aumento da temperatura provocou a perda gradativa de intensidade do sinal em 230 nm, correspondente aos resíduos de triptofano presentes na proteína p23.

Na presença de 10 mM NBS a p23 clivada com trombina não apresentou alterações no perfil de CD em função do aumento de temperatura (figura 52).



Figura 52: Desenovelamento térmico monitorado por CD da proteína p23 clivada na presença de 10 mM NBS. O aumento da temperatura praticamente não alterou o espectro de CD da proteína p23 na presença de NBS.

6.1.4 Caracterização da proteína p23 humana por fluorescência

O espectro de emissão de fluorescência da p23-His apresentou um máximo de emissão em aproximadamente 336 nm (figura 53). Já os dados da proteína p23 clivada foram praticamente idênticos (não mostrados).



Figura 53: Espectro de emissão de fluorescência da proteína p23-His. O espectro apresentou um máximo em 336 nm e intensidade relativa de aproximadamente 5,5 (unidades arbitrárias). Condições experimentais: proteína p23-His a 50 μg/mL no tampão HEP100 e excitação em 295 nm.

6.1.5 Desenovelamento químico da proteína p23 monitorado por dicroísmo circular e fluorescência

A proteína p23 humana clivada com trombina foi submetida ao desenovelamento químico por uréia e cloreto de guanidina, visando estabelecer seus parâmetros de estabilidade. Para realizar este experimento utilizamos aproximadamente 0,2 mg/mL de proteína e soluções estoque de uréia a 8 Molar e cloreto de guanidina a 7 Molar diluídas no próprio tampão de diálise (TNB100). A concentração destes agentes desnaturantes foi aferida em um refratômetro, o que permitiu que erros na diluição dos mesmos fossem corrigidos posteriormente. A figura 54 apresenta o desenovelamento por uréia da proteína p23 com pontos em intervalos de 0,5 Molar de agente desnaturante, abrangendo desde 0 até 7 Molar de uréia, monitorados por CD. Os dados representam o desenovelamento acompanhado por dois comprimentos de onda distintos, 230 e 222 nm. Já a figura 55 apresenta os dados de desenovelamento da co-chaperona p23 com cloreto de guanidina e, da mesma forma que os dados obtidos para uréia, foram monitorados os comprimentos de onda de 230 e 222 nm.



Figura 54: Desenovelamento por uréia da proteína p23 clivada. A) Espectro de dicroísmo circular em cada concentração de uréia analisada. B) Desenovelamento por uréia, acompanhado por dois comprimentos de onda distintos: 222 e 230 nm. A proteína começou a perder estrutura secundária a partir de 2,5 Molar de uréia, perdendo praticamente toda a estrutura em aproximadamente 5 Molar do agente desnaturante. Dados foram coletados no tampão TNB100.



Figura 55: Desenovelamento por cloreto de guanidina da proteína p23 clivada. A) Espectro de dicroísmo circular em cada concentração de cloreto de guanidina analisada. B) Desenovelamento por cloreto de guanidina, acompanhado por dois comprimentos de onda distintos: 222 e 230 nm. Dados foram coletados no tampão TNB100.

Foram obtidos também dados de desenovelamento por fluorescência estática da proteína p23 utilizando uréia e cloreto de guanidina, ambos nas mesmas concentrações observadas por CD. Os triptofanos foram excitados a 295 nm e o espectro de emissão coletado no intervalo entre 300 e 400 nm. Cada conjunto de dados representa a acumulação de 3 leituras e a média de dois experimentos independentes. A figura 56 apresenta o deslocamento do máximo de emissão de fluorescência provocado pelo desenovelamento da proteína por uréia, e a figura 57 mostra os mesmos dados para o desenovelamento com cloreto de guanidina.



Figura 56: Desenovelamento por uréia da proteína p23 clivada monitorado por fluorescência. A) Espectro de emissão de fluorescência em função da concentração de uréia. O aumento da concentração de uréia provocou um deslocamento do máximo de emissão de fluorescência, indicando a exposição dos triptofanos ao solvente. B) Variação do centro de massa espectral em função do aumento da concentração de uréia, partindo da condição sem uréia, em que a proteína se apresenta enovelada, até 7 Molar de uréia, condição em que se encontra desenovelada. Dados coletados no tampão TNB100.



Figura 57: Desenovelamento por cloreto de guanidina da proteína p23 clivada monitorado por fluorescência. A) Espectro de emissão de fluorescência em função da concentração de cloreto de guanidina. O aumento da concentração deste agente desnaturante provocou um deslocamento do máximo de emissão de fluorescência, indicando a exposição dos triptofanos ao ambiente do solvente. B)

Variação do centro de massa espectral em função do aumento da concentração do cloreto de guanidina, de 0 a 7 Molar. Dados coletados no tampão TNB100.

Através da variação de sinal de CD e de fluorescência, tanto para uréia como para o cloreto de guanidina, foi possível comparar os efeitos destes agentes desnaturantes sobre a proteína p23 clivada. A figura 58 apresenta os dados normalizados de CD (230 nm) e fluorescência para os dois agentes desnaturantes.



Figura 58: Variação do sinal de CD e fluorescência em função da concentração de uréia e cloreto de guanidina. A) Diferença de sinal no desenovelamento por uréia monitorado por CD a 230 nm e fluorescência. O valor de C_m , concentração de uréia no ponto médio do desenovelamento, foi de 3,6 M para a fluorescência e 3,3 M para os dados de CD (valores aproximados). B) Diferença de sinal no desenovelamento por cloreto de guanidina monitorado por CD a 230 nm e fluorescência. O valor de C_m , concentração de cloreto de guanidina monitorado por CD a 230 nm e fluorescência. O valor de C_m , concentração de cloreto de guanidina monitorado por CD a 230 nm e fluorescência. O valor de C_m , concentração de cloreto de guanidina no ponto médio do desenovelamento, foi de 1,4 M para a fluorescência e 1,1 M para os dados de CD (valores aproximados).
6.1.6 Análise hidrodinâmica da proteína p23

A proteína p23-His foi submetida a análises por espalhamento dinâmico de luz para a determinação de seu coeficiente de difusão. A figura 59 apresenta os valores de D experimental para três concentrações diferentes da proteína, bem como os respectivos $D_{20,W}$ (coeficiente de difusão em água e 20°C) e o ajuste linear dos dados, com o qual foi possível calcular o valor de $D_{20,W}^{o} = 8.8 \times 10^{-7} \pm 0.05 \text{ cm}^2/\text{seg}.$



Figura 59: Coeficientes de difusão obtidos para a proteína p23-His por espalhamento dinâmico de luz. A proteína p23, no tampão HEP500, foi submetida a experimentos de EDL em três concentrações distintas, o que permitiu calcular o valor de $D^{o}_{20,W}$ por meio de uma regressão linear. Pontos pretos: dado experimental; pontos vermelhos: valores de $D_{20,W}$ calculados a partir dos respectivos dados experimentais; retas vermelhas: ajuste dos dados.

A proteína p23-His também foi submetida a experimentos de velocidade de sedimentação (VS) por ultracentrifugação analítica. A figura 60 mostra é possível notar que a proteína apresenta uma fração monomérica predominante e uma pequena fração oligomérica (setas). A análise dos dados de $S_{20,W}$ permitiu inferir os valores de $S_{20,W}^{o}$ para o monômero (2,30 ± 0,07 S) e para o oligômero (4,6 ± 0,02 S) da

proteína p23-His. A figura 61 mostra os dados de massa molecular em função da concentração de proteína para o monômero e também para o oligômero.



Figura 60: Experimentos de velocidade de sedimentação (VS) da proteína p23-His. A) Ajuste dos dados em várias concentrações de proteína, no tampão HEP500, utilizando o programa SEDFIT. É possível

notar, além do monômero, uma pequena fração oligomérica com aproximadamente 4 S. B) Análise do $S_{20,W}$ em função da concentração para cada uma das espécies encontradas, permitindo o calculo do valor de $S_{20,W}^0$. Os experimentos foram realizados a uma velocidade de 50000 RPM, temperatura de 20 °C e coleta de dados em 280 nm.



Figura 61: Análise da massa molecular por experimentos de velocidade de sedimentação (VS) da proteína p23-His. O monômero apresentou uma massa molecular de 23 kDa e o oligômero de 51 kDa (valores aproximados), em ambos os casos a variação da massa da partícula sofreu influência da concentração.

A proteína p23 clivada com trombina também foi analisada por velocidade de sedimentação no tampão HEP500 em uma velocidade de 50000 RPM, temperatura de 20 °C e coleta de dados em 280 nm. Os dados foram tratados utilizando o programa SedFit utilizando como informações prévias o $D_{20,W}^{0}$, estimado pelo EDL (D = 8.4 x10⁻⁷ ± 0,1 cm²/seg), para o cálculo da c(s). A figura 62 apresenta o gráfico da c(s) do conjunto de dados nas diferentes concentrações medidas, sendo possível notar, assim como para proteína p23-His, a presença de duas espécies. A figura 63 mostra os dados de massa molecular das espécies observadas em função da concentração de p23 clivada.

A)





Figura 62: Experimentos de velocidade de sedimentação (VS) da proteína p23 clivada com trombina. A) Ajuste dos dados em várias concentrações de proteína, no tampão HEP500, utilizando o programa SEDFIT. A p23 clivada apresentou, assim como a p23-His, duas espécies distintas (setas). B) Valores de S_{20,W} em função da concentração para cada uma das espécies. O valor de S⁰_{20,W} para o monômero foi de 2,23 ± 0,08 e para o oligômero de 4,26 ± 0,05.



Figura 63: Análise da massa molecular por experimentos de velocidade de sedimentação (VS) da proteína p23 clivada. A) Ajuste dos dados em várias concentrações de proteína, no tampão TNB100, utilizando o programa SEDFIT. A p23 clivada apresentou, assim como a p23-His, duas espécies distintas (setas). B) Valores de massa molecular em função da concentração para cada uma das espécies. O monômero apresentou uma massa molecular de 19,5 ± 0,8 kDa, e o oligômero, que sofreu influência devido da concentração, de 47 ± 4 kDa.

A)

6.2 Discussão p23

Foi obtido o clone da proteína p23, a qual foi purificada em grandes quantidades e não apresentou nenhum problema de solubilidade.

6.2.1 Dicroísmo circular

A proteína p23 é um importante componente do complexo da Hsp90, participando diretamente da atividade chaperona de proteínas tais como receptores esteróides e DNA polimerase do vírus da hepatite B (Ritcher e Buchner, 2001; Young et al., 2001; Pratt e Toft, 2003). A estrutura do N-terminal (resíduos 1 a 110) foi resolvida por cristalografia de raios-X e mostrou a formação de um domínio rico em folhas β antiparalelas (Weaver et al., 2000). Baseado em dados de CD e de sensibilidade a proteases, têm se afirmado que os aminoácidos da região C-terminal, composta pelos últimos 50 resíduos, não apresenta estrutura definida (Weikl et al., 1999). Os espectros de CD da p23 realizados na ausência de N-bromosuccinamida (NBS), reagente que provoca o apagamento da banda correspondente aos triptofanos por CD (Nery et al., 2006), foram similares aos encontrados na literatura para esta proteína (Weikl et al., 1999). Porém, na presença de 10 mM NBS há um aumento do sinal na região entre 220 e 210 nm, indicando que o sinal dos triptofanos em 230 nm mascara os mínimos observados na presença de NBS. Weikl e colaboradores (1999) também analisaram por CD um peptídeo sintético correspondente aos últimos 37 resíduos da p23 completa, não encontrando nenhum elemento de estrutura secundária. Entretanto, a região N-terminal rica em folhas beta pode ser fundamental para a estruturação destas hélices localizadas no C-terminal da proteína.

6.2.2 Fluorescência

A proteína p23 possui cinco resíduos de triptofano localizados em seu domínio rico em folhas beta. A fluorescência apresentou um máximo de emissão em 336 nm, indicando que os resíduos de triptofano se encontram parcialmente protegidos do solvente. Como o esperado, a clivagem da cauda de Histidina não alterou o perfil do espectro de emissão, uma vez que os triptofanos se encontram reunidos no domínio N-terminal rico em folhas beta e não possuem relação estrutural com a cauda adicionada.

6.2.3 Estabilidade medida por CD e fluorescência

No que se refere ao desenovelamento térmico acompanhado por CD, a principal característica observada para a proteína sem NBS foi o apagamento da banda dos triptofanos, o que indicaria a exposição dos mesmos ao solvente. Além disto, o aquecimento até 90°C da p23 humana não foi suficiente para desenovelar a proteína segundo análise por CD. Esta estabilidade não foi observada no trabalho de Weikl e colaboradores (1999), pois a proteína selvagem começou a perder estrutura em torno de 50°C e o desenovelamento foi completado por volta de 70°C. No entanto, foi utilizado um tampão de fosfato de potássio (pH 6,8) enquanto que os dados apresentados aqui foram coletados em tampão Hepes 40 mM pH 7.5, NaCl 500 mM, β -mercaptoetanol 5 mM, e tampão Tris-HCl 25 mM pH 7.5, NaCl 100 mM, β -mercaptoetanol 5 mM. A presença de sal, bem como a do agente redutor β -mercaptoetanol, parecem aumentar sobremaneira a estabilização da proteína p23. Também é importante destacar que a adição do reagente NBS não alterou a estabilidade da p23.

Como os dados de desenovelamento térmico por CD não foram conclusivos, pois a proteína não foi completamente desenovelada mesmo a 90 °C, foram realizados desenovelamentos químicos por uréia e cloreto de guanidina. Não é de nosso conhecimento outro estudo sobre o desenovelamento químico desta proteína na literatura, apenas o desenovelamento térmico. Através da análise da variação de sinal foi possível notar que a proteína p23 apresentou dois estados na via de enovelamento, nativo e desenovelado. O cálculo de C_m , valor de concentração do agente desnaturante no ponto médio da curva de desenovelamento, permitiu observar que foi necessária uma maior concentração de agente desnaturante para afetar o ambiente dos triptofanos monitorado por fluorescência do que o sinal de estrutura secundária monitorado por CD. O esperado seria a proteína perder primeiro sua estrutura terciária e depois a secundária, portanto o valor de C_m para a sonda terciária deveria ser inferior ao da sonda secundária. Porém, como a banda em 230 nm é muito influenciada por resíduos aromáticos, esta aparente inversão poderia se dever a perturbações causadas por estes resíduos (9 fenilalaninas, 2 tirosinas e 5 triptofanos).

6.2.4 Ultracentrifugação analítica e espalhamento dinâmico de luz

Utilizando dados de velocidade de sedimentação e espalhamento dinâmico de luz, foi possível avaliar o estado oligomérico desta proteína. Tanto a p23 com cauda de histidina, como a p23 clivada com trombina, apresentaram uma fração monomérica predominante e um discreto oligômero. Ensaios de ultracentrifugação analítica (Weaver et al., 2000), espectrometria de massas e SAXS (McLaughlin et al., 2006), mostraram que mesmo na presença de agentes redutores a fração dimérica está presente. Além disto, a estrutura cristalográfica da p23∆c-term se apresentou como um dímero ligado por uma ponte dissulfeto intermolecular entre as cisteínas 58 (Weaver et al., 2000). Portanto, os dados referentes à oligomerização e massa molecular do monômero e do dímero da p23 aqui apresentados são compatíveis com o conhecimento estabelecido atualmente.

6.3 Conclusões p23 humana

- As análises de CD indicaram que a proteína se encontra estruturada e que é constituída majoritariamente por folhas betas, com uma pequena fração correspondente a hélices alfa.
- O emprego do reagente NBS provocou o apagamento do sinal relativo aos resíduos de aminoácido aromáticos, permitindo observar um ganho de sinal na região entre 220 e 210 nm (correspondente a hélice alfa).
- A proteína p23 apresentou um máximo de emissão de fluorescência de 336 nm, sendo que este não foi alterado em relação à proteína sem a cauda de histidinas (clivada).
- A proteína p23 apresentou, nas condições utilizadas, uma maior estabilidade térmica monitorada por CD do que aos reportados pela literatura até o momento.
- Foram realizados estudos inéditos de desenovelamento químico monitorados por CD e fluorescência, resultando nos seguintes valores de C_m: 3,6 M e 3,3 M uréia (fluorescência e CD), 1,4 M e 1,1 M cloreto de guanidina (fluorescência e CD).
- Os experimentos de espalhamento dinâmico de luz permitiram determinar o valor do coeficiente de difusão para a proteína p23 (8,8 x 10⁻⁷ ± 0,05 cm²/seg).
- Foram observadas duas espécies em todos os experimentos de velocidade de sedimentação para a proteína p23, uma fração majoritária monomérica e um discreto oligômero. Os seguintes valores de coeficiente de sedimentação foram encontrados: 2,3 S para o monômero e 4,6 S para o oligômero (p23-His), 2,2 S para o monômero e 4,3 S para o oligômero (p23 clivada).
- Utilizando os dados de velocidade de sedimentação e os dados de EDL foi possível estimar a massa molecular do monômero como sendo de 23 kDa (p23-His) e 19 kDa (p23 clivada) e do oligômero 51 kDa (p23-His) e 47 kDa (p23 clivada).

7 Referências

1. Agorogiannis El, Agorogiannis Gl, Papadimitriou A, Hadjigeorgiou GM. (2004) Protein misfolding in neurodegenerative diseases. Neuropathol Appl Neurobiol 30(3): 215-24.

2. Akoev V, Gogol EP, Barnett ME, Zolkiewski M. (2004) Nucleotide-induced switch in oligomerization of the AAA+ ATPase ClpB. Protein Sci 13: 567-74.

3. Ali MUM, Roe MS, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C, Pearl LH. (2006) Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. Nature 440: 1013-1017.

Ammelburg M, Frickey T, Lupas AN. (2006) Classification of AAA+ proteins. J Struct Biol. 156(1): 2 11.

5. Anfinsen CB. (1973) Principles that govern the folding of protein chains. Science 181: 223-30.

6. Appenzeller-Herzog C, Ellgaard L. (2008) The human PDI family: versatility packed into a single fold. Biochim Biophys Acta. 1783(4): 535-48.

7. Baldwin RL. (1989) How does protein get started? Trends Biochem. Sci. 14: 291-4.

8. Barnett ME; Zolkiewska A; Zolkiewski M. (2000) Structure and activity of ClpB from Escherichia coli. Role of the amino-and -carboxyl-terminal domains. J Biol Chem. 275(48): 37565-71.

9. Belssinger M, Buchner J (1998) How chaperones fold proteins. Biol. Chem. 379: 245-59.

10. Ben-Zvi AP, Goloubinoff P.(2002) Proteinaceous infectious behavior in non-pathogenic proteins is controlled by molecular chaperones. J Biol Chem. 277(51): 49422-7.

11. Borges JC, Fischer H, Craievich AF, Ramos CH. (2005) Low resolution structural study of two human HSP40 chaperones in solution. DJA1 from subfamily A and DJB4 from subfamily B have different quaternary structures. J Biol Chem. 280: 13671-81.

Borges JC, Ramos CH. (2005) Protein folding assisted by chaperones. Protein Pept Lett. 12: 257 61.

13. Borges JC, Cagliari TC, Ramos CH. (2007) Expression and variability of molecular chaperones in the sugarcane expressome. J Plant Physiol. 164: 505-513

14. Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S. (1989) Hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. Mol Cell Biol. 9: 3919-3930.

15. Bose S, Weikl T, Bügl H, Buchner J. (1996) Chaperone function of Hsp90-associated proteins. Science. 274(5293): 1715-7.

16. Bosl B, Grimminger V, Walter S. (2005) Substrate binding to the molecular chaperone Hsp104 and its regulation by nucleotides. J Biol Chem. 280: 38170-6.

17. Buchner J. (1999) Hsp90 & Co. - a holding for folding. Trends Biochem Sci. 24: 136-41.

18. Cagliari TC, Tiroli AO, Borges JC, Ramos CH. (2005) Identification and in silico expression pattern analysis of Eucalyptus expressed sequence tags (ESTs) encoding molecular chaperones. Genetics and Mol Biol. 28: 520-528.

19. Caplan AJ, Mandal AK, Theodoraki MA. (2007) Molecular chaperones and protein kinase quality control. Trends Cell Biol. 17(2): 87-92.

20. Chadli A, Ladjimi MM, Baulieu EE, Catelli MG. (1999) Heat-induced oligomerization of the molecular chaperone Hsp90. Inhibition by ATP and geldanamycin and activation by transition metal oxyanions. J Biol Chem. 274: 4133-4139.

21. Chai Y, Koppenhafer SL, Bonini NM, Paulson HL. (1999) Analysis of the role of heat shock protein (Hsp) molecular chaperones in polyglutamine disease. J Neuroscience. 19: 10338-47.

22. Chen S, Smith DF. (1998) Hop as an adaptor in the heat shock protein 70 (Hsp70) and hsp90 chaperone machinery. J Biol Chem. 273(52): 35194-200.

23. Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A, Bukau B. (1999) Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature. 400(6745): 693-6.

24. Diemand AV, Lupas AN. (2006) Modeling AAA+ ring complexes from monomeric structures. J Struct Biol. 156(1): 230-43.

25. Dittmar KD, Hutchison KA, Owens-Grillo JK, Pratt WB. (1996) Reconstitution of the steroid receptor.hsp90 heterocomplex assembly system of rabbit reticulocyte lysate. J Biol Chem. 271(22): 12833-9.

26. Dollins DE, Warren JJ, Immormino RM, Gewirth DT. (2007) Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. Mol Cell. 28(1): 41-56.

27. Doyle SM, Hoskins JR, Wickner S. (2007) Collaboration between the ClpB AAA+ remodeling protein and the DnaK chaperone system. Proc Natl Acad Sci U S A. 104(27): 11138-44.

28. Doyle SM, Shorter J, Zolkiewski M, Hoskins JR, Lindquist S, Wickner S. (2007) Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity. Nat Struct Mol Biol. 14(2): 114-22.

29. Edelhoch H. (1967) Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. Biochemistry. 6(7): 1948-1954.

30. Emelyanov VV. (2002) Phylogenetic relationships of organellar Hsp90 homologs reveal fundamental differences to organellar Hsp70 and Hsp60 evolution. Gene. 299: 125-133.

31. Felts SJ, Toft DO. (2003) p23, a simple protein with complex activities. Cell Stress Chaperones. 8:108-13.

32. Ferrarini M, Heltai S, Zocchi MR, Rugarli C. (1992) Unusual expression and localization of heatshock proteins in human tumor cells. Int J Cancer. 51: 613-619.

33. Fersht AR, Itzhaki LS, ElMasry NF, Matthews JM, Otzen De. (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. Proc Natl Acad Sci USA. 91: 10426-9.

34. Fink AL. (1999) Chaperone-mediated protein folding. Physiol Rev. 79:425-49.

35. Freeman BC, Yamamoto KR. (2002) Disassembly of transcriptional regulatory complexes by molecular chaperones. Science. 296(5576): 2232-5.

36. Freeman BC, Toft DO, Morimoto RI. (1996) Molecular chaperone machines: chaperone activities of the cyclophilin Cyp-40 and the steroid aporeceptor-associated protein p23. Science. 274(5293): 1718-20.

37. Freeman BC, Felts SJ, Toft DO, Yamamoto KR. (2000) The p23 molecular chaperones act at a late step in intracellular receptor action to differentially affect ligand efficacies. Genes Dev. 14(4): 422-34.

38. Böhm G, Muhr R, Jaenicke R. (1992) Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng. 5(3): 191-5.

157

39. Gallie DR, Fortner D, Peng J, Puthoff D. (2002) ATP-dependent Hexameric Assembly of the Heat Shock Protein Hsp101 Involves Multiple Interaction Domains and a Functional C-proximal Nucleotidebinding Domain. J Bio Chem. 277: 39617–39626.

40. Gamer J, Bujard H, Bukau B. (1992) Physical interaction between heat shock proteins DnaK, DnaJ, and GrpE and the bacterial heat shock transcription factor sigma 32. Cell. 69(5): 833-42.

41. Garcia-Ranea JA, Mirey G, Camonis J, Valencia A. (2002) p23 and HSP20/alpha-crystallin proteins define a conserved sequence domain present in other eukaryotic protein families. FEBS Lett. 529(2-3): 162-7.

42. Garnier C, Lafitte D, Jorgensen TJ, Jensen ON, Briand C, Peyrot V. (2001) Phosphorylation and oligomerization states of native pig brain HSP90 studied by mass spectrometry. Eur J Biochem. 268: 2402-2407.

43. Garnier C, Barbier P, Devred F, Rivas G, Peyrot, V. (2002). Hydrodynamic properties and quaternary structure of the 90 kDa heat-shock protein: effects of divalent cations. Biochem. 41: 11770-11778.

44. Germaniuk A, Liberek K, Marszalek J. (2002) A bichaperone (Hsp70-Hsp78) system restores mitochondrial DNA synthesis following thermal inactivation of Mip1p polymerase. J Biol Chem. 277(31): 27801-8.

45. Gill SC, Von Hippel PH. (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 182:319-26.

46. Glover JR, Lindquist S. (1998) Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell. 94(1): 73-82.

47. Goldberg, R.J. (1953) Sedimentation in the ultracentrifuge. J. Phys. Chem. 57: 194-202.

48. Goldfless SJ, Morag AS, Belisle KA, Sutera VA Jr, Lovett ST. (2006) DNA repeat rearrangements mediated by DnaK-dependent replication fork repair. Mol Cell. 21(5): 595-604.

49. Goloubinoff P, Mogk A, Zvi APB, Tomoyasu T, Bukau B. (1999) Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proc Nat Acad Sci. 96: 13732-7.

50. Grimaud R, Kessel M, Beuron F, Steven AC, Maurizi MR. (1998) Enzymatic and structural similarities between the Escherichia coli ATP-dependent proteases, ClpXP and ClpAP. J Biol Chem. 273(20): 12476-81.

51. Grimshaw JP, Jelesarov I, Siegenthaler RK, Christen P. (2003) Thermosensor action of GrpE. The DnaK chaperone system at heat shock temperatures. J Biol Chem. 278(21): 19048-53.

52. Groemping Y, Reinstein J. (2001) Folding properties of the nucleotide exchange factor GrpE from Thermus thermophilus: GrpE is a thermosensor that mediates heat shock response. J Mol Biol. 314(1): 167-78.

53. Haslberger T, Weibezahn J, Zahn R, Lee S, Tsai FT, Bukau B, Mogk A. (2007) M domains couple the ClpB threading motor with the DnaK chaperone activity. Mol Cell. 25(2): 247-60.

54. Hendrick JP, Hartl FU. (1993) Molecular chaperone functions of heat-shock proteins. Annu Rev Biochem. 62: 349-84.

55. Hendrick JP, Langer T, Davis TA, Hartl FU, Wiedmann M. (1993) Control of folding and membrane translocation by binding of the chaperone DnaJ to nascent polypeptides. Proc Natl Acad Sci U S A. 90(21): 10216-20.

56. Hinnerwisch J, Fenton WA, Furtak KJ, Farr GW, Horwich AL. (2005) Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. Cell. 121(7): 1029-41.

57. Höhfeld J, Cyr DM, Patterson C. (2001) From the cradle to the grave: molecular chaperones that may choose between folding and degradation. EMBO Rep. 2(10): 885-90.

58. Holt SE, Aisner DL, Baur J, Tesmer VM, Dy M, Ouellette M, Trager JB, Morin GB, Toft DO, Shay JW, Wright WE, White MA. (1999) Functional requirement of p23 and Hsp90 in telomerase complexes. Genes Dev. 13(7): 817-26.

59. Hu J, Toft D, Anselmo D, Wang X. (2002) In vitro reconstitution of functional hepadnavirus reverse transcriptase with cellular chaperone proteins. J Virol. 76(1): 269-79.

60. Hubert DA, Tornero P, Belkhadir Y, Krishna P, Takahashi A, Shirasu K, Dangl JL. (2003) Cytosolic HSP90 associates with and modulates the Arabidopsis RPM1 disease resistance protein. EMBO J. 22: 5679-89.

61. Jaenicke R. (1999) Stability and folding of domain proteins. Prog Biophys Mol Biol. 71: 155-241.

62. Jaenicke R. (1987) Folding and association of proteins. Prog Biophys Mol Biol. 49: 117-237.

63. Jakob U, Lilie H, Meyer I, Buchner J. (1995) Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase. Implications for heat shock in vivo. J Biol Chem. 270(13): 7288-94.

64. Johnson JL, Toft DO. (1994) A novel chaperone complex for steroid receptors involving heat shock proteins, immunophilins, and p23. J Biol Chem. 269(40): 24989-93.

65. Johnson JL, Beito TG, Krco CJ, Toft DO. (1994) Characterization of a novel 23-kilodalton protein of unactive progesterone receptor complexes. Mol Cell Biol. 14(3): 1956-63.

66. Johnson ML, Correia JJ, Yphantis DA, Halvorson HR. (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least- squares techniques. Biophys J. 36: 575-588.

67. Kim PS, Baldwin RL. (1990) Intermediates in the folding reactions of small proteins. Annu Rev Biochem 59: 631-60.

68. Kimura Y, Rutherford SL, Miyata Y, Yahara I, Freeman BC, Yue L, Morimoto RI, Lindquist S. (1997) Cdc37 is a molecular chaperone with specific functions in signal transduction. Genes Dev. 11(14): 1775-85.

69. Kitagawa M, Wada C, Yoshioka S, Yura T. (1991) Expression of ClpB, an analog of the ATPdependent protease regulatory subunit in Escherichia coli, is controlled by a heat shock sigma factor (sigma 32). J Bacteriol. 173(14): 4247-53.

70. Kochetkov SN, Rusakova EE, Tunitskaya VL. (1998) Recent studies of T7 RNA polymerase mechanism. FEBS Letters. 440: 264-267.

71. Kosano H, Stensgard B, Charlesworth MC, McMahon N, Toft D. (1998) The assembly of progesterone receptor-hsp90 complexes using purified proteins. J Biol Chem. 273(49): 32973-9.

72. Koyasu S, Nishida E, Kadowaki T, Matsuzaki F, Iida K, Harada F, Kasuga M, Sakai H, Yahara I. (1986) Two mammalian heat shock proteins, HSP90 and HSP100, are actin-binding proteins. Proc Natl Acad Sci U S A. 83(21): 8054-8.

73. Krzewska J, Langer T, Liberek K. (2001) Mitochondrial Hsp78, a member of the Clp/Hsp100 family in Saccharomyces cerevisiae, cooperates with Hsp70 in protein refolding. FEBS Lett. 489(1): 92-6.

74. Kuwajima K. (1989) The molten globule state as clue for understanding the folding and cooperativity of globular-protein structure. Proteins: Structure, function and genetics. 6:87-103.

75. Laemmli UK. (1970) Cleavage of structural proteins during the assembly of the head bacteriophage T4. Nature. 277: 680-5.

76. Lai BT, Chin NW, Stanek AE, Keh W, Lanks KW. (1984) Quantitation and intracellular localization of the 85K heat shock protein by using monoclonal and polyclonal antibodies. Mol Cell Biol. 4: 2802-2810.

77. Langer T, Schlatter H, Fasold H. (2002) Evidence that the novobiocin-sensitive ATP-binding site of the heat shock protein 90 (hsp90) is necessary for its autophosphorylation. Cell Biol Int. 26: 653-657.

78. Laskowska E, Kuczyńska-Wiśnik D, Skórko-Glonek J, Taylor A. (1996) Degradation by proteases Lon, Clp and HtrA, of Escherichia coli proteins aggregated in vivo by heat shock; HtrA protease action in vivo and in vitro. Mol Microbiol. 22(3): 555-71.

79. Laue TM. (2001) Biophysical studies by ultracentrifugation. Curr Opin Struct Biol. 11: 579-583.

80. Lee S, Sowa ME, Watanabe YH, Sigler PB, Chiu W, Yoshida M, Tsai FT. (2003) The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell. 115(2): 229-40.

81. Lee S, Choi JM, Tsai FT. (2007) Visualizing the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB. Mol Cell. 25(2): 261-71.

82. Leskovar A, Wegele H, Werbeck ND, Buchner J, Reinstein J. (2008) The ATPase cycle of the mitochondrial Hsp90 analog Trap1. J Biol Chem. 283(17): 11677-88.

83. Levchenko I, Yamauchi M, Baker TA. (1997) ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. Genes Dev. 11: 1561-72.

84. Lewandowska A, Gierszewska M, Marszalek J, Liberek K. (2006) Hsp78 chaperone functions in restoration of mitochondrial network following heat stress. Biochim Biophys Acta. 1763(2): 141-51.

85. Lewandowska A, Matuszewska M, Liberek K. (2007) Conformational properties of aggregated polypeptides determine ClpB-dependence in the disaggregation process. J Mol Biol. 371(3): 800-11.

86. Li GC, Werb Z. (1982) Correlation between synthesis of heat shock proteins and development of thermotolerance in Chinese hamster fibroblasts. Proc Natl Acad Sci U S A. 79(10): 3218-22.

87. Liberek K, Georgopoulos C. (1993) Autoregulation of the Escherichia coli heat shock response by the DnaK and DnaJ heat shock proteins. Proc Natl Acad Sci U S A. 90(23): 11019-23.

88. Liberek K, Georgopoulos C, Zylicz M. (1988) Role of the Escherichia coli DnaK and DnaJ heat shock proteins in the initiation of bacteriophage lambda DNA replication. Proc Natl Acad Sci U S A. 85(18): 6632-6.

89. Liu Y, Burch-Smith T, Schiff M, Feng S, Dinesh-Kumar SP (2004) Molecular chaperone Hsp90 associates with resistance protein N and its signaling proteins SGT1 and Rar1 to modulate an innate immune response in plants. J Biol Chem 279: 2101-2108.

90. Lu R, Malcuit I, Moffett P, Ruiz MT, Peart J, Wu AJ, Rathjen JP, Bendahmane A, Day L, Baulcombe DC. (2003) High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. Embo J. 22: 5690-5699.

91. Lum R, Tkach JM, Vierling E, Glover JR. (2004) Evidence for an unfolding/threading mechanism for protein disaggregation by Saccharomyces cerevisiae Hsp104. J Biol Chem. 279(28): 29139-46.

92. Maurizi MR, Xia D. (2004) Protein binding and disruption by Clp/Hsp100 chaperones. Structure. 12: 175-83.

93. Mayer MP, Brehmer D, Gässler CS, Bukau B. (2001) Hsp70 chaperone machines. Adv Protein Chem. 59: 1-44.

94. McLaughlin SH, Ventouras LA, Lobbezzo B, Jackson SE. (2004) Independent ATPase activity of Hsp90 subunits creates a flexible assembly platform. J Mol Biol. 344: 813-826.

95. McLaughlin SH, Smith HW, Jackson SE. (2002) Stimulation of the weak ATPase activity of human hsp90 by a client protein. J Mol Biol. 315(4): 787-98.

96. McLaughlin SH, Sobott F, Yao ZP, Zhang W, Nielsen PR, Grossmann JG, Laue ED, Robinson CV, Jackson SE. (2006) The co-chaperone p23 arrests the Hsp90 ATPase cycle to trap client proteins. J Mol Biol. 356: 746-58.

97. Mogk A, Tomoyasu T, Goloubinoff P, Rüdger S, Röder D, Langen H, Bukau B. (1999) Identification of thermolabile Escherichia coli proteins: prevention and reversion of agregation by DnaK and ClpB. EMBO J. 18(24): 6934-49.

98. Mosser DD, Ho S, Glover JR. (2004) Saccharomyces cerevisiae Hsp104 enhances the chaperone capacity of human cells and inhibits heat stress-induced proapoptotic signaling. Biochemistry. 43: 8107-15.

99. Motohashi K, Watanabe Y, Yohda M, Yoshida M. (1999) Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones. Proc Natl Acad Sci USA. 96:7184-9.

100. Nair SC, Toran EJ, Rimerman RA, Hjermstad S, Smithgall TE, Smith DF. (1996) A pathway of multichaperone interactions common to diverse regulatory proteins: estrogen receptor, Fes tyrosine kinase, heat shock transcription factor Hsf1, and the aryl hydrocarbon receptor. Cell Stress Chaperones. 1(4): 237-50.

101. Nery FC, Bressan GC, Alborghetti MR, Passos DO, Kuniyoshi TM, Ramos CH, Oyama S Jr, Kobarg J. (2006) A spectroscopic analysis of the interaction between the human regulatory proteins RACK1 and Ki-1/57. Biol Chem. 387: 577-82.

102. Nollen EA, Morimoto RI. (2002) Chaperoning signaling pathways: molecular chaperones as stresssensing 'heat shock' proteins. J Cell Sci. 115: 2809-2816.

103. Obermann WMJ, Sondermann H, Russo AA, Pavletich NP, Hartl FU. (1998) In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. J Cell Biol. 143: 901-910.

104. Panaretou B, Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. (1998) ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. EMBO J. 17(16): 4829-36.

105. Panaretou B, Siligardi G, Meyer P, Maloney A, Sullivan JK, Singh S, Millson SH, Clarke PA, Naaby-Hansen S, Stein R, Cramer R, Mollapour M, Workman P, Piper PW, Pearl LH, Prodromou C. (2002) Activation of the ATPase activity of hsp90 by the stress-regulated cochaperone aha1. Mol Cell. 10(6): 1307-18.

106. Park SH, Bolender N, Eisele F, Kostova Z, Takeuchi J, Coffino P, Wolf DH. (2007) The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin-proteasome system. Mol Biol Cell. 18(1): 153-65.

107. Parsell DA, Kowall AS, Lindquist S. (1994) Saccharomyces cerevisiae Hsp104 Protein. J Biol Chem. 269: 4480-7.

108. Pasta SY, Raman B, Ramakrishna T, Rao ChM. (2002) Role of the C-terminal extensions of alphacrystallins. Swapping the C-terminal extension of alpha-crystallin to alphaB-crystallin results in enhanced chaperone activity. J Biol Chem. 277(48): 45821-8. 109. Pearl LH, Prodromou C. (2006) Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu Rev Biochem. 75: 271-94.

110. Pelham HR. (1986) Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell. 46(7): 959-61.

111. Picard, D. (2002) Heat-shock protein 90, a chaperone for folding and regulation. Cell Mol. Life Sci.59: 1640-1648.

112. Pratt WB, Toft DO. (1997) Steroid receptor interactions with heat shock protein and immunophilin chaperones. Endocr Rev. 18(3): 306-60.

113. Pratt WB, Toft DO. (2003) Regulation of signaling protein function and trafficking by the Hsp90/Hsp70-based chaperone machinery. Exp Biol Med. 228: 111–133.

114. Pratt WB, Silverstein AM, Galigniana MD. (1999) A model for the cytoplasmic trafficking of signalling proteins involving the hsp90-binding immunophilins and p50cdc37. Cell Signal. 1(12): 839-51.

115. Price NC. (2000) Conformational issues in the characterization of proteins. Biotechnol Appl Biochem. 31: 29-40.

116. Privalov PL. (1996) Intermediate states in protein folding. J Mol Biol. 258: 707-25.

117. Prodromou, C., and Pearl, L. H. (2003) Structure and functional relationships of Hsp90 Curr Cancer Drug Targets. 3: 301-323.

118. Prodromou C, Panaretou B, Chohan S, Siligardi G, O'Brien R, Ladbury JE, Roe SM, Piper PW, Pearl LH. (2000) The ATPase cycle of Hsp90 drives a molecular 'clamp' via transient dimerization of the N-terminal domains. EMBO J. 19: 4383-92.

119. Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. Cell. 90(1): 65-75.

120. Ptitsyn OB. (1991) How does protein synthesis give rise to the 3D-structure? FEBS Lett. 285: 176-81.

121. Queitsch C, Hong SW, Vierling E, Lindquist S. (2000) Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis. Plant Cell. 12(4): 479-92.

164

122. Queitsch C, Sangster TA, Lindquist S. (2002) Hsp90 as a capacitor of phenotypic variation. Nature.417: 618-624.

123. Radanyi C, Renoir JM, Sabbah M, Baulieu EE. (1989) Chicken heat-shock protein of Mr=90,000, free or released from progesterone receptor, is in a dimeric form. J Biol Chem. 264: 2568-2573.

124. Ramos CH, Ferreira ST. (2005) Protein folding, misfolding and aggregation: evolving concepts and conformational diseases. Protein Pept Lett. 12: 213-22.

125. Richter K, Buchner J. (2001). Hsp90: chaperoning signal transduction. J Cell Physiol. 188: 281-290.

126. Richter K, Reinstein J, Buchner J. (2002) N-terminal residues regulate the catalytic efficiency of the Hsp90 ATPase cycle. J Biol Chem. 277(47): 44905-10.

127. Richter K, Walter S, and Buchner J (2004) The Co-chaperone Sba1 Connects the ATPase Reaction of Hsp90 to the Progression of the Chaperone Cycle. J Mol Biol. 342: 1403-1413.

128. Richter K, Buchner J. (2006) hsp90: twist and fold. Cell. 127(2): 251-3.

129. Richter K, Soroka J, Skalniak L, Leskovar A, Hessling M, Reinstein J, Buchner J. (2008) Conserved conformational changes in the ATPase cycle of human Hsp90. J Biol Chem. 283(26): 17757-65.

130. Roe SM, Prodromou C, O'Brien R, Ladbury JE, Piper PW, Pearl LH. (1999) Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. J Med Chem. 42(2): 260-6.

131. Ruddon RW, Sherman SA, Bedows E. (1996) Protein folding in the endoplasmic reticulum: lessons from the human chorionic gonadotropin beta subunit. Protein Sci. 5: 1443-52.

132. Ruddon RW, Bedows E. (1997) Assisted protein folding. J Biol Chem. 272: 3125-8.

133. Rutherford SL, Lindquist S. (1998) Hsp90 as a capacitor for morphological evolution. Nature. 396: 336-342.

134. Sambrook J, Fritsch EF, Maniatis T (1989) In: Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press (Ed.) New York, United States of America.

135. Sanchez Y, Lindquist S. (1990) HSP104 required for induced thermotolerance. Science. 248: 1112-5.

136. Schaupp A, Marcinowski M, Grimminger V, Bösl B, Walter S. (2007) Processing of proteins by the molecular chaperone Hsp104. J Mol Biol. 370(4): 674-86.

137. Scheibel T, Weikl T, Buchner J. (1998) Two chaperone sites in Hsp90 differing in substrate specificity and ATP dependence. Proc Natl Acad Sci U S A. 95(4): 1495-9.

138. Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, Hartl FU, Moarefi I. (2000) Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell. 101: 199-210.

139. Schirmer EC, Glover JR; Singer MA; Lindquist S. (1996) HSP104/Clp proteins: a common mechanism explains diverse functions. Trends Biochem Sci 21:289-96.

140. Schirmer EC, Queitsh C, Kowal AS, Parsell DA, Lindquist S. (1998) The ATPase Activity of Hsp104, Effects of Environmental Conditions and Mutations J Biol Chem. 273(25): 15546-52.

141. Schlatter H, Langer T, Rosmus S, Onneken ML, Fasold H (2002). A novel function for the 90 kDa heat-shock protein (Hsp90): facilitating nuclear export of 60 S ribosomal subunits. Biochem J. 362: 675-684.

142. Schlieker C, Weibezahn J, Patzelt H, Tessarz P, Strub C, Zeth K, Erbse A, Schneider-Mergener J, Chin JW, Schultz PG, Bukau B, Mogk A. (2004) Substrate recognition by the AAA+ chaperone ClpB. Nat Struct Mol Biol. 11(7): 607-15.

143. Schmitt M, Neupert W, Langer T. (1996) The molecular chaperone Hsp78 confers compartmentspecific thermotolerance to mitochondria. J Cell Biol. 134(6): 1375-86.

144. Sharp S, Workman P. (2006) Inhibitors of the HSP90 molecular chaperone: current status. Adv Cancer Res. 95: 323-48.

145. Shiau AK, Harris SF, Southworth DR, Agard DA. (2006) Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. Cell. 127(2): 329-40.

146. Siligardi G, Panaretou B, Meyer P, Singh S, Woolfson DN, Piper PW, Pearl LH, Prodromou C. (2002) Regulation of Hsp90 ATPase activity by the co-chaperone Cdc37p/p50cdc37. J Biol Chem. 277(23): 20151-9.

147. Skowyra D, Georgopoulos C, Zylicz M. (1990) The E. coli dnaK gene product, the hsp70 homolog, can reactivate heat-inactivated RNA polymerase in an ATP hydrolysis-dependent manner. Cell. 62(5): 939-44.

148. Smith DF, Baggenstoss BA, Marion TN, Rimerman RA. (1993) Two FKBP-related proteins are associated with progesterone receptor complexes. J Biol Chem. 268(24): 18365-71.

149. Smith DF. (1993) Dynamics of heat shock protein 90-progesterone receptor binding and the disactivation loop model for steroid receptor complexes. Mol Endocrinol. 7(11): 1418-29.

150. Spence J, Georgopoulos C. (1989) Purification and properties of the Escherichia coli heat shock protein, HtpG. J Biol Chem. 1264: 4398-403.

151. Spence J, Cegielska A, Georgopoulos C. (1990) Role of Escherichia coli heat shock proteins DnaK and HtpG (C62.5) in response to nutritional deprivation. J Bacteriol 172 (12): 7157–7166.

152. Squires CL, Pedersen S, Ross BM, Squires C. (1991) ClpB is the Escherichia coli heat shock protein F84.1. J Bacteriol. 173(14): 4254-62.

153. Stafford WF 3rd. (1994) Boundary analysis in sedimentation velocity experiments. Methods Enzymol. 240: 478-501.

154. Studier FW, Moffatt BA. (1986) Use of bacteriophage T7 RNA polymerase to direct selective highlevel expression of cloned genes. J Mol Biol. 189(1): 113-30.

155. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185: 60-89.

156. Sullivan WP, Owen BA, Toft DO. (2002) The influence of ATP and p23 on the conformation of hsp90. J Biol Chem. 277(48): 45942-8.

157. Sullivan W, Stensgard B, Caucutt G, Bartha B, McMahon N, Alnemri ES, Litwack G, Toft D. (1997) Nucleotides and two functional states of hsp90. J Biol Chem. 272(12): 8007-12.

158. Teter SA, Houry WA, Ang D, Tradler T, Rockabrand D, Fischer G, Blum P, Georgopoulos C, Hartl FU. (1999) Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell. 97(6): 755-65.

159. Tetu B, Lacasse B, Bouchard HL, Lagace R, Huot J, Landry J. (1992) Prognostic influence of HSP-27 expression in malignant fibrous histiocytoma: a clinicopathological and immunohistochemical study. Cancer Res. 52: 2325-8.

160. Thomas PJ, Qu BH, Pedersen PL. (1995) Defective protein folding as a basis of human disease. Trends Biochem Sci. 20: 456-9.

161. Thor A, Benz C, Moore D2d, Goldman E, Edgerton S, Landry J, Schwartz L, Mayall B, Hickey E, Weber LA. (1991) Stress response protein (srp-27) determination in primary human breast carcinomas: clinical, histologic, and prognostic correlations. J Natl Cancer Inst. 83: 170-8.

162. Vale RD. (2000) AAA proteins. Lords of the ring. J Cell Biol. 150(1): F13-9.

163. van Montfort RL, Basha E, Friedrich KL, Slingsby C, Vierling E. (2001) Crystal structure and assembly of a eukaryotic small heat shock protein. Nat Struct Biol. 8(12): 1025-30.

164. Sullivan WP, Toft DO. (1993) Mutational analysis of hsp90 binding to the progesterone receptor. J Biol Chem. 268(27): 20373-9.

165. Wawrzynow A, Wojtkowiak D, Marszalek J, Banecki B, Jonsen M, Graves B, Georgopoulos C, Zylicz M. (1995) The ClpX heat-shock protein of Escherichia coli, the ATP-dependent substrate specificity component of the ClpP-ClpX protease, is a novel molecular chaperone. EMBO J. 14(9): 1867-77.

166. Weaver AJ, Sullivan WP, Felts SJ, Owen BAL, Toft DO. (2000) Crystal structure and activity of human p23, a heat shock protein 90 co-chaperone. J Biol Chem. 275: 23045–23052.

167. Weibezahn J, Tessarz P, Schlieker C, Zahn R, Maglica Z, Lee S, Zentgraf H, Weber-Ban EU, Dougan DA, Tsai FT, Mogk A, Bukau B. (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. Cell. 119: 653-65.

168. Weikl T, Muschler P, Richter K, Veit T, Reinstein J, Buchner J. (2000) C-terminal regions of Hsp90 are important for trapping the nucleotide during the ATPase cycle. J Mol Biol. 303(4): 583-92.

169. Weikl T, Abelmann K, Buchner J. (1999) An unstructured C-terminal region of the Hsp90 cochaperone p23 is important for its chaperone function. J Mol Biol. 293: 685–691.

170. Westra DF, Welling GW, Koedijk DG, Scheffer AJ, The TH, Welling-Wester S. (2001) Immobilised metal-ion affinity chromatography purification of histidine-tagged recombinant proteins: a wash step with a low concentration of EDTA. J Chromatogr B Biomed Sci Appl. 760(1): 129-36.

171. Whitesell L, Mimnaugh EG, De Costa B, Myers CE, Neckers LM. (1994) Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. Proc Natl Acad Sci U S A. 91(18): 8324-8.

172. Wickner S, Gottesman S, Skowyra D, Hoskins J, McKenney K, Maurizi MR. (1994) A molecular chaperone, ClpA, functions like DnaK and DnaJ. Proc Natl Acad Sci U S A. 91(25): 12218-22.

173. Wiech H, Buchner J, Zimmermann R, Jakob U. (1992) Hsp90 chaperones protein folding in vitro. Nature. 358: 169-170.

174. Wojtkowiak D, Georgopoulos C, Zylicz M. (1993) Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of Escherichia coli. J Biol Chem. 268(30): 22609-17.

175. Young JC, Moarefi I, Hartl FU. (2001) Hsp90: a specialized but essential protein-folding tool. J Cell Biol. 154(2): 267-73.

176. Ziemienowicz A, Skowyra D, Zeilstra-Ryalls J, Fayet O, Georgopoulos C, Zylicz M. (1993) Both the Escherichia coli chaperone systems, GroEL/GroES and DnaK/DnaJ/GrpE, can reactivate heat-treated RNA polymerase. Different mechanisms for the same activity. J Biol Chem. 268(34): 25425-31.

177. Zietkiewicz S, Krzewska J, Liberek K. (2004) Successive and synergistic action of the Hsp70 and Hsp100 chaperones in protein disaggregation. J Biol Chem. 279: 44376-44383.

178. Zietkiewicz S, Lewandowska A, Stocki P, Liberek K. (2006) Hsp70 chaperone machine remodels protein aggregates at the initial step of Hsp70-Hsp100-dependent disaggregation. J Biol Chem. 281(11): 7022-9.

179. Zietkiewicz S, Krzewska J, Liberek K. (2004) Successive and synergistic action of the Hsp70 and Hsp100 chaperones in protein disaggregation. J Biol Chem. 279: 44376-83.

180. Zolkiweski M. (1999) ClpB cooperates with DnaK, DnaJ and GrpE in suppressing protein aggregation. A novel multichaperone system from Escherichia coli. J Biol Chem. 274: 28083-6.

181. Zolkiewski M. (2006) A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. Mol Microbiol. 61(5): 1094-100.

182. Zzaman S, Reddy JM, Bastia D. (2004) The DnaK-DnaJ-GrpE chaperone system activates inert wild type pi initiator protein of R6K into a form active in replication initiation. J Biol Chem. 279(49): 50886-94.

8 Anexos: artigos

8.1 Artigo: Genetics and Molecular Biology, 28, 3 (suppl), 520-528 (2005)

Identification and in silico expression pattern analysis of Eucalyptus expressed sequencing tags (ESTs) encoding molecular chaperones

Thiago C. Cagliari^{1,2#}, Ana O. Tiroli^{1,2#}, Júlio C. Borges^{1,2} and Carlos H.I. Ramos^{1,2}

¹Laboratório Nacional de Luz Síncrotron, Centro de Biologia Molecular Estrutural, Campinas, SP, Brazil. ²Departamento de Bioquímica, Instituto de Biologia, Universidade Estadual de Campinas. Campinas, SP, Brazil.

Abstract

Expressed Sequence Tags (ESTs) sequencing provides reliable and useful information concerning gene expression patterns in the genomic context. Our group used bioinformatics to identify and annotate 5'EST-contigs belonging to the molecular chaperones within the Eucalyptus Genome Sequencing Project Consortium (FORESTs) database. We found that 1,959 5'EST-contigs, or approximately 1.6% of the total 5'EST-contigs, encoded chaperones, emphasizing their biological importance. About 55% of the chaperones that we found were Hsp70 chaperones and its co-chaperones, 18% were Hsp90 chaperones, 15% were Hsp60 and its co-chaperone, 8% were Hsp100 chaperones, and 4% were Small Hsps. We also investigated the digital expression profile of the chaperone genes to gain information on gene expression levels in the different libraries and we found that molecular chaperones may have differential expression. The results discussed here give important hints about the role of chaperones in Eucalyptus cells.

Key words: chaperones, heat shock proteins, genome, expressed sequence tags.

Send correspondence to Carlos H.I. Ramos. Caixa Postal 6192, 13084-971 Campinas, SP, Brazil. E-mail: cramos@lnls.br.

[#]These authors contributed equally to this work.

Introduction

Environmental temperature fluctuations and other forms of stress can provoke partial or complete protein unfolding. This unfolding induces the exposition of hydrophobic residues that are usually buried in the native state, leading to intracellular protein aggregation and its dreadful consequences. Molecular chaperones are able to bind to the exposed hydrophobic residues in unfolded proteins, preventing incorrect folding and aggregation (for recently reviews see Wang *et al.* 2004, and Borges and Ramos, 2005). Due to their proper response to stress conditions, chaperones were first classified as heat shock proteins (Hsp), however some of them are also constitutively expressed because they are required for substrate protein maturation, protein transport, and other functions (Ellis and Hartl, 1996; Fink, 1999). Molecular chaperones are divided into families according to their approximate molecular weight (Fink, 1999; Borges and Ramos, 2005). The main features of each family, such as state of oligomerization, presence of ATPase activity, and interaction with co-chaperones, which are proteins that somehow assist their activities, are described in Table 1.

Advances in large-scale cDNA sequencing provided a reliable source of information concerning quality and quantity of expressed mRNAs under very diverse developmental and environmental conditions. This information can lead to new insights on specific protein importance and activity in response to environmental conditions sensed by the organism. EST genome projects indicate trustworthy patterns of gene expression on specific tissues or even in entire organisms (Mayer and Mewes, 2002). Through the use of bioinformatics tools, such as annotation by homology search, valuable information about specific target molecules may be accessed.

Usually, plants are more exposed to unfavorable environmental changes than animals because of their sessile existence. Therefore, knowledge about the network of stress proteins in plants is of great interest not only to improve the agriculture production but also to enhance our understanding about the function of the chaperones and the protein folding process. Here we describe the *in silico* search and identification of molecular chaperone transcripts in the Brazilian *Eucalyptus* Genome Sequencing Project Consortium (FORESTs – https://forests.esalq.usp.br/) by EST homology comparison between sequences available at public databases and sequences generated by FORESTs. Assuming that the whole 5'EST set from FORESTs database is in agreement with the real *Eucalyptus* mRNA expression profile, a reasonable

amount of putative sequences related to molecular chaperones were identified, indicating the importance of these proteins to the cell. A digital northern blot was also performed showing that molecular chaperones may have differential expression. The relevance of these findings is also discussed.

Materials and Methods

The EST library preparation, sequencing, clustering, and other important data are available at the FORESTs home page (https://forests.esalq.usp.br) and in accompanying manuscripts in this issue. The strategy to annotate molecular chaperones in the FORESTs project was as follows. The translated amino acid sequences of specifics mRNAs from chaperones and stress-related proteins were chosen as described by Borges et al. (2001). The translated sequences were compared with the cluster consensus generated by the FORESTs database utilizing the software program tBLASTn (Altschul et al., 1997), which is basically an alignment research tool (Figure 1). The FORESTs cluster that had a statistically significant match (*i.e.* an E-value lower than 1e-5) was considered to be a molecular chaperone. The homology between protein sequences was analyzed using the alignment program LALIGN (http://www.ch.embnet.org/software/LALIGN form). In order to determine the amount of sequences associated with molecular chaperone genes, the data were expressed as a percentage of all the 5'EST belonging to annotated chaperone families in the FORESTs. The annotation process was improved by increasing the number of data mining rounds, which was done by submitting the predicted protein family encountered in the first round to the following rounds (Figure 1). Each additional round increases the probability of finding a cluster that was incorrectly excluded in previous rounds.

A digital northern blot was also performed in order to determine the chaperone expression levels for each FORESTs library. The libraries with fewer than one thousand sequence tags were not considered in our analysis because the abundance of cDNA not picked up among one thousand clones is unlikely (about 5% chance) to be larger than 3.7/1000 (Audic and Claverie, 1997). The molecular chaperones expression level was calculated by dividing the number of the 5'EST sequences annotated to a specific chaperone in each particular library, by the total number of 5'EST sequenced in that library. The resulting number allows a quantitative and qualitative expression analysis of the molecular chaperones in different *Eucalyptus* tissues.

Results and Discussion

The data mining approach used here found 1,959 5'ESTs belonging to the molecular chaperone proteins category. The FORESTs database contains a total of 123,889 5'ESTs, which means that about 1.6% of the 5'ESTs sequenced by the FORESTs consortium were associated to chaperone genes (Table 2). Considering that the number of 5'EST in each library is related to the number of mRNAs produced by the cell and therefore to the plant gene expression profile, the chaperone genes have a remarkable expression level in *Eucalyptus* cells. This result is expected because the cell requires proper protein folding to be viable. About 55% of all 5'EST annotated as belonging to the molecular chaperones category were homologous to the Hsp70 and its co-chaperones), 18% were homologous to the Hsp90 family, 13% were homologous to the Hsp60/TCP-1 family, 8% to the Hsp100 family, 4% to the Small Hsp family, and about 2% to the Hsp10 family (Figure 2).

Hsp70 and co-chaperones

Hsp70 acts as a pivot chaperone, interacting with other chaperones and exchanging unfolded polypeptides with them. Hsp70 transports proteins to the Hsp60 system for *de novo* folding (Siegers *et al.*, 1999), participates in aggregates recovery together with Hsp100 chaperones (Mogk *et al.*,1999), cooperates with the small Hsps to refold heat-denatured polypeptides (Lee and Vierling, 2000), and controls the activity of regulatory proteins as heat-shock factors, the transcriptional factors of Hsps genes (Morimoto, 1998). In plants, members of the Hsp70 family are expressed in response to environmental or abiotic stress conditions such as heat, cold, drought, etc. (Guy and Li, 1998; Sung *et al.*, 2001). Despite the determination of thermotolerance acquisition upon Hsp70 overexpression, which also enhanced salt and water tolerance (Sugino *et al.*, 1999; Alvim *et al.*, 2001), the cellular mechanisms of Hsp70 upon stress are not fully understood (Wang *et al.*, 2004).

Hsp70 requires two or more cofactors, and the most important among them are the co-chaperones Hsp40 and GrpE. Sequences related to Hsp70 and Hsp40 were responsible for more than half of the 5'EST identified as chaperones in the *Eucalyptus* (Figure 2). Hsp70 sequences comprise about 25% of all identified chaperone ESTs and Hsp40 sequences comprise about 30% (Figure 2). The approximated 1:1 expression profile between Hsp70 and Hsp40 in *Eucalyptus* is in agreement with the results obtained

from the analyses of 5EST libraries from other organisms (Borges *et al.*, 2001). Hsp70 and Hsp40 chaperones were highly expressed in *Eucalyptus*, a result also observed for *Arabidopsis* (Miernyk, 2001), sugar cane (Borges *et al.*, 2001) and human genome databases. The data from the FORESTs analysis reinforced the importance of the Hsp70 and Hsp40 proteins for the protein-folding homeostasis in all organisms.

Chaperones from the Hsp70 family are classified in three subfamilies based on their nucleotide dissociation properties: the DnaK subfamily, the Hsc70 subfamily, and the HscA subfamily (Brehmer *et al.*, 2001). Chaperones belonging to the three Hsp40 subfamilies are able to bind all Hsp70 subfamily members, and have different functions and structures (Lu and Cyr, 1998; Borges *et al.*, 2005), which is in agreement with the high diversity of this protein family in the cell. Unlike Hsp40, GrpE binds only to the DnaK subfamily (Brehmer *et al.*, 2001), which may explain why GrpE is less expressed than Hsp40 in *Eucalyptus* (Table 1) and also in sugar cane (Borges *et al.*, 2001). The lower expression level of GrpE genes compared to Hsp40 genes is probably caused by the compartmentalization of the first, since it was described only in prokaryotes, and in both mitochondria and chloroplast eukaryotic organelles (Netzer and Hartl, 1998). Another explanation for this lower expression is the restricted and transient functions of GrpE. GrpE participates in the folding reaction only during the release of ADP, while Hsp70 and Hsp40 remain bound to the substrate protein during the whole process (Hartl, 1996).

Hsp70 homologues from all cellular compartments were found in FORESTs database, but the cytoplasmic Hsp70 (Hsc70 subfamily) possesses 55% of all reads related to Hsp70 family or 10% of all reads related to chaperones at the FORESTs database (Table 2). It is characteristic of a single plant species to display multiple genes for Hsc70 encoding proteins that are more than 90% identical to each other at the primary structure (Boston *et al.*, 1996), which is probably due to the fact that most of the nascent proteins are produced in the cytoplasm. Hsc70 has also the most abundant expression (about 65%) among Hsp70 proteins in sugar cane cells (Borges *et al.*, 2001), pointing to a universal pattern of expression for this protein in plants. These results are very important because Hsc70 plays an important role in the net response to low temperature in sugar cane (Nogueira *et al.*, 2003) and it may be playing the same role in *Eucalyptus* (see discussion below). The 37 Hsp70 Clusters found within FORESTs are far more diverse than the 18 genes found in *Arabidopsis* (Lin *et al.*, 2001) and the 12 found in spinach (Guy and Li, 1998). This difference cannot be solely explained by genetic distance. The Hsp70 Clusters encountered may be related to diverse mRNA processing and even clustering process failures, since FORESTs is an EST genome project.

The Hsp70 and Hsp40 families were encountered in all the FORESTs libraries examined (Figure 3), which indicates their relevance for protein folding in various tissues as indicated by several works with these proteins (Boston *et al.*, 1996; Fink, 1999). Worthwhile mentioning is the high expression identified in the *E. grandis* seedling tissues cultivated in the dark (SL7 library) (Figure 3). This is an important result from which the relevance of the chaperones for the cell function maintenance is verified (see discussion below).

Hsp60 and Hsp10

Chaperones belonging to the Hsp60 family play an important role in the cell because they assist the newly synthesized and newly translocated proteins to achieve native conformation, and function as pairs of stacked rings with seven subunits each (Borges and Ramos, 2005). Hsp60 proteins are necessary for the folding of many chloroplastic proteins such as Rubisco, the enzyme that makes up about 50% of the total chloroplast protein content and is responsible for the CO₂ fixation reaction.

Approximately 15% of the total reads annotated as molecular chaperones belonged to the Hsp60 family and its co-chaperone Hsp10 (Figure 2). The abundance of ESTs annotated as belonging to the Hsp60 family was almost equally distributed among the three intracellular compartments: mitochondria, chloroplast and cytoplasm. The expression of cytoplasmic Hsp60 (subfamily TCP-1) was the most significant, with 38% of the total 5'EST annotated as Hsp60 chaperone family, followed by other Hsp60 proteins located in the chloroplast (33%), and mitochondria (22%) (Table 2). The higher expression of the subfamily TCP-1 compared to the other Hsp60 subfamilies may be explained because the folding of newly synthesized or imported proteins assistance occurs mainly in the cytoplasm (Grantcharova *et al.*, 2001). It is noteworthy that all the three Clusters related to chloroplastic Hsp60, one cluster related to the mitochondrial Hsp60 and four Clusters belonging to the TCP-1 subfamily had the sequence coding for the whole protein, reinforcing the confidence of the annotation results showed here. The number of *Eucalyptus* Clusters related to mitochondrial Hsp60 and TCP-1 were quite similar to the number of mitochondrial Hsp60 and TCP-1 genes found in the *Arabidopsis* genome (Hill and Hemmingsen, 2001). Both organisms have 9 TCPs and *Arabidopsis* has 4 genes related to mitochondrial Hsp60 while 3 mitochondrial Hsp60 Clusters were annotated in the FOREST. However, only three Clusters encoding plastid Hsp60 members were found in *Eucalyptus* while *Arabidopsis* has 6 genes belonging to this category (Hill and Hemmingsen, 2001). This difference may be explained by a hypothetical very low expression of these genes in *Eucalyptus*, which would result in the absence of their ESTs in the final library, or by discrepancy between *Arabidopsis* predicted Cpn60 genes and *Eucalyptus* hypothetical Cpn60 genes, with the relatively low identity between these proteins increasing the difficulty in finding new Cpn60 orthologues (Hill and Hemmingsen, 2001). The nineteen Hsp60-like sequences annotated for *Eucalyptus* (Table 2) may be highly divergent Cpn60 genes which were not identified due to low homology.

The Hsp60 family was encountered in all the FORESTs libraries examined (Figure 3), which indicates its relevance for protein folding in various tissues as indicated by many works with these proteins. The higher expression levels of the Hsp60 family were registered for the SL6 (*E.urophylla* seedlings) and for the RT3 (seedlings root) libraries (Figure 3), which represent tissues in development, suggesting that Hsp60 members are important to *Eucalyptus* developmental stages (see discussion below). This is compatible with the fact that mutated species of *Arabidopsis* chloroplasts Cpn60α exhibit defects in embryo and seedlings development (Apuya *et al.*, 2001). The expression of Hsp60 genes in leaves colonized by the *Thyrinteina* grub for seven days is also worthy of note (LV3 library), and may indicate a differential expression of Hsp60 under biotic stress conditions.

Hsp90

Hsp90 is a dimeric chaperone that interacts with more than 40 protein substrates including transcription factors, protein kinases and unrelated proteins (Pearl and Prodromou, 2001), and forms cytoplasmic chaperone complexes or foldosomes with others chaperones, co-chaperones, and cellular proteins (Zhao *et al.*, 2005). The main task of Hsp90 is to assist protein folding (Buchner, 1999) but it also influences other activities related to signal transduction, cell cycle control, and protein degradation and trafficking

(Pearl and Prodromou, 2001). Hsp90 has important functions in plant cells. It acts as a buffer for cryptic genetic variations in *Arabidopsis* (Queitsch *et al.*, 2002), which indicates an important role in morphological evolution, and is a key player in biotic stress because it interacts with resistance gene products (Hubert *et al.*, 2003; Liu *et al.*, 2004; Lu *et al.*, 2003).

The Hsp90 family represented about 18% of all 5'EST annotated to molecular chaperone category in *Eucalyptus* (Figure 2). This result was in agreement with the fact that Hsp90 is one of the most abundantly expressed cytoplasmic proteins in several organisms, even in the absence of stress (Boston *et al.*, 1996; Buchner, 1999; Borges *et al.*, 2001). The cytoplasmic Hsp82 is by far the most expressed protein within the Hsp90 chaperone family with 78% of the total 5'EST belonging to this family and more than 10% of all 5'EST belonging to the chaperone category (Table 2). Hsp82 has also the most abundant expression (about 70%) among Hsp90 proteins in sugar cane cells (Borges *et al.*, 2001), pointing to a universal pattern of expression for this protein in plants. This result is very important because Hsp82 is likely to be involved in a net response to low temperature in sugar cane (Nogueira *et al.*, 2003) and it may be playing the same role in *Eucalyptus*. Since Hsc70, another protein involved in response to low temperatures in eucalyptus, and for extension to several other chaperone category (Table 2, and discussion above), we suggest that these two proteins are likely to be directly involved in the response to low temperatures in eucalyptus, and for extension to several other plants. Therefore, we suggest that Hsc70 and Hsp82 are good candidates for *in vivo* studies, in order to better characterize the role of these proteins in plant tolerance under stress conditions.

The *Arabidopsis* genome presents four cytoplasmic Hsp90 genes while the FORESTs database presented ten cytoplasmic Hsp90 related Clusters. This difference may be attributed to alternative RNA splicing, not detected in the *Arabidopsis* genome project. *Arabidopsis* has one Hsp90 gene for endoplasmic reticule, mitochondria, and chloroplast, while our data mining found one and two Hsp90 Clusters for endoplasmic reticule and chloroplast respectively, and failed to find an Hsp90 cluster in mitochondria. The absence of a mitochondrial Hsp90 in *Eucalyptus* is intriguing since plant Hsp90 sequences are very conserved, sharing 60-70% identity with animal and yeast Hsp90 representatives (Krishna and Gloor, 2001). However, the *Eucalyptus* mitochondrial Hsp90 gene was probably present among the 13 Hsp90-like

Clusters found but which were not annotated due to their low confidence since the average of the sequences in each cluster was about one.

The 5'EST related to the Hsp90 family was found in all libraries (Figure 3), which is in agreement with its indispensable necessity for viability of eukaryotic organisms (Buchner, 1999). In general, the number of transcripts related to the Hsp90 family was very significant among all the libraries, but CL2 (callus) and BK1 (bark, duramen, pith, and alburnum) libraries exhibited an especially high number of Hsp90 transcripts while RT6 (root) and WD2 (wood) presented a lower number of transcripts (Figure 3). The important Hsp90 expression observed in tissues representing distinct developmental steps, such as adult (BK1) and callus (CL2), underlined the importance of this chaperone family for diverse conditions of tissue development in *Eucalyptus grandis* (see discussion below). Tissues under growth conditions demand, theoretically, more expression of Hsp90 because it activates many kinases that regulate the mitotic cell cycle (Schulte *et al.*, 1996).

Hsp100 family

Hsp100 proteins are specialized molecular machines that promote protein disaggregation and/or protein degradation, removing potentially harmful polypeptides arising from misfolding, denaturation or aggregation (Mogk *et al.*, 1999). The Hsp100 family, also named Clp (caseinolytic protease), solubilizes the aggregate protein and releases it in a conformation that can be refolded by the Hsp70 chaperone system (Wang *et al.*, 2004).

The Hsp100 family represented approximately 8% of the total 5'EST found to be molecular chaperones (Figure 2) and about 56% of these belonged to the ClpX subfamily (Table 2). ClpX belongs to the Class II of Clp ATPases (Gottesman *et al.*, 1990 and 1993), which has only one ATP-binding domain that has significant homology with the second ATP-binding domain, from Class I Clp ATPases (Horwich, 2002). This low expression of Hsp100 proteins was spread almost indistinguishably throughout all the libraries investigated (Figure 3). The lack of heat shock induction, which is required for expression of many Hsp100 genes (Schirmer *et al.*, 1994), may explain the lower expression of Hsp100 representatives compared to other families. Unlike other chaperone families, Hsp100 neither promotes the initial folding nor prevents the aggregation of proteins (Wang *et al.*, 2004). For instance, the disassemble of protein

aggregates by Hsp101 takes place mainly during the post-stress phase (see Agarwal and colleagues, 2003, for an example in plants). In other words, the methodology applied for certain libraries (*i.e.* exposition to light during three hours) was not the most effective in triggering Hsp100 differential expression in *Eucalyptus* cells.

Small heat shock proteins

Small Hsp proteins have monomeric masses ranging from 12 to 43 kDa, are characterized by a α crystallin domain in its C-terminus (De Jong *et al.*, 1998), and are responsible for maintaining the solubility of unfolding proteins during heat stress (Kim *et al.*, 2003). Small Hsps seem to form large heteroligomeric aggregates that undergo conformational changes with temperature increasing, promoting the display of hydrophobic amino acids residues that bind to proteins denaturated by heat (Ehrnsperger *et al.*, 1998). Small Hsps are highly induced in mitochondria and chloroplasts by oxidative stress (Lee and Vierling, 2000).

Of all the chaperone families investigated here, Small Hsp was the least abundant (Figure 2), which is in accordance with most small Hsps not being detected in the vegetative tissues, under normal growth conditions (Sun *et al.*, 2002), although they are capable of being produced and accumulated in response to environmental stress and developmental stimuli. Only one of its classes (cytoplasmic Class I) accounted for about 90% of the Clusters found (Table 2). These results are consistent with those from the digital expression analysis of small Hsps in sugar cane cells (Borges *et al.*, 2001) and may indicate a universal pattern of expression for this family in plants.

The expression profile analysis of the Small Hsp family in *Eucalyptus* tissues was marked by the high expression of this protein found in the BK1 library (Figure 3). This library originated from bark tissues that are in general exposed to daily stress conditions, which indicated the participation of Small Hsps in the folding process of synthesized protein in this tissue. Our analyses identified a very small or even lack of expression of Small Hsps in leaf (LV2 and LV3) and root (RT3 and RT6) libraries (Figure 3). These results are in accordance with previous works that show that Small Hsps are not detected in the absence of stress in leaf or root tissues (DeRocher *et al.*, 1991; Hsieh *et al.*, 1992).

Small Hsps are usually found in reproductive organs during several stages of plant development (Waters *et al.*, 1996) and after heat (or other) shock induction (DeRocher *et al.*, 1991), in agreement with the high levels of Small Hsp 5'ESTs found in libraries (ST6 and ST7) originating from *Eucalyptus* stem tissues susceptible to hydric deficit (Figure 3). Similar results were found for other plants such as sunflowers and *A. thaliana*, in which an increase in Small Hsp expression correlates with osmotic stress (Sun *et al.*, 2002). The dehydration can contribute to aggregation caused by the increasing concentration of proteins inside
the cells. Thus, the presence of molecular chaperones, especially Small Hsps, is important because they bind partially denatured proteins and prevent their aggregation (Ehrnsperger *et al.*, 1998). The high diversification of plant Small Hsps shows a molecular adaptation to stress conditions that seems to be exclusively from plants. Therefore, the abundance and diversity of small Hsp family suggest their important function in plant stress tolerance (Sun *et al.*, 2002; Wang *et al.*, 2003).

Tissue expression analysis

The digital northern blot analysis performed here is based on the assumption that the number of EST clones is directly related to the abundance of messenger RNA and therefore to the protein expression level in each library (Audic and Claverie, 1997). Although there is general agreement that variations may exist, this approach is generally accepted as a reasonable approximation to the true protein level of expression. The expression profile of the molecular chaperone mRNAs provided by the digital northern blot indicated that chaperones were differently expressed in the *Eucalyptus* tissues (Figure 3). The libraries with the highest amount of 5'EST sequences annotated as chaperones were SL7 (*E. grandis* seedlings), SL6 (*E. urophylla* seedlings), and ST7 (stem tissue), with 2.8%, 2.3%, and 2% of sequences, respectively. The libraries with the lowest amount of 5'EST sequences annotated as chaperones, approximately 1% each, were RT6 (root) and WD2 (wood).

The SL libraries (4, 5, 6, 7, and 8) represent five different species of *Eucalyptus* seedlings submitted to the same experimental conditions – totally cultivated in the dark: *E. grandis* (2.8% of the total 5'EST from SL7 library were related to molecular chaperone category), *E. urophylla* (2.3% of the total 5'EST from SL6), *E. saligna* (1.7% of the total 5'EST from SL5), *E. globulus* (1.5% of the total 5'EST from SL4) and *E. camaldulensis* (1.3% of the total 5'EST from SL8) (Figure 3). These results show that the expression of chaperones was species-dependent, one species with twice as much chaperone expression as the other. As far as we know, a comparison of stress resistance among these species has not yet been experimentally performed. It would be interesting to find out if this character is associated with the chaperone gene expression profile found here.

There are other observations regarding the general expression of chaperone proteins in the *Eucalyptus* that deserve to be mentioned. Two different libraries originating from *E. grandis* seedlings were created.

The SL1 library originated from seedlings cultivated in the dark and exposed to light for three hours before RNA extraction. The SL7 library originated from seedlings totally cultivated in the dark. About 2.8% of the total ESTs of the SL7 library were associated to chaperone genes, but only 1.4% were found in the SL1 library (Figure 3). These results showed that these tissues had differential chaperone expression that was dependent on the exposition to light during their development. Closer inspection of the expression profile of these two libraries showed that the increase in chaperone expression is mainly due to the increase in the expression of Hsp70 and its co-chaperones (Figure 3). There are two other libraries, originating from callus tissues, which differ from their treatment with regard to exposition to light. The CL2 library originated from callus exposed to light and the CL1 library originated from callus tissues developed in the absence of light. About 1.9% of the total ESTs of the CL2 library were associated to chaperone genes, whereas about 1.3% were found in the CL1 library (Figure 3). These results showed that these tissues also had differential chaperone expression that was dependent on the exposition to light during their development. A close inspection of the expression profile of these two libraries showed that the increase in chaperone expression is mainly due to the increase in the expression of Hsp70 and Hsp90 genes (Figure 3). Hsp70 and Hsp90 are protein families involved with basic cellular process that do not necessarily concern protein folding (for reviews see Buchner, 1999; Jensen and Johnson, 1999; Borges and Ramos, 2005) and that explains their high expression not only during stress conditions but also whenever high cellular activities take place. The diverse function of these two proteins may be the explanation for one tissue presenting higher chaperone expression when exposed to light than in dark conditions, and another tissue presenting an opposite behavior. Since seedlings are tissues in development, they may need the "over-assistance" of chaperones in basic cell functions (Wang et al., 2004). In the case of the callus tissue, the chaperones may be necessary for decreasing the effects caused by external conditions that stressed the cell. Therefore, since these two functions are difficult to distinguish and they may be taking place at the same time, more experimental work has to be done in order to support the hypothesis discussed above.

By comparing the library LV3, which was originated from leaf tissues that were colonized by the *Thyrinteina* grub, wich a leaf library not exposed to these insects (LV2) we found an increased level of chaperone expression (1.9% vs. 1.5% of the total ESTs associated to chaperone genes) (Figure 3). This

result suggested that this biotic stress condition could affect the chaperone levels in *Eucalyptus* cells. The increase in chaperone gene expression was mainly restricted to Hsp70 and Hsp60 families, which are mainly involved in the folding of newly synthesized proteins. The likely explanation is that the expression of defense proteins increases in these conditions and Hsp70 and Hsp60 are need for their proper folding (Wang *et al.*, 2004). It would be interesting to cross the data found here with the data collected for the defense proteins expression profile to search for expression patterns. In addition, Hsp90 has been implicated in biotic stress tolerance in plants, since it interacts with proteins which confer resistance to pathogens (Liu *et al.*, 2004; Lu *et al.*, 2003). Our analyses showed that Hsp90 expression was not significantly altered by the presence of *Thyrinteina* grub, indicating that the nature of the biotic stress influences Hsp90 expression.

Figure 3 also shows other libraries that have an increased expression profile of proteins associated to chaperones: roots from seedlings (RT3 had 1.8% of the total ESTs annotated as chaperones), stem tissues from six-month seedlings susceptible to hydric deficit (ST2 had 1.6%) and from seedlings susceptible to hydric-deficit (ST6 had 2.0%). Since these tissues are either under development (and therefore require an increased rate of protein synthesis and activation) or facing stress conditions, we believe it is important to point out the increased expression of chaperones that are requested for nascent protein proper folding (Hsp60 family) or protein activation (Hsp90 family) as expected for members of theses families (Boston *et al.*, 1996; Buchner, 1999).

Final Considerations

The data generated by the Brazilian *Eucalyptus* Genome Sequencing Project Consortium (FORESTs) allowed the evaluation of the most relevant molecular chaperone gene expression levels within a library context, which helped to give clues about the molecular response to diverse physiological and environmental conditions. The genes belonging to the chaperones category had a considerable expression in *Eucalyptus*, which was expected because these proteins have high importance to the cell function. The cytoplasm compartment displays the most representative expression profile concerning molecular chaperones when compared to others. The Hsp70 and Hsp40 families were the most

abundant, followed by the Hsp90 family, the Hsp60 family, the Hsp100 family, and the Small Hsp family. In conclusion, we found that stress-related proteins are abundantly expressed and diverse in *Eucalyptus*.

Acknowledgments

We would like to thank the Fundação de Amparo do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for their financial support. TCC and JCB wish to thank FAPESP for fellowships and AOT is grateful to CNPq for a fellowship.

References

Agarwal M, Sahi C, Katiyar-Agarwal S, Agarwal S, Young T,Gallie RR, Sharma VM, Ganesan K and Grover A (2003) Molecular characterization of rice Hsp 101: Complementation of yeast Hsp104 mutation by disaggregation of protein granules and differential expression in *indica* and *japonica* rice types. Plant Mol Biol 51:545-553.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller Wand Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Alvim FC, Carolino SM, Cascardo JC, Nunes CC, Martinez CA, Otoni WC and Fontes EP (2001) Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress. Plant Physiol 126:1042-1054.

Apuya NR, Yadegari R, Fischer RL, Harada JJ, Zimmerman JL and Goldberg RB (2001) The *Arabidopsis* embryo mutant *schlepperless* has a defect in the chaperonin-60 gene. Plant Physiol 126:717-730.

Audic S and Claverie J (1997) The significance of digital gene expression profiles. Gen Res 7:986-995.

Borges JC, Fischer H, Craievich AF and Ramos CHI (2005) Low resolution structural study of two human Hsp40 chaperones in solution: DjA1 from subfamily A and DjB4 from subfamily B have different quaternary structures. J Biol Chem 280, in press.

Borges JC, Peroto MC and Ramos CHI (2001) Molecular chaperone genes in the sugarcane expressed sequence database (SUCEST). Genetics and Molecular Biology 24:85-92. Borges JC and Ramos CHI (2005) Protein folding assisted by chaperones. Protein Pep Lett 12:257-261.

Boston RS, Viitanen PV and Vierling E (1996) Molecular chaperones and protein folding in plants. Plant Mol Biol 32:191-222.

Brehmer D, Rüdiger S, Gässler CS, Klostermeier D, Packschies L, Reinsten J, Mayer MP and Bukau B (2001) Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. Nat Struct Biol 8:427-432.

Buchner J (1999) Hsp90 & Co. - A holding for folding. Trends Biochem Sci 24:136-141.

De Jong WW, Caspers GJ and Leunissen JA (1998) Genealogy of the alpha-crystallin – Small heat-shock protein superfamily. Int J Biol Macromol 22:151-62.

DeRocher AE, Helm KW, Lauzon LM and Vierling E (1991) Expression of a conserved family of cytoplasmic low molecular weight heat shock proteins during heat stress and recovery. Plant Physiol 96:1038-1047.

Ehrnsperger M, Gaestel M and Buchner J (1998) Structure and function of small heat shock proteins. In: Fink AL and Goto Y (eds) Molecular Chaperones in the Life Cycle of Proteins. Dekker, New York, pp 533-575.

Ellis RJ and Hartl FU (1996) Protein folding in the cell: Competing models of chaperonin function. FASEB J 10:20-26.

Fink AL (1999) Chaperone mediated protein folding. Physiol Ver 79:425-449.

Gottesman S, Clark WP, de Crecy-Lagard V and Maurizi MR (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*: Sequence and *in vivo* activities. J Biol Chem 268:22618-22626.

Gottesman S, Squires C, Pichersky E, Carrington M, Hobbs M, Mattick JS, Dalrymple B, Kuramitsu H, Shiroza T and Foster T (1990) Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. Proc Natl Acad Sci USA 87:3513-3517.

Grantcharova V, Alm EJ, Baker D and Horwich AL (2001) Mechanisms of protein folding. Curr Opin Struct Biol 11:70-82.

Guy CL and Li QB (1998) The organization and the evolution of the spinach stress 70 molecular chaperone gene family. Plant Cell 10:539-556.

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381:571-580.

Hill JE and Hemmingsen SM (2001) *Arabidopsis thaliana* type I and II chaperonins. Cell Stress Chap 6:190-200.

Horwich A (2002) Protein folding in the cell. In: Richards FM, Eisenberg DS and Kuriyan J (eds) Advances in Protein Chemistry. Academic Press, California, pp 175-180.

Hsieh M-H, Chen J-T, Jinn T-L, Chen Y-M and Lin C-Y (1992) A class of soybean low molecular weight heat shock proteins. Immunological study and quantitation. Plant Physiol 99:1279-1284.

Hubert DA, Tornero P, Belkhadir Y, Krishna P, Takahashi A, Shirasu K and Dangl JL (2003) Cytosolic Hsp90 associates with and modulates the *Arabidopsis* RPM1 disease resistance protein. EMBO J 22:5679-5689.

Jensen RE and Johnson AE (1999) Protein translocation: Is Hsp70 pulling my chain? Curr Biol 9:779-782.

Kim R, Lai L, Lee HH, Cheong GW, Kim KK, Wu Z, Yokota H, Marqusee S and Kim SH (2003) On the mechanism of chaperone activity of the small heat-shock protein of *Methanococcus jannaschii*. Proc Natl Acad Sci USA 14:8151-8155.

Krishna P and Gloor G (2001) The Hsp90 family of proteins in *Arabidopsis thaliana*. Cell Stress Chap 6:238-246.

Lee GJ and Vierling E (2000) A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. Plant Physiol 122:189-197.

Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, Barakat A and Delseny M (2001) Genomic analysis of the Hsp70 superfamily in *Arabidopsis thaliana*. Cell Stress Chap 6:201-208.

Liu Y, Burch-Smith T, Schiff M, Feng S and Dinesh-Kumar SP (2004) Molecular chaperone Hsp90 associates with resistance protein N and its signaling proteins SGT1 and Rar1 to modulate an innate immune response in plants. J Biol Chem 279:2101-2108.

Lu Z and Cyr DM (1998) Protein folding activity of Hsp70 is modified differentially by the Hsp40 cochaperones Sis1 and Ydj1. J Biol Chem 273:27824-27830.

Lu R, Malcuit I, Moffett P, Ruiz MT, Peart J, Wu AJ, Rathjen JP, Bendahmane A, Day L and Baulcombe DC (2003) High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. EMBO J 22:5690-5699.

Mayer K and Mewes HW (2002) How can we deliver the large plant genomes? Strategies and perspectives. Curr Opin Plant Biol 5:173-177.

Miernyk JA (2001) The J-domain proteins of *Arabidopsis thaliana*: An unexpectedly large and diverse family of chaperones. Cell Stress Chaperones 6:209-218.

Mogk A, Tomoyasu T, Goloubnoff P, Rüdger S, Röder D, Langen H and Bukau B (1999) Identification of thermolabile *Escherichia coli* proteins: Prevention and reversion of aggregation by DnaK and ClpB. EMBO J 18:6934-6949.

Morimoto RI (1998) Regulation of heat shock transcriptional response: Cross talking between a family of transcription factors, molecular chaperones and negative regulators. Genes Dev 12:3788-3796.

Netzer WJ and Hartl FU (1998) Protein folding in the cytosol: Chaperonin-dependent and independent mechanisms. Trends Biochem Sci 23:68-73.

Nogueira FTS, De Rosa Jr VE, Menossi M, Ulian EC and Arruda P (2003) RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiol 132:1811-1824.

Pearl LH and Prodromou C (2001) Structure, function, and mechanism of the Hsp90 molecular chaperone. Adv Protein Chem 59:157-186.

Queitsch C, Sangster TA and Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417:618-624.

Schirmer EC, Lindquist S and Vierling E (1994) An *Arabidopsis* heat-stress protein complements a thermotolerance defect in yeast. Plant Cell 6:1989-1909.

Schulte TW, Blagosklonny MV, Romanova L, Mushinski JF, Monia BP, Johnston JF, Nguyen P, Trepel J and NeckersLM (1996) Destabilization of Raf-1 by geldanamycin leads to disruption of the Raf-1-MEK-mitogen-activated protein kinase-signalling pathway. Mol Cell Biol 16:5839-5845.

Siegers K, Waldmann T, Leroux MR, Grein K, Shevchenko A, Schiebel E and Hartl FU (1999) Compartmentation of protein folding *in vivo*: Sequestration of non-native polypeptide by the chaperonin-GimC system. EMBO J 18:75-84.

Sugino M, Hibino T, Tanaka Y, Nii N, Takabe T and Takabe T (1999) Overexpression of DnaK from a halotolerant cyanobacterium *Aphanotece halophytica* acquires resistance to salt stress in transgenic tobacco plants. Plant Sci 146:81-88.

Sun W, Van Montagu M and Verbruggen N (2002) Small heat shock proteins and stress tolerance in plants. Bioch et Bioph Acta 1577:1-9.

Sung DY, Vierling E and Guy CL (2001) Comprehensive expression profile analysis of the *Arabidopsis* Hsp70 gene family. Plant Physiol 126:789-800.

Wang W, Vinocur B and Altman A (2003) Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. Planta 218:1-14.

Wang W, Vinocur B and Altman A (2004) Role of plant heatshock proteins and molecular chaperones in the abiotic response. Trends Plant Sci 9:244-252.

Waters ER, Lee GJ and Vierling E (1996) Evolution, structure and function of the small heat shock proteins in plants. J Exp Bot 47:325-338.

Zhao R, Davey M, Hsu YC, Kaplanek P, Tong A, Parsons AB, Krogan N, Cagney G, Mai D, Greenblatt J, Boone C, Emili A and Houry WA (2005) Navigating the chaperone network: An integrative map of physical and genetic interactions mediated by the Hsp90 chaperone. Cell 120:715-27.

Figures and Tables

Table 1 - Molecular chaperones classification.

Chaperone family	Co-chaperone	Molecular mass (kDa)	n-mers	Comments
Hsp100	Unknown	≈100	6 or 7	ATPase activity; recover proteins from aggregates, form proteolytic complexes
Hsp90	HIP, HOP, p23 and others	≈90	2	Weak intrinsic ATPase activity; prevent protein aggregation; in- teract with a large set of proteins like transcription factors and pro- tein kinase
Hsp70	Hsp40, GrpE and others	≈70	1 /2 /2	ATPase activity; bind nascent peptide chains to help many pro- cesses such as folding, transport through membranes, degrada- tion, escape from aggregation
Hsp60 (GroEL)	Hsp10 (GroES)	≈60	14/7	ATPase activity; assist protein to fold efficiently
SmHsp	Unknown	15-30	9 to24	ATP-independent, involved in thermotolerance and escape from aggregation by keeping proteins in a refoldable state

Molecular chaperones			Mining results				
Family	Class	Intracellular location	Annotated clusters	Annotated 5' EST sequences	Total Annotated 5' EST sequences (family only)		
Small Hsp	Class I Class II Class III Class IV	cytoplasm cytoplasm mitochondria chloroplast	24 - 1 3	69 - 2 11	82		
Hsp60 /TriC ¹	TCP-1 ² Hsp60-like	mMitochondria chloroplast cytoplasm -	3 3 9 -	58 89 100 19	266		
Chaperonin-like co-chaperone	Hsp10	chloroplast	4	27	27		
Hsp70	Hsc70 Mt-Hsp70 Cp-Hsp70 Bip ³ /Grp78 Hsp110	cytoplasm mitochondria chloroplast ER ⁴	21 3 3 5 5	272 25 56 89 53	495		
Hsp70 co-chaperones	Hsp40	cytoplasm, mitochondria, ER mitochondria	91	568	578		
Hsp90	Hsp82 Grp94 Cp-Hsp82 Hsp90-like	cytoplasm ER chloroplast	10 1 2 13	279 41 25 15	360		
Clp ⁶	Hsp100/ClpB ClpA/C ClpX	cytoplasm, mitochondria chloroplast	4 13 10	13 53 85	151		
Total			232	1,959	1,959 (1.6%)		

Table 2 - The main chaperone families and classes, their cellular location and the mining results. Our data mining approach found 1,959 5'ESTs belonging to the molecular chaperones category, which is about 1.6% of the total 5'ESTs sequenced by the FORESTs consortium.

¹TriC: TCP ring Complex; ²TCP1: Chaperonin-containing T-Complex; ³Bip: Binding protein; ⁴ER: endoplasmic reticulum; ⁵GrpE: GroP-like gene E; ⁶Clp: Caseinolytic protease.



Figure 1 - Data mining and annotation strategy flowchart. Translated amino acid sequences of specific mRNAs were chosen from chaperones and stress-related proteins as described by Borges *et al.* (2001). The prototype sequences were compared with the first level cluster consensus generated by the *Eucalyptus* Genome Sequencing Project Consortium (FORESTs) using the basic local alignment search tool tBLASTn program (Altschul *et al.*, 1997). The tBLASTn program was used to compare amino acid sequences from public databases with the FORESTs database. The FORESTs cluster sequences with an E-value lower than 1e-5 were considered to belong to the chaperone category. To improve the prediction accuracy, matches found in the first round of mining went through a second or even more rounds of mining, allowing precise identification of more matches and/or new Clusters not previously found.



Figure 2 - Relative abundance of the chaperone families in *Eucalyptus*. The number of reads of each chaperone family was normalized as the percentage of the total chaperones annotated. The Hsp60 family slice represented here also contains the Hsp10 family, which represented nearly 2% of all annotated chaperones in FORESTs.



Figure 3 - Expression of the chaperone families in the different tissues of eucalyptus. The percentages represent the number of 5'EST associated to chaperone genes in each library. The colors indicate the contribution of each chaperone family. The FORESTs libraries are: BK1 (bark, duramen, pith, and alburnum); CL1 (*E. grandis* callus formed in the dark), CL2 (*E. grandis* callus formed in the light); FB1 (flower buds, flowers, and fruits); LV2 (leaves of efficient and inefficient trees in Phosphorus and Boron utilization); LV3 (leaves colonized by the *Thyrinteina* grub for seven days); RT3 (roots from seedlings); RT6 (roots from trees resistant and susceptible to frost); SL1 (*E. grandis* seedlings cultivated in the dark and exposed to light for three hours before the RNA extraction); SL4 (*E. globulus* seedlings cultivated in the dark); SL5 (*E. saligna* seedlings cultivated in the dark); SL6 (*E. urophylla* seedlings cultivated in the dark); ST2 (stem from six month seedlings susceptible to hydric deficit); ST6 (seedlings stem susceptible to hydric deficit); and WD2 (*E. grandis* wood).

8.2 Artigo: Journal of Plant Physiology 164: 505-513 (2007)

Expression and variability of molecular chaperones in the sugarcane expressome

Julio C. Borges^{a,b,*}, Thiago C. Cagliari^{a,b}, Carlos H.I. Ramos^a

^aLaboratório Nacional de Luz Síncrotron, Caixa Postal 6192, 13084-971 Campinas SP, Brasil ^bDepartamento de Bioquímica, Instituto de Biologia, Universidade Estadual de Campinas, Campinas SP, Brasil

Summary

Molecular chaperones perform folding assistance in newly synthesized polypeptides preventing aggregation processes, recovering proteins from aggregates, among other important cellular functions. Thus their study presents great biotechnological importance. The present work discusses the mining for chaperone-related sequences within the sugarcane EST genome project database, which resulted in approximately 300 different sequences. Since molecular chaperones are highly conserved in most organisms studied so far, the number of sequences related to these proteins in sugarcane was very similar to the number found in the Arabidopsis thaliana genome. The Hsp70 family was the main molecular chaperone system present in the sugarcane expressome. However, many other relevant molecular chaperones were found in every sugarcane library, despite their heterogeneous expression profiles. The results presented here suggest the importance of molecular chaperones to polypeptide metabolism in sugarcane cells, based on their abundance and variability. Finally, these data have being used to guide more in deep analysis, permitting the choice of specific targets to study.

Abbreviations: AD, SUCEST library AD1; AM, SUCEST libraries AM1 and AM2; CL, SUCEST libraries CL1-7; CAP3, Contig assembly program; cpn60, chloroplast Hsp60; EST, expressed sequence tag; FL, SUCEST libraries FL1-6 and FL8; HR, SUCEST library HR1; Hsp, heat shock protein; LB, SUCEST libraries LB1 and

LB2; LR, SUCEST libraries LR1 and LR2; LV, SUCEST library LV1; Phrap, phragment assembly program; PDIase, protein disulfide isomerase; PPIase, peptidyl-prolyl cis-trans isomerase; RT, SUCEST libraries RT1, RT2 and RT3; RuBisCo, Ribulose bisphosphate carboxylase; RZ, SUCEST libraries RZ1, RZ2 and RZ3; SB, SUCEST library SB1; SD, SUCEST libraries SD1 and SD2; SmHsp, small heat shock proteins; ST, SUCEST libraries ST1, ST2 and ST3; SUCEST, sugar cane EST genome project; TCP, tric complex proteins *Corresponding author. Present address. Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas, UNESP, Rua Cristóvão Colombo, 2265, Jd. Nazareth, CEP: 15054-000 São José do Rio Preto, Brasil.

E-mail addresses: borgesjc@yahoo.com (J.C. Borges), cramos@lnls.br (C.H.I. Ramos).

Introduction

Sugarcane is one of the most important plants cultivated in tropical and subtropical areas worldwide and it has mainly been used for the sugar production. Nowadays, sugarcane plantations have increased their economical role in Brazil for sugar production and as energy source (Vettore et al., 2003). The sugarcane is also used in the production of ethanol, as an alternative fuel to Brazilian vehicles. In order to attend the Kyoto protocol, many countries have bought ethanol from Brazil and are considering this technology for the future, as the number of petroleum reservoirs constantly decreases. The main goal of the sugar cane EST genome project (SUCEST) is to improve the sugarcane production through a genomic investigation approach. This genome project is based on expressed sequence tags (ESTs), which are also described as expressome or transcriptome, and may indicate trustworthy patterns of gene expression in specific tissues or even in entire organisms (Mayer and Mewes, 2002).

Molecular chaperones are proteins that assist the folding of other proteins in vivo without interfering in the final native structure of their substrate (Gething and Sambrook, 1992; Ellis and Hartl, 1996; Miernyk, 1999; Hartl and Hayer-Hartl, 2002; Borges and Ramos, 2005). They may be described as proteins that prevent non-productive interactions involving folding intermediates, avoiding misfolding and aggregation in the pathway of protein folding (Gething and Sambrook, 1992; Borges and Ramos, 2005). Thus, the protein folding study and the role of molecular chaperones is of great importance to the biotechnology and the pharmaceutical industries. For instance, molecular chaperones may be used to improve production of eukaryotic proteins with economic interest that are unable to fold inside bacterial cells and also to understand the aggregation-based amyloid diseases (Ruddon and Bedows, 1997; Westerheide and Morimoto, 2005).

Molecular chaperones are found in all intracellular compartments of plants and may be described in terms of their molecular weight and biochemical functions (Boston et al., 1996; Miernyk, 1999; Westerheide and Morimoto, 2005). The most relevant molecular chaperone families are: (1) The chaperonin or Hsp60 family, which includes GroEL, GroES, TCP (Tric complex protein) and recently, the prefoldin as well, assist protein folding at a late stage (Ellis and Hartl, 1996; Hartl and Hayer-Hartl, 2002). Chloroplast Hsp60 (cpn60) assists the folding of the ribulose bisphosphate carboxylase (RuBisCo), which is the most abundant protein in plant cells.

The cpn60 was previously described as the RuBisCo large-subunit-binding protein (Boston et al., 1996; Miernyk, 1999). (2) Hsp70 and its co-chaperones, GrpE and J-domain proteins (or Hsp40), which bind growing peptide chains while they are synthesized, participate in protein transport through membranes, preventing aggregation processes, recovering proteins from aggregates, among other functions (Fink, 1999; Hartl and Hayer-Hartl, 2002). The Hsp70 system works in a pivotal way receiving from and also delivering substrates to other molecular chaperone systems (Borges and Ramos, 2005). The Hsp70 family is highly present in many plant EST genome projects such as Eucalyptus EST genome (Cagliari et al., 2005). (3) Hsp90 family is overexpressed in response to stress and it plays an important role in signal transduction, being one of the most abundant heat shock proteins inside non-stressed cells (Buchner, 1999). It may interact with other chaperones and co-chaperones to form a chaperone complex in the cytoplasm including proteins from the Hsp70 system, co-chaperones (p23, Hip and Hop), foldases (see below) and a huge range of cytoplasmic proteins (Miernyk, 1999; Wegele et al., 2004). (4) Hsp100 family is able to disassemble and to refold aggregated proteins in cooperation with the Hsp70 system, and participating in proteolysis complexes (Schirmer et al., 1996; Bem-Zvi and Goloubinoff, 2001). (5) The small heat shock proteins family (SmHsp) prevents protein aggregation process by a temperaturedependent mechanism, and it is conserved to a lesser extent than other molecular chaperones are though (Jakob and Buchner, 1994; Waters et al., 1996).(6) Calnexin and calreticulin are calcium-binding proteins located in the rough endoplasmic reticule and are involved in protein folding, aggregation avoidance and oligomerization, among other functions (Miernyk, 1999). (7) Foldases: protein disulfide isomerase (PDIase) and peptidyl-prolyl cistrans isomerase (PPIase) are folding catalysts enzymes that accelerate slow chemical reactions, which are rate limiting in the protein folding process (Boston et al., 1996; Miernyk, 1999). Many proteins are described as co-chaperones because they assist molecular chaperones in their function. Many co-chaperones interacts with more than one chaperone system, such as Hip and Hop, which participate both in Hsp70 and Hsp90 systems, revealing the complementary aspects of different molecular chaperone systems within the cells (Miernyk, 1999).

This work aims to comprehend the abundance and distribution of molecular chaperones in sugarcane cells through an in silico identification of the transcripts related to them in the SUCEST database, providing a remarkable source to guide the choice of specific targets to other studies. We compared the chaperone-related sequences present in the SUCEST database to the Arabidopsis thaliana genome and the result was a very similar content, despite the methodological differences between these two genome projects. We also searched for molecular chaperone expression profiles in the sugarcane tissues by a digital RNA blot approach, allowing the evaluation of their expression levels in several different conditions and tissues. Our results revealed that molecular chaperones were significantly diverse and abundantly expressed in the sugarcane cells. Moreover, some differences were observed when considering the specific libraries as followed below.

Materials and methods

The SUCEST consortium (Vettore et al., 2001; Vettore et al., 2003 – sucest.lbi.ic.unicamp.br/en/) performed the preparation and the large-scale sequencing of sugarcane ESTs. The description of the SUCEST libraries, which were built from different sugarcane tissue samples obtained from different Saccharum hybrid cultivars, is reported by Vettore et al. (2001). The bioinformatics tools used are described by Telles et al. (2001). The molecular chaperone annotation process was performed as previously described by Borges et al. (2001) and Cagliari et al. (2005). Summarizing, prototype sequences were compared with the cluster consensus, in the database generated by the software, Contig assembly program (CAP3 – Huang and Madan, 1999), utilizing the program tBLASTn (Altschul et al., 1997), which is an alignment driven research tool. Any CAP3 cluster that exhibited a statistically significant match (i.e. an

E-value lower than 1e–5) was considered to be a molecular chaperone. In order to improve the accuracy of the prediction, the open reading frames (ORFs) found in the first round of mining were resubmitted through a second or even more rounds of mining. This process allowed a precise confirmation of the previously found sequences and also the identification of new clusters that were not previously found (Borges et al., 2001; Cagliari et al., 2005). Therefore, each additional round increases the probability to find clusters that were incorrectly excluded in previous rounds.

In order to determine the number of potential chaperone-related genes in the SUCEST database, the following analyses were considered: (1) analysis of chaperone-related sequences in both 5' and 3' EST senses in different CAP3 clusters, which represent the nucleotide sequence from either end of the same cDNA; (2) identification of the related sequences among different CAP3 clusters (when the identity between sequences was greater than 95%); (3) the sequencing quality of CAP3 clusters formed by only one EST. These procedures increased the accuracy of the numeric prediction of chaperone-related sequences in the SUCEST database, while minimizing the redundancy of the clustering process (see below). The ORFs were identified by comparison, with the best overall hit in the BLASTx NR search and by the ORF Finder software (Wheeler et al., 2003 –www.ncbi.nlm.nih.gov/gorf/gorf.html). The potential chaperone related sequences in the Arabidopsis thaliana genome were found by tBLASTn sequence analysis of the Arabidopsis thaliana genome database (The Arabidopsis Genome Initiative, 2000 – www.tigr.org/tdb/e2k1/ath1/) and also in the literature (Table 1).

The global chaperone expression level in the SUCEST database was calculated based on the number of all 5'ESTs annotated to a certain molecular chaperone family divided by the total number of sugarcane 5'ESTs. Only ESTs from clones that were sequenced from the 50-end and that belonged to non-normalized libraries were scored (Audic and Claverie, 1997). The chaperone expression level in each particular tissue was calculated by dividing the number of the 5'ESTs annotated to a specific chaperone family by the total number of 5'ESTs sequenced in that particular library. The construction of the sugarcane libraries are designated (Vettore et al., 2001) as follows: AD (SUCEST library AD1) – plantlets infected with *Gluconacetobacter diazotroficans*; AM (SUCEST libraries AM1 and AM2) – apical meristem; CL (SUCEST libraries CL1-7) – callus; FL (SUCEST libraries FL1-6 and FL8) –sugarcane flower; HR (SUCEST library HR1) – plantlets infected with Herbaspirilum diazotroficans; LB (SUCEST libraries LB1 and LB2) –

lateral bud; LR (SUCEST libraries LR1 and LR2) – leaf roll; LV (SUCEST library LV1) – leaves; RT (SUCEST libraries RT1, RT2 and RT3) – root; RZ (SUCEST libraries RZ1, RZ2 and RZ3) – leaf–root transition zone; SB (SUCEST library SB1) – stem bark; ST (SUCEST libraries ST1, ST2 and ST3) – stem; SD (SUCEST libraries SD1 and SD2) – seeds.

Results and discussion

The SUCEST clustering process was performed by two distinct approaches. The Phrap (Phragment assembly program) software was used in a preliminary approach, producing an array of 81,223 clusters (Telles et al., 2001; Telles and da Silva, 2001). However, this array presented many problems, such as cluster excess, ribosomal sequences and nearly 41,500 clusters formed by only one EST, which were named singletons (Telles and da Silva, 2001). These problems compelled the SUCEST bioinformatics group to perform a second clustering process employing the CAP3 software (Huang and Madan, 1999; Telles et al., 2001; Telles and da Silva, 2001), which resulted in a more accurate clustering process than the Phrap software (Liang et al., 2000). For instance, the total number of SUCEST clusters was reduced from 81,223 to 43,141, a significant reduction of almost 50%. The number of singletons was decreased from around 41,500 to 16,838, which represented a reduction of about 60% when compared to the first clustering approach (Telles and da Silva, 2001). The analysis from all 43,141 CAP3 clusters indicated that 26,982 of these presented a BLASTx NR match with an E-value below 1.0e-5. No similar proteins were found in the BLASTx NR search for the remaining 16,156 CAP3 clusters in that database. Therefore, the SUCEST expressome may be shared in a preliminary database, constructed with Phrap software, and a posterior database, built with CAP3 software.

Our group already developed an identification of chaperone-related sequences in the Phrap-built SUCEST database (Borges et al., 2001). In that preliminary work, 777 different SUCEST Phrap clusters related to molecular chaperones were identified. However, an accurate prediction of the number of genes of molecular chaperones could not be established, since SUCEST Phrap database allowed only the prediction of subclasses genes (Borges et al., 2001). The search for chaperone-related sequences in the SUCEST database generated by CAP3 software yielded information about 425 CAP3 clusters with an E-value lower than 1.0e-5 in BLASTx NR. Surprisingly, these 425 clusters represented almost half of the previously reported number (Borges et al., 2001), illustrating that CAP3 software substantially decreased the redundancy and facilitating other mining analysis (Telles and da Silva, 2001).

Among the chaperone-related clusters identified, 142 CAP3 clusters were predicted to have the entire coding region when compared to similar proteins, while 283 CAP3 clusters displayed incomplete sequences (Table 1), and were, therefore, considered as putative sequences. The number of clusters

related to molecular chaperones represented around 1.6% of the total CAP3 clusters, emphasizing their significant presence and diversity in the SUCEST database. Analyzing 5'EST-ends in the CAP3 database, approximately 2.2% of them were annotated to molecular chaperones, which represents 4,702 5'ESTs from 213,318 5'EST-end used in clustering process. The abundance of chaperone-related 5'ESTs confirms their high variability and importance to sugarcane protein folding metabolism. In comparison to the first report (Borges et al., 2001) an increase of 0.4% in the chaperone-related sequences was detected and it can be attributed to three main reasons. The first one is the inclusion of other important molecular chaperones in this data mining work, such as p23, Hip, Hop and prefoldin. Second, the improved search for proteins with a high variability, like J-domain proteins and foldases. Finally, the CAP3 clustering process reduced the number of singletons, which presented an E-value greater than 1.00e-5 in the BLASTx NR, although they were chaperone-related sequences.

A total of 71 putative CAP3 clusters can be assembled into a complete CAP3 cluster following the 50–30end analysis and the similarity criteria. Other 55 complete CAP3 clusters can be grouped into another CAP3 cluster using the same approach (see Materials and methods). Therefore, the SUCEST database presented almost 15% of redundant chaperone-related sequences. These additional clustering processes were possible only due to low sequence quality of some ESTs, which might have harmed the clustering process, increasing the redundancy. As discussed by Telles and da Silva (2001), the second approach resulted in an important redundancy decrease of SUCEST sequences when compared to the first approach. However, some technique-generated features could not be completely removed, even from the second clustering process. Thus, the definition "one cluster, one gene" cannot be applied to this databank context and must not be attributed to the SUCEST expressome content. Consequently, the SUCEST database is more appropriately defined as "a set of very similar transcripts" (Telles and da Silva, 2001).

The annotation presented here suggested that the sugarcane expressome contains approximately 300 different cDNAs related to molecular chaperones (Table 1). These sugarcane sequences were compared to the Arabidopsis thaliana genome, resulting in a very similar number of specific chaperones in both organisms (Table 1). These results reinforce the great importance of those proteins to polypeptide metabolism in plant cells and their high conservation degree, if one considers the difference of genome

magnitudes from those two organisms. The small differences between them might be due to the fact that the SUCEST project is an expressome, which did not allow detection of pseudogenes such as in the *Arabidopsis thaliana* genome project. Nevertheless, the SUCEST project allowed detections of introns and exons from messenger RNAs, which the Arabidopsis thaliana genome project did not. These messenger RNAs may be products of the same gene but with different post-transcriptional modifications (Pradet-Balade et al., 2001). The evaluation of the chaperone-related 5'ESTs indicated that 74% were forming almost one third of all CAP3 clusters annotated as chaperones and that contained full length coding region. A total of 8% (71 putative CAP3 clusters) were grouped into any complete CAP3 cluster. The remaining 18% were related to putative CAP3 clusters that represented half of all chaperone-annotated SUCEST CAP3 clusters. Therefore, around 80% of all 5'ESTs indicating those sequences that are the mostly transcribed mRNAs, and subsequently the more abundantly expressed proteins. In addition, those less scored 5'ESTs might be either tissue or time-specifically expressed. However, these data need to be carefully considered since the lifespan of mRNA and intracellular protein are different and a direct correlation between their levels is not often reasonable (Pradet-Balade et al., 2001).

Digital RNA blot

Fig. 1 represents a digital RNA blot analysis of chaperone-related 5'ESTs from different tissues that were used in the building process of the SUCEST expressome (see Materials and methods). The molecular chaperone expression level observed for each tissue, considering the ratio of 5'EST annotated to molecular chaperone in that tissue by the total number of sequenced 5'EST was constant for half of the tissues. The chaperone-related expression level in the LB, LR and CL tissues were above the average (95% of the confidence interval of the mean) and the Hsp70 family was the most expressed chaperone system in these tissues (see below). Data concerning high expression levels of molecular chaperones may be explained by the following: LB libraries ought to contribute with mRNAs from initials organ differentiation stages; LR libraries were constructed from leaves of immature plants, and CL grouped libraries were submitted to cold and heat treatments (Vettore et al., 2001). The expression analysis of AM, HR, RTand SD tissues revealed an expression level below average (Fig. 1). The relative low

expression in RT and SD grouped libraries may be explained by root and seeds physiological characteristics, in minor environmental stress exposition and tissue storage. Despite the apical meristem (AM library) presents characteristic of a differentiation tissue, its expression profile was unexpectedly and strangely inferior when compared to LB grouped libraries, if one considers that these tissues share similar characteristics.

While the Hsp70 family (J-domain proteins and Hsp70 chaperones) was heavily expressed in the LB tissue, its expression in the AM tissue was not significant when compared to others (Fig. 1). Plantlets infected with Herbaspirilum diazotroficans (HR library) exhibited low expression of genes from the Hsp70 family. Additionally, the 5'ESTs from Hsp100, cpn21 and SmHsp proteins were absent, while the expression of foldases was elevated (Fig. 1). Specific molecular chaperone analysis indicated the presence of chaperone-related 5'ESTs annotated to every families studied in all libraries, except by families with low overall expression level (Hsp100, cpn21 and SmHsp were not identified in HR; p23 was not identified in SB; Cpn21 was not identified in LV and SmHsp was not identified in AD libraries). A total of 4,702 chaperone-related 5'ESTs from the SUCEST database were annotated to different molecular chaperone families (Table 1). About 42% were related to the Hsp70 family and its J-domain co-chaperone proteins, 12% to the Hsp60/TCP, 9% to the Hsp90 and 15% to the PPIase families. Despite the usage of a new database, these numbers were similar to those exposed in the preliminary data mining work and they confirmed its conclusion (Borges et al., 2001), if one consider that the total 5'ESTs sequenced by SUCEST project is in good accordance with mRNA expression profile in sugarcane cells.

Considering the libraries, the Hsp70 family (J-domain and Hsp70 chaperones) was the most expressed molecular chaperone system (Fig. 1). These data indicate their relevance for protein folding in various types of cell, as previously shown (Boston et al., 1996; Fink, 1999; Miernyk, 2001; Lin et al., 2001; Cagliari et al., 2005). Nonetheless, some particularities between the Hsp70 and the J-domain proteins expression patterns were noticed. While only 19 sequences were predicted to code to Hsp70 proteins, 87 possible J-domain proteins were identified. These data suggest that J-domain proteins are more diverse than Hsp70 in sugarcane cells. Therefore, it is reasonable to suppose that the J-domain proteins, which are structurally unrelated (Borges et al., 2005), may act as a scanning factor, selecting substrate to the less stringently Hsp70 chaperone (Ru⁻⁻diger et al., 2001). When considering the Hsp70 family in detail, three

different varieties of Hsp70 and members of the Hsp110 sub-family were found, and their locations predicted to the cytoplasm, mitochondria, chloroplast and lumen (data not shown). A CAP3 cluster coding a member of the Hsc70 sub-family, which was predicted to be located in the cytoplasm, was responsible for 75% of all 5'ESTs related to Hsp70 chaperone and a single J-domain protein sequence was responsible for 44% from all 5'EST annotated as J-domain members (data not shown). These data suggested the main Hsp70 and J-domain proteins expressed in sugarcane cells.

ESTs attributed to the Hsp60 family were found in all the SUCESTs libraries (Fig. 1). In the cell proliferative environment, the Hsp60-related 5'EST content in AM and CL tissues were above average. Together, this data indicates that Hsp60 is relevant for protein folding in several tissues. The Hsp60-related 5'ESTs were mainly located in the chloroplast (75%) when compared to the mitochondria (25%). The cpn60 is responsible for the folding of the RuBisCo enzyme and other important proteins, which might explain their reasonable expression. The 5'ESTs related to cytoplasm Hsp60 homologue, TCP-1, represented 7% from all chaperone-annotated 5'ESTs. Higher expression levels of the TCP-1 sub-family, when compared to other Hsp60 subfamilies, may be explained for the folding of newly synthesized proteins in the cytoplasm (Grantcharova et al., 2001).

Proteins from Hsp90 family are indispensable for viability in eukaryotic organisms (Buchner, 1999). Hsp90-related 5'ESTs were found in all libraries, displaying a considerably higher expression level in LV library (Fig. 1). The majority of 5'ESTs (72%) related to the Hsp90 family were attributed to the cytoplasmic protein, followed by the mitochondrial protein (19%). The foldases (PDIase and PPIase) seemed to be indispensable for sugarcane cells, since they were found in all libraries, as well as calnexin (Fig. 1).

Conclusions

A reasonable number of chaperone-related sequences were estimated for the SUCEST expressome, which was in good agreement to the amount of sequences found in the Arabidopsis thaliana genome. The mining work resulted in the identification of approximately 300 different molecular chaperones in the SUCEST CAP3 database, confirming that the CAP3 software produced a less redundant database than the Phrap software did. The SUCEST database was submitted to a digital RNA blot analysis, leading to the

conclusion that these proteins were found in all sugarcane tissues, despite their heterogeneous expression. These data corroborates to the high abundance and variability of molecular chaperones and their importance to sugarcane protein folding activities. These results also make possible other data mining analyses and rendered new insights about sugarcane molecular chaperones using SUCEST CAP3 database as information source. Besides the interesting features found, a careful, deep and individual analysis is needed in order to relate these data to functional mechanisms of molecular chaperones. Finally, this data has been used to guide other analyses and permitted the choice, among the chaperone-annotated sequences, of specific targets for structure-function studies.

Acknowledgments

We sincerely thank FAPESP and MCT/CNPq (Ministério da Ciência e Tecnologia/Conselho Nacional de Desenvolvimento Científico e Tecnológico) for financial support. TCC and JCB thank FAPESP for fellowships. We are in great debt to all the SUCEST members for the data construct what helped us to accomplish this work.

References

Agarwal M, Katiyar-Agarwal S, Sahi C, Gallie DR, Grover A. Arabdopsis thaliana Hsp100 proteins: kith and kin. Cell Stress Chaperones 2001;6:219–24.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389–402.

Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res 1997; 7:986–95.

Bem-Zvi AP, Goloubinoff P. Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. J Struct Biol 2001; 135:84–93.

Borges JC, Ramos CHI. Protein folding assisted by chaperones. Protein Pept Lett 2005;12:256–61.

Borges JC, Peroto MC, Ramos CHI. Molecular chaperone genes in the sugarcane expressed sequence database (SUCEST). Genet Mol Biol 2001;24:85–92.

Borges JC, Fischer H, Craievich AF, Ramos CHI. Low-resolution structural study of two human Hsp40 chaperones in solution. DjA1 from subfamily A and DjB4 from subfamily B, have different quaternary structures. J Biol Chem 2005;280:13671–81.

Boston RS, Viitanen PV, Vierling E. Molecular chaperones and protein folding in plants. Plant Mol Biol 1996;32:191–222.

Buchner J. Hsp90 & Co. – a holding for folding. Trends Biochem Sci 1999;24:136–41.

Cagliari TC, Tiroli AO, Borges JC, Ramos CHI. Identification and in silico expression pattern analysis of eucalyptus expressed sequencing tags (ESTs) encoding molecular chaperones. Genet Mol Biol 2005;28:520–8.

Ellis RJ, Hartl FU. Protein folding in the cell: competing models of chaperonin function. FASEB J 1996;10:20–6.

Fink AL. Chaperone-mediated protein folding. Physiol Rev 1999; 79:425–49.

Gething MJ, Sambrook J. Protein folding in the cell. Nature 1992; 355:33–45.

Grantcharova V, Alm EJ, Baker D, Horwich AL. Mechanisms of protein folding. Curr Opin Struct Biol 2001; 11: 70–82.

Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 2002; 295:1852–8.

Hill JE, Hemmingsen SM. Arabidopsis thaliana type I and type II chaperonins. Cell Stress Chaperones 2001; 6:219–24.

Huang X, Madan A. CAP3: a DNA sequence assembly program. Genome Res 1999; 9:868–77. Jakob U, Buchner J. Assisting spontaneity: the role of Hsp90 and small Hsps as molecular chaperones. Trends Biochem Sci 1994; 19:205–11.

Krishna P, Gloor G. The Hsp90 family of proteins in Arabidopsis thalina. Cell Stress Chaperones 2001; 6:238–46.

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J. An optimized protocol for analysis of EST sequences. Nucleic Acids Res 2000; 28:3657–65.

Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, Barakat A, et al. Genomic analysis of the Hsp70 superfamily in Arabidopsis thaliana. Cell Stress Chaperones 2001; 6:201–8.

Mayer K, Mewes HW. How can we deliver the large plant genomes? Strategies and perspectives. Curr Opin Plant Biol 2002; 5:173–7.

Miernyk JA. Protein folding in the plant cell. Plant Physiol 1999; 121:695–703.

Miernyk JA. The J-domain proteins of Arabidopsis thaliana: an unexpectedly large and diverse family of chaperones. Cell Stress Chaperones 2001; 6:209–18.

Nover L, Bharti K, Döring P, Mishra SK, Ganguli A, Scharf KD. Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones 2001; 6:177–89.

Pradet-Balade B, Boulme' F, Beug H, Müllmer EW, Garcia-Sanz JA. Translation control: bridging the gap between genomics and proteomics? Trends Biochem Sci 2001; 26:225–9.

Ruddon RW, Bedows E. Assisted protein folding. J Biol Chem 1997; 272:3125–8.

Rüdiger S, Scheneider-Mergener J, Bukau B. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. EMBO J 2001; 20:1042–50.

Scharf KD, Siddique M, Vierling E. The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing *a*-crystallin domains (Acd proteins). Cell Stress Chaperones 2001; 6:225–37.

Schirmer EC, Glover JR, Singer MA, Lindquist S. HSP100/Clp proteins: a common mechanism explains diverse functions. Trends Biochem Sci 1996; 21:289–96.

Telles GP, da Silva FR. Trimming and clustering sugarcane ESTs. Genet Mol Biol 2001; 24:17–23.

Telles GP, Braga MDV, Dias Z, Lin T-L, Quitzau JAA, da Silva FR, et al. Bioinformatics of the sugarcane EST project. Genet Mol Biol 2001; 24:9–15.

The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 2000; 408:796–815.

Vettore AL, da Silva FR, Kemper EL, Arruda P. The libraries that made SUCEST. Genet Mol Biol 2001; 24:1–7.

Vettore AL, Da Silva FR, Kemper EL, Souza GM, Da Silva AM, Ferro MI, et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 2003; 13:2725–35.

Waters ER, Lee GJ, Vierling E. Evolution structure and function of the small heat shock protein in plants. J Exp Bot 1996; 47:325–38.

Webb MA, Cavaletto JM, Klanrit P. Orthologs in Arabidopsis thaliana of the Hsp70 interacting protein Hip. Cell Stress Chaperones 2001; 6:247–55. Wegele H, Müller L, Buchner J. Hsp70 and Hsp90 – a relay team for protein folding. Rev Physiol Biochem Pharmacol 2004; 151:1–44.

Westerheide SD, Morimoto RI. Heat shock response modulators as therapeutic tools for diseases of protein conformation. J Biol Chem 2005; 280:33097–100.

Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, et al. Database resources of the National Center for Biotechnology. Nucleic Acids Res 2003; 31:28–33.

Figures and Tables

Table 1. Molecular chaperones are abundant and diverse in the SUCEST database and in the Arabidopsis thaliana genome

Molecular chaperones	SUCEST analysis						
Family	CAP3 clusters		Possible different proteins		Percent in		
		Complete	Putative	SUCEST	Arabidopsis thaliana	the socest	
Hsp100	Hsp101/ClpB	_	6	4	8 ^a	1.3	
	ClpA/C/X	1	19	9	7 ⁶	3.0	
Hsp90		4	21	9	7 ^c	3.0	
Hsp90 co-chaperone p23		4	_	3	1 ^b	1.0	
Hsp70		9	25	19	18 ^d	6.3	
Hsp70 <i>co</i> -chaperones	J-domain proteins	30	82	87	89 ^e	29.0	
	GrpE	1	2	3	3 ^b	1.0	
	HOP	4	13	13	2 ^b	4.3	
	HIP	_	3	2	2 ^f	0.7	
Hsp60/TCP Hsp60 <i>co</i> -chaperone Prefoldin		13	21	24	20 ^{g,h}	8.0	
		6	_	6	5 ^g	1.7	
		7	2	7	6 ^g	2.3	
SmHsp		25	6	24	19 ⁱ 20 ^b	8.0	
Calnexin – calreticulin		5	10	5	8 ^b	1.7	
Foldases	PPlase	25	30	42	46 ^b	14.0	
	PDIase	9	26	23	11 ^b	7.7	
Heat shock transcription fa	1	24	22	21 ^j	7.3		
Trigger factor	-	1	1	1 ^b	0.3		
Total		142 (425)	283	301	273	100	

The table displays the most relevant data about chaperone-related sequences found in SUCEST CAP3 and Arabidopsis thaliana genome databases.

Both organisms presented similar number of predicted genes/cDNAs.

Around 1.5% of SUCEST CAP3 clusters, that had a BLASTx NR hit, were annotated as a molecular chaperone.

These results demonstrate that molecular chaperones genes are diverse in the sugarcane genome and in the Arabidopsis thaliana genome.

^aAgarwal et al. (2001), two are putative amino acids sequences.

^bMining in Arabidopsis thaliana genome database (The Arabidopsis Genome Initiative, 2000).

^cKrishna and Gloor (2001).

^dLin et al. (2001).

^eMiernyk (2001).

^fWebb et al. (2001).

^gHill and Hemmingsen (2001).

^hTwo pseudogenes for Hsp60.

ⁱScharf et al. (2001).

^jNover et al. (2001).



Figure 1. Expression profile of molecular chaperones is heterogeneous in the SUCEST tissue database. The expression levels were calculated as the percentage of the number of 50EST related to a specific molecular chaperone in each particular tissue library. The continuous line represents the mean of chaperone-related 50EST identified in SUCEST database and the dashed lines are the confidence intervals. UCIL – upper confidence interval of the mean and LCIL – lower confidence interval of the mean.

9 Curriculum Vitae

Thiago Carlos Cagliari (Cagliari TC)

Nascido em 23 de Setembro de 1977, em Valinhos/SP.

Formação acadêmica

- 2003 2009: Doutorando em Biologia Funcional e Molecular com ênfase em Bioquímica (Instituto de Biologia - UNICAMP) sob a orientação do Prof. Dr. Carlos H.I. Ramos (livre-docente, Instituto de Química - UNICAMP). Auxílio financeiro: Bolsa de doutorado-direto FAPESP.
- 1998 2002: Graduação em Ciências Biológicas (Licenciatura) pela Universidade Estadual de Campinas (UNICAMP).

Outras atividades:

1999 - 2001: Bolsista de iniciação científica do programa PIBIC/CNPq no Laboratório Nacional de Luz Síncrotron sob orientação do prof. Dr. Jörg Kobarg.

2001 - 2002: Bolsista de iniciação científica da FAPESP no Laboratório Nacional de Luz Síncrotron sob orientação do prof. Dr. Jörg Kobarg.

Áreas de interesse

Química de Macromoléculas – Proteínas; Bioquímica – Estrutura e função de chaperonas moleculares; Biofísica – Enovelamento de proteínas e estabilidade.

Publicações Científicas

LIRA, C. B. ; SIQUEIRA NETO, J. L. ; KHATER, L. ; CAGLIARI, T. C. ; PERONI, L. A. ; REIS, J. R. ; RAMOS, C. H. I. ; CANO, M. I. LaTBP1: a Leishmania amazonensis DNA-binding protein that associates in vivo with

telomeres and GT-rich DNA using a Myb-like domain. Archives of Biochemistry and Biophysics, v. 465, p. 399-409, 2007.

BORGES, J. C. ; CAGLIARI, T. C. ; RAMOS, C. H. I. Expression and variability of molecular chaperones in the sugarcane expressome. Journal of Plant Physiology, v. 164, p. 505-513, 2006.

SEBOLLELA, A. ; CAGLIARI, T. C. ; LIMAVERDE, G. S. C. S. ; CHAPEAUROUGE, A. ; SORGINE, M. H. F. ; COELHO-SAMPAIO, T. ; RAMOS, C. H. I. ; FERREIRA, S. T. Heparin-binding sites in Granulocyte-Macrophage Colony-stimulating Factor. The Journal of Biological Chemistry, v. 280, p. 31949-31956, 2005.

CAGLIARI, T. C. ; TIROLI, A. O. ; BORGES, J. C. ; RAMOS, C. H. I. Identification and in silico expression pattern analysis of Eucalyptus expressed sequencing tags (ESTs) encoding molecular chaperones. Genetics and Molecular Biology, Brasil, v. 28, n. 3, p. 520-528, 2005.

Publicações em Anais de congressos e reuniões científicas

Publicou 9 trabalhos em Congressos e reuniões científicas, sendo 8 em forma de painel e uma em forma oral.