Universidade Estadual de Campinas Instituto de Física Gleb Wataghin Campinas, 26 de Fevereiro de 2003 Tese de Doutorado:

Estudos estruturais sobre a β -manosidase de Trichoderma reesei e a Triose fosfato isomerase de músculo de coelho

RICARDO APARICIO

Banca examinadora:IFSC/USPOrientador: Prof. Dr. Igor PolikarpovIFSC/USPProf. Dr. Richard Charles GarrattIFSC/USPProf. Dra. Heloise de Oliveira PastoreIQ/UNICAMPProf. Dra. Maria Cristina dos SantosIFGW/UNICAMPProf. Dr. Carlos Manuel Giles A. de MayoloIFGW/UNICAMP

FICHA CATALOGRÁFICA ELABORADA PELA

BIBLIOTECA DO IFGW - UNICAMP

Ap12e

Aparicio, Ricardo Estudos estruturais sobre a β-manosidase de Trichoderma reesei e a Triose fosfato isomerase de músculo de coelho / Ricardo Aparicio. – Campinas, SP : [s.n.], 2003.

Orientador: Igor Polikarpov. Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Física "Gleb Wataghin".

1. Cristalografia de proteínas. 2. Raios X – Espalhamento a baixo ângulo. 3. Biologia estrutural. I. Polikarpov, Igor. II. Universidade Estadual de Campinas. Instituto de Física "Gleb Wataghin". III. Título.

(vsv/ifgw)



iv

Este trabalho foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP, através do projeto 98/06761-1.

Documento preparado 100% em Linux usando ${\rm I\!AT}_{\rm E}\!{\rm X}\,2_{\mathcal{E}}.$



Prefácio

A história desta tese começou bem antes de eu ingressar no Curso 51 da UNICAMP, o "cursão", em 1994. Desde criança eu dizia que queria ser cientista, mesmo sem saber direito o que de fato significava e, de alguma forma acabei indo por este caminho durante a adolescência, quando montei um pequeno laboratório dentro do meu quarto (como bico de bulsen, uma lamparina a querosene, como beckers e bastões de vidro, copos e talheres desapropriados da cozinha). Lá brinquei de químico, o que, mais tarde, me levou a ser micro-empresário por alguns anos com relativo sucesso. Embora eu tivesse algum dinheiro no bolso, havia um enorme vazio de conhecimento. Eu sempre quisera saber que poder tem um Físico de, sentado a uma mesinha, formular teorias e fazer previsões. Como tal maravilha seria possível? Além disso, como seria possível traduzir a natureza em termos matemáticos?! Isso foi o que me levou a escolher a Física. Muitas outras razões eu viria a descobrir depois e até hoje eu continuo escolhendo a Física mesmo com as minhas limitações.

Houve momentos muito importantes durante a graduação. O primeiro foi o contato com o Método Científico no início do curso, introduzido de forma contundente pelo Prof. René Brenzikofer através de experimentos simples em Mecânica Clássica. Outro foi quando decidi fazer Iniciação Científica no IMECC, até hoje lembro-me e sou muito grato, em especial, ao Prof. Alcebíades Rigas pelo suporte na aquisição de um ferramental matemático que ampliou muito minha percepção das relações entre a Física e a Matemática. Foi também aí que tive oportunidade de conhecer gênios da Matemática durante um curso de verão no IMPA (RJ) e que entendi que eu não era nem gênio e nem seria um Matemático, apesar de admirar os Matemáticos e gostar muito de Matemática.

Muitos professores tiveram papel decisivo na minha vida como cientista. Lembro do carinho pela Física demonstrado pelo Prof. Jorge Ivan Cisneros, do qual fomos a última turma antes de ele se aposentar. Agradeço ao Prof. Roberto Martins por saber melhor porque a Física é como é hoje, de ter a clara consciência de que temos modelos apenas. Tenho carinho também pelo Prof. Bernardo Laks, um Físico excelente que não tem preconceito de também conversar sobre Renoir. Me lembro de como passei dezenas de fins de semana fazendo as listas de exercícios que a Profa. Kyoko Furuya nos dava e... de como me apaixonei por Mecânica Quântica! Muito obrigado, Profa.. E os amigos? Puxa, que turma dez nós fomos! Tive muita sorte de ter companhia tão agradável e competente durante minha graduação, muitos são hoje colegas de profissão. Deixo aqui um caloroso abraço a todos.

Eu não poderia deixar de lembrar como ingressei no doutorado. Foi no último semestre da graduação que tive contato com a "Transformada de Fourier" numa disciplina dada pela Profa. Iris Torriani, que sempre esteve presente desde então. Fiquei pasmo quando pude suavizar uma figura de jornal, que era composta por pontinhos, através de experimentos em Óptica usando raio *laser*. Quando o Prof. Igor Polikarpov, que seria meu futuro orientador de doutorado, apresentou um seminário sobre Cristalografia de Proteínas no IFGW eu não tive dúvidas de que aquele seria o caminho a seguir. Era uma área de pesquisa interdisciplinar que não apenas estava diretamente relacionada ao uso da Teoria de Fourier como também à Biologia, assunto pelo qual eu estava muito interessado. Para completar, o trabalho seria realizado no Laboratório Nacional de Luz Síncrotron (LNLS), um centro de excelência em pesquisa. A cada boa lembrança dessas uma outra é esquecida e, para não cometer mais injustiças, deixo meus profundos agradecimentos a todo o corpo docente do IFGW, onde tenho a mais grata satisfação de ter estudado durante todo este tempo.

Quando entrei no LNLS, logo no início do doutorado, senti muito orgulho de ter feito Física, de ser brasileiro e de saber que, pelo menos em um lugar, o dinheiro público era tratado com o devido cuidado. O LNLS é uma prova da capacidade dos brasileiros. Se eu fosse decorrer sobre minhas impressões positivas daquele laboratório, seriam muitas páginas mas me atenho a estas poucas linhas. Há muito o que agradecer ao LNLS, desde o corpo administrativo, pelo apoio necessário, até o pessoal da limpeza. Sempre troquei muitos sorrisos - gratuitos de todos os lados - com todos dali.

Isto me traz ao "Grupo CPR", um time interdisciplinar de primeira que o Igor formou no LNLS, quase todos vindos da UNICAMP. Eu precisaria de um livro para contar as nossas histórias e, como até hoje estamos próximos, me limito a uma palavra de carinho para todos, dizendo muito obrigado por ter conhecido todos vocês. Entretanto, não posso deixar de agradecer especialmente a dois amigos. Primeiro ao amigo José Brandão Neto, companheiro de coletas de dados, de inovações instrumentais, profissional competente e criativo, mais um que nosso país deixou escapar. Em seguida, ao meu amigo Basílio (que também está no exterior), que me incentivou a procurar ajuda para os complicados problemas que eu vinha enfrentando em minha pesquisa. Disto resultou, com apoio do Igor, o período que passei em York, na Inglaterra, onde obtive os resultados positivos apresentados nesta tese. Lá, tive oportunidade de conhecer pessoas maravilhosas que foram fundamentais na minha formação. Agradeço a todos do York Structural Biology Laboratory pela ajuda gratuita e carinho que me deram. Especialmente, gostaria de deixar um terno agradecimento aos Drs. Eleanor e Guy Dodson e aos Drs. Garib Murshudov e Marek Brzozowski. Posteriormente, mudamos para o IFSC, na USP de São Carlos, onde conheci novos amigos. Em especial, vai um abraço e o meu muito obrigado para o Hannes, pessoa que admiro e com quem trabalhei por algum tempo.

Quero agradecer também aos amigos que colaboraram indiretamente para que eu chegasse até aqui. Me sinto feliz por ter como amigo o José L. Saunders, que agradeço pelo apoio e incentivo constantes desde quando resolvi ser Físico até hoje. Obrigado pelas opiniões, pelas longas madrugadas de diálogos. Tenho gratas lembranças da Renata, que esteve ao meu lado por tantos anos, desde os tempos daquele Corcel II com o qual eu ia até o laboratório. Que tantas vezes foi paciente, por tantas vezes esperou as longas coletas de dados no LNLS, suportou os fins de semana em que eu trabalhei e o desgaste que isso nos trazia. Obrigado pelo companheirismo. Agradeço à minha irmã Cristina por sempre ter se preocupado com meu bem-estar, por todo seu carinho comigo e ao Márcio pelas idéias, por apontar alternativas e outros possíveis pontos de vista durante nossas longas conversas.

Apesar dos caminhos tortuosos que se apresentaram durante o projeto, talvez este tenha sido o melhor doutorado que eu poderia fazer, à moda russa, como diz o Igor, a quem agradeço pelo suporte e apoio em tudo que foi preciso. Eu faria tudo outra vez se fosse necessário. Espero ter dado uma contribuição ao meu país e ter colocado algumas pequeninas peças no grande quebra-cabeças da natureza. Se ainda não tenha sido como eu gostaria, foi o melhor possível que pude fazer.

Por fim, agradeço à Dna. Lourdes e ao Sr. Olival pelo seu amor, pessoas únicas que, quanto mais o tempo passa, mais me orgulho de ter como pais.

> Ricardo Aparicio Agosto/2003

Resumo

A presente tese resume os resultados obtidos durante o programa de doutorado do autor. O projeto principal consistiu em estudos estruturais das enzimas β -manosidase do fungo *Trichoderma reesei* e Triose fosfato isomerase (TIM) de músculo de coelho. Cristalografia de Proteínas e Espalhamento a Baixo Ângulo foram as técnicas mais utilizadas. Os principais resultados foram a determinação da estrutura de baixa resolução e o provável enovelamento da β -manosidase e três estruturas cristalográficas de TIM, resolvidas em duas formas cristalinas e a diferentes temperaturas. Na estrutura de TIM de maior resolução (1.5 Å), o *loop* do sítio ativo de uma das subunidades encontra-se na conformação fechada, apesar da ausência de ligante, fato nunca antes observado. Em ambos os projetos de pesquisa, informação biológica nova e relevante foi obtida. Como um resultado dos trabalhos de doutorado, um total de cinco artigos foram publicados em periódicos internacionais e dois encontram-se em fase final de preparação.

xii

Abstract

This PhD thesis summarizes the results obtained during the PhD program of the author. The main project consisted in structural studies of β -mannosidase from *Trichoderma reesei* and rabbit muscle Triose phosphate isomerase (TIM). Protein Crystallography and Small-Angle X-ray Scattering have been the mostly used techniques. The main results are the low resolution structure and a putative fold of β -mannosidase and three cristallographic TIM structures, determined in two crystal forms and different temperatures. One of the subunits in the highest resolution TIM structure (1.5 Å) has the active site loop in the closed conformation despite the absence of a ligand in the active site, what had not been observed before. Both projects have led to new and important biological information. As a result of the doctorate program, five articles have been published and two further are in preparation.

 \mathbf{xiv}

Índice

Pı	refác	io		vii
Re	esum	0		xi
A	bstra	ct	2	xiii
Ín	dice			xv
Li	Lista de Figuras xix			xix
Li	sta d	le Tab	elas	xxi
1	Intr	oduçã	0	1
	1.1	β -mar	nosidase	1
		1.1.1	A degradação de glicoproteínas e a β -manosidose	2
		1.1.2	A doença provocada pela ausência de $\beta\text{-manosidase}$	3
		1.1.3	Degradação de hemiceluloses	5
		1.1.4	$\beta\text{-manosidases}$ em plantas, alcalofílicos e termo fílicos $% \beta$.	6
		1.1.5	$\beta\text{-manosidase}$ em pacientes diabéticos	7
		1.1.6	A β -manosidase de Trichoderma reesei	7
	1.2	Triose	fosfato isomerase	9
		1.2.1	TIM de músculo de coelho	13
	1.3	A cris	talografia de Proteínas por difração de Raios X	15

ÍNDICE

2	Me	todologia 1	17	
	2.1	Cristalização	18	
	2.2	Criocristalografia	23	
	2.3	Derivados de átomos pesados 2	25	
		2.3.1 Critérios de escolha dos compostos utilizados 2	25	
		2.3.2 Preparação de cristais derivados	28	
	2.4	Coleta de dados	29	
		2.4.1 Fontes de Raios X	30	
		2.4.2 Aparato experimental e montagem dos cristais 3	31	
		2.4.3 Redução de dados	35	
	2.5	O problema das fases	37	
		2.5.1 Métodos diretos	38	
		2.5.2 Substituição Molecular	39	
		2.5.3 Átomos pesados	40	
	2.6	Refinamento do modelo	52	
		2.6.1 Free R_{factor} ou R_{free}	53	
		2.6.2 Validação e deposição da estrutura	56	
3	β -m	anosidase: estudos cristalográficos 5	59	
	3.1	Purificação e seqüenciamento	59	
	3.2	Cristalização da β -manosidase 6	34	
	3.3	Coletas de dados no LNLS	38	
		3.3.1 Otimização das condições de cristalização 6	38	
		3.3.2 Preparação para a coleta de cristais nativos e derivados 6	39	
		3.3.3 Coleta a temperaturas criogênicas	70	
		3.3.4 Análise preliminar dos conjuntos coletados até 1999 7	70	
		3.3.5 Análise preliminar dos conjuntos coletados em 2000 7	74	
	3.4	Resultados do estágio no YSBL		
		3.4.1 Diagnóstico para os dados coletados no LNLS 8	31	

		3.4.2	Coletas de dados no ESRF	. 95
4	β - m	anosio	lase: estudos a baixa resolução	101
	4.1	Novos	s resultados cristalográficos	. 109
		4.1.1	Faseamento no grupo $P2_12_12_1$. 109
		4.1.2	Conteúdo de solvente da ASU	. 111
		4.1.3	A configuração do eixo helicoidal nos dois grupos	. 111
		4.1.4	A falta de isomorfismo entre os cristais e inexistência	
			de perfeita simetria	. 114
		4.1.5	O espalhamento anisotrópico	. 115
		4.1.6	Últimas tentativas de Substituição Molecular	. 118
5	TIN	/I: cris	talização e coleta de dados	119
	5.1	Coleta	as de dados no LNLS	. 119
	5.2	Coleta	as de dados no ESRF	. 124
6	TIN	/I: nov	o paradigma para o <i>loop</i> do sítio ativo	129
	6.1	Suppo	prting information	. 168
7	Cor	nclusão)	173
	7.1	Outra	s publicações	. 175
A	Teo	ria: de	efinições elementares	177
В	Cur	csos, co	ongressos e atividades acadêmicas	181
	B.1	Ativid	lades acadêmicas	. 181
	B.2	Curso	s e congressos	. 183
Re	e fer ê	ncias l	Bibliográficas	189

Lista de Figuras

1.1	Reação química catalisada pela β -manosidase
1.2	Reação química catalisada pela TIM
1.3	O barril-TIM e <i>loop</i> do sítio ativo
2.1	Poço de cristalização
2.2	Fotos da linha CPR, no LNLS
2.3	Aparato experimental
2.4	Posicionamento de cristal para coleta
3.1	Cristais de β -manosidase
3.2	Imagens do conjunto bm200600
3.3	Escalonamento das imagens do conjunto bm200600 85
3.4	Conjunto bm200600: seções do espaço recíproco
3.5	Uma das imagens do conjunto bm Eu 040200 91
3.6	Mapa de densidade eletrônica
3.7	SDS-PAGE gel: amostras de β -manosidase contaminadas 96
3.8	Seção hk0 do conjunto do grupo $P2_12_12_1$ 99
4.1	Eixo helicoidal
4.2	Empacotamento cristalino
5.1	Cristais de TIM (LNLS)

LISTA DE FIGURAS

5.2	Cristais de TIM (YSBL)
6.1	Fatores de temperatura para estrutura de TIM (A1) 169
6.2	Fatores de temperatura para estrutura de TIM (A2) 170
6.3	Fatores de temperatura para estrutura de TIM (B) 171

Lista de Tabelas

2.1	$f' \in f''$ para dois comprimentos de onda	29
3.1	$\beta\text{-manosidase:}$ conjuntos coletados em 1999 (LNLS)	71
3.2	$\beta\text{-manosidase:}$ conjuntos coletados em 2000 (LNLS)	75
3.3	Sítios de átomos pesados: comparação de coordenadas	79
3.4	Estatísticas do conjunto bm200600	87
3.5	Figuras de mérito.	94
3.6	β -manosidase: conjuntos coletados no ESRF	97
5.1	TIM: conjunto coletado no ESRF	125

Capítulo 1

Introdução

A presente tese resulta dos trabalhos desenvolvidos durante o programa de doutorado do autor, cujo principal objetivo foram estudos estruturais das enzimas β -manosidase de *Trichoderma reesei* e Triose fosfato isomerase (TIM) de músculo de coelho através de técnicas de Raios X.

1.1 β -manosidase

A β -manosidase (β -D-manosídio manohidrolase, EC 3.2.1.25) é uma exoglicosidase que catalisa a hidrólise de resíduos de β -D-manose em β -Dmanosídios. Sua existência foi comprovada experimentalmente em 1943, onde fazia parte de preparações de invertase (β -D-frutofuranosídio frutohidrolase, EC 3.2.1.26) (Adams *et al.*, 1943). A enzima foi isolada de mais de 30 fontes diferentes, compreendendo plantas (Houston *et al.*, 1974; Li & Lee, 1972), fungos (Bouquelet *et al.*, 1978; Neustroev *et al.*, 1992; Kulminskaya *et al.*, 1999), animais (Iwasaki *et al.*, 1989; Cavanagh *et al.*, 1985; Percheron *et al.*, 1992), e bactérias (Akino *et al.*, 1988; Oda *et al.*, 1993).

Grupos de cientistas, em várias partes do mundo têm estudado β manosidases com diferentes propósitos mas não havia nenhuma informação experimental direta sobre sua estrutura tri-dimensional. Sua estrutura atômica pode ser de grande valia na compreensão dos mecanismos de sua ação enzimática nos vários processos expostos abaixo, com destaque para a degradação lisossômica de glicoproteínas e a β -manosidose. Neste sentido, é interessante notar que as β -manosidases de fungos, baixos eucariontes, apresentam várias semelhanças com as de eucariontes mais complexos, como o homem, por ex. (Neustroev *et al.*, 1992; Kulminskaya *et al.*, 1999; Percheron *et al.*, 1992). O sítio de ligação à galactomanana, cuja função é desconhecida, e uma estrutura compreendendo, possivelmente, mais de um domínio, características particulares da β -manosidase de *T. reesei*, também constituem fatos interessantes sobre a enzima. Por estas razões, os estudos estruturais descritos aqui foram iniciados.

A β -manosidase participa de vários processos biológicos, sumarizados abaixo. Apresentamos, também, as principais propriedades e características da enzima secretada pelo fungo *T. reesei*, objeto de nosso estudo.

1.1.1 A degradação de glicoproteínas e a β -manosidose

A maioria das proteínas secretadas e associadas à membrana das células de eucariontes são glicosiladas. As cadeias de oligossacarídeos podem estar ligadas à cadeia polipeptídica de duas formas (Voet & Voet, 1995):

- ligação N-glicosídica: uma GlcNAc¹ encontra-se ligada ao nitrogênio do grupo amida de um resíduo de Asn;
- ligação O-glicosídica: os oligossacarídeos estão ligados ao grupo OH de um resíduo de Ser ou Thr.

Apesar da grande diversidade, todos os oligossacarídeos ligados Nglicosidicamente (Kornfeld & Kornfeld, 1985) possuem o núcleo comum

¹ Abreviações utilizadas: NeuAc: ácido N-acetilneuramínico (siálico); GlcNAc: N-acetilglucosamina; Man: manose; Gal: galactose; Fuc: fucose.



A degradação de glicoproteínas ocorre predominantemente nos lisossomos celulares por ação seqüencial de endo e exoglicosidases. A deficiência de alguma enzima participante do processo resulta em acúmulo de uma ou mais classes de macromoléculas contendo oligossacarídeos. O esquema da Figura 1.1 ilustra o processo para um oligossacarídeo complexo². A β -manosidase é a exoglicosidase que cliva a ligação $\beta(1 \rightarrow 4)$ em $Man\beta(1 \rightarrow 4)GlcNAc$ do núcleo acima mencionado (Scriver *et al.*, 1989).

1.1.2 A doença provocada pela ausência de β -manosidase

Defeitos nos genes que codificam glicosidases são responsáveis por muitos males geneticamente herdados e estão entre as mais freqüentes síndromes de base genética no homem, entre elas, o grande grupo de doenças provocadas por acúmulos lisossômicos (*lisossomal storage deseases*) (Watts & Gibbs, 1986). A β -manosidose é uma doença hereditária, de caráter autossômico recessivo, devido a uma deficiência isolada de β -manosidase que provoca o acúmulo, em vários tecidos, do dissacarídeo $Man\beta(1 \rightarrow 4)GlcNAc$, do trissacarídeo $Man\beta(1 \rightarrow 4)GlcNAc\beta(1 \rightarrow 4)GlcNAc$ e de outros oligossacarídeos

² Um oligossacarídeo ligado N-glicosidicamente é classificado como complexo quando contém números variáveis de unidades de GlcNAc, de resíduos de ácido siálico e/ou fucose, ligados ao núcleo comum.



Figura 1.1: Atividades enzimáticas envolvidas no catabolismo lisossômico de um oligossacarídeo complexo ligado à cadeia polipeptídica de uma glicoproteína por meio de uma ligação N-glicosídica. A figura indica a ligação química sobre a qual a β -manosidase atua.

mais complexos (Percheron *et al.*, 1992). A doença foi descrita primeiramente em bodes (Jones & Dawson, 1981) e, após, no homem (Cooper & Sardharwalla, 1986; Wenger *et al.*, 1986) e em bovinos (Bryan *et al.*, 1990).

Em bodes e bois, à parte a sintomatologia clínica, a profunda deficiência de β -manosidase no plasma e em vários tecidos provoca extensiva vacuolação citoplasmática (acúmulo de vacúolos lisossômicos em todos os tipos de células), desmielinação no sistema nervoso central e hipotireoidismo. Nos tecidos e urina, ocorre o acúmulo, principalmente, do trissacarídeo associado à deficiência enzimática. Sem intervenção, a doença leva o animal à morte no período neonatal, com um ou dois dias de vida. Alguns países requerem teste negativo para β -manosidose em bodes para transações comerciais (Mason, 1986). No homem, as manifestações clínicas são menos severas que nos ruminantes e os casos reportados exibem considerável heterogeneidade clínica, incluindo vários graus de retardamento mental, problemas na fala e audição, neuropatias periféricas, dismorfologia e infecções recorrentes de pele e do trato respiratório (Rodriguez-Serna *et al.*, 1996).

A base molecular da β -manosidose humana começou a ser entendida a partir do seqüenciamento do cDNA para a β -manosidase humana e identificação de mutações relacionadas à doença (Alkhayat *et al.*, 1998). A seqüência completa da região codificando o cDNA da β -manosidase caprina foi determinada (Leipprandt *et al.*, 1996). A comparação com cDNA bovino (Chen *et al.*, 1995) revela que 96,3% dos nucleotídios e 95,2% dos aminoácidos deduzidos são idênticos. A identidade das seqüências de aminoácidos das β manosidases humana, caprina e bovina é de 75% (Alkhayat *et al.*, 1998).

1.1.3 Degradação de hemiceluloses

Hemiceluloses são cadeias curtas de heteropolissacarídeos ramificados compostos de hexoses e pentoses que, em conjunto com a celulose e lignina, são os principais componentes da madeira (as hemiceluloses mais importantes na madeira são xilanas e glucomananas). A associação das hemiceluloses com a celulose e lignina contribui para a rigidez e flexibilidade da fibra das paredes celulares em plantas. Hemiceluloses estão presentes em resíduos agrícolas, representando, em média, 30% de sua composição (Kuhad *et al.*, 1997).

A β -manosidase secretada por fungos e bactérias capazes de degradar hemiceluloses participa da hidrólise de 1, $4 - \beta - D$ -manooligossacarídeos produzidos da polpa de hemicelulose por endoenzimas; daí a β -manosidase ser dita, também, uma mananase (as hemicelulases mais importantes são as xilanases e as mananases) (Suurnakki *et al.*, 1997). A ausência desta enzima nas preparações de mananases resulta no acúmulo de oligossacarídeos durante a hidrólise de mananas (Kuhad *et al.*, 1997). A biotecnologia baseada nas hemicelulases tem aplicações na indústria de polpa e papel, de alimentos e têxtil, na bioconversão de açucares poliméricos para produção de etanol e outros químicos, dentre outras.

1.1.4 β -manosidases em plantas, alcalofílicos e termofílicos

A β -manosidase está presente em várias plantas e participa do processo de germinação de sementes que têm galactomananas como carboidrato de reserva, onde converte manooligossacarídeos produzidos pela β -mananase em monossacarídeos (Dey & del Campillo, 1984).

Pouco é conhecido a respeito de β -manosidases e β -mananases $(1, 4 - \beta - D - manana manohidrolase, EC 3.2.1.78)$ produzidas por microorganismos alcalofílicos, embora uma β -manosidase com pH ótimo de 7.0 tenha sido isolada de uma linhagem (AM001) de *Bacillus* sp., uma bactéria alcalofílica (Akino *et al.*, 1988).

Microorganismos termofílicos vivem entre 70 °C e 100 °C, produzindo enzimas termoestáveis que apresentam considerável potencial biotecnológico. Como exemplo, β -manosidases foram isoladas de duas espécies de microorganismos termofílicos. Nestes organismos, a enzima poderia participar na degradação de galactomananas extracelulares com fins nutricionais (Duffaud *et al.*, 1997; Bauer *et al.*, 1996) ou na biossíntese intracelular de proteínas, componentes de membrana e outros compostos (Bauer *et al.*, 1996).

1.1.5 β -manosidase em pacientes diabéticos

As atividades de quatro glicosidases lisossômicas, incluindo a β manosidase, foram avaliadas em 64 pacientes diabéticos. De acordo com o tipo de *diabetes mellitus*, três atividades estavam levemente aumentadas. A atividade β -manosidase, entretanto, mostrou um decréscimo altamente significativo no soro dos pacientes, qualquer que fosse o tipo de diabetes (Bernard *et al.*, 1985). Até o momento, não houve prosseguimento desta pesquisa.

1.1.6 A β -manosidase de Trichoderma reesei

Propriedades da enzima

A β -manosidase secretada pelo fungo *Trichoderma reesei* é uma glicoproteína de 105 kDa purificada a partir de um filtrado da cultura do fungo (Kulminskaya *et al.*, 1999). Ela apresenta atividade específica para PNPM (p-nitrofenil- $\beta - D$ -manopirosídio), um substrato artificial, de 3.2 U (mg de proteína)⁻¹ no pH ótimo de 3.5, com um $K_m = 0.12$ mM; seu PI é de 4.85. Chitina, xilana, amido crú e celulose microcristalina não têm afinidade pela enzima. A β -manosidase hidrolisa $\beta - 1$, 4-manooligossacarídeos ($4-O-\beta-D$ -manopiranosil-D-manopiranosídio, entre outros) com cinética dependente do comprimento da cadeia e libera manose de frações solúveis e insolúveis de β -galactomanana (Kulminskaya *et al.*, 1999).

Organização estrutural compreendendo vários domínios

É sabido que algumas proteínas apresentam múltiplos domínios em suas estruturas onde, além do domínio catalítico, existe um domínio de ligação ao substrato que, em vários casos, intensifica a ação enzimática. Há exemplos onde o sítio de ligação ao substrato pode ser eliminado por proteólise limitada, influenciando a atividade específica da enzima (Teramoto *et al.*, 1989; Kuhad et al., 1997). Várias celulases, representantes bem estudadas deste tipo de enzima (Kuhad *et al.*, 1997), possuem domínios de ligação à celulose (*cellulose-binding domain* - CBD) que, embora não sejam essenciais para a atividade catalítica, modulam as atividades específicas das enzimas para substratos celulósicos solúveis e insolúveis. No entanto, as interações entre tais domínios e a celulose não são claras. Alguns microorganismos capazes de degradar lignocelulose produzem sistemas complexos de celulases extracelulares e pode ser vantajoso excretar enzimas ligantes ao substrato e difusíveis, já que alguns destes (por ex., Trichoderma reesei) são organismos relativamente estáticos (Gilkes et al., 1991). Os CBDs existem também em hemicelulases. Um sítio de ligação a xilanas e celulose, encontrado na xilanase D de *Cellulomonas fimi*, aumenta a afinidade da enzima por xilana insolúvel (Black et al., 1995). Além das xilanases, várias mananases apresentam estruturas organizadas em multidomínios (Suurnakki et al., 1997).

Um outro exemplo é o da glucoamilase (EC 3.2.1.3) secretada pelo fungo Aspergillus awamori (usada industrialmente para hidrólise de amido e produção de xarope de glicose, etanol e adoçantes), onde a presença de um sítio de ligação a ciclodextrinas permite a ligação da enzima às paredes celulares do fungo, aumentando localmente a concentração enzimática e melhorarando o fluxo de glicose para as células, criando, assim, condições mais favoráveis para o crescimento do fungo (Neustroev *et al.*, 1993). Além de glicosidases, outros exemplos existem, como a fosforilase participante da quebra do glicogênio (EC 2.4.1.1) (Voet & Voet, 1995). Por meio de estudos cinéticos, descobriu-se que a β -manosidase de *T. reesei* porta um sítio não-catalítico de ligação à β -galactomanana (β galactomannana-binding site - GMBS) (Kulminskaya et al., 1999). Ainda não está claro o papel desempenhado pelo GMBS e se este sítio está contido num domínio estruturalmente separado do domínio que contém o sítio ativo.

Estudos teóricos sobre o provável enovelamento da β -manosidose também foram descritos. Com o uso de uma classificação das glicosidases (O-glicosídio hidrolases, EC 3.2.1.x) baseada em similaridades seqüenciais, uma modelagem teórica para a β -manosidase bovina (Chen *et al.*, 1995) e outras quatro glicosidases implicadas em doenças provocadas por acúmulos lisossômicos foi feita com o objetivo de identificar características estruturais e funcionais de seus domínios catalíticos (Durand *et al.*, 1997).

1.2 Triose fosfato isomerase

A glicólise é a via metabólica pela qual a glicose é convertida em piruvato com a geração de dois mols de ATP por mol de glicose, fornecendo parte da energia utilizada pela maioria das células. A Triose fosfato isomerase (TIM, D-gliceraldeído-3-fosfato cetol-isomerase, EC 5.3.1.1) catalisa a quinta reação da via glicolítica (Figura 1.2), o processo de interconversão, provavelmente através de um intermediário enediol (ou enediolato), de gliceraldeído-3-fosfato (GAP) em dihidroxiacetona fosfato (DHAP), dois isômeros cetose-aldose (Yüksel & Gracy, 1991; Rieder & Rose, 1959; Rose, 1962). Estruturas tri-dimensionais de TIM de várias espécies foram descritas, incluindo TIM de procariontes (Noble *et al.*, 1993b; Delboni *et al.*, 1995; Alvarez *et al.*, 1998; Maes *et al.*, 1999; Walden *et al.*, 2001), organismos unicelulares (Lolis *et al.*, 1990; Wierenga *et al.*, 1991a; Velanker *et al.*, 1997; Maldonado *et al.*, 1998; Williams *et al.*, 1999) e eucariontes mais desenvolvidos (Banner *et al.*, 1975; Mande *et al.*, 1994). Com excessão das enzimas de *Pyrococ*-



Gliceraldeido-3-fosfato

Figura 1.2: Reação química catalisada pela TIM.

cus woesei e Methanothermus fervidus, que são homotetrâmeros (Kohlhoff et al., 1996), em todas as outras espécies estudadas até hoje, um dímero de subunidades iguais é a espécie fisiologicamente ativa. Por outro lado, já foram obtidos vários mutantes de TIM que exibem atividade catalítica na forma monomérica (Borchert et al., 1993; Mainfroid et al., 1996; Schliebs et al., 1996, 1997). Um grande número de estudos estruturais, de cinética enzimática e mecanismo de reação para TIM de várias espécies, empregando uma variedade de técnicas, resultou numa considerável quantidade de informações disponíveis sobre a enzima (Albery & Knowles, 1976; Alber et al., 1981; Nickbarg & Knowles, 1988; Knowles, 1991).

TIM é considerada uma enzima perfeita no que tange à formação de produto, que ocorre tão rapidamente quanto a enzima e substrato entram em contato na solução. Nas concentrações de substrato que ocorrem *in vivo*, a velocidade da catálise é essencialmente controlada por difusão (Knowles, 1991; Knowles & Albery, 1977), num processo susceptível a variações na viscosidade do meio.

O barril α/β (Figura 1.3) é uma estrutura supersecundária na qual oito tiras β , alternadas seqüencialmente com oito hélices α , estão arranjadas de tal modo que as oito tiras, em paralelo, compõem uma folha β que se curva num cilindro exteriormente circundado pelas oito hélices α . Esta estrutura foi primeiramente reconhecida na estrutura da TIM de músculo de coelho (Banner *et al.*, 1975) e por isso ficou conhecida também como *TIM barrel* (Branden & Tooze, 1991). Este tipo de enovelamento é encontrado num grande número de outras proteínas, na maioria, enzimas com variadas atividades (Lesk *et al.*, 1989). O sítio ativo de quase todas as enzimas que possuem um barril α/β está localizado, supreendentemente, na "boca" do barril (carboxi-terminal). Apesar de que poucas destas enzimas exibem homologia seqüencial significativa, pensa-se que elas tenham evoluído de um ancestral comum. Por outro lado, tem-se demonstrado que este tipo de estrutura supersecundária é particularmente estável e que a ele se chegou independentemente em várias ocasiões. As evidências para as duas linhas de raciocínio (evolução divergente e convergente, respectivamente) não são conclusivas e o assunto continua em discussão (Farber, 1993).

O loop do sítio ativo da triose fosfato isomerase, também chamado loop flexível ou *loop* 6 é formado por cerca de 10 resíduos e exibe um movimento de tampa-dobradiça entre duas conformações bem definidas (Figura 1.3). Estas conformações, conhecidas como aberta e fechada, foram descritas em grande detalhe (Joseph et al., 1990; Alber et al., 1981; Davenport et al., 1991; Lolis et al., 1990; Lolis & Petsko, 1990; Noble et al., 1991, 1993a; Verlinde et al., 1991; Wierenga et al., 1991a, b, 1992a). Normalmente, a conformação aberta foi encontrada em estruturas sem ligantes e a conformação fechada quando a estrutura foi resolvida na presença de análogos do substrato contendo grupos fosfato (tanto a partir de co-cristalização quanto soaking) (Joseph et al., 1990; Mande et al., 1994). Os resíduos que formam as dobradiças nas bordas N-terminal (166-168) and C-terminal (174-176) do loop são altamente conservados dentre TIMs de mais de 80 espécies (Xiang et al., 2001). O loop assemelha-se a uma tampa rígida apoiada nas bordas por resíduos que funcionam como dobradiças. Uma mudança de posição de até 8 A (resíduo central do *loop*) é observada entre as duas conformações. Ao mesmo tempo, a posição do principal resíduo catalítico, Glu165, sofre um deslocamento de cerca de 7 Å quando do fechamento do *loop*. Com este mecanismo, resíduos chave para a catálise são adeqüadamente reposicionados. Vários membros da família TIM barrel possuem um loop no sítio ativo cujas funções são similares ao loop 6 da TIM (Joseph et al., 1990; McMillan et al., 2000; Brzovic et al., 1993; Lundqvist & Schneider, 1991).

TIM é também importante em vista de seu potencial para o desenvolvimento de drogas contra alguns parasitas tropicais, uma vez que estes dependem da glicólise como fonte de energia. No caso da esquistossomose, causada pelo Schistosoma mansoni, foram feitas tentativas de se obter uma vacina a partir da TIM do parasita (Doenhoff, 1998). A estrutura de TIM de Trypanosoma brucei brucei, causador da doença do sono, foi profundamente investigada com o objetivo de desenvolver inibidores que pudessem ter valor terapêutico contra a infecção (Wierenga et al., 1991a, 1992b). Também no caso de Trypanosoma cruzi, a enzima do parasita tem sido alvo de estudos para desenho racional de drogas (Maldonado et al., 1998). Outro aspecto clinicamente importante, que também tem sido pesquisado, é que a deficiência de TIM no organismo humano constitui uma doença genética de herança autossômica dominante (locus 12p13) cujos sintomas incluem miopatia, falhas cardíacas e anemia (Daar et al., 1986; Watanabe et al., 1996).

1.2.1 TIM de músculo de coelho

A TIM de músculo de coelho (*Oryctolagus cuniculus*) tem 248 aminoácidos, com um peso molecular de 2×26 kDa. Os resíduos catalíticos são Lys13, His95 e Glu165 e o *loop* do sítio ativo inclui os resíduos de número 166 a 176. As TIMs de coelho e humana têm 98% de identidade seqüencial (ambas com 248 aminoácidos). As 4 mutações estão localizadas em áreas de superfície e nenhuma delas está situada na interface dimérica ou é implicada durante a catálise. O mecanismo de ação da TIM humana já foi descrito em detalhes (Mande *et al.*, 1994).

Apesar de ser uma das primeiras enzimas glicolíticas a ser purificada (Krietsch *et al.*, 1970; Scopes, 1977b,a), a estrutura de TIM de músculo de coelho não foi ainda descrita. Sua seqüência primária é conhecida (Corran & Ley, 1975) e suas propriedades cinéticas tem sido estudadas extensivamente (Lambeir *et al.*, 1987; Garza-Ramos *et al.*, 1992; Fernandez-Velasco *et al.*, 1995) Estudos biofísicos sobre a enzima também foram realizados através de espectroscopia no infravermelho (Castresana *et al.*, 1988) e microscopia



(a) *Loop* do sítio ativo (colorido) na conformação aberta.



(b) *Loop* do sítio ativo (colorido) na conformação fechada, na presença de ligante.

Figura 1.3: TIM: enovelamento e conformações usuais do *loop* do sítio ativo. O modelo utilizado foi o da TIM de *Trypanosoma brucei brucei*, entrada 6TIM do PDB (Wierenga *et al.*, 1991a), resolvida em complexo (obtido por *soaking*) com o ligante gliceraldeído-3-fosfato (G3P). Uma das subunidades (cadeia A), mostrada na figura (a), possui o *loop* na conformação aberta, não havendo ligação do substrato. Na outra subunidade (cadeia B), o *loop* está na conformação fechada e uma molécula de G3P é observada no sítio ativo.
eletrônica (Tanaka *et al.*, 1994). A cinética de re-enovelamento da TIM de coelho após denaturação induzida por hidrocloreto de guanidina (GndHCl) foi investigada por Zabori e colaboradores (Zabori *et al.*, 1980). A fim de relacionar estabilidade conformacional à estabilidade química e funcional da TIM, foram realizados estudos de enovelamento e desenovelamento de TIM de músculo de coelho usando pressão hidrostática e agentes caotrópicos (GdnHCl) como agentes perturbadores (Rietveld & Ferreira, 1996). Investigações do ponto de vista energético foram feitas em seqüência (Rietveld & Ferreira, 1998). Um dos resultados destes trabalhos foi a observação de que a enzima exibe heterogêneidade persistente, com a proposição da existência de um *ensemble* heterogêneo de dímeros de TIM em solução. Além da importância da estrutura de TIM de músculo de coelho por si só, os estudos cristalográficos foram iniciados também com o intuito de prover uma possível base estrutural para a heterogeneidade persistente observada em solução.

1.3 A cristalografia de Proteínas por difração de Raios X

A cristalografia de proteínas por difração de Raios X é a aplicação de técnicas de difração de Raios X a cristais de proteínas e é fundamental na obtenção de informação sobre estruturas tridimensionais de macromoléculas (Kleywegt & Jones, 2002). O comprimento médio da cadeia de aminoácidos das estruturas de proteína determinadas por Raios X de 1990 a 1994 era de cerca de 260 aminoácidos, com exemplos de até 1000 resíduos. Além de proteínas, um crescente número de outras estruturas têm sido analisadas: vírus, ácidos ribonucléicos (RNA), ácidos desoxirribonucléicos (DNA) e complexos de nucleoproteínas (McPherson, 1982).

Embora certa atenção seja dada ao aspecto de que a incorporação de uma

proteína dentro de um cristal possa modular sua estrutura ou dinâmica, há muitos exemplos mostrando que uma dada proteína (ou proteínas estreitamente relacionadas), cristalizada em diferentes células unitárias, têm estruturas cristalinas muito parecidas (McPherson, 1982). Também, a comparação de estruturas cristalográficas com as determinadas por NMR multidimensional, mostra grande semelhança entre a estrutura em solução e em forma cristalina.

Hoje, dez novas estruturas são determinadas a cada dia, em média, e o número de novas estruturas tem crescido exponencialmente nos últimos anos, com um total de vinte mil estruturas já resolvidas (*Protein Data Bank*, http://www.rcsb.org/pdb). A cristalografia de Raios X também tem ajudado a elucidar muitos aspectos da estabilidade, enovelamento e função das proteínas. O emprego desta técnica em conjunto com a mutagênese sítiodirigida tornou possível engendrar proteínas mais estáveis e alterar a atividade catalítica de uma enzima. O desenvolvimento racional de drogas é uma das aplicações mais promissoras da Biologia Estrutural e a cristalografia por difração de Raios X contribuiu substancialmente para o projeto de drogas, algumas das quais já disponíveis no mercado.

Capítulo 2

Metodologia

Nos estudos estruturais sobre a β -manosidase de *T. reesei* e a TIM de músculo de coelho, a técnica mais utilizada foi a cristalografia por difração de Raios X, embora experimentos de espalhamento a baixo ângulo (SAXS) e dicroísmo circular (CD) tenham sido realizados para obtenção de informação estrutural no caso da β -manosidase.

As técnicas que foram empregadas estão muito bem estabelecidas e há farto material disponível, em mídia impressa ou eletrônica (gratuitamente), que apresenta o assunto de forma didática e em diferentes níveis de profundidade e complexidade matemática. Uma vez que o trabalho de doutorado não foi centrado no desenvolvimento de novas metodologias, preferimos uma apresentação mais geral das técnicas empregadas, incluindo referências onde cada aspecto é tratado em maior profundidade. Nos estudos realizados por SAXS e CD, no caso da β -manosidase, e nos estudos cristalográficos sobre TIM, a metodologia empregada é descrita nos artigos que fazem parte desta tese.

O objetivo é descrever os métodos de uma forma geral. Nos próximos capítulos nos referiremos especificamente a quais variações foram adotadas no caso da β -manosidase e da TIM.

Uma descrição detalhada dos processos normalmente seguidos na análise da estrutura de um cristal de proteína e dos métodos empregados pode ser encontrada, por ex., em Blundell & Johnson, 1976; McPherson, 1982; Carter Jr. & Sweet, 1997a; Carter Jr. & Sweet, 1997b e Rossmann & Arnold, 2001. Além destes, material de ótima qualidade sempre pode ser encontrado nos *Proceedings* dos CCP4 *Study Weekends* (http://www.ccp4.dl.ac.uk; CCP4, 1994; Winn *et al.*, 1997). Vários cursos estão disponíveis na internet, por ex., em http://www.ccp4.dl.ac.uk, http://www-structure. llnl.gov/Xray ou através da própria página da *International Union of Crystallography*, http://www.iucr.org (teaching pamphlets).

É importante mencionar que existem diferenças fundamentais entre a cristalografia de proteínas e a cristalografia de pequenas moléculas, desde o processo de cristalização até os passos para resolução da estrutura, o que não será discutido aqui.

2.1 Cristalização

De forma simplificada, uma molécula de proteína é uma longa cadeia linear de aminoácidos (desde algumas dezenas até centenas ou milhares) que se enovela de forma bem determinada e adquire uma estrutura tri-dimensional cujos principais componentes são as chamadas hélices α e folhas β . Uma descrição detalhada sobre a composição química e propriedades estruturais das proteínas pode ser encontrada em Branden & Tooze, 1991 ou Voet & Voet, 1995.

Em sua grande maioria, as proteínas apresentam-se como substâncias amorfas, nas quais as moléculas estão desordenamente distribuídas no espaço. A cristalização é o primeiro passo para obtenção da estrutura atômica de uma proteína por difração de Raios X, fazendo com que as moléculas arranjem-se ordenadamente no espaço, numa estrutura periódica - um cristal. A proteína a ser cristalizada deve ser suficientemente pura, homogênea química e estruturalmente e em quantidade adequada para os experimentos de cristalização ($\sim mg$). Sua estabilidade e atividade devem ser asseguradas desde a purificação e durante todo o processo de cristalização, até a coleta de dados.

Os procedimentos para cristalização de proteínas (e outra macromoléculas biológicas como DNA e vírus) diferem radicalmente daqueles adotados no caso de pequenas móleculas, assim como os cristais, que possuem características típicas:

- dimensões macroscópicas: décimos de milímetros (ex.: $0, 1mm \times 0, 5mm \times 0, 7mm$), com volume da ordem de $1mm^3$, contendo $\sim 10^{20}$ moléculas, o que implica uso de microscópio óptico para sua análise;
- célula unitária: dezenas a centenas de Å;
- alto conteúdo de solvente: entre 20% e 80% do volume do cristal é constituído de solvente, principalmente água. O solvente encontra-se adsorvido nas superfícies moleculares ou nos espaços intermoleculares e é necessário para manter a integridade do cristal;
- interações fracas: as interações moleculares que mantêm o cristal coeso são fracas. Em geral, a maior parte da superfície molecular está exposta ao solvente, em comparação com a superfície de contato intermolecular;
- fragilidade: o alto conteúdo de solvente e as fracas interações responsáveis pela coesão cristalina fazem com que os cristais sejam extremamente frágeis e sensíveis às condições externas, exigindo manipulação cuidadosa em todas as etapas até a coleta de dados;
- formas cristalinas: a mesma proteína pode se cristalizar em mais de uma forma cristalina, o que não constitui um fato raro.

Basicamente, o processo de cristalização pode ser dividido nas etapas de nucleação, crescimento e interrupção do crescimento. A nucleação, onde aparecem os primeiros agregados ordenados de moléculas é induzida a partir de um estado de super-saturação (excesso de soluto) que, por sua vez, é provocado por alterações na solubilidade da proteína.

A proteína em solução, na presença de sais, solventes e aditivos, constitui um sistema de alta complexidade do ponto de vista físico-químico, para o qual não existe uma descrição teórica capaz de prever as condições de cristalização e possibilitar o controle dos mecanismos de nucleação e crescimento. Desta forma, a cristalização de uma proteína tem um forte componente experimental e empírico, onde os resultados muitas vezes podem não ser completamente reprodutíveis.

Os principais fatores que afetam a solubilidade da proteína em solução são a concentração e espécie dos sais presentes e precipitantes¹, pH e temperatura. No caso do uso de um sal como precipitante, a teoria eletrostática de Debye-Hückel (Blundell & Johnson, 1976) para íons pequenos pode ser usada para modelar a solubilidade da proteína (que pode ser considerada um grande íon polivalente já que possui estrutura bem definida, com uma superfície contendo grupos carregados e polares) de forma aproximada.

Numa abordagem racional, procura-se obter experimentalmente um diagrama de fase para a proteína, que determina a influência das diversas variáveis em sua solubilidade. Condições iniciais para cristalização são escolhidas partindo-se de informações previamente disponíveis sobre a proteína (PI, características bioquímicas, etc.). Os principais parâmetros são man-

¹ Chama-se *precipitante* um composto químico que promove, numa situação específica, a diminuição da solubilidade da proteína em solução aquosa, levando o sistema a um estado de super-saturação que, ocasionalmente, dá origem a precipitado ou a cristais. Alguns compostos que podem atuar como precipitantes são: sais, solventes orgânicos e polietilenoglicóis (PEG) de vários pesos moleculares.

tidos sob controle: pH, concentrações de precipitante, proteína e sais, além da temperatura que, em geral, é mantida constante (18 °C ou 4 °C). De acordo com os resultados obtidos, os parâmetros são variados em busca de melhores cristais ou sinais de cristalização, num processo cíclico. Geralmente, os parâmetros são variados dentro de determinados limites, que vão sendo restringidos conforme o refinamento das condições iniciais avança.

Dada a diversidade e quantidade de fatores envolvidos na cristalização de uma proteína, nem sempre se conseguem cristais de modo "racional". Uma alternativa bem sucedida em muitos casos, é submeter a proteína a condições as mais variadas possíveis na esperança de que alguma delas leve aos desejados cristais. Este método é conhecido como *sparse-matrix method* (Jancarik & Kim, 1991) e, atualmente, é facilmente implementado a partir de soluções comercialmente disponíveis. Uma vez que sejam observados sinais de cristalização em alguma das condições pré-determinadas incluídas no *kit* de cristalização utilizado, estas são refinadas de maneira similar ao exposto no parágrafo anterior.

Vários métodos foram desenvolvidos para fazer com que a solução de proteína atinja o estado de super-saturação. O aparato experimental utilizado na grande maioria dos casos está baseado no chamado Método da Gota Suspensa (*hanging-drop method*).

Neste método, que pertence à classe dos métodos por difusão de vapor, prepara-se uma solução aquosa contendo precipitante, sais e coadjuvantes, nas condições desejadas para cristalização. A Figura 2.1 ilustra o aparato utilizado. Um volume em torno de 0.5-1.0 ml é colocado no reservatório (poço) a ser selado por uma lamínula de vidro que conterá a gota na qual os cristais serão formados.

Geralmente, a gota, que tem um volume da ordem de 2-20 μ l, é preparada misturando a solução do poço e a solução de proteína em partes iguais. Após selar o poço, a gota fica dependurada na lamínula durante o restante do pro-



Figura 2.1: Montagem utilizada para cristalização de proteínas pelo hanging drop method. Uma gota de ~10 μ l, contendo solução do reservatório e solução de proteína, geralmente em partes iguas, fica dependurada numa lamínula de vidro selada contra o reservatório, que contém ~1 ml de solução precipitante.

cesso, cujo objetivo é elevar a concentração de precipitante na gota até que o ponto de supersaturação da proteína seja atingido. Uma vez que a concentração de precipitante do poço é maior que a concentração de precipitante na gota e volume de solução no poço pode ser considerado infinitamente maior que o volume da gota, esta tende a perder água, ou seja, troca de vapor ocorrerá até que o equilíbrio termodinâmico do sistema seja atingido e as concentrações de precipitante sejam iguais no poço e na gota. Idealmente, a concentração no poço é ligeiramente menor que a do ponto de precipitação da proteína, levando a uma situação de supersaturação, instável, que pode levar a precipitado ou cristais.

2.2 Criocristalografia

Cristais de proteína são bastante sensíveis a danos causados por radiação, um problema parcialmente controlado submetendo o cristal de proteína a temperaturas criogênicas (abaixo de 120 K) durante a coleta de dados. O conjunto de técnicas empregadas para este fim é designado *Criocristalografia* (Garman & Schneider, 1997), atualmente empregada na maioria absoluta dos casos por trazer importantes benefícios:

- grande redução nos danos causados pela radiação, o que permite coletar um conjunto de dados completo de um único cristal e possibilita o uso de doses mais altas de radiação;
- possível aumento no limite de resolução devido à redução nos movimentos atômicos (desordem dinâmica) com uma diminuição dos parâmetros de temperatura;
- dispensa o uso de capilares para montagem de cristais, o que diminui o background advindo do espalhamento do líquido em excesso no cristal e do vidro do capilar;
- permite trabalhar com cristais muito pequenos ou muito frágeis porque exige menos manipulação;
- possibilita longo tempo de estocagem e reuso de cristais;
- viabiliza o estudo de estados biológicos metaestáveis, o que é útil, por ex., na pesquisa de processos enzimáticos.

O mecanismo pelo qual os Raios-X afetam o cristal de proteína apresenta um aspecto dependente da dose e outro aspecto dependente do tempo.

No primeiro caso, a ionização provocada por fótons de Raios-X causam danos imediatos às moléculas, iniciando reações químicas que destroem a ordem cristalina. Tentativas de controle por meios químicos (uso de glutaraldeído e estireno, por ex.) perturbam as frágeis forças que mantém o cristal coeso. Por este ser um efeito inerente ao processo de difração, requer-se que o cristal tenha um tamanho mínimo para resistir à coleta de dados.

O aspecto dependente do tempo está relacionado à grande quantidade de água contida no cristal, normalmente entre 20% e 80%. O solvente presente no cristal, além de ser fonte de radicais livres que reagem com as macromoléculas causando danos à estrutura cristalina, facilita a própria difusão dos radicais, estendendo seu campo de ação. Além disso, o aquecimento da amostra em feixes intensos de Raios-X acelera tal difusão e, por consequência, a velocidade de degradação do cristal, podendo alterar, ainda, a solubilidade da macromolécula, com danos para o cristal.

A difração a baixa temperatura torna mais lenta ou quase inexistente a difusão de radicais dentro do cristal. Além disso, o solvente cristalino forma uma matriz rígida que impede que desarranjos locais espalhem-se por "efeito-dominó". Desta forma, os danos causados pela radiação tornam-se muito pequenos, trazendo os benefícios citados acima.

Em contrapartida, o uso de baixas temperaturas pode levar à formação de gelo. O gelo cristalino (cúbico ou hexagonal) é indesejável porque seu padrão de difração, constituído por anéis bem definidos, provocaria a perda dos dados advindos do cristal de macromolécula nas camadas de resolução coincidentes com os anéis do gelo. Entretanto, a água pura apresenta um diagrama de fase no qual é possível, a pressão ambiente, obter-se uma forma amorfa de gelo, chamada gelo vítreo.

O gelo cristalino pode ser evitado pelo uso do que se chama crioprotetor, um agente químico (glicerol, etilenoglicol e outros) capaz de alterar propriedades físico-químicas do solvente (água) de modo a reduzir a velocidade de formação do gelo cristalino (isto é, a velocidade de nucleação e crescimento dos cristais de gelo são alteradas). Desse modo, se houver um resfriamento do cristal rápido o suficiente, haverá formação apenas do gelo vítreo, cuja presença não inviabiliza o experimento.

Assim como no caso da cristalização, é necessário que condições ideiais para crioproteção sejam encontradas, basicamente através da escolha de crioprotetor e concentração adequados. Além da ausência de gelo, espera-se que a transferência do cristal da solução-mãe da gota de crescimento para uma nova solução contendo crioprotetor não altere as propriedades do cristal, por exemplo, aumentando sua mosaicidade ou mesmo provocando trincamentos, de modo que o padrão de difração não seja prejudicado.

Uma vez encontradas as condições de crioproteção, o cristal pode ser resfriado diretamente em nitrogênio líqüido (para posterior transferência para o local de coleta) ou diretamente no fluxo de nitrogênio gasoso (neste caso, no próprio local de coleta).

2.3 Derivados de átomos pesados

Chamamos de derivados os cristais onde átomos pesados foram incorporados e estão ligados às moléculas de proteína de forma ordenada, em contrapartida aos cristais chamados nativos, onde estão presentes apenas moléculas de proteína. Alguns aspectos da escolha e preparação dos derivados propriamente ditas são dados a seguir (informações na internet podem também ser encontradas em http://www.bmm.icnet.uk/had/heavyatom.html).

2.3.1 Critérios de escolha dos compostos utilizados

Os reagentes contendo átomos pesados podem ser classificados de acordo com as suas interações com a molécula de proteína, por ex., quanto à preferência por ligantes que contém carboxilatos ou outros grupos onde há oxigênio ou ligantes mais leves como enxofre, nitrogênio e haletos. Mais de uma classificação pode ser feita mas, em linhas gerais, a classificação abaixo, em 6 grupos, é adequada²:

- Metais da Classe A: incluem metais alcalinos (X^+) , terras raras (X^{2+}) , lantanídeos (X^{3+}) e o actnídeo UO_2^{2+} . Além destes, alguns outros que apresentam características mistas com a Classe B abaixo, são Tl^+ , Pb^{2+} e Cd. Ligam-se a grupos contendo F^- , OH^- , H_2O carboxilatos e fosfatos. Devem ser utilizados em pHs ácidos. Em pHs mais altos (> 7.5) há problemas de solubilidade.
- Metais da Classe B: os compostos mais utilizados pertencem a esta classe, que inclui Hg^{2+} , Ag^+ , Pt^{2+} , Au^{3+} , Pd^{2+} , Ir^{3+} , $Os^{4+} \in Cd^{2+}$. Preferem ligantes que contém N, S e haletos como, por ex., histidinas, cisteínas e metioninas (no caso da Pt). Não devem ser usados em pHs ácidos (< 6) e, em pHs mais altos, apresentam menos problemas do que os metais da Classe B.
- Aniônicos: neste caso, a interação dá-se através de sua carga negativa total, ligando-se a regiões básicas da proteína. A ligação será favorecida em pHs baixos, onde a proteína está mais carregada positivamente, facilitando a protonação de histidinas. O grupo inclui iodeto e complexos de metais aniônicos e estáveis: I^- , $(HgI_3)^-$, $(Pt(CN)_4)^{2-}$, $(IrCl_6)^{3-}$, $(Au(CN)_2)^-$.
- Catiônicos: em oposição aos compostos aniônicos, a interação predominante ocorre através de sua carga total positiva, ligando-se a regiões ácidas da proteína. Seu uso é favorecido por pHs mais altos, onde a proteína estará mais negativamente carregada, havendo também maior chance de deprotonação de histidinas. Alguns exemplos de com-

²http://eagle.mmid.med.ualberta.ca/tutorials/HA/

postos com estas características são $(Pt(NH_3)_4)^{2+}$, $(Ir(NH_3)_6)^{3+}$ e $(Hg(NH_3)_2)^{2+}$.

- Hidrofóbicos: em geral, os reagentes que contém os grupos acima ligam-se à superfície da molécula protéica, tanto devido ao tamanho, que impede a entrada no interior da proteína, quanto por serem carregados. Os principais exemplos desta classe são os gases nobres Xe e Kr, capazes de penetrar a molécula, interagindo em regiões hidrofóbicas.
- Selenometionina: atualmente, uma estratégia largamente empregada é a utilização, durante a expressão da proteína, de resíduos de metionina onde o elemento S tenha sido substituído por Se. O espalhamento anômalo advindo dos átomos de Se permite resolver o problema das fases pelos métodos MAD ou SAD (pág. 44).

Além das propriedades químicas dos reagentes, outros fatores devem ser levados em consideração durante o planejamento do experimento de derivatização. Os compostos contendo átomos pesados são bastante reativos, podendo romper as frágeis ligações que mantém coeso o cristal de proteína. Em geral, são utilizadas quantidades muito pequenas e tempos de exposição (banho do cristal em solução contendo átomo pesado) que variam de horas até dias, permitindo que os átomos pesados liguem-se à proteína. A escolha dos compostos a serem utilizados pode ser limitada pelas próprias condições de cristalização da proteína nativa, por exemplo, faixa de pH na qual é possível manter os cristais coesos e presença de precipitantes como sulfato de amônio ou outros sais que possam competir com os átomos pesados por sítios de ligação à proteína.

Outro aspecto de fundamental importância está na possibilidade de uso do sinal anômalo que pode resultar quando da inclusão de átomos pesados no cristal. O espalhamento anômalo, que é uma função do comprimento de onda utilizado durante a coleta, pode ser otimizado através da escolha de átomos pesados apropriados para o aparato experimental disponível. No caso de difratômetros convencionais, com comprimento de onda fixo, a escolha fica mais limitada. No caso de radiação síncrotron, existe a possibilidade de selecionar o comprimento de onda desejado, havendo uma maior liberdade de escolha. Entretanto, a intensidade da radiação síncrotron depende do comprimento de onda e, na prática, é utilizada uma faixa limitada de comprimentos de onda para os quais há intensidade suficiente para coleta de dados cristalográficos. Esta faixa depende do síncrotron onde a coleta é feita.

No caso da β -manosidase, conjuntos de dados foram coletados em dois síncrotrons muito diferentes, o LNLS e o ESRF. No primeiro, foram utilizados comprimentos de onda próximos a 1.5 Å, bastante perto do comprimento de onda da linha Cu K_{α} (1.54 Å) dos difratômetros convencionais, enquanto que no ESRF, o comprimento de onda onde a intensidade é maior está em torno de 1 Å (o que maximiza o sinal anômalo de átomos como Se e viabiliza o emprego de selenometioninas).

A Tabela 2.1 mostra o cálculo teórico (CROSSEC, CCP4, 1994; Winn et al., 1997) das componentes real (f') e imaginária (f'') do fator de espalhamento atômico (Apêndice A) dos átomos pesados presentes nos derivados de β -manosidase que foram coletados (Capítulo 3). No ESRF, apenas cristais "nativos" (contendo Cd) foram usados para coleta. A Tabela ilustra a variação do sinal anômalo para dois comprimentos de onda. Se foi incluído para comparação apenas.

2.3.2 Preparação de cristais derivados

Normalmente, o processo de derivatização é feito a partir de cristais nativos, crescidos da maneira usual, nas condições de cristalização previamente determinadas. Os cristais são transferidos para uma solução onde foi acrescentado o reagente que contém o átomo pesado. Uma vez que não tenham

	f'		f''	
	$\lambda = 1.0 \mathring{A}$	$\lambda = 1.5 \mathring{A}$	$\lambda = 1.0 \mathring{A}$	$\lambda=1.5 \mathring{A}$
Cd	-0.4468	-0.0825	2.2092	4.4480
Se	-3.4770	-0.9357	0.5185	1.0845
Au	-9.3039	-5.2288	9.5197	6.9941
Ag	-0.4981	-0.0696	2.0274	4.0920
Eu	-0.8273	-6.2647	6.5228	12.5103
Hg	-12.1619	-5.1081	10.0235	7.3676
\mathbf{Pb}	-9.0076	-4.9095	4.2712	8.1549
\mathbf{Pt}	-8.4785	-5.3970	9.0510	6.6373
Sm	-0.6970	-4.7558	6.1361	11.8122
U	-5.9026	-5.2950	6.8472	12.8715

Tabela 2.1: $f' \in f''$ para dois comprimentos de onda.

sido detectados visualmente danos ao cristal, o modo de avaliar se o objetivo principal, que é a ligação de átomos pesadas à molécula de proteína, foi atingido, é através da coleta de dados e posterior busca de sítios de ligação. O intervalo de tempo do banho (*soaking*) e a concentração de reagente ideais são determinados experimentalmente.

As condições do banho em solução contendo átomos pesados devem ser compatíveis com as condições de crioproteção pré-determinadas para os cristais nativos, um fator primordial para a coleta de dados.

2.4 Coleta de dados

A coleta de dados é uma parte importante na resolução de uma estrutura por cristalografia de proteínas. Dados de má qualidade podem compremeter todos os passos seguintes e equívocos durante a coleta podem trazer danos irreparáveis aos dados.

A estratégia de coleta depende do método escolhido para resolução do problema das fases (pág. 37). Quando não é possível resolver a estrutura por

Substituição Molecular, as exigências quanto a acuidade e precisão dos dados são muito maiores, uma vez que a obtenção de fases será necessariamente feita a partir de pequenas diferenças nos fatores de estrutura, usualmente através da coleta de dados de cristais derivados de átomos pesados.

Aspectos importantes da coleta de dados são comentados abaixo.

2.4.1 Fontes de Raios X

Uma descrição das fontes de Raios X, difratômetros convencionais comumente encontrados nos laboratórios, pode ser encontrada, por ex., em Cullity, 2001 e Azàroff, 1968. No caso da β -manosidase e TIM, radiação síncrotron foi utilizada. A radiação emitida pelos síncrotrons (máquina baseada na aceleração de elétrons em campo magnético e cujo produto é a radiação eletromagnética (Jackson, 1975) apresenta várias vantagens sobre as fontes comuns de Raios X no que diz respeito à cristalografia de proteínas, entre elas maior intensidade e espectro contínuo, o que possibilita selecionar o comprimento de onda desejado.

Dados cristalográficos foram coletados em dois síncrotrons de performance bastante diferentes, o Laboratório Nacional de Luz Síncrotron (LNLS), localizado em Campinas, SP e o *European Synchrotron Radiation Facility* (ESRF), situado em Grenoble, França.

O LNLS, a única fonte de luz síncrotron do hemisfério sul, é um síncrotron de média energia e que ainda não possui dispositivos de inserção (*wigglers* ou onduladores). O anel de armazenamento tem um raio médio de 28 m onde os elétrons circulam com uma energia de 1.37 GeV, com o comprimento de onda crítico dos fótons sendo 6.0 Å. A corrente em uso normal está em torno de 150 mA e o tempo de vida do feixe está em torno de 13h. O LNLS possui uma linha exclusiva para cristalografia de proteínas, cujo detector é uma placa de imagem MAR345 (recentemente substituída por um CCD) (Polikarpov *et al.*,

1998a,b) equipada com um soprador de nitrogênio da Oxford Cryosystems (Aparicio *et al.*, 1999).

O ESRF é um síncrotron de terceira geração que conta com onduladores e wigglers. O anel principal tem 844.4 m de circunferência (\sim 134 m de raio) ao passo em que o booster (síncrotron primário que acelera os elétrons de baixa energia gerados num acelerador linear) tem um percurso de 300 m; a energia dos elétrons que circulam no anel principal é de 6 GeV (Lindley, 1999). O ESRF conta com linhas para várias aplicações, em particular, cristalografia de macromoléculas. A linha utilizada nas medidas que realizamos foi a ID14-2, equipada com um detector CCD (Westbrook & Naday, 1997; Lindley, 1999).

Abaixo descrevemos em maior detalhe a linha CPR do LNLS, que ilustra o *setup* de uma linha de difração de Raios X dedicada à Cristalografia de Proteínas.

2.4.2 Aparato experimental e montagem dos cristais

No início do projeto, não havia equipamento disponível para coleta de dados a baixa temperatura e, durante o primeiro ano de doutorado, uma solução foi buscada em conjunto com o departamento de engenharia do LNLS³. Um soprador de nitrogênio começou a ser projetado com base em modelos disponíveis na literatura e alguns protótipos para o bico soprador chegaram a ser desenvolvidos. Entretanto, dada a complexidade do projeto e prioridades do departamento de engenharia, a idéia foi abandonada e a solução final foi a aquisição de um soprador disponível comercialmente, o OXFORD Cryosystems 600 Series Cryostream Cooler. Após a instalação do soprador, problemas de ventilação dentro do prédio do anel do LNLS, que não raro provocavam a formação de gelo nas amostras, foi resolvido com o projeto

³Agradecimentos especiais ao eng. Régis T. Neuenschwander.

e construção de uma cabana de acrílico englobando a amostra e detector⁴ (Aparicio *et al.*, 1999).

A linha de cristalografia de proteínas do LNLS é mostrada na Figura 2.2. A aquisição de imagens era feita através de uma placa de imagem MAR345 (recentemente substituída por um detector tipo CCD). A Figura 2.3 mostra os equipamentos em maior detalhe, incluindo o detector e o bico do soprador de nitrogênio.

Montagem dos cristais

Normalmente, vários cristais crescem em uma mesma gota (pág. 21). Após os procedimentos de crioproteção e derivatização (caso necessário), os cristais são selecionados para coleta de dados. Os principais critérios são os seguintes:

- o cristal deve ser um mono-cristal, ou seja, cristais geminados ou múltiplos ou com outros defeitos de crescimento são descartados de início;
- o cristal deve apresentar dimensões mínimas para que a coleta seja possível, entretanto, placas de dimensões tão pequenas quanto algumas dezenas de μm podem ser suficientes conforme o síncrotron utilizado;
- padrão de difração de qualidade satisfatória. Este fator em geral é avaliado após a coleta de algumas imagens. Aspectos como intensidade média do padrão de difração, resolução, anisotropia e mosaicidade são averiguados nesta etapa.

No caso de coleta a baixa temperatura, o cristal é montado em um *loop* de nylon, onde fica suspenso por tensão superficial na solução de onde foi re-

 $^{^4\}mathrm{Agradecimentos}$ ao amigo e mestre J. R. Brandão Neto e ao depto. de engenharia do LNLS.



(a) Foto mostrando a linha CPR e setor do anel de armazenamento de elétrons.



(b) Perspectiva da linha CPR, mostrando a mesa com o detector.

Figura 2.2: Fotos da linha CPR, no LNLS. Fotos por J. R. Brandão Neto.



(a) Foto da mesa móvel onde se apóia a placa de imagem MAR345 (utilizada para as coletas aqui descritas mas recentemente trocada por um detector CCD). Uma cabana de acrílico foi construída para evitar as correntes de ar internas do prédio do anel, capazes de provocar turbulência no fluxo de nitrogênio e ar seco, resultando em formação de gelo na amostra. (b) Uma vista ampliada da figura anterior, onde vê-se o goniômetro, as fendas de Raios X e o bico soprador de nitrogênio.

Figura 2.3: Aparato experimental. Fotos por J. R. Brandão Neto.

tirado (o próprio *loop* é usado para capturar o cristal diretamente da gota). Uma vez capturado da gota, o cristal é imediatamente resfriado, ou em nitrogênio líqüido ou no fluxo de nitrogênio gasoso diretamente na linha de cristalografia. Atualmente, não se realizam mais coletas a temperatura ambiente (a não ser por algum motivo particular ou no caso de não ser encontrado um crioprotetor adequado), caso em que o cristal seria montado num capilar de quartzo selado nas extremidades com cera de abelha ou carnaúba.

A Figura 2.4 ilustra o cristal sendo posicionado para coleta. Pode-se ver o goniômetro, a extremidade de onde os Raios X são emitidos e o bico do soprador de nitrogênio. O cristal é montado num *loop* de nylon colado na extremidade de uma haste metálica. Na Figura, a haste contendo um cristal, foi recém-removida de seu armazenamento em nitrogênio líqüido e está sendo posicionada para coleta através de um arco removível.



(a) A haste em cuja extremidade foi colado o *loop* de nylon onde encontra-se o cristal (não visível), recém-removida de um recipiente contendo nitrogênio líqüido, está na posição vertical. (b) A haste desliza pelo arco até atingir a posição de coleta, horizontal e fixa no goniômetro. Após, o arco pode ser removido.

Figura 2.4: Cristal sendo posicionado para coleta com auxílio de um arco removível, acoplado ao goniômetro apenas durante este procedimento. Fotos por J. R. Brandão Neto.

2.4.3 Redução de dados

O método da oscilação é usualmente empregado no caso de cristais de macromoléculas, de modo que o resultado experimental é um conjunto de imagens bi-dimensionais que, após processado, deve resultar num arquivo final contendo os fatores de estrutura relativos ao cristal em questão. A este processamento chamamos redução de dados. A estratégia de coleta é escolhida caso a caso levando em conta o objetivo da coleta, as qualidades do próprio cristal e os parâmetros obtidos após indexamento. De forma simplificada, a redução de dados divide-se da seguinte forma (Rossmann, M. G. and van Beek, C. G., 1999):

- Indexamento automático: nesta etapa, picos de difração são selecionados em uma ou mais imagens. A partir de uma análise de suas posições, o retículo de Bravais, as dimensões da cela unitária e a orientação do cristal são determinados;
- Pré-refinamento: um primeiro refinamento de parâmetros da câmera (por ex. distância cristal-detector e desvio da posição do detector em relação à normal ao feixe de Raiox X) e do cristal (orientação do cristal e mosaicidade) é feito;
- Integração das intensidades: utilizando os valores calculados com os parâmetros pré-refinados, as posições das reflexões são preditas e as intensidades correspondentes são integradas por *profile fitting*. O grau de parcialidade das reflexões é calculado, os *overlaps* e *overloads* são tratados adeqüadamente. Principalmente, os erros das intensidades de cada reflexão são estimados.
- Redução das reflexões simetricamente relacionadas: as reflexões são reduzidas a uma única unidade assimétrica no espaço recíproco, de acordo com o grupo de Laue previamente determinado.
- Escalonamento das imagens: fatores como tempo de exposição e diferentes regiões do cristal expostas levam a imagens que encontram-se em diferentes escalas de intensidade. Por isso, é necessário colocar todas as imagens numa única escala, determinando os fatores de escala, que

serão aplicados às intensidades das reflexões e aos erros associados. Em seguida, é feita a média das reflexões com os mesmos indíces de Miller, já reduzidos. A detecção e rejeição de *outliers* é uma parte importante desta etapa.

Análise estatística: o processo de integração e promediação das intensidades das reflexões envolve uma análise cuidadosa das estatísticas resultantes, que podem indicar problemas com o cristal (por ex., desordem e anisotropia) e com o aparato experimental (por ex., oscilações na intensidade do feixe, problemas com o detector), erros de estratégia durante a coleta, problemas diversos (como existência de gelo ou cristal desalinhado), além da indição prematura da presença de sinal anômalo. Alguns itens importantes para avaliação são o comportamento do número de rejeições, dos fatores de escala, R_{merge} (Apêndice A) e $\overline{I/\sigma(I)}$ em função do número das imagens. As duas últimas estatísticas devem também ser avaliadas contra as faixas de resolução.

2.5 O problema das fases

Uma vez que o padrão de difração (fatores de estrutura) é uma função da densidade eletrônica do cristal (matematicamente relacionados através de uma Transformada de Fourier), deveria-se esperar que a partir deste poderse-ia obter imediatamente a densidade eletrônica e, em seguida, o modelo atômico da molécula cristalizada (Taylor, 1981).

Infelizmente, entretanto, não é possível medir diretamente as fases dos fatores de estrutura, ou seja, não é possível obter informação sobre as fases relativas com as quais os átomos espalham no interior do cristal, ao que se chama o Problema das Fases, um problema central em cristalografia. As razões pelas quais não é possível medir as fases diretamente são as seguintes

(Taylor, 1981; Warren, 1969):

- resolução temporal: medidas de fase implicam medir frações de um período. Para um comprimento de onda de 1.5 Å, seria necessário medir intervalos de tempo da ordem de 10^{-19} s, uma impossibilidade técnica na atualidade;
- resolução espacial: a distância entre o ponto em que se mede a onda de referência e o sítio de difração deve ser medida com precisão de décimos de Å, o que é inviável na prática;
- incoerência dos Raios X: mesmo a radiação síncrotron não é coerente, o que inviabiliza de antemão qualquer tentativa de medição de fases.

Embora não seja possível medir diretamente as fases, há mais de um modo de solucionar o problema. Os métodos atualmente mais utilizados para este fim estão descritos abaixo.

2.5.1 Métodos diretos

Os Métodos Diretos consistem no emprego de relações matemáticas e estatísticas entre os módulos dos fatores de estrutura na determinação direta das fases. A fórmula da tangente é um exemplo de relação matemática que teve grande impacto nos métodos diretos e até hoje é muito empregada (Gilmore, 2000; Karle & Hauptman, 1956). Ao final, não apenas o mapa de densidade eletrônica é obtido mas também o modelo atômico. Exemplos de programas atualmente utilizados são Shake&Bake (Weeks & Miller, 1999) e Shelx (Sheldrick, 1998). Estes métodos são largamente empregados no caso de moléculas pequenas mas dificilmente são empregados no caso de estruturas de macromoléculas uma vez que os cristais difratam até um limite de resolução insuficiente para sua aplicação. Entretanto, podem ser úteis para determinar a subestrutura de átomos pesados incorporados ao cristal de proteína nativa (pág. 46).

2.5.2 Substituição Molecular

As fases podem ser estimadas a partir do modelo atômico de uma proteína previamente resolvida (o "modelo de busca" ou *search model*) e que seja parecida com a proteína de interesse. Esta abordagem vem ganhando importância à medida em que crescem os bancos de dados de estruturas resolvidas, o que aumenta as chances de que se encontre um modelo de busca adequado.

Para obter estimativas das fases da proteína cristalizada, é necessário posicionar corretamente o modelo de busca na cela unitária. A Substituição Molecular consiste, basicamente, no processo de orientar (através de um operador de rotação) e posicionar (através de um operador de translação) o modelo de busca. Este método é preferencialmente escolhido pela facilidade de uso, rapidez e funcionalidade. Entretanto, há dois fatores que limitam sua aplicação:

- fração da proteína de interesse representada pelo modelo de busca: o modelo de busca deve corresponder a uma fração razoavelmente grande da molécula alvo. Quanto maior a porção da proteína de interesse que pode ser substituída pelo modelo de busca, ou seja, quanto mais ele é completo, maior a chance de sucesso;
- 2. nível de semelhança entre as duas estruturas: a estrutura terciária das proteínas está correlacionada à sua seqüência primária. Isto fornece um critério para escolha do modelo de busca. Assim, quanto maior a identidade seqüencial, maior a semelhança estrutural tri-dimensional e maiores as chances de que o método funcione.

Um terceiro ponto a ser considerado é o de que os algoritmos mais utilizados para Substituição Molecular, baseados em métodos de Patterson (com o cálculo de vetores intra- e inter-moleculares), podem não funcionar bem no caso de proteínas de forma muito alongada.

Para se ter uma idéia, requerimentos mínimos para que o método funcione são um modelo de busca que compreende em torno de 70% da molécula cristalizada, com uma identidade seqüencial de 30%. Às vezes, é possível concatenar domínios de mais de uma proteína para formar um modelo de busca que corresponda a uma fração maior da molécula alvo.

A Substituição Molecular é preferencialmente utilizada quando:

- uma proteína homóloga já foi resolvida;
- a mesma proteína está disponível em outro grupo de espaço;
- o objetivo é determinar mutantes ou complexos da proteína nativa. Este último, um procedimento muito utilizado em desenho racional de drogas;
- há disponibilidade de modelos obtidos por outras técnicas, por ex., NMR, microscopia eletrônica ou modelos teóricos.

No caso em que haja um modelo de busca adequado, as operaçõs de rotação e translação serão encontradas com maior facilidade e as fases calculadas estarão mais próximas das fases verdadeiras. Em caso de baixa fracionalidade ou identidade, o mapa obtido pode ser não-interpretável mesmo que as soluções sejam encontradas.

Um ponto importante a ser lembrado é que, além fornecer um mapa de densidade eletrônica para a proteína de interesse, o método resulta também num primeiro modelo a ser posteriormente refinado.

2.5.3 Átomos pesados

A Substituição Molecular não é um método geral para resolução de estruturas de proteínas, estando limitada a casos particulares onde uma estrutura homóloga já tenha sido previamente determinada. Esta seção engloba vários métodos que têm em comum a utilização de derivados de átomos pesados (a preparação de derivados foi descrita na seção 2.3, pág. 25) e que são de aplicação geral.

As técnicas utilizadas estão muito bem descritas em várias fontes e não repetiremos os detalhes neste texto (referências são dadas no início deste Capítulo). Ao invés, nossa intenção é dar subsídios de forma geral que permitam o entendimento dos próximos capítulos. Com isso em mente, iniciamos introduzindo idéias importantes sobre as técnicas.

Sinal isomorfo

Um "derivado" é um cristal de proteína onde átomos pesados foram introduzidos. O termo "nativo" é usado para designar o cristal de proteína em sua forma original, sem a incorporação de átomos pesados. Dizemos que dois cristais são *isomorfos* quando possuem essencialmente a mesma estrutura, ou seja, o arranjo das moléculas de proteína no cristal derivado é essencialmente o mesmo que o do cristal nativo. Na prática, sempre existe um mínimo grau de não-isomorfismo, com pequenas mudanças nas posições das moléculas de proteína dentro do cristal e nos parâmetros de rede. Informação sobre fases pode ser obtida a partir das diferenças entre os fatores de estrutura do cristal nativo e derivado, diferenças essas que são chamadas "diferenças isomorfas", a cujo conjunto chamamos de "sinal isomorfo" entre nativo e derivado.

A mudança média nas intensidades das reflexões devido à incorporação dos átomos pesados ao cristal é estimada através da fórmula

$$\frac{\langle \Delta I^2 \rangle^{1/2}}{\langle I \rangle} = \left(\frac{2N_H}{N_P}\right)^{1/2} \left(\frac{f_H}{f_P}\right), \qquad (2.1)$$

onde N_H e N_P são o número de átomos pesados e de átomos da molécula de

proteína, respectivamente, com f_H e f_P sendo os correspondentes fatores de espalhamento atômico (medidos em elétrons).

Como exemplo, consideremos a β -manosidase, para a qual $N_P \cong 7400$ átomos não-hidrogênio. O f_P médio para proteínas é $f_P = 6.7e^-$. Para um único sítio de U, $N_H = 1$ com $f_H = 92e^-$ Substituindo os valores na fórmula acima, temos que a estimativa da mudança média na intensidade das reflexões é de 22%.

Sinal anômalo

O segundo modo importante a partir do qual a introdução de átomos pesados pode levar à solução do Problema das Fases é através de espalhamento anômalo. O fator de forma atômica é uma função do comprimento de onda usado durante a coleta, de modo que as diferenças entre F_{PH}^+ e F_{PH}^- podem ser maximizadas por uma escolha apropriada de comprimento de onda. A partir destas diferenças, chamadas "diferenças de Bijvoet" (Apêndice A) ou sinal anômalo, é possível obter-se informação de fase. É importante observar que estas diferenças ocorrem para fatores de estrutura de um mesmo cristal, ao contrário das diferenças isomorfas, decorrentes da comparação de fatores de estrutura de cristais diferentes.

A razão média entre as diferenças de Bijvoet e a amplitude total devida à proteína é uma função do comprimento de onda usado durante a coleta e pode ser estimada através da fórmula

$$\frac{\langle |F_{PH}^{+} - F_{PH}^{-}| \rangle}{\langle |F_{P}| \rangle} = \left(\frac{2N_{H}}{N_{P}}\right)^{1/2} \left(\frac{f''}{Z_{eff}}\right), \qquad (2.2)$$

onde F_P é o fator de estrutura da proteína nativa (sem adição de átomos pesados), N_H e N_P são o número de átomos pesados e de átomos da molécula de proteína, respectivamente, f'' é a componente imaginária do fator de espalhamento atômico (resultado do espalhamento anômalo ressonante; Apêndice A) e Z_{eff} é o número atômico efetivo médio para os átomos de proteína. $F_{PH}^+ - F_{PH}^-$ são as diferenças de Bijvoet (Apêndice A).

Adotemos o exemplo anterior, um cristal de β -manosidase com um átomo de U por molécula de proteína. Neste caso, $N_P \cong 7400$, $N_H = 1$ e $Z_{eff} =$ $6.7e^-$. O sinal anômalo depende do comprimento de onda. Utilizando o valor f'' = 12.8715 calculado para $\lambda = 1.5$ Å (Tabela 2.1), a razão média estimada das diferenças de Bijvoet para a amplitude total da proteína é de 3 %, muito menor que as diferenças isomorfas. É importante observar que a magnitude do sinal anômalo independe da resolução, ao contrário das intensidades, que descrescem exponencialmente com a resolução. Isso significa que o sinal anômalo passa a ser um componente importante para as reflexões de mais alta resolução.

Os métodos mais utilizados

Os sinais isomorfo e anômalo podem ser combinados de diferentes maneiras, dando origem aos métodos abaixo:

MIR (Multiple Isomorphous Replacement)

Este é um método tradicional, cuja aplicação à cristalografia de proteínas iniciou-se na década de 1950 (Green *et al.*, 1954). O método utiliza o sinal isomorfo obtido a partir de um cristal nativo e, no mínimo, dois derivados. Como o nome diz, uma condição necessária é que o cristal derivado seja isomorfo ao cristal nativo, ou seja, que o átomo pesado seja incorporado ao cristal de proteína sem perturbar o arranjo molecular;

MIRAS (Multiple Isomorphous Replacement with Anomalous Scattering)

Variação do método MIR onde, além dos derivados, emprega-se o sinal anômalo advindo dos átomos pesados em um ou mais derivados;

SIR (Single Isomorphous Replacement)

Variação do método MIR onde apenas um derivado e o nativo são empregados. A falta de um segundo derivado leva a uma distribuição bimodal de fases, onde cada fase tem dois valores possíveis. Esta ambigüidade é resolvida através de técnicas de modificação de densidade eletrônica (pág. 47);

SIRAS (Single Isomorphous Replacement with Anomalous Scattering)

Variação do método MIR onde apenas o nativo e um derivado são necessários. Além do sinal isomorfo, o sinal anômalo do derivado permite resolver a ambigüidade de fase presente no método SIR;

MAD (Multi-wavelength Anomalous Dispersion)

Esta técnica consiste no uso de um único cristal utilizado repetidamente para a coleta de conjuntos de dados a diferentes comprimentos de onda, de forma que não há o problema de não-isomorfismo comumente encontrado quando se utiliza MIR. As fases são obtidas a partir das pequenas diferenças introduzidas nos fatores de estrutura pela mudança de comprimento de onda durante a coleta. Utiliza-se tanto as diferenças de Bijvoet, cujo sinal é proporcional à componente imaginária f" do fator de espalhamento atômico (eq. 2.2), quanto as "diferenças dispersivas" (calculadas a partir dos fatores resultantes da coleta a comprimentos de onda diferentes e proporcionais à componente real do fator de espalhamento atômico). A técnica ganhou impulso com o advento dos *Projetos Genoma*, sendo empregada principalmente na determinação de estruturas de proteínas através da incorporação de selenometioninas (seção 2.3.1, pág. 27). O método não pode ser utilizado com difratômetros convencionais, com comprimento de onda fixo.

SAD (Single-wavelength Anomalous Dispersion)

Neste método, apenas um cristal é utilizado e apenas um conjunto de dados com dispersão anômala é necessário, ou seja, as fases são obtidas apenas a partir das diferenças de Bijvoet (Apêndice A; Dauter *et al.*, 2002). De modo análogo ao método SIR, existe uma ambigüidade de fase que é resolvida posteriormente durante a modificação de densidade eletrônica. Este método tem sido cada vez mais empregado por sua simplicidade e eficácia. É mais simples que o método MAD e apresenta as mesmas vantagens, com o adicional de que pode ser utilizado com difratômetros convencionais pois a coleta de sinal anômalo a um único comprimento de onda é suficiente. No caso do uso de selenometioninas, é necessário coleta em síncrotron porque não há fonte de Raios X comercialmente disponível cujo feixe tenha intensidade suficiente no comprimento de onda adequado (~ 1 Å).

A escolha do método geralmente é determinada pelas limitações impostas pelo cristal e aparato experimental disponível. Como exemplo, no caso de β manosidase, o método SAD foi utilizado, uma vez que não havia isomorfismo entre os cristais (Capítulo 3). Na direção inversa, há casos que permitem que mais de uma abordagem seja utilizada e o método de escolha será o que levar ao melhor mapa de densidade eletrônica. Por ex., embora conjuntos de MAD possam estar disponíveis, muitas vezes um mapa obtido utilizando um único comprimento de onda - SAD, neste caso - apresenta melhor qualidade final. Ainda, vários derivados isomorfos podem estar disponíveis mas nem todos serem utilizados. Qualquer que seja o método adotado, medidas muito precisas e acuradas são essenciais para permitir a detecção dos pequenos sinais isomorfo ou anômalo esperados (cuja magnitude é estimada através das eqs. 2.1 e 2.2), o que exige aparato experimental e procedimento de coleta adequados, além de cristais de qualidade suficientemente boa. Em geral, um grande número de imagens é coletado em cada conjunto, aumentando a multiplicidade das medidas, diminuindo o erro associado às reflexões e melhorando a coerência interna dos dados ($R_{merge/sym}$, Apêndice A).

O processo de determinação de fases propriamente dito ocorre a partir da determinação das posições dos átomos pesados, por ex., através de métodos baseados no uso mapas de Patterson, como implementado no CNS (Brünger *et al.*, 1998), RSPS (CCP4, 1994; Winn *et al.*, 1997) e SOLVE (Terwilliger & Berendzen, 1999) ou através de Métodos Diretos, como nos programas Shake&Bake (Weeks & Miller, 1999), ShelxD (Sheldrick, 1998; Sheldrick & Schneider, 2001) e RANTAN (CCP4, 1994; Winn *et al.*, 1997), este último também utilizado no pacote Auto-SHARP (de La Fortelle & Bricogne, 1997). Uma vez que os primeiros sítios tenham sido localizados, a tarefa de encontrar outros sítios fica facilitada porque fases, mesmo que inacuradas, podem ser calculadas e utilizadas na busca de novos sítios, no mesmo conjunto ou em outros conjuntos isomorfos.

Além do erro experimental presente nas intensidades das reflexões, erros nos parâmetros dos sítios de átomos pesados (por ex., coordenadas, ocupância, e fator de temperatura) e falta de isomorfismo entre os cristais exigem um tratamento estatístico bastante complexo. O processo de faseamento, que inclui o refinamento de parâmetros globais (por ex., falta de isomorfismo, fatores de escala entre os vários conjuntos de dados utilizados e fatores de temperatura globais) e específicos para cada sítio de ligação, além do cálculo das fases (e determinação da mão correta do conjunto-solução de coordenadas de átomos pesados), é controlado através de várias estatísticas que indicam a qualidade dos derivados e sítios de ligação, bem como a qualidade das fases obtidas. Infelizmente, apesar de haver estes indicadores, a única forma segura de saber se o processo teve sucesso é através da avaliação dos mapas de densidade eletrônica resultantes.

Modificação de densidade eletrônica

O objetivo do faseamento é levar a um mapa de densidade eletrônica que seja interpretável, ou seja, que exiba elementos de estrutura secundária a partir dos quais se possa iniciar o traçado da cadeia de aminoácidos que constitui a molécula de proteína. As fases obtidas pelos métodos expostos nesta seção geralmente contém erros que, em muitos casos, inviabilizam a interpretação direta do mapa de densidade resultante, uma vez que elementos de estrutura secundária podem não estar aparentes. Dizer se um mapa é interpretável ou não a partir de uma avaliação gráfica envolve um bom nível de empirismo. Além disso, pequenas diferenças qualitativas podem ser responsáveis por fazer com que um mapa seja interpretável ou não.

Qualquer que seja o caso, antes de iniciar a construção do modelo, o melhor mapa possível deve ser obtido (evitando trabalho desnecessário posteriormente), o que é feito através das técnicas de modificação de densidade eletrônica, também aplicadas para resolver ambigüidade de fase⁵ (métodos SAD e SIR). Durante a modificação de densidade eletrônica, um número maior de vínculos é imposto sobre as fases, eliminando possibilidades errôneas e levando a melhores mapas ao final do processo. O conjunto de técnicas comumente empregadas inclui:

• Solvent Flatenning: O cristal de proteína contém solvente (água e demais aditivos de cristalização) que preenche as lacunas entre as

⁵ a ambigüidade de fase também pode ser resolvida através de métodos matemáticos como o implementado no programa OASIS, do CCP4, por exemplo.

moléculas de proteína de forma desordenada de cela unitária para cela unitária, de modo que sua contribuição para os espalhamento será uma constante (espalhamento difuso). Esta técnica faz uso desta característica e consiste em "achatar" a região de solvente (Wang, 1985; Leslie, 1988);

- Histogram Matching: enquanto Solvent Flatenning atua apenas nas regiões de solvente, esta técnica atua apenas na região onde há proteína. As proteínas têm proporções similares dos átomos que as compõem e que estão distribuídos na cela cristalina de acordo com vínculos conhecidos de espaçamento atômico. Isto faz com que seja possível predizer um histograma para os valores de densidade eletrônica nas regiões de proteína. A técnica modifica o mapa de densidade eletrônica de modo a torná-lo consistente com o histograma predito (Zhang & Main, 1990). Embora imponha vínculos mais fracos às fases, em relação ao Solvent Flatenning, o volume envolvido é, em geral, maior que o volume de solvente e a técnica ganha importância no processo de extensão das fases até mais alta resolução;
- Skeletonisation: as proteínas consistem em cadeias peptídicas lineares, de modo que a conectividade do mapa de densidade eletrônica deve ser linear, justamente a propriedade que permite a identificação de elementos de estrutura secundária (hélices α e folhas β). Esta técnica consiste num processo iterativo de construção de esqueleto nos mapas, de modo a melhorar sua conectividade (Baker *et al.*, 1994; Swanson, 1994);
- Sayre's equation: as técnicas anteriores podem ser aplicadas a qualquer caso. No entanto, a resoluções > 2.0 Å, o mapa de densidade eletrônica começa a apresentar mais detalhes, com os átomos bem resolvidos no limite de alta resolução. Características atômicas bem resolvidas é o que

se chama propriedade de atomicidade, o que constitui um vínculo bastante empregado nos Métodos Diretos aplicados a pequenas moléculas, através da equação de Sayre (Sayre, 1974; Zhang & Main, 1990). Caso haja resolução suficiente, mesmo no caso de proteínas, a equação de Sayre pode ser empregada de modo eficaz na melhoria das fases;

- Averaging: esta é uma técnica originária do processamento de imagens em microscopia eletrônica e diminui a razão sinal/ruído por um fator N^{1/2} quando N imagens independentes são promediadas. Em cristalografia de proteínas, sua aplicação consiste em promediar, no espaço real, a densidade eletrônica de cópias equivalentes de uma mesma molécula (Bricogne, 1974; Schuller, 1996). A técnica é muito poderosa e pode trazer melhoria significativa ao mapa de densidade eletrônica, porém, seu uso é basicamente restrito a dois casos:
 - Simetria não-cristalográfica (NCS non-crystallographic symmetry): a simetria cristalográfica aplica-se globalmente a todo o conteúdo do cristal. Não é raro, entretanto, casos em que a unidade assimétrica contém mais de uma unidade da mesma molécula (o que pode ocorrer, por ex., quando dímeros são cristalizados). A operação que relaciona tais objetos não é global para todas as móleculas, ou seja, tem aplicação local. A existência de móleculas relacionadas por NCS fornece uma fonte de informação independente, permitindo a promediação das porções de densidade eletrônica relacionadas por tais operadores;
 - Diferentes formas cristalinas: não é raro que a mesma molécula seja cristalizada - e mapas de densidade eletrônica sejam obtidos

 em mais de uma forma cristalina, resultando em cópias equivalentes de uma mesma molécula. Cada forma cristalina representa informação independente que pode ser usada para promediação. É

possível, inclusive, utilizar ao mesmo tempo a informação advinda de NCS e múltiplas formas cristalinas.

Todas as técnicas acima necessitam clara distinção entre as regiões de proteína e solvente. Normalmente, é possível distinguir regiões de solvente e construir uma máscara delimitando as fronteiras proteína-solvente a partir do mapa inicial. Esse processo pode ser feito automaticamente ou iterativamente em casos mais complexos (por ex., em casos de oligômeros). As técnicas acima são geralmente empregadas todas ao mesmo tempo e de forma automatizada e estão implementadas em dois pacotes principais: DM/DMMULTI (CCP4, 1994; Winn *et al.*, 1997) e RESOLVE (Terwilliger, 2000).

E interessante notar que cada uma destas técnicas é baseada em um aspecto identificável graficamente nos mapas de densidade eletrônica, quantificando, de certo modo, o que o pesquisador observa visualmente de forma apenas qualitativa.

Construção de modelo atômico

A modificação de densidade eletrônica deve levar a um mapa de densidade interpretável. O passo seguinte é interpretar o mapa, construindo um modelo atômico inicial para a molécula. Mesmo nos casos em que apenas algumas porções do mapa são interpretáveis, a construção de parte do modelo pode levar à melhorias em outras regiões do mapa, permitindo que também elas sejam interpretadas.

Atualmente, a construção do modelo é feita através das formas abaixo:

• Automaticamente: esta abordagem é bastante recente e rapidamente tornou-se a opção preferida, uma vez que a construção manual é um processo laborioso e demorado. Infelizmente, entretanto, ela está restrita a dois casos:
- Alta resolução (> 2 Å): o pacote Arp/Warp (Perrakis *et al.*, 1999) é um conjunto de módulos cuja parte principal é um algoritmo para a construção automática de modelos, propriamente dita. O algoritmo implementado usa a propriedade de atomicidade, exigindo que os mapas exibam características atômicas bem resolvidas, o que só é possível a resolução > 2 Å, com uma razão (dados)/(parâmetros a refinar) igual a 2, no mínimo.
- Média resolução (2.0 3.0 Å): para dados nesta faixa de resolução, foram desenvolvidos algoritmos capazes de encontrar, automaticamente, fragmentos de estrutura secundária no mapa de densidade eletrônica, principalmente hélices α e tiras β. Dependendo do caso, isso pode significar um ponto de partida para a construção manual. Nos melhores casos, o processo pode resultar em grandes blocos de fragmentos que, após, deverão ser concatenados manualmente. Os principais programas disponíveis são RESOLVE (Terwilliger, 2000) e FFFEAR (CCP4, 1994; Winn *et al.*, 1997). RESOLVE, além do algoritmo de busca de fragmentos, implementa também excelente algoritmo de modificação de densidade eletrônica.
- Manualmente: há ferramentas disponíveis que permitem a construção manual de modelos qualquer que seja a resolução do mapa de densidade eletrônica. O processo, que envolve considerável empirismo por parte do pesquisador, consiste, basicamente, nos seguintes passos: 1) um esqueleto que representa a continuidade do mapa é construído automaticamente; 2) o esqueleto é editado iterativamente de modo a refletir o traçado correto da cadeia de aminoácidos; 3) o esqueleto é, finalmente transformado numa cadeia de carbonos onde as distânciaspadrão são impostas; 4) após o traçado da cadeia principal, as cadeias

laterais são adicionadas. Os principais aplicativos gráficos para esta finalidade são O (Jones *et al.*, 1991), XtalView (McRee, 1999) e Quanta (http://www.msi.com), este último comercialmente disponível (MSI -Molecular Simulations Inc.).

É importante observar que, quando um mapa não parece ser interpretável, dificilmente os algoritmos de construção de modelo ou busca de fragmentos terão sucesso e a obtenção de melhores mapas é necessária, mesmo que através da realização de novos experimentos.

2.6 Refinamento do modelo

A resolução do Problema das Fases através de Métodos Diretos ou Substituição Molecular leva, concomitantemente, à obtenção de um modelo inicial para a molécula de interesse. Já os métodos baseados em átomos pesados não resultam diretamente num modelo e sua construção deve ser feita posteriormente.

Quando dados de alta resolução (~ 1 Å) estão disponíveis, Métodos Diretos podem empregados e, geralmente, um modelo inicial muito próximo da realidade é obtido automaticamente. No caso da Substituição Molecular e Átomos Pesados, modelos incompletos ou bastante imperfeitos podem ser o resultado. Além disso, a qualidade do modelo inicial depende diretamente da qualidade do mapa de densidade eletrônica inicial e, por conseqüência, diretamente dos passos anteriores, principalmente a qualidade dos dados coletados. Qualquer que seja o caso, o modelo inicial deve ser modificado, em maior ou menor grau, até refletir de forma satisfatória os dados experimentais, o que é feito através de um processo cíclico de reconstrução/refinamento.

A reconstrução do modelo, que corrige erros locais, é feita iterativamente em aplicativos gráficos (os mesmos utilizados para construção manual de modelos, citados acima) e o resultado é um novo modelo a ser refinado globalmente contra o conjunto de dados. Atualmente, os programas mais utilizados para refinamento cristalográfico de macromoléculas são REFMAC (CCP4, 1994; Winn *et al.*, 1997) e CNS (Brünger *et al.*, 1998).

O refinamento é uma etapa crítica na determinação de uma estrutura cristalográfica e deve ser conduzido de forma criteriosa e sensata, sob pena de distorcer os resultados e levar a uma estrutura não realista, não fiel a realidade. Isto porque envolve um alto grau de subjetividade, uma vez que o mapa de densidade eletrônica deve ser *interpretado* pelo pesquisador, que também deve tomar decisões sobre a estratégia de refinamento mais adequada para cada caso. Mesmo com toda atenção e cuidado que deve permear o processo, ainda assim podem ocorrer equívocos.

2.6.1 Free R_{factor} ou R_{free}

O principal problema no caso de macromoléculas tem origem no próprio tipo de cristal que se obtém. Conforme mencionado na seção 2.1, na pág. 19, os cristais têm alto conteúdo de solvente, com fracas interações moleculares. Isto faz com que haja maior mobilidade (refletida em altos fatores de temperatura), diminuindo a relação sinal/ruído e aumentando o erro associado às reflexões. Um segundo efeito é uma redução da intensidade dos dados de alta resolução, o que leva a uma redução do limite de resolução (que, em última instância, é a quantidade de informação disponível para refinamento). Em conseqüência, a razão entre número de observações e o número de parâmetros a ajustar torna-se muito baixa, impedindo o uso de minimização por Mínimos Quadrados e possibilitando facilmente que distorções sejam introduzidas por um *over-fitting* do modelo. A baixa razão dados/parâmetros é remediada através da introdução de vínculos estereoquímicos e utilização de simetria não-cristalográfica (pág. 49) em alguns casos⁶.

⁶ NCS introduz correlação entre os fatores de estrutura, de modo que reflexões selecionadas para o cáculo de R_{free} podem estar fortemente correlacionadas às reflexões do

O refinamento de macromoléculas foi alvo de extensas discussões durante a década de 1990 e um grande número de trabalhos sobre o assunto foi publicado chamando atenção para os principais aspectos envolvidos. Vários dos assuntos mencionados abaixo foram tratados em detalhe, por exemplo, nos artigos Brünger, 1992, 1997; Dodson *et al.*, 1996; Kleywegt & Jones, 1995; EU 3-D Validation Network, 1998; Kleywegt & Jones, 1997; Kleywegt & Brünger, 1996, além das próprias referências dadas no início deste Capítulo.

Além disso, um grande esforço foi feito para melhoria dos protocolos de refinamento. Como resultado, foram feitos novos programas que utilizam Maximum Likelihood como modelo probabilístico para o cálculo de parâmetros sob refinamento, mais adequadas para refinamento de estruturas de macromoléculas⁷.

Outro progresso marcante foi a introdução de um indicador de qualidade confiável. Conforme mencionado acima, no caso de macromoléculas, o número de parâmetros atômicos sob refinamento pode exceder o número de reflexões medidas. A introdução de novos parâmetros, por exemplo, fatores de temperatura anisotrópicos ou múltiplas conformações, com o intuito de melhorar o ajuste dos dados, provavelmente acabe ajustando ruídos, sem adicionar características genuínas da estrutura. Tradicionalmente, um parâmetro utilizado para indicar o quão bem o modelo ajusta os dados experimentais é o R_{factor}

$$R = \frac{\sum_{\boldsymbol{h} \in U} w_{\boldsymbol{h}} ||F_{obs}(\boldsymbol{h})| - k|F_{calc}(\boldsymbol{h})||}{\sum_{\boldsymbol{h} \in U} w_{\boldsymbol{h}}|F_{obs}(\boldsymbol{h})|},$$
(2.3)

conjunto de trabalho, o que leva a um valor subestimado para o R_{free} e invalida o teste, uma vez que não constituem mais um conjunto independente (Brünger, 1997; Kleywegt & Brünger, 1996).

⁷ O método de Mínimos Quadrados é um caso particular do *Maximum Likelihood* e é empregado quando a solução já está próxima do mínimo, utilizando-se a estatística χ para avaliação do processo. Além disso, é o método de escolha para refinamento no caso de pequenas moléculas.

onde U é conjunto de todas as reflexões, w_h é um fator de peso (geralmente, inversamente proporcional à variância da intensidade da reflexão), k é um fator de escala e $|F_{obs}(\mathbf{h})|$ e $|F_{calc}(\mathbf{h})|$ são as amplitudes dos fatores de estrutura observadas e calculadas, respectivamente. O R_{factor} é uma medida da concordância entre os fatores de estrutura medidos e calculados.

Apesar de amplamente utilizado, mostrou-se que é possível obter um R_{factor} tão baixo quanto se queira com modelos completamente errados do ponto de vista bioquímico (Branden & Jones, 1990; Kleywegt, 2000; Kleywegt & Brünger, 1996). Além disso, uma vez que o modelo é refinado contra o mesmo conjunto sobre o qual o teste é aplicado, uma introdução de novos parâmetros ou, de forma equivalente, a exclusão de reflexões do conjunto original, reduz artificialmente o R_{factor} .

Este problema foi solucionado com a introdução da técnica estatística de cross-validation, que consiste em separar uma fração dos dados (geralmente 5% das reflexões) em um conjunto-teste que não será usado desde ínicio até o final do refinamento. Define-se, então, o R_{free} de forma similar

$$R_{free} = \frac{\sum_{\boldsymbol{h}\in T} w_{\boldsymbol{h}} ||F_{obs}(\boldsymbol{h})| - k|F_{calc}(\boldsymbol{h})||}{\sum_{\boldsymbol{h}\in T} w_{\boldsymbol{h}}|F_{obs}(\boldsymbol{h})|},$$
(2.4)

porém, desta vez, apenas o conjunto teste (T) das reflexões colocadas à parte durante o refinamento são utilizadas para cálculo do R_{free} (as grandezas $w_{\mathbf{h}}$, k, $|F_{obs}(\mathbf{h})| \in |F_{calc}(\mathbf{h})|$ são definidas da mesma forma que na equação (2.3) acima).

O R_{free} é utilizado para previnir *over-fitting*, permitindo avaliar se a introdução de novos parâmetros é justificada pela adição de características genuínas do modelo ou se apenas dá origem a artefatos resultantes da modelagem de ruído presente nos dados. Quanto mais pobres são os dados, maior a chance de que ruídos sejam erroneamente modelados através da introdução de novos parâmetros, o que se reflete num aumento da diferença entre o R_{factor} e o R_{free} (neste caso, apenas o primeiro decai). Vínculos de NCS, ao contrário, diminuem o número de parâmetros e a diferença entre os dois fatores.

Outras aplicações do R_{free} são a indicação de erros introduzidos no modelo durante o processo de reconstrução, a otimização do protocolo de refinamento, guiando a determinação de valores apropriados para parâmetros como o peso dos vínculos estereoquímicos em relação aos dados experimentais e o peso dos vínculos utilizados quando há NCS. Também, uma estimativa da acuidade do modelo pode ser obtida a partir do conjunto-teste utilizado para o cálculo do R_{free} .

2.6.2 Validação e deposição da estrutura

O refinamento resulta num modelo aprimorado e em novas fases. Embora haja alternativas, no protocolo mais amplamente utilizado, dois mapas são calculados a partir das novas fases, contra os quais o novo modelo será confrontado para localizar erros e identificar novas características estruturais:

- 2mF_{obs} DF_{calc}: este mapa resulta de uma série de Fourier cujos coeficientes são 2mF_{obs} DF_{calc} e as fases, as resultantes do refinamento. Os pesos m e D (Murshudov *et al.*, 1997; CCP4, 1994; Winn *et al.*, 1997), definidos no Apêndice A, evitam vícios na densidade decorrentes de um modelo imperfeito, usado no cálculo de F_{calc}. Este mapa é usado como guia para interpretação e reconstrução do modelo;
- $mF_{obs} DF_{calc}$: este mapa, composto a partir dos coeficientes $mF_{obs} DF_{calc}$, é chamado "mapa de diferença", entretanto, tanto este quanto o primeiro são mapas de diferença de Fourier. Os pesos m e D são os mesmos mencionados acima. O mapa de diferença é um instrumento importante pois indica as regiões onde novos átomos são

necessários (picos positivos) e onde átomos foram introduzidos incorretamente no modelo (picos negativos). Os picos positivos são, também, o instrumento a partir do qual moléculas de água são adicionadas ao modelo.

Se tudo correr bem, o processo de refinamento deve abaixar os indicadores R_{factor} e R_{free} , levando-os até valores estacionários aceitáveis, uma indicação de que mais ciclos de refinamento levam a mudanças negligenciáveis nos parâmetros. Ao mesmo tempo, os mapas de diferença tornam-se uniformes, sem picos de densidade significativos (positivos ou negativos). É o momento em que o refinamento é interrompido, já que toda informação disponível foi utilizada para aprimoramento do modelo.

A etapa seguinte é o que se chama validação do modelo, que é a aplicação de critérios de qualidade ao modelo, verificando sua acuidade e precisão. Em cristalografia de proteínas, acuidade diz respeito a $\langle |\Delta \Phi| \rangle$, a média do módulo dos erros das fases, quantidade diretamente proporcional ao R_{free} . A precisão do modelo está relacionada ao nível de detalhe do modelo, cuja descrição é limitada pela quantidade e qualidade dos dados (conteúdo de informação). A validação é uma etapa delicada onde vários aspectos são considerados. Excelentes compêndios a respeito podem ser encontrados, por exemplo, em Kleywegt, 2000 e EU 3-D Validation Network, 1998.

Enquanto R_{factor} e R_{free} são estatísticas globais que avaliam o modelo contra os dados experimentais, uma série de critérios são usados para avaliar localmente a geometria e estereoquímica do modelo, através da comparação com valores "ideais" depositados em bancos de dados (distâncias e ângulos de ligação) de pequenas moléculas. Entre os critérios, um dos mais utilizados é o *plot* de Ramachandran (Kleywegt & Jones, 1996), que mostra a distribuição dos ângulos torsionais $\varphi \in \psi$ da cadeia principal, constituindo um excelente indicador local de qualidade. Resíduos com combinações não usuais de valores (φ, ψ) provavelmente apresentam problemas e devem ser re-examinados.

Uma vez que a qualidade da estrutura seja satisfatória, pode-se utilizála para análise biológica, que é o fim último que justifica o trabalho feito. Geralmente, a estrutura resolvida juntamente com os dados coletados é disponibilizada para a comunidade científica através da deposição num banco de dados, usualmente o *Protein Data Bank* (PDB, http://www.rcsb.org/pdb). O processo de deposição envolve vários passos e não será descrito aqui.

Capítulo 3

β -manosidase: estudos cristalográficos

Os estudos cristalográficos da β -manosidase foram iniciados em Jan/1998. Não havia um modelo disponível no PDB que pudesse ser usado com sucesso para Substituição Molecular, de forma que métodos baseados em átomos pesados foram adotados.

Durante os anos de 1999 e 2000, trinta conjuntos de dados foram coletados na linha CPR do LNLS, entre nativos e derivados de átomos pesados. Uma segunda etapa do doutorado iniciou-se com o estágio realizado no primeiro semestre de 2001, no York Structural Biology Laboratory (YSBL), em York, Inglaterra, quando todos os dados disponíveis foram novamente analisados junto a pesquisadores do YSBL reconhecidos na área.

3.1 Purificação e seqüenciamento

A β -manosidase utilizada foi purificada a partir de um filtrado da cultura do fungo *Trichoderma reesei*. Durante o desenvolvimento do projeto, o trabalho de purificação foi feito pelo grupo da Divisão de Biofísica Molecular e Radiação do Instituto de Física Nuclear Petersburg, em São Petersburgo, Rússia, cujo principal colaborador foi o pesquisador Alexander M. Golubev.

A seqüência da β -manosidase não era conhecida no início do projeto, o que representava um problema, já que ela seria necessária não só na etapa de interpretação dos mapas de densidade eletrônica mas em outros procedimentos, como a determinação do conteúdo da unidade assímétrica e modificação de densidade eletrônica. O problema não foi completamente resolvido porque apenas parte da proteína foi seqüenciada, como segue. Entretanto, essa seqüência parcial mostrou-se útil para alguns procedimentos (Capítulo 4).

O seqüenciamento parcial da proteína foi feito no laboratório do Prof. Dr. H. F. Al El-Dorry, no IQ/Bioquímica - USP/ São Paulo, pelo seu aluno Eric D. Alessandro Bonaccorsi. O Dr. Jörg Kobarg, do LNLS, teve importante participação nas discussões sobre as estratégias de seqüenciamento disponíveis¹.

Para a clonagem do gene da β -manosidase de *Trichoderma reesei*, inicialmente o caminho clássico foi utilizado, onde seqüências de peptídios obtidos por clivagem com proteases seriam utilizadas para o desenho de oligonucleotídeos degenerados, a serem usados em PCR (*polymerase chain reaction*) para amplificar fragmentos de DNA genômico ou de cDNA específicos para a β -manosidase. Bibliotecas genômicas de cDNA de *Trichoderma reesei* foram disponibilizados pelo Prof. El Dorry e utilizadas como DNA de molde.

Esta primeria tentativa foi feita como segue. A proteína foi purificada através de SDS-gel-eletroforese e transferida para uma membrana de PVDF usando o método de Western blotting, corada e a banda correspondente à β -manosidase (~ 100 kDa) foi recortada e enviada à University of British Columbia, Biotechnology Laboratory (NAPS, www.biotech.ubc.ca/ naps/napshome.htm), Vancouver, B.C./Canada para obtenção de fragmen-

 $^{^{-1}}$ A quem agradeço, também, pela colaboração na descrição da metodologia utilizada para seqüenciamento.

tos e sequências de peptídeos por degradação de Edman. Os 12 peptídeos abaixo foram seqüenciados²:

²O ponto de interrogação acusa ambigüidade no aminoácido identificado.

- 1 STGKASIHDLALQKXTVTNEY
- 2 S(H/Q/T/L) A S? M(F) G(T/S) V I(G) I T A(F) G(A) S G(Y/V)
- 3 S $Y(R/D/V) \to K(F/R)$ A K A A I R Q H A G A
- 4 $A(F/I/M) \mathbf{P} \mathbf{V} \mathbf{G} \mathbf{A} \mathbf{D} \mathbf{A} \mathbf{V} \mathbf{L} G(W) I L N W? N? E$
- 5 G(F/Y) Q F(P) V(N) L(N) F(Y) A(G) P(N) L(F) Q(G) S L(A / V) P(S) Q? V
- $6 \quad \mathbf{S} \ \mathbf{A} \ \mathbf{V} \ \mathbf{M} \ \mathbf{S} \ \mathbf{V} \ \mathbf{I} \ \mathbf{Y} \ \mathbf{G} \ \mathbf{T} \ \mathbf{X} \ \mathbf{N} \ \mathbf{G} \ \mathbf{G}$
- 7 A N E L Y S R N T H D(I/G) D A S? X Q
- 8 S Q(E/L) A(V) X G(R/L) D D V M F X I R L F(D)
- 9 K? X N V(P) Y V(A) L(E) R(I) R(T) Y(F/) F(E/N) N(V) L(R) F(R) I(G)
- 10 K? N P E L Y X L N T D G(I/L/Q/E/A) D I G F R K.
- 11 K? X V X N E Y(I) G
- 12 K? PILVXQRRVGFRTI(T)LF

Nem todas essas seqüências acima são úteis para o desenho de oligonucleotídeos (as regiões mais úteis estão em negrito). Uma série de oligonucleotídeos foram desenhados, a maioria deles baseada numa seqüência peptídica derivada do N-terminal da β -manosidase (peptídio 1 acima). Os oligos foram utilizados para amplificar o gene da β -manosidase usando bibliotecas genômicas de *T. reesei* como DNA de molde. Vários produtos de DNA foram amplificados, clonados e seqüenciados. A idéia era obter uma seqüência com alta similaridade com as seqüências já conhecidas de β -manosidase de outras espécies. O DNA do respectivo clone seria, então, utilizado para procurar um clone de cDNA da β -manosidase num banco de cDNA de *T. reesei*. O novo clone seria seqüenciado por inteiro, gerando toda informação seqüencial da enzima.

Na abordagem baseada em PCR, tanto uma biblioteca de T. reesei quanto uma biblioteca de controle de Aspergillus niger foram usadas. Embora alguns produtos de PCR tenham sido obtidos com peso molecular comparável em ambas as reações, o seqüenciamento de DNA do produto vindo de T. reesei mostrou seqüências sem nenhuma relação com as seqüências de β -manosidase descritas nos bancos de dados. Portanto, com esta abordagem não se obteve sucesso, o que poderia estar baseado no fato de que as seqüências peptídicas obtidas eram de baixa qualidade. Além disso, devido à degeneração do código genético, era muito difícil predizer exatamente a seqüência de DNA correspondente, o que resultou num grau de incerteza nos oligonucleotídeos preditos, mais um fator que limitou esta abordagem.

Em conseqüência, uma nova estratégia de clonagem foi adotada, desta vez baseada na comparação entre as seqüências do gene de β -manosidase de dois fungos estreitamente relacionados ao *T. reesei*: *Aspergillus niger* e *Aspergillus aculeatus*. A comparação das seqüências entre os genes de β -manosidase nas duas espécies de *Aspergillus* indicou várias regiões que eram altamente conservadas entre elas, as quais seriam localizadas em regiões afastadas por 1500 bp. A alta similaridade e conservação destas regiões indicava que as mesmas deveriam também ser conservadas no gene de β -manosidase de *T. reesei*.

Subseqüentemente, novos oligonucleotídeos correspondentes a estas regiões foram desenhados e usados em reações de PCR. Novamente, as bibliotecas de T. reesei foram usadas como DNA de molde e, como controle, as bibliotecas de A. niger. Um produto de 1500 bp pôde ser obtido de ambas as reações, o que indicou que a nova estratégia teve sucesso.

O produto de PCR amplificado foi radioativamente marcado e usado num *screen* de hibridização de placas de fagos contendo fragmentos de DNA genômico de *T. reesei*. O *screening* resultou na identificação de vários clones reagentes. Os DNAs de fagos recombinantes foram isolados e digeridos com enzimas de restrição apropriadas. Dois insertos genômicos de ~ 3000 e \sim 4500 bp foram identificados, sub-clonados e submetidos a seqüenciamento de DNA.

O seqüenciamento revelou vários "contigs" que geraram mais de 70% da seqüência do gene codificador de β -manosidase. Buscas em bancos de dados e alinhamentos comparativos de todos os "contigs" obtidos com seqüências conhecidas de genes que codificam β -manosidas
es identificaram sem ambigüidade o gene de T.~reesei .

O seqüenciamento não foi terminado, principalmente porque o laboratório onde estava sendo feito teve participação importante no seqüenciamento da Xylella fastidiosa, o qual foi priorizado na época. De um total de 950 aminoácidos, aproximadamente, esperados para β -manosidase de *T. reesei*, o seqüenciamento levou a uma seqüência final que inclui um bloco de ~ 500 aminoácidos e blocos menores, separados por lacunas:

GDCDLRSILEDPQFCGQHIGNADNQFRQWVFDISEVLASCTSPQIAAEPGQETWPDGINEVY EIPNRWFVRKEQSDFGWYVDLSARFIWGPAFVPVGIWQPAYLVQLDGPGSVYVKNSAFDLYR KGQLNNLPGTIPDEAQIRYTIVDSDSLEEVSSGSLSDVENAGDVITGLATLDLYNVTVEVIS NNKSIASVNKRMGFRTIVLNMEPISELELSQGIIFNCGLG-SRATVHFEVNGHVFYAKGSNF VPPDPFWPRVTFEHIRWSSGAYSPDFMYDLADEMGILLWCEFQFSVSLYPVDPAFLENVRAL WAGGNEMEKDELPAVKSRAPSQYERYLSEYLLLFLDTLLPTVGYLELNFSLPIPFVDRLYNT TPGYLYGDSDFYDYDPSHAFDINKSLETWREAIPEDQLFFNSTMTVLRNHHYPPGSLTTNNT ADPLRGKTDTVANFAAWCHSTQVFQADFYHTKIQYYRAGSGMPQRQLGSLEYDGRWKVTFYA TKDIFQPVIVAPVFNVTTGVLEIFAVSDLWSDVGDPEASVQSQKFTVGPVNSTVLSRINITQ LTTSGGLPASNAILKTYSHSNYFTATPLSKAELIDPGLTIRHDGDAFTVTAEKGVSAFDDNG FWLRQGQSRTLGYRVHGESSGWESRVTVSSIWNNTLPS-FRGRWL-KLAALEE

3.2 Cristalização da β -manosidase

As condições de cristalização e análise preliminar de um cristal nativo de β -manosidase (Figura 3.1) foram publicadas sob a referência

Aparicio, R., Eneiskaya, E. V., Kulminskaya, A. A., Savel'ev, A. N., Golubev, A. M., Neustroev, K. N., Kobarg, J. and Polikarpov, I. (2000).

Crystallization and preliminary X-ray study of β -mannosidase from *Tricho*derma reesei. Acta Cryst. D 56, 342-343.



(a) Porção de uma gota de cristalização com vários cristais da proteína, que têm dimensão máxima de ~ 0.3 mm.

(b) Cristais em uma gota límpida mostrados em maior detalhe.

Figura 3.1: Cristais de β -manosidase de *Trichoderma reesei*.

Este artigo é anexado a seguir. Após, são apresentadas as análises de dados que foram feitas e os resultados obtidos durante o estágio no YSBL. Estas análises revelaram que o projeto da determinação da estrutura cristalográfica da β -manosidase era de uma complexidade ímpar, em grande parte devida à desordem inerente aos cristais da proteína, resultante da glicosilação da própria β -manosidase.

crystallization papers

Acta Crystallographica Section D Biological Crystallography

ISSN 0907-4449

R. Aparicio,^{a,b} E. V. Eneiskaya,^c A. A. Kulminskaya,^c A. N. Savel'ev,^d A. M. Golubev,^c K. N. Neustroev,^c J. Kobarg^a and I. Polikarpov^a*

^aNational Synchrotron Light Laboratory, LNLS, Caixa Postal 6192, CEP 13083-970, Campinas, SP, Brazil, ^bGleb Wataghin Physics Institute, State University at Campinas, UNICAMP, Caixa Postal 6165, CEP 13083-970, Campinas, SP, Brazil, ^cPetersburg Nuclear Physics Institute, Gatchina, St Petersburg, 188350, Russia, and ^dSt Petersburg Technical University, 29 Politechnicheskaya Str., St Petersburg, 195251, Russia

Correspondence e-mail: igor@lnls.br

© 2000 International Union of Crystallography Printed in Denmark – all rights reserved



Figure 1

A typical diffraction pattern from the data set reported above. The crystal-to-detector distance was set to give a resolution limit of 2.65 Å at the outer edge of the image. Other resolution rings were drawn.

 β -Mannosidase from *Trichoderma reesei*, a 105 kDa glycoprotein, has been crystallized. The crystals belong to the space group $P4_12_12$ or $P4_32_12$, with unit-cell dimensions a = b = 165.86, c = 122.46 Å, and diffract beyond 2.75 Å resolution. X-ray diffraction data were collected from a frozen crystal on a synchrotron X-ray source.

Received 18 November 1999 Accepted 20 December 1999

1. Introduction

 β -Mannosidase is an exoglycosidase that catalyzes hydrolysis of non-reducing residues of β -D-mannose in β -D-mannosides. The enzyme has been isolated from plants (Houston et al., 1974; Li & Lee, 1972), fungi (Bouquelet et al., 1978; Neustroev et al., 1992; Kulminskaya et al., 1999), animals (Iwasaki et al., 1989; Cavanagh et al., 1985) and bacteria (Akino et al., 1988; Oda et al., 1993), including thermophilic and alkalophilic microorganisms. The complete human (Alkhayat et al., 1998), caprine (Leipprandt et al., 1996), bovine (Chen et al., 1995) and Aspergillus aculeatus (Takada et al., 1999) β -mannosidase cDNAs have been sequenced, revealing 75% identity between the first three sequences (Alkhayat et al., 1998). The lysosomal degradation of N-glycoproteins is the most important process in which β -mannosidase participates. Glycoprotein degradation

> occurs predominantly within cell lysosomes by the sequential action of endo- and exoglycosidases. β -Mannosidase cleaves mannosidic bonds in the core part of N-linked glycans attached to glycoproteins (Scriver *et al.*, 1989).

> In man and ruminants, a deficiency of β -mannosidase leads to an autosomal recessive inherited lysosomal storage disease (Watts & Gibbs, 1986) called β -mannosidosis. The disease is characterized by storage and excretion of the disaccharide Man- $\beta(1\rightarrow 4)$ -GlcNAc, the trisaccharide Man- $\beta(1\rightarrow 4)$ -GlcNAc, the trisaccharide Man- $\beta(1\rightarrow 4)$ -GlcNAc and other more complex undegraded oligosaccha

rides in various tissues (Percheron et al., 1992). The disorder was first described in Nubian goats (Jones & Dawson, 1981) and later in humans (Cooper & Sardharwalla, 1986; Wenger et al., 1986) and Saler cattle (Bryan et al., 1990). In man, clinical manifestations are heterogeneous and include mental retardation, peripheral neuropathy, skeletal abnormalities and others (Rodriguez-Serna et al., 1996). A theoretical study of structural and functional features of bovine β -mannosidase (Chen *et al.*, 1995) and another four enzymes implicated in lysosomal storage diseases, based on a family classification of glycosidases (EC 3.2.1.x), was performed using the hydrophobic cluster analysis method (Durand et al., 1997). Some likely preserved features of the catalytic domains were deduced, despite the low levels of sequence identity. Another observed pathological variation of the enzyme, which manifests itself in a significantly decreased serum activity in 64 diabetic patients, remains to be investigated (Bernard et al., 1985).

Fungi and bacteria able to degrade hemicellulose also secrete β -mannosidase. This β -mannosidase hydrolyzes $\beta(1\rightarrow 4)$ -Dmannosyl groups from manno-oligosaccharides and mannose-containing glycopepetides produced from the hemicellulose pulp by endoenzymes (Kuhad et al., 1997). In seeds that have galactomannans as storage carbohydrates, the enzyme converts manno-oligosaccharides to monosaccharides (Dey & del Campillo, 1984). Some enzymes have a substrate-binding site in addition to the catalytic site, which may modulate enzymatic action, possibly within a multidomain structure (Kuhad et al., 1997; Neustroev et al., 1993; Voet & Voet, 1995).

A non-catalytic galactomannan-binding site was established in β -mannosidase from *T. reesei* on the base of kinetic studies (Kulminskaya *et al.*, 1999). Knowledge of the threedimensional structure of the enzyme will help in the understanding of the structural basis of its activity and function.

crystallization papers

Table 1Data-collection statistics.

Values in parentheses refer to the last resolution shell.

Number of images	115
Space group	P41212 or P43212
Unit-cell dimensions at 80 K (Å)	
a = b	165.86
с	122.46
Resolution range (Å)	20.00-2.75
Last shell (Å)	2.81-2.75
Number of measured reflections	373335 (17061)
Number of unique reflections	44389 (2693)
Redundancy	8.41 (6.34)
$R_{\rm sym}$ † (%)	0.214 (0.697)
Completeness (%)	99.1 (91.6)
$\langle I/\sigma(I)\rangle$	12.80 (2.13)

† $R_{\text{sym}} = \sum (I - \langle I \rangle) / \sum I.$

2. Crystallization and data collection

 β -Mannosidase was purified to homogeneity from a culture filtrate of T. reesei, as described by Kulminskaya et al. (1999). Initial crystallization conditions were screened by the sparse-matrix method (Jancarik & Kim, 1991) using the macromolecular crystallization reagent kits I and II (Hampton Research). In each trial, a hanging drop of 1 µl of protein solution $(10 \text{ mg ml}^{-1} \text{ in water})$ mixed with 1 µl of precipitant solution was equilibrated against a reservoir containing 500 µl of precipitant solution. Small clusters of needles grew at room temperature in precipitant solution 12 [30%(v/v) PEG 400, 0.1 M CdCl₂, 0.1 M sodium acetate pH 4.6] from the Hampton reagent kit II in one week. The experiment was repeated at 277 K and small crystals were obtained from a drop containing protein solution, precipitant solution and water in a 2:1:1 ratio. Further optimization at 291 K including pH refinement (McPherson, 1995) led to the following conditions. A reservoir solution containing 26% PEG 400, 0.13 M CdCl₂ and 0.1 M sodium acetate pH 4.7 was prepared. A 2 µl drop of the protein solution was mixed with 1 µl of reservoir solution plus 1 µl of water. After equilibration, well formed bipyramidal crystals of dimensions $0.5 \times 0.4 \times 0.3$ mm grew in 3– 15 d. The crystals obtained diffracted poorly and to low resolution. In spite of the fact that the crystals showed good morphology and size, they were rapidly damaged when exposed to the X-ray beam, resisting a few images only. The best resolution of the room-temperature data set was limited to 3.5 Å. Cryo-crystallographic techniques (Garman & Schneider, 1997) were employed to overcome radiation damage. Crystallization conditions allowed flash-freezing of the crystal in a gaseous nitrogen stream (Oxford Cryosystems) without further modifications of the mother liquor. Several diffraction data sets were collected. Data were collected on a 345 mm MAR Research imaging-plate detector at the LNLS Protein Crystallography beamline (Polikarpov, Oliva *et al.*, 1998; Polikarpov, Perles *et al.*, 1998) by the oscillation method from a single native crystal at 80 ± 1 K (Fig. 1). The wavelength was 1.38 Å and the crystal-to-detector distance was 205 mm, with an oscillation range of 1° and an exposure time of ~15 min per image.

3. Results and discussion

Data were processed using the *DENZO* and *SCALEPACK* packages (Otwinowski & Minor, 1997). Results of data processing are summarized in Table 1. Flash-freezing significantly improved the crystal's resistance to X-rays. This allowed a higher X-ray dose per image. An improvement in the $I/\sigma(I)$ ratio along with a high redundancy led to better statistics and a higher resolution cutoff. We aim to solve this structure by the multiple isomorphous replacement method. Screening for heavy-atom derivatives is under way.

This work was supported by grants 98/06761-1 and 99/03387-4 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, and by grant 97-04-50035 from the Russian Foundation for Basic Research (RFBR), Russia. RA and IP gratefully acknowledge the financial support from FAPESP and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

References

- Akino, T., Nakamura, N. & Horikoshi, K. (1988). Agric. Biol. Chem. Tokyo, 52(6), 1459–1464.
 Alkhavat, A. H., Kraemer, S. A., Leipprandt, J. R.,
- Aikinayai, A. H., Kraemer, S. A., Leipprandt, J. K., Macek, M., Kleijer, W. J. & Friderici, K. H. (1998). *Hum. Mol. Genet.* 7(1), 75–83.
- Bernard, M., Brochet, C. & Percheron, F. (1985).
 Clin. Chim. Acta, **152**(1/2), 171–174.
 Bouquelet, S., Spik, G. & Montreuil, J. (1978).
- Biochim. Biophys. Acta, **522**, 521–530. Bryan, L., Schmutz, S., Hodges, S. D., Snyder, F. F.
- (1990). Biochem. Biophys. Res. Commun. 173(2), 491–495.
- Cavanagh, K. T., Fisher, R. A., Legler, G., Herrchen, M., Jones, M. Z., Julich, E., Sewell Alger, R. P., Sinnott, M. L. & Wilkinson, F. E. (1985). *Enzyme*, **34**(2), 75–82.

- Chen, H., Leipprandt, J. R., Traviss, C. E., Sopher, B. L., Jones, M. Z., Cavanagh, K. T. & Friderici, K. H. (1995). J. Biol. Chem. 270(8), 3841–3848.Cooper, A. & Sardharwalla, I. B. (1986). N. Engl.
- J. Med. **315**(19), 1231. Dey, P. M. & del Campillo, E. (1984). Adv.
- Enzymol. Relat. Areas Mol. Biol. 56, 141-249. Durand, P., Lehn, P., Callebaut, I., Fabrega, S.,
- Henrissat, B. & Mornon, J. P. (1997). *Glycobiology*, 7(2), 277–284.
 Garman, E. F. & Schneider, T. R. (1997). *J. Appl.*
- Garman, E. F. & Schneider, T. R. (1997). J. Appl. Cryst. **30**, 211–237.
- Houston, C. W., Latimer, S. B. & Mitchell, E. D. (1974). *Biochim. Biophys. Acta*, **370**(1), 276–282.
- Iwasaki, Y., Tsuji, A., Omura, K. & Suzuki, Y. (1989). J. Biochem. (Tokyo), 106 (2), 331–335.
- Jancarik, J. & Kim, S.-H. (1991). J. Appl. Cryst. 24, 409–411.
- Jones, M. Z. & Dawson, G. (1981). J. Biol. Chem. **256**(10), 5185–5188.
- Kuhad, R. C., Singh, A. & Eriksson, K. L. (1997). Biochem. Eng. Biotechnol. 57, 45–125.
- Kulminskaya, A. A., Eneiskaya, E. V., Isaeva-Ivanova, L. S., Savel'ev, A. N., Sidorenko, I. A., Shabalin, K. A., Golubev, A. M. & Neustroev, K. N. (1999). *Enzyme Microb. Technol.* **25**(3–5), 372–377.
- Leipprandt, J. R., Kraemer, S. A., Haithcock, B. E., Chen, H., Dyme, J. L., Cavanagh, K. T., Friderici, K. H. & Jones, M. Z. (1996). *Genomics*, **37**(1), 51–56.
- Li, Y. T. & Lee, Y. C. (1972). J. Biol. Chem. 247, 3677–3683.
- McPherson, A. (1995). J. Appl. Cryst. 28, 362-365.
- Neustroev, K. N., Krylov, A. S., Firsov, L. M., Abroskina, O. N. & Khorlin, A. Y. (1992). *Biochemistry (Moscow)*, 56(8), 982–986.
- Neustroev, K. N., Valter, S. N., Timchenko, N. V., Firsov, L. M., Golubev, A. M. & Khokhlov, S. E. (1993). *Biochem. Mol. Biol. Int.* **30**(1), 115–120.
- Oda, Y., Komaki, T. & Tonomura, K. (1993). Food Microbiol. 10(4), 353–358.
- Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307-326.
- Percheron, F., Foglietti, M. J., Bernard, M. & Ricard, B. (1992). *Biochimie*, **74** (1), 5–11.
- Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R. C., Arruda, P., Leite, A. & Craievich, A. (1998). Nucl. Instrum. Methods A, 405, 159–164.
- Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998). J. Synchrotron Rad. 5(2), 72–76.
- Rodriguez-Serna, M., Botella-Estrada, R. M. D., Chabas, A., Coll, M., Oliver, V., Febrer, M. & Aliaga, A. (1996). Arch. Dermatol. 132(10), 1219–1222.
- Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (1989). The Metabolic Basis of Inherited Disease, Vol. II, 6th ed. New York: McGraw-Hill, Inc.
- Takada, G., Kawaguchi, T., Kaga, T., Sumitani, J. & Arai, M. (1999). Biosci. Biotechnol. Biochem. 63(1), 206–209.
- Voet, D. & Voet, J. (1995). Biochemistry, 2nd ed. New York: John Wiley & Sons, Inc.
- Watts, R. W. E. & Gibbs, D. A. (1986). Lysosomal Storage Diseases – Biochemical and Clinical Aspects. Philadelphia: Taylor & Francis, Inc.
- Wenger, D. A., Sujanski, E., Fennessey, P. V. & Thompson, J. N. (1986). N. Engl. J. Med. 315(19), 1201–1205.

3.3 Coletas de dados no LNLS

3.3.1 Otimização das condições de cristalização

Na busca de cristais que difratassem a maior resolução e com maior intensidade, procurou-se otimizar as condições de cristalização originais através da variação de parâmetros como pH, concentração da proteína, precipitante e sais e tentativas de troca do Cd por Zn ou Mg. Não houve melhoras significativas dos dados.

Também, durante o primeiro ano do projeto, amostras de β -manosidase foram enviadas para os Estados Unidos para experimentos de cristalização em ambiente de micro-gravidade mas, apesar de todos os cuidados tomados, não houve resultados satisfatórios uma vez que os cristais que retornaram ao Brasil eram bastante pequenos e praticamente não exibiam nenhum padrão de difração.

Durante o processo de otimização, notou-se que o tempo de crescimento é um fator importante para a qualidade do cristal, o que se explica da seguinte forma. A β -manosidase é uma proteína glicosilada, o que implica em heterogeneidade molecular uma vez que as cadeias de açúcares têm comprimentos variáveis. Moléculas menos glicosiladas são menos solúveis e devem cristalizar-se a menores concentrações de precipitante (em comparação a moléculas mais glicosiladas), ou seja, devem cristalizar-se primeiro, já que a concentração de precipitante na gota aumenta com o tempo (pág. 21). Isto faz com que cristais crescidos em menor tempo, formados por moléculas menos glicosiladas e, portanto, mais homogêneas, sejam mais ordenados, exibindo um padrão de difração de melhor qualidade e a mais alta resolução. Assim, sempre que possível, cristais com menor tempo de crescimento eram preferidos para realização dos testes de coleta. Outro fator importante é que o nível de glicosilação muda de amostra para amostra, fazendo com que a solubilidade da proteína e as condições de cristalização variem de acordo com a remessa de proteína.

O mais nefasto efeito da glicosilação, do ponto de vista da cristalografia, é tornar as moléculas heterogêneas, aumentando a desordem do cristal, de modo que os problemas comentadas à pág. 53 tendem a acentuar-se drasticamente. Porisso, outra estratégia em busca de melhores dados foram os ensaios de cristalização de proteína deglicosilada, tanto a partir das condições já conhecidas para a proteína glicosilada como iniciando novo *screening*. Apesar do esforço, cristais de proteína deglicosilada não foram obtidos (conforme dito, a glicosilação interfere na solubilidade da proteína e, portanto nas condições de cristalização).

3.3.2 Preparação para a coleta de cristais nativos e derivados

Para coleta de dados ou obtenção de derivados (metodologia comentada à pág. 28), era necessário transferir o cristal para outra solução que não a solução da gota onde ele havia crescido. Usualmente, uma quantidade extra de precipitante (de 2% a 4%) era adicionada à solução original, de modo a previnir a dissolução do cristal durante a transferência. Este procedimento também dava uma margem de segurança para a crioproteção através do próprio PEG400 (próxima seção). No caso de coletas de derivados, os cristais já crescidos eram transferidos para uma solução semelhante à solução de crescimento mas onde havia sido acrescentada certa quantidade do reagente com o átomo pesado escolhido e feitos os ajustes requeridos (por. ex., acerto de pH).

Após certo tempo de banho, os cristais eram retirados da gota e resfriados diretamente no feixe de nitrogênio gasoso, dando início à coleta de dados. Uma dificuldade extra é que a maior parte dos cristais de β -manosidase, embora tivessem um hábito excelente, não eram adequados para coleta de

dados, exigindo considerável tempo de busca para que se encontrasse um cristal que exibisse um padrão de difração aceitável. Caso não houvesse disponibilidade de feixe de Raios X, os cristais eram guardados imersos em nitrogênio líqüido para reutilização posterior e continuação da coleta.

3.3.3 Coleta a temperaturas criogênicas

Já nos primeiros testes de difração ficou claro que os cristais eram extremamente sensíveis aos Raios X e, a temperatura ambiente, não mais que duas imagens podiam ser coletadas antes da completa deterioração da estrutura interna do cristal. O problema foi solucionado com a coleta de dados a baixas temperaturas (pág. 23), que também levou a um aumento relevante no limite de resolução dos dados (de 3.5 Å a temperatura ambiente para ~ 2.5 Å).

As condições de crescimento da β -manosidase - onde se usa PEG400 como precipitante - permitem o resfriamento dos cristais diretamente no fluxo de nitrogênio gasoso sem adição de crioprotetores (PEG400 atua também como crioprotetor). A maioria dos conjuntos foi coletada à temperatura de $(85 \pm 1)K$, sendo que alguns foram coletados a $(100 \pm 1)K$, temperatura onde era mais fácil obter estabilidade do soprador de nitrogênio. A estratégia de coleta (Dauter, 1999) foi determinada com o programa MARS-TRATEGY³. Para redução de dados, foram usado os programas DENZO e SCALEPACK (Otwinowski & Minor, 1997).

3.3.4 Análise preliminar dos conjuntos coletados até 1999

A Tabela 3.1 apresenta um sumário das principais estatísticas para os conjuntos coletados até 1999.

³ Klein, C. (1997), X-ray Research G. m. b. H., Germany, xray@mail.ppp.de

											Ι	Dimens	ões da	$\begin{bmatrix} 3\\ 3\\ 3\end{bmatrix}$
Cj	Nome do conjunto a,b	Composto contendo	$Soaking^c$	λ^d	Resolução	Imgs^e	$\operatorname{Comp.}^{f}$	$Mult.^{g}$	$\overline{I/\sigma(I)}^h$	$R_{sym}{}^i$		cela ui	nitária	l 0
		átomo pesado		(Å)	(Å)		(%)			(%)			a, b, c	let
Con	juntos do grupo ^{j} $P4_{1/3}$	2 ₁ 2:						-						as
1	cryobman111198	cristal nativo		1.376	20.0 - 3.4	35	78.9	2.6	5.2	19	164.4	164.4	131.9	de
2	bman 250299	cristal nativo		1.380	15.0 - 3.0	54	88.7	3.0	6.8	16	165.7	165.7	123.2	dad
3	bman050599	cristal nativo		1.380	15.0 - 3.0	115	100.0	8.8	12.7	19	165.9	165.9	122.5	sol
4	bman290799	cristal nativo		1.380	30.0 - 3.3	55	96.1	3.4	6.8	18	165.8	165.8	123.4	no
5	bman030899	cristal nativo		1.380	30.0 - 3.4	84	98.7	5.5	7.4	22	164.9	164.9	125.2	E
6	bmanK2PtCl4070599	K_2PtCl_4	01 6.0 12:00	1.380	15.0 - 4.3	32	93.9	2.6	7.3	14	165.2	165.2	127.3	ILS
7	bmanCl6K2Pt070599	K_2PtCl_6	01 6.0 24:00	1.380	15.0 - 4.0	39	90.2	3.3	6.9	18	165.1	165.1	128.2	
8	bmanEMP280799	$(C_2H_5HgO)HPO_2$	$10 \ 6.4 \ 00:22$	1.380	25.0 - 3.5	36	90.2	2.7	9.6	10	165.2	165.2	128.1	
9	bmanHg280799	$HgCl_2$	$05 \ 6.4 \ 00:25$	1.380	30.0 - 3.3	88	95.6	7.0	10.8	15	165.4	165.4	126.0	
10	bmanthimer 300799	Timerosal(Hg)	$10\ 6.4\ 24{:}00$	1.380	25.0 - 4.3	42	96.7	3.3	6.3	21	165.5	165.5	123.6	
11	bmanSmAc060899	$Sm(O_2C_2H_3)_3$	$10 \ 6.4 \ 00:50$	1.380	25.0 - 4.3	47	88.8	3.7	8.4	14	164.9	164.9	127.6	
12	bmanU300799	$UO_2(C_2H_3O_2)_2$	$10 \ 4.7 \ 12:00$	1.380	25.0 - 5.3	39	87.4	2.6	6.7	14	164.9	164.9	127.6	
13	bmanU040899	$UO_2(C_2H_3O_2)_2$	$01 \ 4.7 \ 23:00$	1.380	25.0 - 3.8	54	90.4	3.5	6.8	17	165.2	165.2	121.3	
14	bmanU050899	$UO_2(C_2H_3O_2)_2$	$05 \ 4.7 \ 00:30$	1.380	25.0 - 4.2	40	83.1	3.4	6.7	17	165.0	165.0	126.8	
15	bmanU061099	$UO_2(C_2H_3O_2)_2$	$10 \ 4.5 \ 24:00$	1.380	30.0 - 4.2	229	90.6	16.7	12.4	17	164.7	164.7	127.1	
16	bmanU131099	$UO_2(C_2H_3O_2)_2$	$05 \ 4.5 \ 14:00$	1.387	30.0 - 2.5	142	99.8	9.9	13.8	16	164.0	164.0	133.4	
17	bmanAuCN280999	$KAu(CN)_2$	10 4.5 18:30	1.387	30.0 - 3.4	149	95.7	11.6	19.2	12	165.1	165.1	125.5	

Tabela 3.1: Conjuntos de dados da β -manosidase coletados no LNLS em 1999.

 a A data de coleta está indicada no nome do conjunto.

 b A mosaicidade média dos cristais era de $\sim 0.9^\circ.$

 c Parâmetros do Soaking: concentração (mM), pH, tempo de banho (hh:mm).

 d Comprimento de onda usado para coleta.

 e Número de imagens coletadas.

 f Completeza

 g Multiplicidade

 h Média das razões entre intensidades e desvios padrões.

 $^{i} R_{sym}$ é definido no Apêndice A.

 j Indicações do correto eixo helicoidal serão discutidas adiante.

|71

Mesmo com a coleta de dados realizada a temperaturas criogênicas, a maioria dos cristais de β -manosidase mostraram-se inadequados para difração, ora porque praticamente não havia padrão de difração, ora por altíssima mosaicidade (até 2.0°). Mesmo os melhores cristais, escolhidos para coleta de dados, fornecem um padrão de difração fraco. Quer sejam nativos ou derivados, apresentam baixa intensidade de espalhamento e baixo limite de resolução. Estes são efeitos relacionados à glicosilação da proteína (seção 3.3.1 acima).

A baixa intensidade de espalhamento exigiu tempos de exposição de até 30 min. por imagem, demandando vários dias para coleta de um conjunto completo⁴. As conseqüências podem ser vistas na Tabela 3.1. Em primeiro lugar, vemos que o limite de resolução dos conjuntos é baixo, suficiente para obtenção de fases através de sinal isomorfo mas inadequado para faseamento por SAD, onde o sinal anômalo é bastante pequeno perto dos erros associados às medidas. Em segundo lugar, na maioria dos casos, o número de imagens coletadas ficou abaixo do desejado, já que o tempo disponível para cada coleta era evidentemente limitado. Este problema reflete-se claramente nos baixos valores de multiplicidade e $\langle I/\sigma(I) \rangle$, concomitantemente aos altos R_{sum} apresentados pelos conjuntos.

Outra questão evidente é a variação nos parâmetros de rede, o que caracteriza um forte não-isomorfismo entre os cristais da proteína, também um provável efeito da glicosilação e do alto conteúdo de solvente dos cristais (\sim 70 %) A Tabela 3.1 indica que há variação em todas as direções mas principalmente ao longo do eixo c da cela unitária, que chega a ser de mais de 10% entre alguns cristais. Notadamente, há variação dos parâmetros de rede entre os próprios nativos.

Testes conhecidos foram considerados na tentativa de distinguir derivados verdadeiros de cristais nativos, o que permitiria a interrupção da coleta em casos desfavoráveis, de modo a poupar tempo de linha. Estes testes são baseados na comparação das intensidades das reflexões, como os sugeridos, por ex., no manual do SCALEPACK (Otwinowski & Minor, 1997). Infelizmente, no caso da β -manosidase, o erro associado às reflexões (dada a baixa intensidade de difração) e o não-isomorfismo entre os cristais eram grandes

⁴ Para comparação, o tempo de coleta estava em torno de 3 min. por imagem para um cristal normal, nas mesmas condições utilizadas para coleta da β -manosidase.

o suficiente para impedir que estes métodos funcionassem, sendo necessário coletar todas as imagens de cada conjunto para avaliação posterior.

Outro problema, não visível a partir da Tabela 3.1, é que todos os cristais de β -manosidase, quer sejam nativos ou derivados, apresentam espalhamento anisotrópico (vide também págs. 79, 90). Este fenômeno faz com que determinadas regiões das imagens contenham pontos de difração até um limite de resolução bem mais baixo que outras.

Estas regiões das imagens formam uma região tridimensional do espaço recíproco que corresponde ao eixo c da cela unitária, justamente, o eixo onde há maior variação de comprimento e que, infelizmente, corresponte à direção de mais alta simetria do cristal. Isto significou um obstáculo a mais, dificultando a aquisição de dados em determinadas regiões do espaço recíproco, justamente as que exibiam maior simetria⁵.

A presença de espalhamento anômalo era esperada e foi levada em conta através do tratamento adequado dos pares I^+ e I^- (pares de Bijvoet; Apêndice A) no SCALEPACK (Otwinowski & Minor, 1997). Todos os conjuntos, incluindo os nativos (que por si só já contém Cd em sua condição de cristalização) foram primeiramente processados com esta opção, que foi removida se as diferenças entre I^+ e I^- medidos não fossem significativas perto do erro experimental.

Após a redução de dados, sítios de ligação de átomos pesados foram buscados através de mapas de Patterson, numa abordagem clássica, utilizando programas do CCP4 (CCP4, 1994; Winn *et al.*, 1997). Os conjuntos foram analisados contra o melhor nativo disponível na época (Tabela 3.1, cj. 3) usando SCALEIT e FHSCAL. RSPS foi usado na busca de sítios e

⁵ O que já era indicado pelo programa MARSTRATEGY (Klein, C. (1997), X-ray Research G. m. b. H., Germany, xray@mail.ppp.de) durante a otimização dos parâmetros da coleta de dados. Imagens referentes às regiões de maior simetria exibiam forte anisotropia, invariavelmente.

os parâmetros da substituição eram refinados no MLPHARE e VECREF e plotados no mapa de Patterson correspondente. As estatísticas após refinamento, qualidade dos mapas de Patterson e sua coerência com o refinamento das coordenadas dos sítios foram usados como critério básico de avaliação da qualidade dos conjuntos de derivados coletados.

Não-isomorfismo, assunto discutido acima (pág. 72), é uma função da resolução e, nesta primeira análise, limites de resolução tão baixos quanto 6Å foram usados durante os cálculos, numa tentativa de amenizar o problema. Sinal anômalo, quando identificável, foi utilizado para composição de mapas de Patterson anômalos. Infelizmente, os conjuntos mostraram-se de uso limitado no processo de obtenção de fases, ao menos pela metologia empregada. Além disso, não foi possível identificar com certeza em quais conjuntos havia sítios de ligação verdadeiros. Conforme comentado acima, a qualidade destes conjuntos não era a desejada.

O caminho seguido foi coletar novos conjuntos a partir de cristais nativos e derivados, no intuito de aumentar o limite de resolução dos dados e sua qualidade geral, focalizando esforços nos derivados que pareciam mais promissores, com potenciais sítios de ligação. Além disso, foram adotadas novas estratégias para busca e refinamento de sítios de átomos pesados, incluindo os métodos mais recentes da época, implementados nos novos pacotes SOLVE (Terwilliger & Berendzen, 1999), CNS (Brünger *et al.*, 1998) e SHARP (de La Fortelle & Bricogne, 1997).

3.3.5 Análise preliminar dos conjuntos coletados em 2000

A Tabela 3.2 resume as principais estatísticas dos conjuntos de dados coletados no ano de 2000.

											Dimensões da
Cj	Nome do conjunto a,b	Composto contendo	$Soaking^c$	λ^d	Resolução	Imgs^e	$\mathrm{Comp.}^f$	$Mult.^g$	$\overline{I/\sigma(I)}^h$	$R_{sym}{}^i$	cela unitária
		átomo pesado		(Å)	(Å)		(%)			(%)	a, b, c
Con	juntos do grupo ^j $P4_{1/3}$	$_{3}2_{1}2:$									
18	bm200600	nativo		1.545	20.00-2.00	373	91	11.2	16.0	11	166.2 166.2 121.1
19	bm110400	nativo		1.535	20.00 - 2.25	252	85	10.6	11.9	13	166.3 166.3 120.3
20	bmPb180100	$Pb(CH_3CO_2)_2$	$25 \ 4.5 \ 20s$	1.535	13.00-4.00	63	77	6.3	20.6	9	$165.4 \ 165.4 \ 125.6$
21	bmAg190100	$AgNO_3$	$20 \ 4.5 \ 20s$	1.535	15.00 - 3.30	128	96	8.9	16.4	12	166.1 166.1 119.3
22	bmEu040200	$Eu(NO_3)_3$	$20 \ 4.5 \ 20s$	1.535	13.00-2.90	123	95	6.1	18.1	8	$165.7 \ 165.7 \ 123.0$
23	$\rm bmHg270400$	$HgCl_2$	$05 \ 6.4 \ 01{:}20$	1.535	20.00 - 3.25	92	72	4.3	6.5	15	167.2 167.2 126.9
24	$\mathrm{bmPb010500}$	$Pb(CH_3CO_2)_2$	$25 \ 4.5 \ 35s$	1.535	20.00-3.10	214	93	3.3	8.8	12	$165.5 \ 165.5 \ 123.4$
25	bmPt6040500	K_2PtCl_6	$01 \ 6.0 \ 20:00$	1.535	20.00-3.10	129	85	2.5	6.0	14	166.1 166.1 119.6
Con	juntos do grupo $P2_12_1$	$2_1:$									
26	$\rm bmHg280400$	$HgCl_2$	$05 \ 6.4 \ 00{:}45$	1.535	20.00-3.00	79	62	2.1	6.7	15	166.0 165.3 127.3
27	bmSmCl250700	$SmCl_3$	$10 \ 6.4 \ 01:00$	1.545	15.00 - 3.20	63	70	1.5	5.3	12	$165.5 \ 166.1 \ 121.7$
28	bmEu080800	$Eu(NO_3)_3$	10 6.0 01:10	1.545	20.00-2.30	215	72	3.2	9.3	12	166.4 165.8 123.0
29	bmPt160800	K_2PtCl_6	$01 \ 6.0 \ 20:00$	1.545	20.00-3.40	60	51	1.8	6.4	12	166.3 166.3 117.3
30	bmSm170800	$Sm(O_2C_2H_3)_3 f$	$10 \ 6.4 \ 01:10$	1.545	20:00-2.80	50	62	1.3	1.8	8	166.1 166.1 121.4

Tabela 3.2: Conjuntos de dados da
 β -manosidase coletados no LNLS em 2000.

 a A data de coleta está indicada no nome do conjunto.

 b A mosaicidade média dos cristais era de $\sim 0.9^\circ.$

 c Parâmetros do Soaking: concentração (mM), pH, tempo de banho (hh:mm).

 d Comprimento de onda usado para coleta.

 e Número de imagens coletadas.

 f Completeza

 g Multiplicidade

 h Média das razões entre intensidades e desvios padrões.

 $^{i} R_{sym}$ é definido no Apêndice A.

 j Indicações do correto eixo helicoidal serão discutidas adiante.

3.3 Coletas de dados no LNLS

75

Durante este período, houve falta de proteína para cristalização, nenhuma remessa foi recebida, o que limitou a quantidade e qualidade dos cristais disponíveis para coleta de dados. Os colaboradores responsáveis pela purificação da proteína começaram a ter problemas tanto nos equipamentos quanto nas linhagens de fungos utilizadas. Apenas em Dezembro/2000 uma nova remessa de proteína foi enviada.

A falta de proteína teve conseqüências graves e foi enfrentada através do aproveitamento de cristais mais velhos, de menor qualidade, que haviam sido descartados em coletas anteriores. A última parcela de proteína, enviada ainda em 1999, foi também cristalizada. Para evitar maior deterioração com o passar do tempo e manter a qualidade dos cristais mais novos, tanto cristais nativos quanto potenciais derivados foram armazenados em nitrogênio líquido no início de 2000 e coletados posteriormente, ao longo do ano. Esta estratégia, a melhor alternativa disponível, por um lado conservou os cristais mas por outro impossibilitou o uso de melhores condições de derivatização que deveriam ter sido aprimoradas com o passar do tempo (daí a semelhança entre as condições de *soaking* mostradas na Tabela 3.2).

De forma geral, os novos conjuntos apresentam os mesmo problemas discutidos anteriormente mas alguns resultados importantes merecem destaque. O primeiro foi um aumento significativo na resolução dos dados nativos para 2.0 Å (Tabela 3.2, cj. bm200600), a partir de um cristal resfriado após poucos dias de crescimento⁶. Um bom número de imagens foi coletado a altas doses de radiação, num total de 5 dias (turnos de 24h) de coleta. O novo limite de resolução melhorou as chances de faseamento por SAD e abriu uma possibilidade para construção do modelo atômico através de métodos automáticos (pág. 50).

O segundo foi a descoberta de cristais em um novo grupo de simetria, $P2_12_12_1$. A estratégia de coleta foi feita para o grupo $P4_{1/3}2_12$, que tem maior simetria. Em conseqüência, os conjuntos do grupo $P2_12_12_1$ são incompletos e tem muito baixa multiplicidade. Havia indicação de incorreções já durante a coleta, porém, o problema era atribuído à já conhecida má qualidade dos cristais de β -manosidase e a coleta interrompida. A hipótese de um outro

⁶ A relação entre tempo de crescimento e limite de resolução foi comentada na pág. 68. No entanto, é provável que o aumento de resolução deva-se também a uma amostra de proteína de melhor qualidade, com menor grau de glicosilação e mais homogênea, portanto.

grupo de espaço só foi aventada posteriormente⁷. Os conjuntos deste grupo vieram a ser utilizados muito posteriormente (Capítulo 4).

Novas abordagens na busca e refinamento de sítios

Os conjuntos das Tabelas 3.1 e 3.2 foram analisados simultaneamente. Em busca de sinal isomorfo, tanto foram comparados nativos e derivados como os próprios derivados entre si⁸. Sinal anômalo, quando existente, foi também utilizado.

Uma nova busca por sítios de átomos pesados foi feita comparando resultados de dois programas. O CNS (Brünger *et al.*, 1998), que utiliza estratégias de busca baseada em mapas de Patterson, e o Shake&Bake (Weeks & Miller, 1999), uma implementação de algoritmos baseados em métodos diretos. (págs. 38 e 46). O refinamento dos parâmetros da substituição passou a ser feito principalmente através do programa SHARP (de La Fortelle & Bricogne, 1997), com novos sítios adicionados ao refinamento a partir da análise dos mapas de resíduos gerados pelo programa. No caso de o refinamento ser feito com o CNS (Brünger *et al.*, 1998), novos sítios são indicados pela coincidência entre os picos positivos nos mapas de gradiente e de diferença anômala feitos a partir das fases obtidas durante o refinamento por SAD. Em vários casos, foram obtidos resultados coerentes, onde novos sítios indicados pelos programas de refinamento concordavam com sítios previamente encontrados nos mapas de Patterson.

Técnicas de modificação de densidade eletrônica (pág. 47) foram utilizadas através dos programas DM e SOLOMON. Não havia indicação clara sobre o conteúdo da unidade assimétrica (ASU), variável chave para a com-

⁷ Agradeço ao amigo Dr. Carlos Basílio Pinheiro pelas discussões sobre o assunto, o grupo correto foi determinado após integração no grupo P1 e análise criteriosa das reflexões.

⁸ Pode-se obter informação de fase através de dois derivados que sejam isomorfos.

posição da máscara necessária à modificação de densidade eletrônica a para uma possível promediação através de simetria não-cristalográfica. Supôs-se a existência de NCS⁹ e operadores de NCS foram buscados principalmente através de relações geométricas entre os sítios de ligação de átomos pesados. Alguns programas utilizados para este fim foram NCS6D, IMP e COMA (Kleywegt *et al.*, 2001), da USF (http://alpha2.bmc.uu.se/~gerard) e ENVDER (Spraggon, 1999), que encontra máscaras cujos centros de gravidade possam ser usados para determinar a operação de translação.

Havia indicação de que os derivados compartilhassem os mesmos sítios de ligação (o que era esperado uma vez que as condições de cristalização da proteína incluem Cd, que está presente nos nativos e nos derivados) e, então, a busca de sítios exclusivamente através de sinal anômalo (SAD, pág. 45) mostrou-se a melhor alternativa, inclusive por ser o único meio de lidar com o grave problema de não-isomorfismo entre os cristais¹⁰. A Tabela 3.3 ilustra o compartilhamento de sítios entre o melhor nativo disponível na época (bm200600) e no derivado bmEu040200. Na comparação de coordenadas, deve-se levar em conta as diferenças esperadas tanto devido à própria substituição quanto ao não-isomorfismo entre os dois cristais. Dentre os sítios existentes, o que se identifica com mais clareza em ambos os cristais é o primeiro sítio do bmEu040200 que, no nativo (sítio #7), sofreu um deslocamento de 0.25 na origem da coordenada y (as operações de simetria deste grupo incluem deslocamentos de 1/4, 1/2 e 3/4), além de uma pequena diferença na coordenada z.

Embora alguns resultados fossem convincentes, os efeitos da anisotropia associada ao eixo c (discutida em maior detalhe adiante) já eram eviden-

⁹Indicação em contrário foi posteriormente encontrada através de estudos por SAXS, Capítulo 4.

¹⁰Os prováveis sítios assim encontrados eram posterioriormente usados para busca de sítios em outros derivados, através de técnicas de diferença de Fourier.

sítio ^a	bm	Eu0402	00^{b}	sítio ^{a}	$bm200600^{c}$					
	x	У	\mathbf{Z}		x	У	z			
1	0.294	0.116	0.069	1	0.032	0.175	0.042			
2	0.423	0.147	0.079	2	0.569	0.781	0.018			
3	0.381	0.356	0.000	3	0.794	0.519	0.082			
4	0.420	0.149	0.106	4	0.865	0.671	0.120			
5	0.432	0.139	0.103	5	0.084	0.538	0.044			
6	0.220	0.946	0.056	6	0.650	0.319	0.076			
7	0.403	0.164	0.086	7	0.290	0.362	0.112			
8	0.476	0.177	0.031	8	0.088	0.979	0.070			

Tabela 3.3: Tabela comparativa de coordenadas de sítios localizadas independentemente em dois conjuntos de dados utilizando sinal anômalo

 $^a {\rm Os}$ sítios são apresentados em coordenadas fracionárias (normalizadas pelo respectivo parâmetro).

 $^b\mathrm{Parâmetros}$ de rede: a=165.64 Å, b=165.64 Å, c=122.98 Å

 $^c\mathrm{Parâmetros}$ de rede: a=166.24 Å, b=166.24 Å, c=121.14 Å

tes. Os sítios refinados apareciam como elipsóides nos mapas de gradiente gerados pelo CNS, levando a crer, no primeiro momento, na existência de novos sítios, próximos uns dos outros. Entretanto, testes neste sentido não deram bons resultados e, uma vez que tais elipsóides sempre tinham o eixo maior orientado numa mesma direção (não por coincidência, o eixo c do cristal, parâmetro de rede de maior variação), foi adotada a hipótese de que se tratava de um artifício relacionado à anisotropia dos dados (vide também págs. 73, 90). Ruído advindo de erros grandes nos fatores de estrutura medidos e truncagem da série de Fourier possivelmente confundiam ainda mais a situação, dificultando o refinamento e a determinação de novos sítios.

Apesar de todos os esforços, a análise dos conjuntos das Tabelas 3.1 e 3.2 não levou a um mapa de densidade eletrônica adequado para construção de um modelo para a β -manosidase.

3.4 Resultados do estágio no YSBL

Em busca de uma solução para o problema, um programa de estágio de 7 meses foi realizado no York Structural Biology Laboratory (YSBL), Universidade de York, Inglaterra, sob supervisão do Dr. Garib Murshudov. Entre os objetivos do estágio, estavam a obtenção de um conjunto de fases confiável e de um modelo inicial para a β -manosidase de *T. reesei*. Três linhas de trabalho foram adotadas:

- nova redução de dados foi feita para quase todos os conjuntos de dados coletados no LNLS, que foram novamente analisados em busca de dados úteis na obtenção de fases. O sinal anômalo exibido pelo Cd presente nos cristais nativos abria possibilidade para obtenção de fases por SAD, evitando problemas de não-isomorfismo entre os conjuntos, conforme já comentado. Esta nova análise também permitiu estabelecer claramente quais eram os principais problemas apresentados pelos dados;
- as análises feitas no Brasil já indicavam que os dados tinham problemas que, em boa medida eram devidos ao uso de fonte de Raios X inadequada para os cristais de β-manosidase, que difratavam com baixa intensidade. Dessa forma, novos dados foram coletados no ESRF -Grenoble/França e analisados;
- busca de uma solução para o problema através de Substituição Molecular, mesmo tendo em vista a inexistência de um modelo adequado para esse fim.

A primeira e segunda linhas de trabalho serão comentadas a seguir. Quanto à última, embora a seqüência completa de β -manosidase de *T. ree*sei não estivesse disponível, uma parte já havia sido seqüenciada (seção 3.1, pág. 59) e, como esperado, mostrava boa identidade com a mesma proteína de A. Niger, cuja seqüência acabou sendo usada para buscas de modelos no PDB (Protein Data Bank, http://www.rcsb.org/pdb). O melhor alinhamento foi conseguido para a β -galactosidase de E. coli, modelo 1BGL do PDB. Uma porção de aproximadamente 200 aminoácidos mostrou uma identidade em torno de 20% com a β -manosidase. Claramente este resultado não era satisfatório. Mesmo assim, foram feitas muitas tentativas através do programa MOLREP (CCP4, 1994; Winn et al., 1997), empregando estratégias bem conhecidas, que incluiam modelo de polialaninas, remoção de loops e uso de sub-domínios¹¹. Não foram encontradas soluções por Substituição Molecular, conforme era esperado dada a baixa identidade entre as sequências.

3.4.1 Diagnóstico para os dados coletados no LNLS

Foi feita uma análise mais profunda dos conjuntos coletados no LNLS com os objetivos de buscar melhores fases e entender porque as tentativas de faseamento anteriores falharam. Isto incluiu a inspeção, integração e escalonamento das imagens da maioria dos conjuntos Encontramos poucos conjuntos potencialmente úteis na obtenção de fases. Conforme já mencionado, os cristais de β -manosidase, de forma geral, espalham com baixa intensidade e a baixa resolução (o que obriga o uso de altas doses de radiação), apresentam alta mosaicidade, espalhamento anisotrópico (em maior ou menor medida) e grave problema de não-isomorfismo. Estes fatores impedem que medidas mais acuradas e precisas sejam obtidas. No entanto, esforços para remediar a situação foram feitos durante integração e escalonamento dos conjuntos, procurando ajustar todos os parâmetros da melhor maneira possível.

¹¹ após a coleta de dados de SAXS e obtenção do modelo de baixa resolução para a β -manosidase (Capítulo 4), procedimentos de Substituição Molecular foram repetidos à exaustão (seção 4.1.6, pág. 118) usando os programas AMORE (Navaza, 1994; Winn *et al.*, 1997), BEAST (Read, 2001; Winn *et al.*, 1997) e EPMR (Kissinger *et al.*, 1999).

Os programas escolhidos para integração e escalonamento foram MOS-FLM e SCALA, respectivamente, ambos do pacote CCP4 (CCP4, 1994; Winn *et al.*, 1997), que mostraram-se bastante flexíveis e levaram a uma maior profundidade na análise feita. Em particular, MOSFLM diagnostica, em seu *logfile*, possíveis problemas com o detector e com o aparato experimental de forma geral. SCALA não apenas permite uma análise profunda das estatísticas após a redução dos dados, como também ajusta de forma mais efeciente os erros das intensidades, o que pode ser crucial nos passos seguintes. Os conjuntos das Tabelas 3.1 e 3.2 têm vários problemas em comum. A seguir apresentamos dois conjuntos representativos da situação.

bm200600

Este conjunto nativo é constituído por 370 imagens e foi o mais usado para tentativas de faseamento por SAD. Há picos de difração até resolução de 2 Å. O tempo total de coleta foi de cerca de 5 dias (turnos de 24h). O cristal utilizado foi um dos melhores obtidos desde o início do projeto. Na Figura 3.2, apresentamos duas imagens do conjunto.



(a) Uma imagem do conjunto considerada de qualidade boa.

(b) A imagem, pertencente ao mesmo conjunto, corresponde a uma região ruim do cristal, o qual exibe espalhamento anisotrópico, evidenciado pelos *spots* "borrados". Esta região corresponde à direção paralela ao eixo helicoidal de ordem 4.

Figura 3.2: Imagens do conjunto bm200600, coletado a partir de um único cristal. Círculos foram desenhados a resoluções de 2.5, 3.8, 7.6 Å.

Mesmo exibindo difração a boa resolução, alta dose de Raios X foi necessária durante a coleta. O tempo de exposição foi em torno de 15 minutos por imagem, ou seja, foram coletadas 4 imagens/hora aproximadamente, tempo suficiente para observar um decaimento da intensidade do feixe no LNLS. Na tentativa de corrigir o problema, a dose de radiação utilizada foi constantemente reajustada. Isto impediu um escalonamento suave e constituiu uma influência nefasta para a soma das reflexões parciais. Se o feixe não é estável, regiões do cristal gravadas numa mesma imagem mas expostas a diferentes doses de radiação, resultam em reflexões cujas intensidades apresentam erros não modelados pelos programas de integração/escalonamento, comprometendo os dados de forma irremediável. O resultado está exposto na Figura 3.3, que apresenta o escalonamento final das imagens e na Tabela 3.4, que mostra os R_{merge} (Apêndice A) contra resolução.

Para coleta dos conjuntos apresentados nas Tabelas 3.1 e 3.2, buscou-se o melhor compromisso entre passo de rotação do cristal, limite de resolução, mosaicidade, parâmetros de rede e tempo de coleta disponível. Nos conjuntos de maior resolução, idealmente, coletas para alta e baixa resolução deveriam ter sido feitas em separado, o que era inviável pois demandaria enorme tempo de coleta, indisponível. O resultado, em alguns casos, foi o *overlap* de muitas reflexões a mais alta resolução e alguns *overloads* a baixa resolução. Dada a alta mosaicidade dos cristais ($\sim 1.0^{\circ}$) e as dimensões da cela unitária, a grande maioria dos conjuntos apresenta apenas reflexões parciais. Notamos, ainda, que, em vários conjuntos, um *backstop* menor teria permitido medir reflexões a mais baixa resolução. Também, a integração usando MOSFLM apontou para possíveis problemas no detector e de mau alinhamento do cristal no feixe em mais de um conjunto.



Figura 3.3: Escalonamento das imagens do conjunto bm200600. Escalas resultantes do escalonamento de 354 imagens (16 imagens foram excluídas por má qualidade) coletadas durante cinco dias no LNLS, com tempo de exposição de ~ 15 minutos por imagem. Devido ao decaimento da intensidade do feixe de Raios X durante a coleta, a dose de radiação foi reajustada constantemente, de modo que a integração e escalonamento foram feitas dividindo o conjunto em 19 pequenos blocos de imagens. O comportamento esperado para este gráfico seria uma reta, indicando que as escalas variam suavemente. No caso, as escalas variam bruscamente e em grande amplitude, o que compromete de forma irremediável a qualidade final do conjunto.

 a Indicações do correto eixo helicoidal são discutidas no texto.
Nome do con	junto:	bm200600								
Grupo de esp	eaço ^a :	$P4_{1}2_{1}2$								
Dimensões da	a cela unitária:	a=165.78 Å, b=165.78 Å, c=121.57 Å								
Número de in	nagens:	360	360							
Resolução	Complet	AnomCmpl	AnomFrc	Mlplcty	$\overline{I/\sigma(I)}$	R_{sym}	R _{meas}	Ranom		
	(%)	(%)	(%)							
27.25 - 6.31	92.1	80.2	84.6	9.0	9.9	0.053	0.058	0.036		
6.31 - 4.46	95.5	88.9	91.9	10.3	9.8	0.059	0.064	0.036		
4.46 - 3.64	96.8	92.3	94.5	11.0	8.8	0.069	0.075	0.039		
3.64 - 3.15	97.7	94.9	96.5	11.9	7.5	0.090	0.098	0.044		
3.15 - 2.82	97.9	94.5	95.6	12.4	5.3	0.135	0.146	0.061		
2.82 - 2.57	97.3	94.0	94.8	12.5	3.8	0.192	0.209	0.083		
2.57 - 2.38	96.3	93.4	93.9	12.4	2.6	0.289	0.314	0.123		
2.38 - 2.23	96.0	94.0	94.4	12.1	1.8	0.412	0.450	0.167		
2.23 - 2.10	96.8	94.9	95.2	11.6	1.3	0.573	0.627	0.229		
2.10 - 1.99	93.5	92.0	96.2	11.3	0.9	0.852	0.936	0.336		

Tabela 3.4: Estatísticas por resolução para o conjunto bm200600. R_{sym} é relativamente alto já a baixa resolução.

Legenda:	
Compl	completeza;
Mlplcty	multiplicidade;
AnomCmpl	porcentagem de reflexões acêntricas medidas;
AnomFrc	porcentagem das reflexões acêntricas para as quais foi medido sinal anômalo;
R_{sym}	R_{merge} definido no Apêndice A.
R_{meas}	R factor pesado pela multiplicidade
R_{anom}	Sum —Mn(I^+) - Mn(I^-)— / Sum (Mn(I^+) + Mn(I^-))

No caso específico do conjunto bm200600, observou-se que existe uma ausência sistemática de reflexões ao longo do eixo l, que não foram medidas, como se vê na seção 0kl do espaço recíproco, feita com o programa HKLVIEW (CCP4, 1994; Winn *et al.*, 1997), exposta na Figura 3.4, apesar de o cristal ter sido realinhado mais de uma vez durante a coleta.



(a) Seção hk0, onde é evidente a simetria de ordem 4.

(b) Seção 0kl. Note-se a sistemática falta de reflexões de baixa resolução ao longo do eixo l.

Figura 3.4: Conjunto bm200600: seções do espaço recíproco com reflexões estendendo-se até 2.5 Å.

bmEu040200

Este conjunto tem 113 imagens e foi coletado durante 2 dias. Na época de coleta foi sugerido o uso da técnica de *quick cryo soaking*, que vinha sendo desenvolvida no grupo, esperando-se que isto pudesse levar a derivados mais isomorfos. Assim, o tempo de banho em $Eu(NO_3)_3$ foi de 20 segundos mas a concentração do composto contendo o átomo pesado foi muito baixa, ao contrário do que prevê o método de *quick cryo soaking*. Este erro foi repetido em diversos outros derivados que, provavelmente não contém outro elemento que não Cd.

Na Figura 3.5 expomos uma das fotos da coleta. Nota-se um forte espalhamento anisotrópico. Numa tentativa de melhorar o processo de integração, diferentes limites de resolução para cada direção h, k , l foram usados no MOSFLM, de modo a remover do processamento reflexões inexistentes.

Novas tentativas de faseamento por SAD

A exposição acima permitiu diagnosticar as causas das dificuldades enfrentadas na obtenção de fases de boa qualidade. Mesmo com dados inadequados, prosseguiram os esforços para faseamento da β -manosidase. O trabalho levou em conta os resultados que já haviam sido obtidos. Um total de 15 sítios haviam sido encontrados. Como um método de validação, os mesmos sítios foram localizados, de forma independente, através de outro programa. Assim, 4 dentre os 15 sítios foram usados como ponto de partida no programa SOLVE (Terwilliger & Berendzen, 1999) que, a partir deles, foi capaz de encontrar todos os outros sítios esperados, o que leva crer que, de fato, os sítios são verdadeiros. O problema de anisotropia dos sítios (vide também págs. 73, 79) foi tratado através do refinamento anisotrópico de fatores de temperatura no programa SHARP (de La Fortelle & Bricogne, 1997).



Figura 3.5: Uma das imagens do conjunto bmEu040200, com círculos desenhados a resoluções de 2.8, 3.8, 5.6, 11.3 Å. Observa-se um forte espalhamento anisotrópico. Para evitar a inclusão de reflexões inexistentes durante a integração, foram definidas diferentes resoluções para cada direção no espaço recíproco: 3 Å para os eixos h, k e 5 Å para o eixo l, direção na qual o cristal espalha mal.

As fases obtidas do sinal anômalo de um único conjunto de reflexões apresenta uma distribuição bi-modal, ou seja, há duas fases prováveis para cada fator de estrutura (pág. 45). A adição de sinal isomorfo ou de outro comprimento de onda é capaz de resolver a ambigüidade, que também pode ser resolvida posteriormente através dos métodos de modificação de densidade eletrônica. No nosso caso, após refinamento com SHARP (de La Fortelle & Bricogne, 1997), DM (CCP4, 1994; Winn *et al.*, 1997)e RESOLVE (Terwilliger, 2000) foram usados para resolver a ambigüidade intrínseca de faseamento por SAD e para melhoramento de fases. Ambos levaram a mapas de densidade de qualidade equivalente no caso de β -manosidase¹².

Os fatores de estruturas dos vários conjuntos foram avaliados em busca de diferenças isomorfas que pudessem melhorar as fases obtidas por SAD. Mesmo com toda insistência, não foi possível obter melhora nos mapas de densidade.

Apesar do grande empenho, que também incluiu nova busca por diferenças isomorfas (apesar do já conhecido não-isomorfismo entre os cristais) que pudessem melhorar as fases obtidas por SAD, os mapas obtidos não são interpretáveis, impossibilitando a construção de um modelo atômico para a β -manosidase de *Trichoderma reesei*. Um exemplo da qualidade dos mapas obtidos está na Figura 3.6, que mostra um mapa obtido por SAD a partir de dados do conjunto bm200600. Refinamento anisotrópico de 15 sítios de Cd foi feito usando-se SHARP (de La Fortelle & Bricogne, 1997). DM (CCP4, 1994; Winn *et al.*, 1997) foi utilizado para modificação de densidade eletrônica, com um conteúdo de solvente de 69%. A Figura de Mérito correspondente é dada na Tabela 3.5. Observamos o estiramento evidente do mapa ao longo da direção z¹³.

¹² RESOLVE utiliza uma metodologia mais eficiente que DM e que, em casos normais, deve levar a resultados melhores. DM ficou obsoleto após o surgimento de RESOLVE.

 $^{^{13}}$ O que está relacionado ao espalhamento anisotrópico, conforme discutido neste e no próximo Capítulo.



(a) O plano da figura é paralelo ao eixo c da cela unitária.



(b) O plano da figura é perpendicular ao eixo c da cela unitária.

(c) Um possível fragmento de hélice α .

Figura 3.6: Mapa de densidade eletrônica obtido a resolução de 2.5 Å e desenhado a 1.25 σ da média do mapa. A mesma porção do mapa é exposta em (a) e (b), mostrando o estiramento evidente na direção c, relacionado à anisotropia dos dados. Em (c), um possível fragmento de hélice α foi identificado pelo programa Quanta (http://www.msi.com). No entanto, não há continuidade suficiente para construção de um modelo para a proteína.

Resolução	< FOM >	< FOM >
Å	(SHARP)	(DM)
20.00 - 11.91	0.461	0.544
11.91 - 8.95	0.450	0.711
8.95-7.47	0.438	0.765
7.47 - 6.54	0.425	0.721
6.54-5.89	0.411	0.754
5.89-5.40	0.410	0.756
5.40 - 5.02	0.407	0.769
5.02 - 4.71	0.409	0.789
4.71 - 4.45	0.391	0.796
4.45 - 4.22	0.365	0.778
4.22 - 4.03	0.337	0.760
4.03 - 3.86	0.325	0.748
3.86 - 3.72	0.320	0.764
3.72 - 3.58	0.298	0.721
3.58 - 3.47	0.293	0.599
3.47 - 3.36	0.280	0.666
3.36 - 3.26	0.263	0.664
3.26 - 3.17	0.250	0.670
3.17 - 3.09	0.240	0.655
3.09 - 3.01	0.237	0.664
3.01 - 2.94	0.237	0.665
2.94 - 2.87	0.232	0.655
2.87 - 2.81	0.227	0.645
2.81 - 2.75	0.233	0.668
2.75 - 2.69	0.218	0.660
2.69 - 2.64	0.219	0.662

Tabela 3.5: Figuras de mérito médias $\langle FOM \rangle$ para o mapa da Figura 3.6, antes (SHARP) e depois (DM) da modificação de densidade eletrônica.

3.4.2 Coletas de dados no ESRF

Esta foi a segunda estratégia seguida durante o estágio, que iniciou-se por novos experimentos de cristalização. Conforme mencionado acima, no ano de 2000, novas remessas de proteína foram recebidas apenas no mês de Dezembro e, logo no início do estágio, em Janeiro/2001, as primeiras placas de cristalização de β -manosidase foram preparadas mas cristais não cresceram como era esperado. Foram tomadas as seguintes providências:

- uso dos mesmos produtos químicos anteriormente utilizados no LNLS;
- uso de aditivos (isopropanol, etc.) e manose, substrato da β manosidase, em co-cristalização;
- *seeding* (no caso de gotas que haviam apresentado algum material cristalino suficientemente bom para uso desta técnica);
- novo screening a procura de novas condições de cristalização.

Mesmo após isto, a cristalização não teve sucesso. Verificamos, então, que havia contaminação em todas as amostras, provavelmente por problemas durante a purificação ou a uma posterior degradação da proteína, o que é exemplificado no gel da Figura 3.7. Novas remessas foram solicitadas mas apresentaram o mesmo problema. Tentativas de purificação¹⁴ foram feitas mas os contaminantes continuavam presentes após as colunas de purificação, dando a entender que algum processo de degradação da proteína poderia estar ocorrendo (como medida preventiva, chegou-se a adicionar um inibidor de protease às amostras).

Apesar do problema de contaminação das amostras, os experimentos de cristalização não foram interrompidos e, em Junho/2001, último mês previsto para o estágio, uma das amostras, da qual havia apenas 1.2 mg, levou,

¹⁴Agradecimentos ao Dr. Simon Chernock, na época, no YSBL.



Figura 3.7: SDS-PAGE gel: amostras de β -manosidase contaminadas As pistas de interesse são as 3 últimas (da esquerda para a direita); a antepenúltima pista é um marcador de peso molecular. As duas últimas pistas são de β manosidase e diferem apenas na concentração usada para fazer o gel. Vê-se a presença de contaminantes de pesos moleculares próximos e distantes da β -manosidase, cujo peso molecular é 105 KDa.

providencialmente, a alguns cristais de qualidade razoável que cresceram em condições muito parecidas (mas relativamente distantes das esperadas). A partir deles, dois conjuntos nativos foram coletados pelo autor em viagem ao ESRF/Grenoble - França (linha ID14-2), por intermédio do YSBL.

Análise dos novos conjuntos

As estatísticas para os novos conjuntos são apresentadas na Tabela 3.6.

Nome do con	junto:	bm4						
Grupo de esp	aço ^a :	$P4_{1}2_{1}2$						
Dimensões da	cela unitária:	a=165.78 Å, $b=165.78$ Å, $c=121.57$ Å						
Número de in	nagens:	936						
Resolução	Complet	AnomCmpl	AnomFrc	Mlplcty	$\overline{I/\sigma(I)}$	R_{sym}	R_{meas}	Ranom
	(%)	(%)	(%)					
47.00 - 6.61	99.5	99.7	99.9	23.9	12.4	0.044	0.046	0.016
6.61 - 4.67	100.0	100.0	100.0	25.9	11.0	0.058	0.060	0.014
4.67 - 3.81	100.0	100.0	100.0	26.1	9.5	0.071	0.074	0.014
3.81 - 3.30	100.0	100.0	100.0	25.1	6.4	0.113	0.117	0.020
3.30 - 2.95	100.0	100.0	100.0	22.4	3.5	0.206	0.215	0.041
2.95 - 2.70	99.2	98.6	99.2	18.4	2.1	0.351	0.370	0.086
2.70 - 2.50	71.7	64.7	64.7 89.3 11.4 1.4 0.524 0.567 0.193					
Nome do conjunto: bm1								
Grupo de esp	aço:	$P2_{1}2_{1}2_{1}$						
Dimensões da	a cela unitária:	a=165.16 Å,	b=165.63 Å	, $c=123.56$	βÅ			
Número de in	nagens:	840						
Resolução	Complet	AnomCmpl	AnomFrc	Mlplcty	$\overline{I/\sigma(I)}$	R_{sym}	R_{meas}	Ranom
	(%)	(%)	(%)					
47.00 - 7.90	99.8	99.6	99.6	10.4	9.5	0.057	0.063	0.022
7.90-5.59	100.0	100.0	100.0	11.2	9.6	0.065	0.072	0.023
5.59-4.56	100.0	100.0	100.0	11.3	9.2	0.068	0.074	0.021
4.56 - 3.95	100.0	100.0	100.0	11.4	8.5	0.075	0.082	0.022
3.95 - 3.53	100.0	100.0	100.0	11.4	7.0	0.097	0.106	0.026
3.53 - 3.23	100.0	100.0	100.0	11.4	5.1	0.136	0.149	0.035
3.23 - 2.99	100.0	100.0	100.0	11.4	3.0	0.228	0.250	0.055
2.99 - 2.79	100.0	100.0	100.0	11.0	2.0	0.346	0.380	0.083
2.79 - 2.63	100.0	100.0	100.0	9.5	1.4	0.474	0.531	0.135
2.63 - 2.50	98.0	93.3	95.0	5.8	1.1	0.564	0.669	0.265

Tabela 3.6: Estatísticas contra resolução para os conjuntos coletados no ESRF. O comprimento de onda da coleta: $\lambda=1$ Å. Mesmo apesar do uso de melhor instrumentação, R_{sym} ainda é relativamente alto a baixa resolução.

 a Indicações do correto eixo helicoidal são discutidas no texto.

Legenda:	
Compl	completeza;
Mlplcty	multiplicidade;
AnomCmpl	porcentagem de reflexões acêntricas medidas;
AnomFrc	porcentagem das reflexões acêntricas para as quais foi medido sinal anômalo;
R_{sym}	R_{merge} definido no Apêndice A.
R_{meas}	R factor pesado pela multiplicidade
R_{anom}	Sum —Mn(I^+) - Mn(I^-)— / Sum (Mn(I^+) + Mn(I^-))

Um deles pertence ao grupo $P4_12_12$ (bm4) e o outro ao grupo $P2_12_12_1$ (bm1), com praticamente as mesmas dimensões da cela unitária. O conjunto do grupo $P2_12_12_1$ possui características intrigantes, exibindo parte das ausências sistemáticas do grupo $P4_12_12$. A Figura 3.8 mostra uma seção do espaço recíproco onde é possível ver que o conjunto quase possui um eixo cristalográfico de ordem 4^{15} , também confirmado através do programa ALMN (CCP4, 1994; Winn *et al.*, 1997). Ainda que tal simetria manifeste-se também nas ausências sistemáticas, é impossível processar o conjunto como sendo do grupo $P4_12_12$, as estatísticas tornam-se inaceitáveis, com rejeição de grande parte das reflexões.

Apesar de os cristais apresentarem os problemas já conhecidos, incluindo o espalhamento anisotrópico ao longo da direção do eixo helicoidal de ordem 4 (paralelo à direção z, com o menor parâmetro de cela), a qualidade geral de ambos é muito superior aos dos conjuntos das Tabelas 3.1 e 3.2, graças ao uso de melhor síncrotron e instrumentação. Não foi possível utilizar o comprimento de onda desejado, de modo que ambos os conjuntos mostraram sinal anômalo mais fraco do que o observado no melhor nativo coletado no LNLS (Tabela 3.2, cj. bm200600).

Determinação do eixo helicoidal no grupo tetragonal

O conjunto bm1 apresenta estatísticas inferiores às do bm4, o que é devido à menor simetria do grupo $P2_12_12_1$, ao qual pertence, que resulta em menor multiplicidade (para um mesmo número de imagens coletadas) e maior erro nas medidas. Apesar de grande esforço, não foi possível obter nenhuma informação de fase para este conjunto, já que as buscas por sítios de ligação de Cd não tiveram sucesso. Felizmente, entretanto, tanto este quanto os conjuntos do mesmo grupo incluídos na Tabela 3.2 (comentários à pág. 76)

¹⁵ o que explica a dificuldade enfrentada para reconhecer os primeiros cristais do grupo ortorrômbico, conforme comentários à pág. 76.



Figura 3.8: Seção hk0 para o conjunto do grupo $P2_12_12_1$ (bm1) coletado no ESRF/Grenoble-França, mostrando a existência de simetria de ordem 4. Resolução dos círculos: 2.5, 3.3, 5.0, 10.0 Å.

viriam a ser utilizados posteriormente (Capítulo 4).

No caso do conjunto bm4, do grupo $P4_12_12$, sítios foram encontrados (poucos eram diferentes dos já conhecidos) e fases foram obtidas. O refinamento e faseamento, mais estáveis que de costume, foram feitos usando-se SHARP.

A qualidade dos mapas foi avaliada após modificação de densidade eletrônica através do programa DM. Foi neste ponto que conseguimos esclarecer o correto grupo de espaço da proteína. Funcionando no modo "auto", o programa DM parava após apenas 2 ciclos quando impúnhamos o grupo $P4_32_12$ mas ia adiante por muitos ciclos (~ 20), quando utilizávamos o grupo $P4_12_12$. Já havia indícios anteriores de que este seria o grupo correto mas até aquela época não houvera um resultado tão claro como o encontrado.

Resultados conclusivos

As fases obtidas por SAD, após melhoramento, mostraram-se bastante úteis. Sua utilização para composição de mapas de Fourier de diferença isomorfa e anômala provou ser uma ferramenta confiável para buscar dentre os mais de 30 conjuntos coletados, aqueles que seriam potencialmente úteis para melhorar as fases através de SIRAS (pág. 44), o que levou a resultados conclusivos.

Os potenciais derivados apresentaram praticamente os mesmos sítios de ligação que os sítios de Cd dos cristais nativos, ou seja, não é possivel saber se houve substituição do Cd por outro átomo pesado e se existe de fato um derivado que seja verdadeiro. Por outro lado, foram encontrados sítios de muito baixa ocupância no nativo bm200600 (Tabela 3.2) que não estavam presentes no novo nativo do grupo $P4_12_12$ (Tabela 3.6, conjunto bm4). Também foram encontrados uns poucos sítios diferentes entre alguns conjuntos.

Utilizamos toda informação disponível, incluindo os novos conjuntos que, como mencionado, também apresentavam problemas de anisotropia e baixa intensidade de difração. Inúmeras estratégias e diferentes metodologias foram empregadas. Mesmo apesar de todos este esforços, os mapas de densidade eletrônica continuaram não-interpretáveis e com a mesma aparência daquele mostrado na Figura 3.6.

Capítulo 4

β -manosidase: estudos a baixa resolução

Um modelo de baixa resolução para a β -manosidase foi obtido através de estudos em solução. Dados de espalhamento a baixo ângulo - SAXS, *Small-Angle X-ray Scattering*) - foram coletados no SRS, em Daresbury, Inglaterra, durante o estágio realizado em 2001. Os resultados destes estudos foram publicados no artigo

Aparicio, R., Fischer, H., Scott, D. J., Verschueren, K. H., Kulminskaya, A. A., Eneiskaya, E. V., Neustroev, K. N., Craievich, A. F., Golubev, A. M. and Polikarpov, I. (2002). Structural insights into the β -mannosidase from *T. reesei* obtained by synchrotron small-angle X-ray solution scattering enhanced by X-ray crystallography. *Biochemistry*, **41(30)**,9370–9375.

Neste trabalho, muito bem recebido pelos *referees* da revista, dados de cristalografia foram utilizados para aumentar o conteúdo informacional dos dados de SAXS, resultando num modelo mais acurado e detalhado. A obtenção de um modelo de baixa resolução foi de grande importância, não

apenas do ponto de vista biológico mas também porque esclareceu as causas dos problemas apresentados pelos cristais de β -manosidase, descritos nos capítulos anteriores. Estes aspectos serão comentados em maior detalhe nas próximas seções.

A seguir, anexamos o artigo acima referido.

Biochemistry 2002, 41, 9370-9375

Structural Insights into the β -Mannosidase from *T. reesei* Obtained by Synchrotron Small-Angle X-ray Solution Scattering Enhanced by X-ray Crystallography[†]

Ricardo Aparicio,[‡] Hannes Fischer,[§] David J. Scott,[#] Koen H. G. Verschueren,^{⊥,§} Anna A. Kulminskaya,^{||} Elena V. Eneiskaya,^{||} Kirill N. Neustroev,^{||} Aldo Felix Craievich,[§] Alexander M. Golubev,^{||} and Igor Polikarpov^{*,&}

Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Campinas, SP, Brazil, Instituto de Física, Universidade de São Paulo, São Paulo, SP, Brazil, Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol, United Kingdom, Structural Biology Laboratory, Department of Chemistry, University of York,

York, United Kingdom, Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg, Russia, and Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil

Received March 14, 2002; Revised Manuscript Received May 16, 2002

ABSTRACT: A molecular envelope of the β -mannosidase from *Trichoderma reesei* has been obtained by combined use of solution small-angle X-ray scattering (SAXS) and protein crystallography. Crystallographic data at 4 Å resolution have been used to enhance informational content of the SAXS data and to obtain an independent, more detailed protein shape. The phased molecular replacement technique using a low resolution SAXS model, building, and refinement of a free atom model has been employed successfully. The SAXS and crystallographic free atom models exhibit a similar globular form and were used to assess available crystallographic models of glycosyl hydrolases. The structure of the β -galactosidase, a member of a family 2, clan GHA glycosyl hydrolases, shows an excellent fit to the experimental molecular envelope and distance distribution function of the β -mannosidase, indicating gross similarities in their threedimensional structures. The secondary structure of β -mannosidase quantified by circular dichroism measurements is in a good agreement with that of β -galactosidase. We show that a comparison of distance distribution functions in combination with 1D and 2D sequence alignment techniques was able to restrict the number of possible structurally homologous proteins. The method could be applied as a general method in structural genomics and related fields once protein solution scattering data are available.

 β -mannosidase (β -D-mannoside mannohydrolase, EC 3.2.1.25) is an exoglycosidase that catalyzes hydrolysis of nonreducing residues of β -D-mannose in β -D-mannosides. The enzyme participates in several biological processes, of which lysosomal degradation of N-glycoproteins is probably the most important one. Glycoprotein degradation occurs predominantly within cell lysosomes by the sequential action of endo- and exoglycosidases. β -Mannosidase cleaves mannosidic bonds in the core part of N-linked glycans attached to glycoproteins (1). In humans and ruminants, a deficiency of β -mannosidase leads to an autosomal recessive inherited lysosomal storage disease (2) called β -mannosidosis. In the

case of man, clinical manifestations are heterogeneous and include mental retardation, peripheral neuropathy, and skeletal abnormalities (3). Fungi and bacteria that are able to degrade hemi cellulose also secrete β -mannosidase. Fungal and bacterial β -mannosidases hydrolyze $\beta(1-4)$ -D-mannosyl groups from manno-oligosaccharides and mannose-containing glycopeptides that are produced from the hemi cellulose pulp by endoenzymes (4). In seeds that have galactomannans as storage carbohydrates, the enzyme converts mannooligosaccharides to monosaccharides (5).

 β -mannosidase from *Trichoderma reesei* is an extracellularly secreted 105 kDa glycoprotein (6). Kinetic studies proved the existence in *T. reesei* β -mannosidase of a noncatalytic galactomannan-binding site (6). Although the precise role of this site is not clear, this result might suggest that the β -mannosidase fold is comprised of more than one domain. To provide experimental structural information about *T. reesei* β -mannosidase, X-ray crystallographic (7), circular dichroism (CD), and small-angle X-ray scattering (SAXS) studies were initiated. In work presented here, a lowresolution envelope of the protein was retrieved by an ab initio procedure from the synchrotron SAXS¹ data. The envelope was further improved with the aid of higher

[†] This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, via Grants 98/06761-1, 99/09471-7, 00/03387-4, and Grant 00-04-48878 of Russian Foundation of Basic Research.

^{*} To whom correspondence should be addressed: Instituto de Física de São Carlos, Universidade de São Paulo, CP 369, 13560-970, São Carlos, SP, Brazil. Phone: +55 16 273 9874. Fax: +55 16 273 9881. E-mail: ipolikarpov@if.sc.usp.br.

^{*} Universidade Estadual de Campinas.

[§] Instituto de Física, Universidade de São Paulo.

[#] University of Bristol.

[⊥] University of York.

[&] Instituto de Física de São Carlos, Universidade de São Paulo.

^{\$} Current address: Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

^{II} Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg, 188300, Russia.

¹ Abbreviations: SAXS, small-angle X-ray scattering; CD, circular dichroism; 1D, one-dimensional; 2D, two-dimensional; SAD, single wavelength anomalous dispersion; ASU, asymmetric unit; NCS, noncrystallographic symmetry.

Low Resolution Model of β -Mannosidase from T. reesei

resolution data derived from the crystallographic experiments. Information obtained from solution X-ray scattering experiments is limited by the inherent low resolution of this technique as compared to X-ray crystallographic data. The number of Shannon channels (8, 9) $N_s = D_{\text{max}} S_{\text{max}} / \pi$ in the available SAXS experimental data is equal to 13. To enhance the informational content and thus improve the accuracy of the molecular envelope, we employed X-ray crystallographic data up to 4.0 Å resolution for which the equivalent number of Shannon channels would be equal to 60. Theoretical distance distribution functions were computed from the available crystallographic models of glycosyl hydrolases and compared with the experimentally derived data. Independently, the glycosyl hydrolase crystallographic models were fitted into the experimental molecular envelope. On the basis of these results and the evidence from 2D sequence alignment (10) and CD spectra, we propose that β -mannosidase has a molecular fold similar to the one of β -galactosidase, a member of family 2, clan GHA.

MATERIAL AND METHODS

Materials and Protein Purification. β -Mannosidase was purified to homogeneity from culture filtrate of *T. reesei* as described in ref 6. Crystallization solutions and buffers were prepared using chemicals of an analytical grade purchased from Sigma Chemicals.

Circular Dichroism Spectroscopy. Circular dichroism spectra were acquired in-house on a J715 spectropolarimeter (Jasco Corp., Japan). The protein at a concentration of 0.3 mg/mL was placed in a 10-mm path-length cuvette. The sample chamber was purged with dry nitrogen to remove oxygen from the optical path. Spectra were obtained in sodium acetate 50 mM buffered at pH 4.7, and spectra for the buffer were subtracted from those for the β -mannosidase samples.

Secondary structure content was analyzed using the neural net program K2D (11). Data were submitted online to the DichroWeb website (http://www.cryst.bbk.ac.uk/cdweb/html/ home.html) via the online submission procedure (12). Data over the range 200-241 nm was used. The determined NRMSD goodness of Fit parameter (13) was found to be 0.143.

Small-Angle X-ray Scattering Experiments and Data Analysis. β -Mannosidase samples with the concentration 4 mg/mL were prepared in 50 mM sodium acetate buffer (pH 4.7). X-ray solution scattering data were collected at 4 °C with the low-angle scattering camera at the station 2.1(14)at the Daresbury Synchrotron Radiation Source using a 200 × 200 mm position-sensitive multiwire proportional counter operated at 512 × 512 pixel mode (15). At a sample-todetector distance of 2 m and an X-ray wavelength λ of 1.54 Å, a momentum transfer q interval of 0.03–0.35 Å⁻¹ was covered ($q = 4\pi \sin \theta / \lambda$, θ being half of the scattering angle). The q range was calibrated using an oriented specimen of wet rat tail collagen (based on a diffraction spacing of 670 Å). The sample was held in a temperature-controlled brass cell containing a Teflon ring sandwiched by two mica windows. The cell has a sample volume of 120 μ L and a thickness of 1.5 mm. Buffer and sample SAXS data were collected alternatively, each in frames of 60 s. The systematic data reduction included radial integration of the two-

Biochemistry, Vol. 41, No. 30, 2002 9371

dimensional images, normalization of the subsequent onedimensional data to the intensity of the transmitted beam, correction for detector artifacts, and subtraction of background scattering from the buffer (subtraction of the scattering from the camera and a cell filled with buffer). Data reduction was performed with the OTOKO software package (16), modified, and adapted for the Daresbury NCD software suite. Scattering data analysis was followed by standard procedures (17). Using the indirect Fourier transform method as implemented in the program GNOM (18, 19), the intraparticle distance distribution function $p(\mathbf{r})$, the radius of gyration of the protein, Rg, using Guinier law (20), and integral properties of $p(\mathbf{r})$ and the extrapolated forward scattering intensity, I(q = 0) (in a relative scale), were determined from the experimental scattering data.

Ab Initio Molecular Shape Determination from SAXS Experiments. The scattering intensity function was determined in 90 frames, 60 s each. Since the scattering intensity from 0.10 up to 0.35 Å⁻¹ did not vary with time, all frames were used to construct it in this q range. Because of radiation-induced aggregation effects on the scattering intensity, only the first frame was used to compose the scattering curve in the low-q range.

The resolution of the resulting solution X-ray scattering curve extended to 18 Å. The low resolution particle shape was restored using the ab initio procedure described by Svergun (21) as implemented in the program GASBOR. In this method, a dummy residue (DR) model is generated by a random-walk C_{α} chain and is folded in such a way to minimize a discrepancy between the calculated scattering curve from the model and the experimental data. The program simulates the protein internal structure, which makes it unnecessary to subtract a constant background from the experimental data to ensure the asymptotic Porod's behavior (22). Several runs of ab initio shape determination with different starting conditions lead to consistent results as judged by the structural similarity of the output models, yielding nearly identical scattering patterns and fitting statistics in a stable and self-consistent process. The final shape restoration was performed using 950 dummy residues and 611 waters assuming no molecular symmetry. The resulting 18 Å resolution structure was used to calculate a low resolution envelope of the protein.

Enhancement of SAXS Envelope by using Crystallographic Data. T. reesei β -mannosidase was crystallized in the presence of CdCl₂ as described (7). Crystals diffracting to medium resolution (2.5 Å) grow in space groups $P4_12_12$ and $P2_12_12_1$ under very similar crystallization conditions. Both crystal forms have nearly identical cell dimensions (a $= b \simeq 165$ Å and $c \simeq 121$ Å) and a similar solvent content (70%). The asymmetric unit (ASU) of $P4_12_12$ crystal form contains 1 molecule and crystals belonging to the $P2_12_12_1$ contain 2 molecules/ASU. Due to very low homology with proteins whose structures are known, ab initio crystallographic methods have been considered for structure solution. A severe nonisomorphism problem prevented the use of the multiple isomorphous replacement (MIR) method. Anomalous signals from cadmium ions that were present in the crystallization solution allowed the use of the singlewavelength anomalous dispersion (SAD) approach. Initial SAD electron density maps were obtained in the space group $P4_{1}2_{1}2$ and were improved by density modification tech-

9372 Biochemistry, Vol. 41, No. 30, 2002

niques. Diffraction patterns of β -mannosidase are highly anisotropic in both crystal forms. Although the quality of final electron density maps was not good enough to build a crystallographic model, it was sufficient to determine the overall protein shape by the procedure described bellow.

The calculated electron density maps exhibit a clear protein/solvent boundary, but to distinguish the limits between two protein molecules by visual inspection proved to be difficult. Nevertheless, we were able to determine the correct position of a single molecule in the unit cell of a native crystal belonging to space group $P4_12_12$ using the GASBOR model obtained from SAXS experiments and initial SAD phases. The phased molecular replacement technique (23), as implemented in program AmoRe (24), was used for this purpose. The use of FSEARCH (25) was computationally prohibitively expensive due to the absence of molecular and noncrystallographic symmetry in space group $P4_12_12$. The molecular replacement solution was modified by a cyclic use of Arp warp (26, 27) and Refmac (28), in a procedure similar to that described in Sabini (29). Data extending up to 2.5 Å were included. The process was interrupted when the Free R-factor stopped decreasing; the initial and final values were about 56 and 45%, respectively. The final model contains 4800 free atoms and, because of the poor quality of the available phase information, reflects only the overall shape of the protein as observed in the electron density map.

Further evidence of the accuracy of the final model obtained using electron density maps in space group $P4_{1}2_{1}2$ was found. First, this model was used successfully for molecular replacement in crystals from space group $P2_12_12_1$, for which no previous phase information was available. Despite the fact that β -mannosidase is a monomer in solution, this space group contains two molecules in the asymmetric unit related by a rotation of 90° (as obtained by calculating the self-rotation function map). Molecular replacement solution was unequivocal, with a correlation coefficient of 65.5% after rigid body refinement. Second, a later comparison of crystal packing in both space groups showed that almost exactly the same dimer is generated by applying crystallographic symmetry operations to the free atom model obtained in space group $P4_12_12$. Moreover, the symmetry relating the two molecules in ASU of space group $P2_12_12_1$ is pseudo-crystallographic and is equivalent to the 4-fold screw axis from space group $P4_12_12$.

Third, phases calculated from the resulting dimer in the space group $P2_12_12_1$ were used to compute anomalous difference maps (as mentioned above, crystals grow in the presence of cadmium ions in both space groups), leading to reliable identification of the heavy atom sites. Nevertheless, this crystal form presents the same problems observed for $P4_12_12$ crystals, and the resulting electron density maps were not interpretable even with a help of putative homologous structures (see Discussion) and were not used for further model building.

Fourth, a number of sites found in space group $P2_12_12_1$ by using the anomalous maps were recognized to be previously found by Patterson methods in early attempts to solve the phase problem. Moreover, a number of heavy atoms were found to bind to the same sites of the protein in both space groups, while some other sites are present exclusively in only one space group, which is probably related to packing changes from one space group to the other. And last but not least, electron density maps obtained in space group $P2_12_12_1$ from refinement of heavy atom parameters are in complete agreement with the cell volume occupied by the protein model resulting from molecular replacement procedure.

In short, the procedure for enhancement of the protein envelope included the use of the SAXS envelope and additional crystallographic information. SAXS results combined with available crystallographic data allowed for the location of the molecule inside the unit cell and for the determination of another envelope with a better resolution than that exclusively inferred from SAXS results.

Alignment Procedures and Envelopes Composition. The superposition of crystallographic structures and free atom models was done using the program SUPCOMB (30). To keep the original low-resolution limit of the SAXS data, envelopes (masks) were computed from coordinate files output by GASBOR using NCSMASK (26, 27). For SAXS models, the NCSMASK parameter "radius", which sets the sphere radius for building the mask out of the model atoms, was adjusted to 5.7 Å to remove all internal characteristics and the parameter "smooth" was set to 18.0 Å to eliminate features smaller than 18.0 Å from the surface of the mask. To obtain masks for the crystallographic model, NSCMASK was used with both parameters set to 4.0 Å to take into account only the shape of the protein. As this model contains about 4800 atoms and was obtained using X-ray data to 4.0 Å resolution, the use of a smaller sphere radius was sufficient to eliminate internal characteristics. Atomic models derived $p(\mathbf{r})$ functions were calculated using CRYSOL (31). Following a standard procedure, all waters from crystallographic models were removed, and a 3 Å hydration layer was added prior to $p(\mathbf{r})$ function computation.

RESULTS

Overall Dimensions and Shape of β -Mannosidase. The scattering curve obtained as described in Materials and Methods and the Guinier plot are presented in Figure 1. The Guinier plot exhibits a linear behavior within the interval of 0.03 Å⁻¹ < q < 0.08 Å⁻¹, from which a radius of gyration of 35.6 ± 0.2 Å was calculated (20). The pair distribution functions $p(\mathbf{r})$ computed for the SAXS and crystallographic molecular models are shown in Figure 2. The function obtained from SAXS low-resolution model indicates a maximum dimension of the molecule $d_{\text{max}} = 120 \pm 10$ Å. The peak is centered at 42 Å indicating that the particle has a prolate shape. As can be seen in Figure 2, the same maximum molecular dimension is inferred from the p(r)function corresponding to the model derived using crystallographic data. This function fits well with the p(r) function obtained from SAXS results up to about 40 Å and is lower between 40 and 120 Å. This last result indicates that the crystallographic model leads to an envelope in which large dimensions in some regions are smaller than those of the SAXS model. This is confirmed by the pictures shown in Figure 3. The overall shapes of the models are in good agreement. They are composed of a globular structural core extending continuously toward a smaller and narrower portion, probably as an indication of a multidomain structure. The main differences occur in particular surface areas where the SAXS model extends into the solvent more than the

Low Resolution Model of β -Mannosidase from T. reesei



FIGURE 1: Experimental and calculated small-angle X-ray solution scattering of β -mannosidase. Experimental solution scattering curve (squares with error bars) was obtained as described in the text. The Guinier plot with linear fit is shown in the inset yielding to a radius of gyration of 35.6 ± 0.2 Å. The smooth curve is the corresponding scattering curve obtained from the dummy residue model derived by GASBOR, χ of fitting is 1.55.



FIGURE 2: Distance distribution functions $p(\mathbf{r})$. The curve obtained from experimental scattering curve (1; open squares) was calculated using program GNOM. Corresponding curves evaluated for β -galactosidases from *E. coli* (2; dotted line) and *Penicillium* sp. (3; solid line) and for the envelope enhanced by using X-ray crystallographic data (4; dashed line) are shown. $p(\mathbf{r})$ functions are normalized to the maximum of each curve individually. For further details, see the text.

crystallographic model, probably as an effect of the intrinsic low resolution of SAXS data or as a consequence of the protein glycosilation. As expected, more structural details can be seen in the envelope enhanced by crystallographic data.

DISCUSSION

The Putative β -Mannosidase Fold. A number of glycosyl hydrolases with molecular masses around 100 kDa were considered as possible templates for β -mannosidase from *T*. reesei. A good fit to the $p(\mathbf{r})$ function calculated from the



FIGURE 3: Stereoviews of β -mannosidase envelopes. The semitransparent envelope in gray represents the dummy residue model derived by GASBOR using only SAXS data. The envelope in black represents the more detailed model, enhanced by using X-ray crystallographic data. The middle and bottom rows are rotated clockwise by 90° around the y-axes and counterclockwise by 90° around x-axes, respectively. See text for further details. Drawings were made using PyMOL (DeLano Scientific, San Carlos, CA, http://www.pymol.org) under Linux.

free atom model was obtained with the family 2 glycosyl hydrolase β -galactosidase from *Escherichia coli* (32), which will be referred to as 1BGLa (PDB entry 1BGL, chain A). The p(r) function calculated for 1BGLa, shown in Figure 2, coincides with p(r) functions calculated for the SAXS and crystallographic free atom models up to 60 and 40 Å, respectively, and has the same maximum dimension of 120 Å. This is a surprisingly good fit in view of expected structural divergence between β -galactosidase and β -mannosidase, differences in the glycosylation state, and the higher molecular weight of the former.

 β -Galactosidase from *E. coli* has a TIM-barrel-like core surrounded by four other largely β domains. A ribbon representation of 1BGLa superposed onto the surface representation of the crystallographically enhanced molecular shape derived for β -mannosidase is shown in the Figure 4. This superposition suggests that the globular core of the protein possibly contains an $(\alpha/\beta)_8$ barrel and two other 9374 Biochemistry, Vol. 41, No. 30, 2002



FIGURE 4: Stereoviews showing the superposition of a cartoon representation of β -galactosidase from *E. coli* (PDB entry 1BGL, chain A) on the crystallographically enhanced β -mannosidase envelope. The same perspectives of Figure 3 are shown, with middle and bottom rows rotated clockwise by 90° around the y-axes and counterclockwise by 90° around x-axes, respectively. The clipping plane allows for an inside view of envelope contents. Domains 1, 2, 3, 4, and 5 of the β -galactosidase monomer are colored in red, green, yellow, lime, and blue, respectively. The domain 3 (TIM barrel) forms the core of the monomer. The other domains surrounding the core are composed mainly of β -structures. This figure was prepared using PyMOL (DeLano Scientific, San Carlos, CA, http://www.pymol.org) under Linux.

possible domains, while the narrower portion extending from the core can accommodate two more substructures. The structural similarity provides support to the hypothesis that the additional domains are composed mainly by β strands. Some of the β strands, mainly located in the 1BGLa domains 4 and 5, protrude out of the β -mannosidase envelope. They could be simply inexistent or this might be an effect of a different orientation of the domains 4 and 5, what is indicated by the empty portions seen close to those domains. This view

Table 1: Comparison of Secondary Structure Composition of β -Galactosidases from *E. coli* (PDB entry 1BGL, chain A) and *Penicillium* sp. and β -Mannosidase from *T. reesei^a*

secondary structure composition	<i>E. coli</i> β -galactosidase	Penicillium sp. β -galactosidase	<i>T. reesei</i> β -mannosidase
α(%)	14	16	28
$\beta(\%)$	41	39	30
others (random, turns) (%)	45	45	42
a. 1. 1. 1.1			

^{*a*}As obtained by circular dichroism spectroscopy.

is supported by a comparison between 1BGLa and the structure of β -galactosidase from *Penicillium* sp., recently solved by our group (unpublished results), which has a very similar structure, where the $(\alpha/\beta)_8$ and the other four largely β domains have slightly different sizes and orientations in comparison with those corresponding in 1BGLa. The *p*(r) function calculated from *Penicillium* sp. β -galactosidase, shown in Figure 2, is closer to the crystallographic *p*(r) than 1BGLa. The superposition of this model to the β -mannosidase envelopes is remarkable and a striking indication that the latter has a great structural similarity with β -galactosidase from *Penicillium* sp.

Further support to our hypothesis comes from the theoretical prediction of the catalytic domain structure of bovine β -mannosidase done by Durand and collaborators (10). In that work, based on known three-dimensional structures from some families of clan GH-A, a group comprising more than 200 proteins sharing low sequence identity, structural and functional features shared by those proteins were deduced. Using the 2D hydrophobic cluster analysis (HCA) method (33), which can be applied to analyze proteins sharing very low identity (typically 15-25%, ref 34), the authors found the probable fold as well as likely functional amino acids of the catalytic domains of five lysosomal glycosyl hydrolases from families 1, 2, 5, 10, and 17 and, in particular, bovine β -mannosidase. It was suggested that bovine β -mannosidase, which has about 65% amino acid sequence identity with β -mannosidase from T. reesei, could have a $(\alpha/\beta)_8$ barrel catalytic domain. Our results are in excellent agreement with this theoretical prediction. Moreover, the β -mannosidase secondary structure assessed by CD technique is in a close agreement with the secondary structure composition of β -galactosidases from both *E. coli* and *Penicillium* sp. (Table 1). This further supports our hypothesis of the close structural similarity of these proteins.

We have shown that the distance distribution functions obtained from SAXS results and crystallographic data provided insights into the arrangements of β -mannosidase domains by comparison with other known structures. We believe that this approach combined with molecular modeling and bioinformatics techniques may be applied as a general tool in protein shape recognition. Although it cannot provide a unique solution in search for protein fold, this method is able to restrict the possible folding space of an unknown protein once solution scattering data and its secondary structure composition obtained for example by CD or infrared spectroscopy are available. This method provides an information independent of the protein amino acid sequence based on the 1D and 2D alignment algorithms. The proposed procedure seems particularly useful within a scope of Low Resolution Model of β -Mannosidase from *T. reesei*

structural genomics programs, particularly for application to poorly crystallizable proteins. Use of SAXS models for molecular replacement when initial phase information is available could also be extended to difficult cases in protein crystallography, when model building proves to be a problem. In addition, we have shown that low-resolution models derived from X-ray solution scattering experiments can actually be improved by addition of low-resolution crystallographic data.

REFERENCES

- Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D. (1989) *The Metabolic Basis of Inherited Disease*, Vol. II, 6th ed., McGraw-Hill, Inc., USA.
- Watts, R. W. E., and Gibbs, D. A. (1986) Lysosomal Storage Diseases – Biochemical and Clinical Aspects, Taylor & Francis, Inc., Philadelphia, USA.
- Rodriguez-Serna, M., Botella-Estrada, R. M. D., Chabas, A., Coll, M., Oliver, V., Febrer, M., and Aliaga, A. (1996) Angiokeratoma corporis diffusum associated with beta-mannosidase deficiency. *Arch. Dermatol.* 132 (10), 1219–1222.
- Kuhad, R. C., Singh, A., and Eriksson, K. L. (1997) Microorganisms and enzymes involved in the degradation of plant fiber cell walls. *Biochem. Eng. Biotechnol.* 57, 45–125.
- Dey, P. M., and del Campillo, E. (1984) Biochemistry of the multiple forms of glycosidases in plants. *Adv. Enzymol. Relat. Areas Mol. Biol.* 56, 141–249.
- Kulminskaya, A. A., Eneiskaya, E. V., Isaeva-Ivanova, L. S., Savel'ev, A. N., Sidorenko, I. A., Shabalin, K. A., Golubev, A. M., and Neustroev, K. N. (1999) Enzymatic activity and betagalactomannan binding property of beta-mannosidase from *Trichoderma reesei*. *Enzyme Microb. Technol.* 25 (3-5), 372-377.
- Aparicio, R., Eneiskaya, E. V., Kulminskaya, A. A., Savel'ev, A. N., Golubev, A. M., Neustroev, K. N., Kobarg, J., and Polikarpov, I. (2000) Crystallization and preliminary X-ray study of betamannosidase from *Trichoderma reesei*. Acta Crystallogr. D 56, 342–343.
- Shannon, C. E., and Weaver, W. (1949) *The Mathematical Theory* of *Communication*, University of Illinois Press, Urbana.
- Moore, P. B. (1980) Small-angle scattering. Information content and error analysis. J. Appl. Crystallogr. 13, 168–175.
- Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B., and Mornon, J. P. (1997) Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology* 7 (2), 277–284.
- Andrade, M. A., Chacon P., Merelo, J. J., and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Protein Eng.* 6, 383–390.
- Lobley, A., Whitmore, L., and Wallace, B. A. (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18, 211– 212.
- Mao, D., Wachter, E., and Wallace B. A. (1982) Folding of the mitochondrial proton adenosinetriphosphatase proteolipid channel in phospholipid vesicles. *Biochemistry* 21 (20), 4960–4968.
- 14. Towns-Andrews, E., Berry, A., Bordas, J., Mant, G. R., Murray, P. K., Roberts, K., Sumner, I., Worgan, J. S., Lewis, R., and Gabriel, A. (1989) Time-resolved X-ray diffraction station: X-ray optics, detectors, and data acquisition. *Rev. Sci. Instrum.* 60, 2346–2349.

- Lewis, R. (1994) Multiwire gas proportional counters: decrepit antiques or classic performers? J. Synchrotron Rad. 1, 43–53.
- Boulin, C., Kempf, R., Koch, M. H. J., and McLaughlin, S. M. (1986) Data appraisal, evaluation and display for synchrotron radiation experiments – hardware and software. *Nucl. Instrum. Methods Phys. Res.* A249, 399–407.
- Grossmann, J. G., Crawley, J. B., Strange, R., Patel, K. J., Murphy, L. M., Neu, M., Evans, R. W., and Hasnain, S. S. (1998) The nature of ligand-induced conformational change in transferrin in solution. An investigation using X-ray scattering, XAFS and sitedirected mutants. J. Mol. Biol. 279, 461–472.
- Svergun, D. I., Semenyuk, A. V., and Feigin, L. A. (1988) Smallangle-scattering-data treatment by the regularization method. *Acta Crystallogr. A* 44, 244–251.
- Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. 25, 495–503.
- 20. Guinier, A., and Fournet, G. (1955) Small-Angle Scattering of X-rays, John Wiley & Sons, New York.
- Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) Determination of Domain Structure of Proteins from X-ray Solution Scattering. *Biophys. J.* 80, 2946–2953.
- Porod, G. (1982) General Theory in Small-Angle X-ray Scattering, (Glatter, O., and Kratky, O., Eds.) pp 17–51, Academic Press, London.
- 23. Bentley, G. A. (1992) Some applications of the phased translation function using calculated phases in Molecular Replacement, *Proceedings of the Daresbury Study Weekend*, DL/SCI/R33.
- Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A50*, 157–163.
- 25. Hao, Q. (2001) Phasing from an envelope. *Acta Crystallogr. D* 57, 1410–1414.
- Collaborative Computational Project (1994) No. 4, Acta Crystallogr. D 50, 760–763.
- Winn, M., Dodson, E. J., and Ralph, A. (1997) Collaborative Computational Project, Number 4: Providing Programs for Protein Crystallography. *Methods Enzymol.* 277, 620–633.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D* 53, 240–255.
- Sabini, E., Schubert, H., Murshudov, G., Wilson, K. S., Siika-Aho, M., and Penttilä, M. (2000) The three-dimensional structure of a *Trichoderma reesei* beta-mannanase from glycoside hydrolase family 5. *Acta Crystallogr. D* 56, 3–13.
- Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J. Appl. Crystallogr.* 34, 33-41.
- Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *J. Appl. Crystallogr.* 28, 768–773.
- Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) Three-dimensional structure of beta-galactosidase from *E. coli. Nature*, 369 (6483), 761–766.
- Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J. P. (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett.* 224(1), 149–55.
- 34. Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A., and Mornon, J. P. (1990) Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D-representation of protein sequences. *Biochimie* 72(8), 555–74.

BI025811P

4.1 Novos resultados cristalográficos

Há registros na literatura do uso bem sucedido de modelos de SAXS em cristalografia, com resultados interessantes (Hao, 2001). No caso da β manosidase, através de uma abordagem nova, o modelo de SAXS foi utilizado com sucesso para localizar uma molécula de proteína dentro da cela unitária do cristal, por meio de Substituição Molecular com uso das fases obtidas por SAD (pág. 45) para conjuntos do grupo $P4_12_12$.

A partir daí, foi possível estudar o empacotamento da proteína nos dois grupos e responder perguntas formuladas desde as primeiras coletas de dados. Foi possível explicar a semelhança entre os grupos $P4_12_12$ e $P2_12_12_1$ e a ineficiência dos processos de promediação. O conhecimento do empacotamento molecular levou a hipóteses muito razoáveis para a anisotropia associada ao eixo z. O grave não-isomorfismo entre os cristais também pôde ser explicado. Os resultados estão sumarizados a seguir.

4.1.1 Faseamento no grupo $P2_12_12_1$

Informação de fase para o grupo $P2_12_12_1$, não obtida até esta etapa do projeto, poderia ser de grande ajuda na resolução da estrutura. Primeiro, porque haveria a possibilidade de promediação de densidade eletrônica entre os dois grupos. Segundo, porque deveria haver 2 moléculas de β -manosidase na ASU deste grupo (o dobro que no grupo $P4_12_12$, já que os parâmetros de rede são praticamente iguais em ambos os grupos), ou seja, NCS que poderia ser utilizada como mais uma fonte de informação durante a promediação de densidade. Terceiro, havia esperança de que o problema de anisotropia pudesse ser menos restritivo neste novo grupo, com uma possível melhora na distorção do mapa na direção z e menor ruído em torno das posições dos átomos pesados.

A Substituição Molecular no grupo $P4_12_12$ e o aperfeiçoamento do mo-

delo obtido por SAXS estão descritos no artigo em anexo. O modelo de átomos livres final, resultado da Substituição Molecular do modelo de SAXS no grupo $P4_12_12$, foi usado para nova Substituição Molecular, desta vez em conjuntos do grupo $P2_12_12_1$. Foram encontradas 2 mol/ASU. As soluções após refinamento de corpo rígido no Amore,

	Alpha	Beta	Gamma	Tx	Ty	Tz	$Corr_F$	Rfac	Corr_I
$F1_1$	117.53	55.82	157.76	0.0046	0.0787	0.2050	65.5	36.3	69.7
$F2_1$	27.78	55.89	157.80	0.3289	0.2450	0.9544	65.5	36.3	69.7

são inequívocas. Note-se que o primeiro ângulo indica uma rotação relativa de 90° entre as 2 moléculas e que os outros dois são idênticos (dentro da margem de erro) e que a diferença fracionária de translação na direção z ("Tz") é de ~ 0.75, indicações da existência de NCS que consiste num eixo helicoidal de ordem 4.

Uma vez que o modelo de átomos livres foi derivado a partir de mapas que não exibiam elementos de estrutura secundária (apenas uma clara fronteira proteína-solvente), ele não reflete a estrutura interna da proteína. Desta forma, após a Substituição, fases calculadas a partir do modelo de átomos livres foram usadas apenas para localizar sítios de ligação de átomos pesados confiáveis. Os sítios encontrados foram usados para faseamento por SAD (pág. 45). O resultado foram mapas de densidade eletrônica no novo grupo que, infelizmente, apresentavam os mesmos problemas do grupo tetragonal. Adição de sinal isomorfo advindo de algum outro conjunto pertencente ao grupo $P2_12_12_1$ (Tabela 3.2) também foi considerada mas os mapas de densidade eletrônica não melhoraram.

Promediação das densidades eletrônicas obtidas nos dois grupos de espaço, incluindo o uso da NCS presente no grupo ortorrômbico, foi feita à exaustão, empregando inúmeras estratégias. Infelizmente, não houve alteração significativa na qualidade geral dos mapas, que continuaram nãointerpretáveis. Observamos, apenas, que os mapas do grupo ortorrômbico pareciam ser um pouco melhores que os do grupo tetragonal no que se referia ao estiramento na direção z, contudo, continuavam não-interpretáveis.

4.1.2 Conteúdo de solvente da ASU

O número de moléculas na ASU é uma informação importante para vários procedimentos durante a determinação de uma estrutura cristalográfica, principalmente durante a promediação e a modificação de densidade eletrônica por *solvent flattening* (pág. 47).

Até a realização dos estudos a baixa resolução, não havia certeza sobre o conteúdo de solvente dos cristais de β -mannosidase. Apenas durante o processo de Substituição Molecular, este parâmetro foi inequivocamente determinado. Os conjuntos do grupo $P4_12_12$ contém uma molécula de proteína na ASU e os grupo $P2_12_12_1$, duas moléculas. Como os parâmetros de rede são praticamente iguas, o conteúdo de solvente em ambos é da ordem de 70 %.

4.1.3 A configuração do eixo helicoidal nos dois grupos

Nas Figura 4.1 pode-se observar como os dois monômeros da ASU do grupo $P2_12_12_1$ interagem. O grupo $P4_12_12$ contém apenas um monômero na ASU, entretanto, para comparação com o primeiro grupo, um segundo monômero foi desenhado (obtido por aplicação de simetria *cristalográfica*). Claramente, o arranjo dimérico formado é praticamente o mesmo, a menos de uma translação do eixo de simetria, quando comparamos os dois grupos. Isto significa que as moléculas estão orientadas praticamente da mesma forma nos dois grupos, ou seja, os mapas obtidos em cada grupo não apresentam diferenças significativas que possam ajudar a vencer a distorção resultante do forte espalhamento anisotrópico associado ao eixo z, paralelo ao eixo helicoidal de ordem 4, ou levem a melhoras significativas nas fases.



Figura 4.1: Empacotamento cristalino e eixo helicoidal. Cada figura abrange quatro celas unitárias e duas moléculas de proteína. (a). Plano XY paralelo à página. No grupo $P4_12_12$ (acima), o eixo é cristalográfico, paralelo a Z, passando por (0.5, 0, 0). No grupo $P2_12_12_1$ (abaixo), o eixo é pseudo-cristalográfico ("NCS1"), quase perfeitamente paralelo a Z passando por (0.25, -0.07, 0). (b). Vista lateral dos eixos nos dois grupos.

No caso do grupo $P2_12_12_1$, o eixo de simetria é pseudo-cristalográfico. Isto explica as ausências sistemáticas do grupo tetragonal. Além disso, a promediação de densidade eletrônica é de grande ajuda quando o operador de NCS introduz informação independente, impondo novos vínculos entre os fatores de estrutura. Entretanto, no caso de NCS pseudo-cristalográfica, a informação que viria do operador já está preliminarmente incluída nos fatores de estrutura, de modo que não há melhoria nas fases e nos mapas de densidade. Isto explica porque o uso desta técnica não foi capaz de melhorar os mapas no grupo ortorrômbico.

4.1.4 A falta de isomorfismo entre os cristais e inexistência de perfeita simetria

O não-isomorfismo entre os cristais de β -manosidase não é causado por um único fator. É possível tecer algumas considerações sobre este problema a partir da informação que foi acumulada.

Inicialmente, deve-se lembrar que a proteína é glicosilada, com cadeias de açucares em quantidade e comprimento variáveis (uma vez que o processo de glicosilação não é controlado por DNA), o que implica na existência de variável grau de heterogeneidade. Além disso, durante o projeto, mais de uma linhagem de fungo e protocolo de purificação foram utilizados, o que levou a amostras diferentes. Os dois aspectos ficam claros quando se avalia os vários conjuntos coletados e as condições de cristalização ao longo do tempo. Os cristais apresentam diferentes qualidades e as condições de cristalização apresentam mudanças significativas, o que implica diferentes níveis de glicosilação, conforme já mencionado (pág. 68).

Este aspecto em conjunto com o alto conteúdo de solvente do cristal, em torno de 70%, abre margem para uma indesejável variabilidade no empacotamento cristalográfico, suficiente para inviabilizar a utilização dos métodos de átomos pesados baseados em sinal isomorfo (pág. 41). Mudanças no empacotamento das moléculas levam às mudanças observadas nos parâmetros de rede e, por vezes, são capazes até mesmo de mudar o grupo espacial do cristal. Há variação, também, nos sítios de ligação de Cd em cristais do mesmo grupo espacial, que, muito provavelmente sofram com tamanha variabilidade do ambiente cristalográfico ao redor. Isto compromete todos os parâmetros do sítio, notadamente, os fatores de temperatura e ocupância, e a introdução de anisotropia.

Com estas considerações, podemos explicar a impossibilidade de encontrar dois conjuntos, mesmo nativos, que sejam suficientemente isomorfos para permitir uso de SIR(AS) ou MIR(AS). E mais, podemos duvidar que exista um cristal que exiba perfeitamente as simetrias do grupo $P4_12_12$. É bem provável que distorções, em maior ou menor grau, existam em todos os conjuntos onde foi possível o processamento dos dados neste grupo.

4.1.5 O espalhamento anisotrópico

Pôde-se, também, aventar uma hipótese que explicasse o espalhamento anisotrópico associado ao eixo z da cela unitária. Como sabido, o espalhamento anisotrópico está relacionado à desordem estática ou dinâmica ao longo de uma direção. A Figura 4.2 mostra o empacotamento molecular nos dois grupos de espaço. Em ambos, o empacotamento visto ao longo da direção z mostra claros vãos de solvente ao mesmo tempo em que, olhando-se o plano XY em direção ao eixo z, nota-se um empacotamento mais compacto.

Embora cálculos quantitativos não tenham sido feitos, aparentemente as moléculas têm mais liberdade de movimento na direção z, o que poderia levar a um maior grau de desordem estática/dinâmica nesta direção, resultando na anisotripia observada. O efeito é mais aparente no grupo $P4_12_12$, o que pode ser a razão para que neste grupo os mapas tenham qualidade inferior aos do grupo $P2_12_12_1$, com o efeito de estiramento na direção z ainda mais aparente (conforme comentado à pag. 110).



(a) Grupo $P4_12_12$.

(b) Grupo $P2_12_12_1$.

Figura 4.2: Empacotamento cristalino. As celas da Figura 4.1 foram preenchidas aplicando-se operações de simetria cristalográfica ao monômero (no caso do grupo $P4_12_12$) ou dímero (no caso do grupo $P2_12_12_1$) presentes na ASU, num raio de ~ 100 Å a partir da origem. Em (a) e (b), a figura de cima tem o plano XY paralelo ao plano da página e a figura de baixo tem o eixo z paralelo ao plano da página.

4.1.6 Ultimas tentativas de Substituição Molecular

Como uma última alternativa de determinar a estrutura de alta resolução da β -manosidase, com base nos resultados estruturais do artigo em anexo, o método da Substituição Molecular foi novamente utilizado, desta vez, tendo como modelos de busca as β -galactosidases de *Trichoderma reesei* e *Penicillium* sp..

Este modelos foram empregados de várias formas. Primeiramente, o modelo completo. Depois sub-estruturas compreendendo tanto o domínio *timbarrel* isoladamente quanto em conjunto com outros domínios. Lançamos mão, também, de modelos teóricos obtidos por homologia através do programa MODELLER¹ (Marti-Renom *et al.*, 2000). Os programas AMoRe (Navaza, 1994; CCP4, 1994; Winn *et al.*, 1997) e BEAST (Read, 2001; CCP4, 1994; Winn *et al.*, 1997), que implementa o método *Maximum Likelyhood*, foram utilizados para a Substituição Molecular.

As centenas de soluções resultantes foram averiguadas quanto à sua localização na cela unitária e ocorrência de *overlap* por simetria cristalográfica, mesmo que não houvesse indicação clara de correção². Em vários casos, soluções chegaram a ser refinadas.

Apesar de todo o esforço, não se conseguiu o resultado esperado, o que pode ser justificado se notarmos que, embora espera-se a β -manosidase e as β -galactosidases tenham enovelamentos muito parecidos, a baixa identidade seqüencial entre elas provavelmente leva a diferenças na estrutura interna dos domínios que as compõem, o que impossibilita seu uso para Substituição Molecular.

¹ Agradecimentos ao Prof. Dr. Richard C. Garratt, do IFSC/USP, São Carlos/SP, pela sempre gentil colaboração.

² Vários procedimentos foram automatizados através de *scripts* em linguagem *Shell* que encadeiam as entradas e saídas de vários programas diferentes, aumentando significativamente a velocidade para obtenção de resultados e o volume de resultados analisados.

Capítulo 5

TIM: cristalização e coleta de dados

Os trabalhos para determinação da estrutura cristalográfica da Triose fosfato isomerase de músculo de coelho podem ser dividos em duas etapas. Na primeira, duas estruturas foram resolvidas a partir de dados coletados no LNLS. A segunda etapa consistiu na coleta de novos dados durante o estágio no YSBL (mencionado na Introdução do Capítulo 3), que levaram a uma nova estrutura.

5.1 Coletas de dados no LNLS

A cristalização, coleta de dois conjuntos de dados no LNLS e resolução de duas estruturas de TIM por Substituição Molecular foram descritos no artigo anexado a seguir, publicado sob a referência

Aparicio, R., Ferreira, S. T., Leite, N. R. and Polikarpov, I. (2000). Preliminary X-ray diffraction studies of rabbit muscle triose phosphate isomerase (TIM). *Acta Cryst. D* 56, 1492-1494.

crystallization papers

Acta Crystallographica Section D Biological Crystallography ISSN 0907-4449

Preliminary X-ray diffraction studies of rabbit muscle triose phosphate isomerase (TIM)

Ricardo Aparicio,^{a,b} Sergio T. Ferreira,^c Ney R. Leite^d and Igor Polikarpov^a*

^aLaboratório Nacional de Luz Síncrotron (LNLS), Caixa Postal 6192, CEP 13083-970, Brazil, ^bInstituto de Física Gleb Wataghin, UNICAMP, Brazil, ^cDepartmento de Bioquímica Médica, UFRJ, RJ, Brazil, and ^dUEPG, PR, Brazil

Correspondence e-mail: igor@lnls.br

Triose phosphate isomerase (TIM) is responsible for the interconversion between GAP and DHAP in the glycolytic pathway. Two crystal forms belonging to space group $P2_12_12_1$ were obtained by the hanging-drop method and were designated A and B. Synchrotron X-ray diffraction data were collected for both forms. Form A had unit-cell parameters a = 65.14, b = 72.45, c = 93.24 Å and diffracted to 2.25 Å at 85 K, whereas form B had unit-cell parameters a = 73.02, b = 79.80, c = 172.85 Å and diffracted to 2.85 Å at room temperature. Molecular replacement was employed to solve the structures, using human TIM as a search model. Further refinement of both structures is under way and is expected to shed light on the recently reported conformational studies for rabbit TIM.

1. Introduction

Glycolysis is the biochemical pathway by which glucose is converted to pyruvate with the generation of two moles of ATP per mole of glucose, providing part of the energy utilized by most cells. Triose phosphate isomerase (TIM, D-glyceraldehyde-3-phosphate ketolisomerase; E.C. 5.3.1.1) catalyzes the fifth reaction of the glycolytic pathway, namely the interconversion, probably via an enediol (or enodiolate) intermediate, of glyceraldehyde-3phosphate (GAP) and dihydroxyacetone phosphate (DHAP), two ketose-aldose isomers (Yüksel & Gracy, 1991). Rabbit muscle TIM is a dimer of identical subunits, each containing 248 residues and with a molecular weight of 26 627 Da. Although the dimer is the physiologically active species, a number of monomeric TIM mutants possessing catalytic activity have been described (Borchert et al., 1993; Mainfroid et al., 1996; Schliebs et al., 1996, 1997). At in vivo concentrations of substrate, the rate of reaction catalyzed by TIM (Knowles, 1991) is essentially limited by diffusion. The α/β -barrel structural motif, a supersecondary structure in which eight β -strands alternate in sequence with eight α -helices forming a parallel β -sheet closed into a cylindrical topology, was first recognized in the structure of chicken muscle TIM (Banner et al., 1975). Similar structures are now known to occur in many other proteins and this folding motif is frequently named the TIM barrel (Branden & Tooze, 1991). The active site of almost all known α/β -barrel enzymes is located at the carboxyl end of the barrel. The fact that most of the members of α/β -barrel proteins are enzymes serves as an argument for

Received 29 May 2000 Accepted 7 August 2000

their divergent evolution from a common ancestor (Farber, 1993). In Schistosoma mansoni, TIM is being considered as a possible target in attempts to obtain a vaccine against schistosomiasis (Doenhoff, 1998). In addition, TIM from Trypanosoma brucei brucei (the aetiologic agent of sleeping sickness) has been actively investigated from a structural point of view (Wierenga et al., 1991, 1992), with the goal of developing drugs that might inhibit its function and be of therapeutic value against infection. Furthermore, TIM deficiency in humans results in a genetic disease of autosomal dominant inheritance (locus 12p13), causing developmental retardation, myopathy, cardiac failure and anaemia (Daar et al., 1986; Watanabe et al., 1996). However, despite being one of the first glycolytic enzymes to be purified (Krietsch et al., 1970; Scopes, 1977), the structure of rabbit muscle TIM has not yet been reported. The amino-acid sequence of this protein is known (Corran & Ley, 1975) and its kinetic properties have been extensively studied (Garza-Ramos et al., 1992; Fernandez-Velasco et al., 1995) and compared with those of T. brucei brucei and yeast (Saccharomyces cerevisiae) TIMs (Lambeir et al., 1987). Biophysical studies of the enzyme have been carried out by infrared spectroscopy (Castresana et al., 1988) and electron microscopy (Tanaka et al., 1994). Recently, in order to relate conformational stability to the chemical stability and function of TIM, unfolding and refolding studies of rabbit muscle TIM using hydrostatic pressure and chaotropic agents as perturbing agents were monitored by fluorescence resonance energy-transfer (FRET) measurements (Rietveld & Ferreira, 1996, 1998). An interesting result from the latter

© 2000 International Union of Crystallography Printed in Denmark – all rights reserved

crystallization papers

Table 1

Data-collection	statistics	for	the	Α	and	В	crystal
forms.							

Parameter	A form	B form
No. of images	75	50
Space group	$P2_{1}2_{1}2_{1}$	$P2_{1}2_{1}2_{1}$
Unit-cell parameters (Å)		
a	65.14	73.02
b	72.45	79.80
с	93.24	172.75
Resolution range (Å)	13.5-2.25	14.00-2.85
Last shell (Å)	2.30-2.25	2.92-2.85
No. of measured reflections	49411 (2645)	34316 (1344)
No. of unique reflections	19851 (1270)	19037 (1012)
Multiplicity	2.49 (2.08)	1.80 (1.33)
$R_{\rm sym}$ (%)	10.7 (36.6)	12.4 (26.4)
Completeness (%)	93.1 (91.8)	78.8 (63.8)
$\langle I/\sigma(I)\rangle$	8.39 (2.14)	5.51 (1.99)

studies was the finding that rabbit muscle TIM exhibits persistent conformational/ energetic heterogeneity in solution. This led to the proposal of the existence of a heterogeneous ensemble of TIM dimers, probably arising from long-lived hetero-







Figure 1

(a) A representative diffraction pattern from the cryo-collected data set (crystal type A); (b) enhanced contrast close-up of the marked region.

geneity of subunit interactions (Rietveld & Ferreira, 1996). This surprising observation has been subsequently investigated from an energetic point of view (Rietveld & Ferreira, 1998), but a detailed investigation of the structural basis of the heterogeneity of TIM dimers is still lacking. Determination of the crystallographic structure of rabbit muscle TIM would contribute to the abovementioned functional and physicochemical studies, as well as adding to knowledge of the structures of triose phosphate isomerases from various organisms. Here, we report the crystallization, data collection and molecular-replacement solutions obtained for two crystalline forms of TIM, one determined at 85 K and the other at 298 K. Subsequent comparison of the refined structures in the two crystal forms will provide information on the influence of crystal contacts on the protein structure. Furthermore, analysis of the temperature factors of the models determined at room temperature and at 85 K might contribute

> to our understanding of the nature of the disorder (*i.e.* static or dynamic) of the flexible parts of the molecule, with possible implications regarding the above-described conformational heterogeneity of TIM dimers.

2. Crystallization and data collection

Rabbit muscle TIM (type X) was purchased from Sigma Chemical Co. The sample was >98% pure as judged by SDS-PAGE analysis. Initial crystallization conditions were screened at room temperature using the sparse-matrix method (Jancarik & Kim, 1991) and the hangingdrop method. Some conditions produced small crystals. These crystallization conditions were used as a guide for further trials. Two crystalline forms have been obtained by growing TIM crystals for 1-2 weeks in closely related conditions. A single data set was collected for each crystal form. Data were collected on 345 mm MAR Research а imaging-plate detector at the LNLS Protein Crystallography beamline (Polikarpov, Oliva et al., 1998; Polikarpov, Perles et al., 1998) by the oscillation method and were processed using the *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) packages. We describe below the growth conditions and details of the data collection and reduction for both crystal forms, which will henceforth be referred to as types A and B.

2.1. Crystal type A

This crystal form grew when reservoir solution containing 24% PEG 4000, 0.2 M MgCl₂, 0.1 M Tris-HCl pH 8.5 was mixed with protein solution (10 mg ml⁻¹ in water) in equal amounts and equilibrated against the reservoir solution. The crystals were needles measuring $0.6 \times 0.2 \times 0.2$ mm. To prevent radiation damage, the crystal was briefly soaked in a cryoprotectant solution containing 10%(v/v) ethylene glycol and rapidly frozen. A data set was collected from a single flash-cooled crystal at 85 K. The X-ray wavelength employed was 1.31 Å with a crystal-to-detector distance of 210 mm, giving an outer-edge resolution of 1.94 Å. The oscillation range was 1° and the exposure time per image (Fig. 1) was about 3 min. Data processing and reduction led to the statistics summarized in Table 1.

2.2. Crystal type B

This crystal form grew under similar conditions as type A, except that the concentrations of PEG 4000 and MgCl₂ in the reservoir solution were 23% and 0.33 M, respectively. These crystals formed clusters of plates and grew to dimensions of 1.0×0.5 \times 0.2 mm. Unfortunately, a suitable cryoprotectant solution for these crystals was not easy to obtain, as the crystals dissolved almost instantaneously when placed in the cryosolution described above. Therefore, we collected X-ray data for this crystal form at room temperature (298 K). Data were collected from a single crystal at a wavelength of 1.37 Å and a crystal-to-detector distance of 205 mm. The oscillation range was 1° and the exposure time was about 3 min per frame (Fig. 2). The results of data processing are shown in Table 1.

3. Molecular-replacement solutions

Phasing was carried out by the molecularreplacement method, as implemented in the program *AMoRe* (Navaza, 1994). The primary sequence search and sequence alignments were made using *ENTREZ* and *BLAST* (Altschul *et al.*, 1997). Human (Maquat *et al.*, 1985) and rabbit TIM

crystallization papers



Figure 2

A diffraction pattern from the data set collected for crystal type B.

(Corran & Ley, 1975) show 98% sequence identity and 99% similarity. Therefore, the human TIM X-ray dimer structure (PDB entry 1hti; Mande et al., 1994) was used as a search model. For the cryo-data set (crystal type A), the Matthews coefficient (Matthews, 1968) for a dimer in the asymmetric unit is $V_{\rm M} = 2.07 \text{ Å } \text{Da}^{-1}$, with a solvent content of 41%. For the type Bcrystal data set, $V_{\rm M} = 2.37$ Å Da⁻¹ indicated two dimers in the asymmetric unit and a solvent content of 48%. In the first case, calculations using 10-3.5 Å resolution limits and default parameters led to a single clear solution with a correlation coefficient of 73.5% and an R factor of 30.8% after fitting. The molecular-replacement procedure for the second (room-temperature) data set included a search for two dimers in the asymmetric unit cell and resulted in a clear solution with a correlation coefficient of 72.1% and an R factor of 33.0%. Further refinement of both structures is under way.

This work has been performed at the Laboratório Nacional de Luz Síncrotron (LNLS). The authors acknowledge financial support from FAPESP (*via* grants 99/03387-4 and 98/06761-1), CNPq, CAPES (Brazil) and the Howard Hughes Medical Institute (USA).

References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Nucleic Acids Res. 25, 3389–3402.
- Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E.,
- Priddle, J. D. & Waley, S. G. (1975). *Nature* (*London*), **255**(5510), 609–614.
- Borchert, T. V., Pratt, K., Zeelen, J. P., Callens, M., Noble, M. E., Opperdoes, F. R., Michels, P. A. & Wierenga, R. K. (1993). *Eur. J. Biochem.* 211(3), 703–710.
- Branden, C. & Tooze, J. (1991). Introduction to Protein Structure. New York: Garland Publishing.
- Castresana, J., Muga, A. & Arrondo, J. L. R. (1988). Biochem. Biophys. Res. Commun. 152(1), 69–75.
- Corran, P. H. & Ley, S. G. (1975). *Biochem. J.* **145**(2), 335–344.
- Daar, I. O., Artymiuk, P. J., Phillips, D. C. & Maquat, L. E. (1986). Proc. Natl Acad. Sci. USA, 83(20), 7903–7907.
- Doenhoff, M. J. (1998). Parasitol. Today, 14(3), 105–109.
- Farber, G. K. (1993). Curr. Opin. Struct. Biol. 3, 409–412.
- Fernandez-Velasco, D. A., Sepulveda-Becerra, M., Galina, A., Darszon, A., Tuena de Gomez-Puyou, M. & Gomez-Puyou, A. (1995). *Biochemistry*, **34**(1), 361–369.

- Garza-Ramos, G., Tuena de Gomez-Puyou, M., Gomez-Puyou, A. & Gracy, R. W. (1992). *Eur. J. Biochem.* 208(2), 389–95.
- Jancarik, J. & Kim, S.-H. (1991). J. Appl. Cryst. 24, 409–411.
- Knowles, J. R. (1991). *Nature (London)*, **350**, 121–124.
- Krietsch, W. K., Pentchev, P. G., Klingenburg, H., Hofstatter, T. & Bucher, T. (1970). *Eur. J. Biochem.* 14(2), 289–300.
- Lambeir, A. M., Opperdoes, F. R. & Wierenga, R. K. (1987). *Eur. J. Biochem.* **168**, 69–74.
- Mainfroid, V., Terpstra, P., Beauregard, M., Frere, J. M., Mande, S. C., Hol, W. G., Martial, J. A. & Goraj, K. (1996). J. Mol. Biol. 257(2), 441–456.
- Mande, S. C., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A. & Hol, W. G. (1994). *Protein Sci.* 3(5), 810–821.
- Maquat, L. E., Chilcote, R. & Ryan, P. M. (1985). J. Biol. Chem. 260(6), 3748–3753.
- Matthews, B. W. (1968). J. Mol. Biol. 33, 491-497.
- Navaza, J. (1994). Acta Cryst. A50, 157-163.
- Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307-326.
- Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R. C., Arruda, P., Leite, A. & Craievich, A. (1998). Nucl. Instrum. Methods A, 405, 159–164.
- Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998). J. Synchrotron Rad. 5(2), 72–76.
- Rietveld, A. W. M. & Ferreira, S. T. (1996). *Biochemistry*, 35, 7743–7751.
- Rietveld, A. W. M. & Ferreira, S. T. (1998). Biochemistry, **37**, 933–937.
- Schliebs, W., Thanki, N., Eritja, R. & Wierenga, R. (1996). Protein Sci. 5(2), 229–239.
- Schliebs, W., Thanki, N., Jaenicke, R. & Wierenga, R. K. (1997). *Biochemistry*, **36**(32), 9655–9662.
- Scopes, R. K. (1977). Biochem. J. 161(2), 253-263.
- Tanaka, T., Kimura, H., Hayashi, M., Fujioshi, Y., Fukuhara, K. & Nakamura, H. (1994). *Protein Sci.* 3(3), 419–427.
- Watanabe, M., Zingg, B. C. & Mohrenweiser, H. W. (1996). Am. J. Hum. Genet. 58(8), 308–316.
- Wierenga, R. K., Noble, M. E. & Davenport, R. C. (1992). J. Mol. Biol. 224(4), 1115–26.
- Wierenga, R. K., Noble, M. E., Vriend, G., Nauche, S. & Hol, W. G. (1991). J. Mol. Biol. 220(4), 995–1015.
- Yüksel, K. U. & Gracy, R. W. (1991). In A Study of Enzymes, Vol. II, edited by S. A. Kuby. Boca Raton: CRC Press.
Conforme o artigo descreve, duas formas cristalinas de TIM de músculo de coelho haviam sido obtidas em condições semelhantes. Os cristais eram como agulhas ou agrupamentos de placas e cresciam em condições muito parecidas, inclusive numa mesma gota (Figura 5.1) Uma das formas foi coletada a temperatura criogênica e a outra a temperatura ambiente. No entanto, uma comparação estrutural (particularmente, fatores de temperatura) mais adequada poderia ser feita se a mesma forma cristalina fosse coletada a duas temperaturas diferentes.



(a) Forma cristalina com 2 dímeros/ASU.

(b) Forma cristalina que contém 1 dímero/ASU

Figura 5.1: Cristais de TIM de músculo de coelho obtidos no LNLS. Cada divisão equivale a 40 μ m, aproximadamente.

5.2 Coletas de dados no ESRF

Durante o estágio realizado na Inglaterra, experimentos de cristalização levaram a cristais que cresceram na mesma forma cristalina para a qual um conjunto havia sido coletado a temperatura ambiente. Como comentamos, havia o interesse em coletar uma mesma forma cristalina a duas tempertaturas diferentes. Ao mesmo tempo, melhores cristais poderiam levar a dados de melhor qualidade e resolução.

Assim, durante o estágio no YSBL, diferentes aditivos foram utilizados em novos ensaios de cristalização e bons resultados foram conseguidos adicionando-se DMSO (dimetil-sulfóxido, aditivo de reconhecida eficácia para melhoramento de cristais) às condições de cristalização originais. Estes cristais (Figura 5.2), levaram a um novo conjunto de dados, coletado no ESRF/Grenoble - França (linha ID14-2).

A partir de diferentes regiões de um único cristal, três conjuntos de dados foram coletados com estratégias de coleta diferentes para alta e baixa resolução. Após cuidadosa redução de dados, obtivemos um conjunto de qualidade geral superior aos anteriores, inclusive com um significativo aumento no limite de resolução para 1.50 Å. Não havia similar para este conjunto, de forma que o par desejado para comparação foi conseguido, levando a resultados muito interessantes, descritos no próximo capítulo, onde também serão comparados os três conjuntos de dados, bem como a resolução da terceira estrutura por Substituição Molecular.

A Tabela 5.1 apresenta em maior detalhe as estatísticas mais importantes para o conjunto coletado no ESRF. Alguns comentários são pertinentes. Em primeiro lugar, o limite de resolução subiu de 2.25 Å no melhor conjunto coletado no LNLS, para 1.50 Å neste novo conjunto. Este aumento também é devido a um empacotamento mais ordenado resultante da presença de DMSO mas não exclusivamente. O uso de síncrotron e instrumentação melhores fo-

Reso	No.ref	$I/\sigma(I) > 3(\%)$	$< I/\sigma(I) >$	Complet	$R_{merge}(I)(\%)$
33.00-4.38	6300	98.3	38.3	96.7	0.024
4.38 - 3.48	5990	98.8	36.5	95.7	0.028
3.48 - 3.04	6160	98.3	31.6	99.6	0.039
3.04 – 2.76	6142	97.2	22.3	99.5	0.057
2.76 – 2.56	5333	96.1	14.4	86.7	0.073
2.56 - 2.41	6097	96.1	16.6	99.9	0.084
2.41 – 2.29	6119	95.7	16.7	99.9	0.093
2.29 – 2.19	6077	94.1	16.5	99.8	0.098
2.19 – 2.11	6076	92.2	12.5	99.8	0.114
2.11 – 2.04	6057	92.1	12.3	99.9	0.122
2.04 – 1.97	6067	88.8	10.0	99.6	0.128
1.97 – 1.92	5965	85.4	8.5	98.2	0.136
1.92 – 1.87	5689	79.5	7.2	94.3	0.153
1.87 - 1.82	5432	77.4	6.4	89.6	0.160
1.82 - 1.78	5480	74.1	6.0	90.5	0.170
1.78 - 1.74	5319	72.6	5.1	87.7	0.192
1.74 – 1.71	5197	70.0	4.7	85.2	0.201
1.71 – 1.67	4902	68.1	4.3	81.6	0.223
1.67 - 1.64	4712	68.2	3.9	78.1	0.239
1.64 - 1.62	4599	67.7	3.8	76.4	0.241
1.62 - 1.59	4553	64.3	3.5	75.4	0.265
1.59 - 1.57	4404	64.1	3.2	72.7	0.267
1.57 - 1.54	4334	62.7	3.0	71.7	0.270
1.54 - 1.52	4323	60.0	2.7	71.8	0.279
1.52 - 1.50	3949	58.8	2.2	65.8	0.327
To dos os hkl	135276	81.0	18.3	88.7	0.066

Tabela 5.1: Estatísticas para o conjunto de TIM coletado no ESRF.



(a) Cristais crescidos sem o uso de aditivos, com dimensão máxima de ~ 0.4 mm

(b) Cristais da mesma forma cristalina exposta em (a) mas crescidos na presença de DMSO. É possível identificar monocristais de tamanho máximo em torno de 0.15 mm

Figura 5.2: Cristais de TIM de músculo de coelho obtidos durante o estágio no YSBL.

ram fatores importantes e não apenas contribuiram para um aumento de resolução mas para a obtenção de dados de maior qualidade geral. Na camada de mais baixa resolução, que inclui dados desde 33 Å até 4.38 Å de resolução, o R_{merge} é de 2.4 %, considerado excelente, com um surpreendente $< I/\sigma(I) >$ de 38.3. Até 1.9 Å o conjunto apresenta praticamente 100% de completeza. Entre 1.9 Å e 1.5 Å, a completeza diminiu mas as outras estatísticas mantém a qualidade e comportam-se da maneira esperada com o aumento de resolução.

A este respeito, deve-se um comentário. Toda a preparação para a coleta no ESRF, que incluiu testes de cristais com a coleta de imagens preliminares (incluindo indexamento) e armazenamento dos melhores cristais em nitrogênio líqüido, foi feita pelo autor que, por motivos financeiros não foi até o ESRF, na França, para realizar a coleta, permanecendo no YSBL, Inglaterra. Os testes realizados no difratômetro *in house* disponível indicavam que um limite resolução em torno de 2.0 Å seria o resultado. A estratégia de coleta, foi então desenhada para este fim. Durante a coleta, os pesquisadores responsáveis perceberam em tempo que seria possível estender um pouco este limite mas, mesmo assim, as reflexões a 1.5 Å só foram medidas nas regiões de canto do CCD (detector da linha ID14-2 do ESRF), que tem a forma de uma pastilha quadrada. A menor completeza nesta faixa de resolução, não compremeteu, entretanto, a qualidade da estrutura final, já que até 1.52 Å, ela ainda é maior que 70%, um valor razoável.

Capítulo 6

TIM: novo paradigma para o *loop* do sítio ativo

As três estruturas de TIM, cuja determinação foi descrita no Capítulo anterior, foram refinadas e estudas em profundidade à luz do farto conhecimento que se produziu sobre Triose fosfato isomerases nas últimas décadas. O resultado mais surpreendente foi a inesperada conformação fechada do *loop* do sítio ativo de um dos monômeros presentes na estrutura de mais alta resolução. Este fato, inédito na literatura, vai em direção contrária ao paradigma aceito até então de que é o ligante que induz a conformação fechada do *loop*. Porém, em nossa estrutura, o *loop* aparece fechado na ausência de análogos do sítio ativo ou quaisquer outros ligantes.

O artigo anexado a seguir, a ser submetido ao *Journal of Molecular Biology* para publicação, descreve as três estruturas (em processo de deposição no PDB) enfocando a discussão na inusitada observação e sugere hipóteses para explicar os fatos. Também foi incluída uma seção com o material extra (figuras) citado no artigo e que deve ser publicado apenas na internet, como *supporting information*.

Closed conformation of the active site loop of rabbit muscle triosephosphate isomerase in the absence of substrate: evidence of conformational heterogeneity

Ricardo Aparicio^a, Sérgio T. Ferreira^b and Igor Polikarpov^{a,*}

^aInstituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador SãoCarlense, 400, CEP13560-970, São Carlos, SP, Brazil

^bLaboratório Nacional de Luz Síncrotron, Campinas, SP and Depto. de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Abstract

The active site loop of triosephosphate isomerase (TIM) exhibits a hinged-lid motion, alternating between the two well defined "open" and "closed" conformations. Until now the closed conformation had only been observed in protein complexes with substrate analogues. Here, we present the first rabbit muscle apo TIM structure, refined to 1.5 Å resolution, in which the active site loop is either in the open or in the closed conformation in different subunits of the enzyme. In the closed conformation described here, the lid loop residues participate in stabilizing hydrogen bonds characteristic of holo TIM structures, whereas chemical interactions observed in the open loop conformation are similar to those found in the apo structures of TIM. In the closed conformation, a number of water molecules are observed at the projected ligand atom positions, that are hydrogen bonded to the active site residues. Additives used during crystallization (DMSO and TRIS molecules and Mg atoms) were modeled in the electron density maps. However, no specific binding of these molecules is observed at, or close to, the active site and the lid loop. To further investigate this unusual closed conformation of the apo enzyme, two more rabbit muscle TIM structures, one in the same and another in a different crystal form, were determined. These structures present the open lid conformation only, indicating that the closed conformation cannot be explained by crystal contact effects. To

rationalize why the active site loop is closed in the absence of ligand in one of the subunits, extensive comparison with previously solved TIM structures was carried out, supported by the bulk of available experimental information about enzyme kinetics and reaction mechanism of TIM. The observation of both open and closed lid conformations in TIM crystals might be related to a persistent conformational heterogeneity of this protein in solution.

Key words:

Triosephosphate isomerase, TIM structure, lid loop conformation, conformational heterogeneity, Protein Crystallography.

1 Introduction

Glycolysis is the biochemical pathway by which glucose is converted to pyruvate with the generation of two mols of ATP/mol of glucose, providing part of the energy utilized by most cells. Triosephosphate isomerase (TIM, Dglyceraldehyde-3-phosphate ketol-isomerase, EC 5.3.1.1) catalyzes the fifth reaction of the glycolytic pathway, namely the interconversion, probably via an enediol (or enodiolate) intermediate, of glyceraldehyde-3-phosphate and dihydroxyacetone phosphate, two ketose-aldose isomers^{1,2,3}. Three-dimensional structures of TIM from various species have been reported, including TIM from prokaryotes^{4,5,6,7,8}, unicellular eukaryotes^{9,10,11,12,13} and higher eukaryotes^{14,15}. With the exception of TIM from *Pyrococcus woesei* and *Methanothermus fer*vidus, which are homotetramers 16 , in all other species so far studied a dimer is the physiologically active species. In addition, a number of monomeric TIM mutants possessing catalytic activity have been described^{17,18,19,20}. Extensive studies of the enzyme kinetics, reaction mechanism and structures of TIM from different sources have resulted in a considerable body of currently available knowledge on structure-function relationships in this enzyme^{21,22,23,24}. TIM is considered a perfect enzyme in the sense that at *in vivo* concentrations of substrate the rate of catalysis is essentially limited by diffusion 24,25 . As expected,

* Corresponding author:

Email address: ipolikarpov@if.sc.usp.br (Igor Polikarpov).

it has been shown that this diffusion-controlled process is sensitive to viscosity ^{25,26}.

The α/β barrel structural motif, frequently named TIM barrel²⁷, a supersecondary structure in which eight β -strands alternate in sequence with eight α -helices forming a parallel β -sheet closed in a cylindrical topology, was first recognized in the structure of chicken muscle TIM¹⁴. The TIM-barrel fold is found in a number of other enzymes exerting a variety of functions. The active site of almost all known α/β barrel enzymes is located at the carboxyl end of the barrel. The fact that most of the members of α/β barrel proteins are enzymes serves as an argument for their divergent evolution from a common ancestor²⁸.

TIM is also important for its potential in drug design, as several parasites causing tropical diseases depend on glycolysis as a source of energy. For example, TIM is being targeted in attempts to obtain a vaccine against schistosomiasis ²⁹. The structure of TIM from *Trypanosoma brucei brucei* (the etiologic agent of sleeping sickness) has also been actively investigated, aiming to develop inhibitors that might be of therapeutic value against infection^{10,30}. Another aspect of medical importance is TIM deficiency in humans, a genetic disease of autosomal dominant inheritance (locus 12p13), causing developmental retardation, myopathy, cardiac failure and anemia^{31,32}.

The active site loop of triosephosphate isomerase (also called the flexible loop, the lid loop or loop 6) comprises about ten residues and exhibits a hinged-lid motion between two well defined conformations. These conformations, termed open and closed, have been described in detail^{9,30,33,34,35,36,37}. Acting as a rigid lid which rests on border residues that function as hinges, the loop moves up to 8 Å upon swinging from the open to the closed conformation. Concomitantly, the position of the main catalytic residue, Glu165, is shifted by about 2 Å. In this fashion, bulk solvent is excluded from the active site and key residues are repositioned for catalysis. Various members of the TIM barrel family possess a flexible loop at the active site with functions similar to those of loop 6 in TIM^{33,38,39,40}.

Rabbit muscle TIM, a dimer of identical subunits with a molecular weight

of 2×26.6 kDa, was one of the first glycolytic enzymes to be purified 41,42,43 . We have now determined three structures of apo rabbit muscle TIM, named A1, A2 and B and refined them to 1.50 Å, 2.85 Å and 2.25 Å, respectively. The A1 and A2 structures contain two dimensions in the asymmetric unit (ASU) displaying a molecular packing which had not been previously described. As expected, the overall structure of rabbit muscle TIM is similar to other known TIM structures. Intriguingly, however, analysis of the highest resolution structure (A1) reveals that the active site loop in one of the subunits of TIM is in the closed conformation. This is a surprising finding as no ligand of TIM was present in the crystallization conditions and, indeed, no ligand was observed at the active site of TIM in that structure. Analysis of the two other structures, A2 and B, showed that the lid loop is not constrained in the closed conformation by crystal contacts. We have investigated the possible reasons for the active site loop closure and its stabilization in the absence of ligands and discuss these results in light of previously reported structural, kinetic and thermodynamic studies. This analysis leads to the conclusion that persistent conformational heterogeneity of TIM in solution might be at the origin of the observed closed loop conformation in the apo form of the enzyme. This also offers a plausible explanation for the peculiar behavior of the active site loop in previously determined TIM structures.

2 Results

2.1 Overall structural features of rabbit muscle TIM

A monomer of rabbit (*Oryctolagus cuniculus*) muscle TIM has 248 amino acids, with Lys13, His95 and Glu165 being the key residues responsible for catalysis. The active site loop segment comprises residues Pro166 to Ala176. Structural studies have shown that this loop constitutes a rigid lid structure resting on flange residues that function as hinges, capable of assuming at least two well-defined conformations. Usually, the open conformation has been described in apo structures and the closed conformation in the holo structures of complexes with substrate analogues, phosphate-containing ligands^{9,15,30,33,34,35,36,37}. The N-terminal hinge (residues 166-168) and C-terminal hinge (174-176) are highly conserved across 81 species⁴⁴. Glu165, the catalytic base, moves by about 2 Å upon the loop closure. Its ideal position for catalysis is in the "swung in" conformation (Chi1 = 60°). The "swung out" conformation, in which Chi1 = -60° and Glu165 is not properly positioned for catalysis, has also been reported^{10,15,30}. Rabbit muscle and human TIM share 98% sequence identity and the same number of amino acids. The four mutations are located in surface regions of the molecule and none of them is situated at the dimer interface or is implicated in catalysis.

2.2 Crystal packing

Crystal form B contains a dimer in the ASU, in a molecular packing similar to that observed for human TIM¹⁵. The tetramers occupying the ASU of the A1 and A2 structures cannot be obtained by application of space group symmetry operations to the dimer of crystal form B. In other words, the dimeric interfaces in crystal forms A and B are not identical. In crystal form B, the largest contact areas are found between the A chains of crystallographically related dimers, whereas in crystal form A, the same occurs for chains B and D of the same ASU. It turns out that in the latter crystal form there are four subunits with distinct crystallographic environments, while only two subunits exist in the former. In line with the usual nomenclature, subunits 1 and 2 will be referred to as chains A and B, respectively. Form A contains two dimers in the ASU arranged in a novel molecular packing. Subunit 2 (chain B) of crystal structure A1 has the active site loop in the closed conformation (as discussed below) and forms a dimer with subunit 1 (chain A), which exhibits an open lid conformation. The lid loops of both subunits of the second dimer in the ASU of the A1 structure (chains C and D) are in the open conformation. In contrast, all four active site loops of the four subunits in the A2 structure, determined in the same crystal form, are in the open conformation. Although A1 and A2 belong to the same crystal form, subtle differences in molecular packing can be observed between them. The molecular packing of the A1 crystal is denser and the solvent content is lower than in the A2 crystal. The four chains of the A1 and A2 structures have a comparable number of intermolecular contacts within the ASU. The number of intermolecular contacts with molecules outside the ASU (space group symmetry matches) is, however, very different. The largest difference is observed for the B chain, which makes 194 contacts in A1 against only 17 contacts in the A2 structure. In addition, the number of crystal contacts of the active site loop in subunit 2 (chain B) of the A2 structure is significantly lower than in the other three subunits (Supplementary Material).

2.3 Structural comparisons

Structural comparison of the three crystal structures described here with previously reported structures of apo TIM shows that the active site loop is in the open conformation in all subunits of the present three structures, except in subunit 2 (chain B) of the A1 structure (Figure 1). Furthermore, comparison with previous structures of TIM complexed with several substrate analogues shows that the active site loop in chain B of our A1 structure is indeed in the closed conformation. In addition, a conformational change of the catalytic residues is observed in chain B of the A1 structure, as compared to chains A, C and D.

A1 is not the first TIM structure in which the active site loops are found in different conformations in different subunits. In the structure of human TIM, which was complexed with 2-phosphoglycolate and solved in a crystalline form identical to the crystal form B described here (PDB entry 1HTI¹⁵), ligand binding and active site loop closure were observed only in subunit 2. Another example is trypanosomal TIM, whose structure was determined from crystals grown in the presence of 2.4 M ammonium sulfate¹⁰. In this apo structure (PDB entry 5TIM), the lid loop of subunit 1 is in the open conformation, whereas in subunit 2 it is in an "almost closed" conformation¹⁰. In that case, the loop closure was related to binding of a sulfate ion near the active site of subunit 2¹⁰. Soaking in glycerol-3-phosphate (G3P), a substrate analogue, in a sulfate-free mother liquor, led to a complex (PDB entry 6TIM¹⁰) where the loop is found in the open and closed conformations in subunits 1 and 2, respectively. Comparison between the conformations of subunit 2 in the G3P complex and in the current A1 apo structure shows that the closed and

"almost closed" conformations are superposable within about 0.43 Å RMS deviation (Table 1). Significant conformational difference is observed for the side chain of the catalytic residue Glu167, for which the Chi1 angle is equal to -70° in the G3P complex (closed conformation) in contrast to a Chi1 = 69° in the "almost closed" conformation of the apo structure. Yet another example of conformational diversity of the active site loop is the structure of trypanosomal TIM in complex with phosphoglycolohydroxamate (PGH), a tight-binding inhibitor structurally analogous to an enediolate intermediate in the reaction pathway. Although the crystal ASU contains the functional TIM dimer, inhibitor binding and the closed lid conformation are observed only in subunit 2 (PDB entry 1TRD⁴⁵).

2.4 Closure of the lid loop: chemical interactions and conformational variability

To determine whether the active site loop conformation of subunit 2 (chain B, Figure 2a) in our A1 structure indeed corresponds to the observed closed conformation in the G3P complex, structural alignment between these two TIM structures was carried out. The active site loop regions in both structures are superposable within an average RMS value of 0.39 Å (Table 1 and Figure 2b). For comparison, the subunits of the G3P and PGH trypanosomal TIM complexes (PDB entries 6TIM and 1TRD, respectively) where the loop is observed in the closed conformation, have also been superposed. Table 1 summarizes the RMS deviations per residue for the active site loop region, and indicates that comparable values were obtained for both superpositions. Higher deviations were obtained in the superposition of subunits from other TIM complexes (including superpositions of different complexes of the same TIM species) where the active site loop is in the closed conformation (not shown). This conformational variability⁴⁶ is likely to be a result of the inherent flexible nature of the loop, as well as different crystallographic environments. In addition to a low overall RMS value, the superposition of the subunits 2 of the A1 apo structure and of the TIM-G3P holo structure shows that the catalytic base Glu165 (in rabbit muscle TIM) adopts very similar "swung out" conformations in both the B chain of A1 apo structure (Chi1 = -50°) and in

the TIM-G3P complex (Glu167, Chi1 = -70°). Indeed, structures of several TIM complexes with substrate analogs have been reported in which the catalytic base assumes Chi1 values ranging from -79° (PDB entry 2YPI³⁵) to -29° (PDB entry 1TRD⁴⁵) when the lid loop is in the closed conformation.

An additional conformational change of the catalytic base Glu165 in the A1 structure accompanies closure of the B chain active site loop. In the A chain, the side chain oxygen $O_{\epsilon 1}$ of Glu165 is at hydrogen bonding distance (2.7 Å) from the backbone nitrogen of Ser96. In the B chain, this hydrogen bond is broken. In comparison to chain A, Glu165 C_{δ} is shifted by 2.7 Å while the backbone carbons move on average about 0.5 Å, distances which are in excellent agreement with reported values for other TIM structures^{9,34,35}. The new positions of the catalytic residues in chain B allow for additional water molecules to be hydrogen bonded to the active site residues. These water molecules occupy positions of the ligand atoms observed in the G3P complex (Figure 3). In this regard, it is noteworthy that water molecules have been shown to be important for the catalytic efficiency of triosephosphate isomerase ⁴⁷.

Another remarkable fact is that the loop B in the A1 structure exhibits the hydrogen bonds characteristic of a closed conformation (Figure 2a). Thus, in chains A, C and D, Trp168 has its indole nitrogen hydrogen bonded to the hydroxyl group of Tyr164, a particular feature of the open conformation which does not occur in the loop region of chain B. Instead, the same indole nitrogen participates in a new hydrogen bond to the carboxyl group of Glu129 (2.88 Å distance). Concomitantly, the side chain of Tyr164 is rotated as compared to the other chains and occupies a position above the active site loop. Additional hydrogen bonds, specific for a closed conformation and absent in chains A, C and D, are observed in the flexible loop region of chain B between the Ala176 amide nitrogen and the Tyr208 hydroxyl group (2.96 Å distance) and between Gly173 and Ser211 (3.06 Å distance).

The B chain of the A1 structure was also superposed, with similar results, to a number of other complexes in which the loop was found in the closed conformation, including subunit 2 of human TIM complexed with 2-phosphoglycolate (PDB entry 1HTI¹⁵). In summary, our results strongly suggest that the active

site loop of the B chain in the A1 structure assumes neither the open nor the "almost closed" conformation previously described for TIM structures¹⁰. Instead, the catalytic residues and the active site loop are in the very same closed conformation as they are found in other structures in the presence of substrate. For these reasons, the conformation of the B chain lid loop in the A1 crystal structure will be referred to as a closed conformation in the following sections.

3 Discussion

3.1 Loop conformations and crystal packing

In various complexes of trypanosomal TIM with different phosphate and phosphonate inhibitors, obtained by soaking of previously grown crystals or by co-crystallization, the active site loop is found in different conformations in subunits 1 and 2. By contrast, in the complexes of yeast TIM obtained by co-crystallization with 2-phosphoglycolate (PDB entry 2YPI³⁵) and phosphoglycolohydroxamate (PDB entry 7TIM³⁴), the active site loop is found in the closed conformation in both subunits. Interestingly, in the apo yeast structure (PDB entry 1YPI⁹), determined in the same space group of the former two complexes, the loops of both subunits 1 and 2 exhibit the open conformation and have the same crystalline contacts. Yet another example of the possible influence of the crystallization process on the loop conformations is the structure of trypanosomal TIM complexed with N-hydroxy-4-phosphonobutanamide (PDB entry 1TSI⁴⁸). This structure was determined by soaking of the crystals (grown in the same crystalline form as those of the TIM-G3P complex mentioned above) with the ligand. Remarkably, both active site loops of the dimer are observed in open conformations even though inhibitor is observed in both active sites. Moreover, in contrast with other complexes in which ligand binding occurs only in subunit 2, in this structure the active site is fully occupied by ligand only in subunit 1 and has much lower occupancy in subunit 2.

It is, therefore, difficult to state exactly how the crystallization conditions interfere with the protein conformation in solution, leading to different crystalline structures. However, careful analysis of the available data confirms that, in the A1 structure, the subunit 2 lid loop is not stabilized in the closed conformation by crystal contacts. The following observations support this statement. First, the contact distances of the side or main chains of the lid loop residues to the atoms of the nearest subunit are larger than 4.3 Å. Second, replacement of the B chain by a subunit in which the lid loop is in the open conformation (chains A, C or D) does not cause steric clashes with neighboring chains, indicating that the loop could potentially adopt an open conformation (Figure 4 and Supplementary Material). Furthermore, the active site loop of subunit 2 is closed in the A1 structure but is open in the A2 structure, determined in the same crystal form and grown under very similar conditions. Taken together, these observations leave very little doubt that crystal contacts are not responsible for the active site loop closure.

3.2 Timescale of loop motion and crystal growth rate

To understand the possible cause of the lid loop closure and its stabilization in the absence of ligand, let us consider the crystallization process itself and compare the time scales for lid loop opening and closing and the period of time necessary for incorporation of a TIM dimer into the growing crystal. McDermot and collaborators⁴⁹ recently showed that the frequency of TIM active site loop opening and closure is on the order of 3000–5000 s^{-1} near room temperature (with a rate that strongly depends on temperature). The influence of G3P on the loop opening/closing rate has also been examined. Open and closed conformations are nearly equally populated at subsaturating concentrations of G3P (1.2 mM). In saturated solutions (10 mM G3P), approximately 85% percent of the enzyme is bound to the ligand. Computer simulations in both cases confirmed the measured data and showed that in the case of a saturated solution approximately 1% of the protein molecules stay as an encounter complex (with the ligand bound, but the lid loop open). It is interesting to note that the open loop with ligand bound previously observed by Verlinde and colaborators⁴⁸ most probably corresponds to the encounter complex⁴⁹.

On the other hand, the molecule intake speed during protein crystal growth has also been estimated. According to McPherson and collaborators⁵⁰, the tangential growth speed of trypsin layers during crystallization is approximately 5 nm/s (as estimated from eq. 1 and Figure 5 of that work). This value is in excellent agreement with studies carried out for lysozyme, for which dimeric layers grow at a tangential speed of about 5 nm/s and with a much smaller speed of 0.28 nm/s in the direction normal to the surface⁵¹. Although these growth rates are likely dependent on the very protein under study, the concentration of the protein solution, the supersaturation state⁵², irregular geometries of protein molecules, different orientational probabilities and other factors, a rough estimate of the time necessary for one TIM dimer to be incorporated by the crystalline mass ($\sim 0.5s-1.0s$) is much larger than the time scale for the opening/closing motions of the flexible loop. Moreover, regarding the growth mechanism, it is also possible that TIM tetramers are formed in solution and only afterwards are incorporated in the crystal. Experimental evidence for this type of mechanism was found in the case of lysosyme, for which tetrameric growth along the crystallographic screw axis was observed⁵¹. Clustering of protein molecules while still in solution, with the formation of large crystallization units occurring randomly in the bulk solution, is considered to be the most plausible growth mechanism⁵¹.

3.3 Lid loop conformers and ligand binding

Is the ligand indispensable for the active site loop closure of TIM? In a very interesting experimental study, McDermot and collaborators have shown that the active site loop motion is not ligand-gated but is, in fact, a natural motion of the protein, and the loop motion rate between the open and closed states matches the catalytic turnover rate⁵³. The affinity for ligands is significantly smaller in the open state than in the closed state, and the transition

state has an intermediate affinity. These data suggest that ligands have a pronounced effect on the relative populations of the two loop conformations. As a consequence, the apo enzyme is mostly in the open conformation and the conformation of the holo protein is mostly closed. Thus, although the crystal growth intake rates for the different populations of TIM molecules in solution are likely to be somewhat different, the much larger fraction of apo TIM in the open conformation and of holo TIM in the closed conformation determine the outcome of the crystallization process.

In the crystal, therefore, an apo TIM molecule mostly appears in the open conformation and its complexes with diverse ligands display a closed conformer of the active site loop. However, under particular circumstances, a higher crystal growth intake rate of one of the species present in solution at lower concentrations (for example, apo protein in the closed conformation or holo protein in the open conformation) may be significant during crystallization and thus counterbalance their smaller proportions in solution. In such case, the crystal may work as a "filter", selectively incorporating specific conformers from the overall ensemble of TIM molecules in solution. This may explain the observation of a closed conformation of the B chain of the apo A1 structure presented here, the "almost closed" conformation previously reported for apo TIM from *Trypanosoma brucei brucei* (PDB entry 5TIM¹⁰), and the open conformation of trypanosomal TIM complexed with N-hydroxy-4-phosphono-butanamide (PDB entry 1TSI⁴⁸).

3.4 The role of DMSO

The A1 and A2 structures were determined using crystals grown in identical crystal forms under very similar conditions (*Materials and Methods*). The only significant difference between them is related to the presence of DMSO in the A1 crystallization solution. This led to a slight shrinkage of the A1 crystal cell, its denser packing, smaller solvent content (Table 2) and alteration of intermolecular contacts with symmetry related molecules (Supplementary Material). Since a closed loop conformation was observed in the A1 structure, but not in the A2 structure, it was important to investigate the influence of DMSO on the crystallization process and on the lid loop stabilization. A1 is not the first TIM structure for which DMSO was used as a crystallization agent and in which DMSO molecules were observed in the electron density maps. The trypanosomal TIM structure (PDB entry 1TPF⁵⁴) was determined from crystals grown in the presence of 1.4 M of DMSO (a concentration similar to the 1.1 M used in the present work). In that structure, each TIM subunit had two equivalent DMSO binding sites⁵⁴. Neither of those two DMSO binding sites were observed in our A1 structure. DMSO is an organic solvent that lowers the bulk dielectric constant of the solvent through hydrogen bonding to water molecules. The importance of the bound solvent in the TIM active site was discussed by Zhang and collaborators⁴⁷. DMSO has a large dipole moment for a small molecule and the ability to attach to accessible hydrophobic portions of the protein. Its effects on lysozyme crystals (soaked in 15% (v/v) DMSO, compared to 10% used during crystallization of form A1) have been studied with single-crystal neutron diffraction⁵⁵ and hydrophobic interactions seemed to be an essential factor for DMSO binding to the protein surface. Although there is no DMSO directly bound to the loop in the closed conformation of structure A1, such molecules might interfere with the active site conformation by other mechanisms. Shan and Herschlag⁵⁶ have recently proposed that hydrogen bonding to groups that undergo charge rearrangement during enzymatic catalysis is strengthened proportionally to the strengthening of the hydrogen bonds in the corresponding solution reactions. Part of this effect was attributed to the relatively lower effective dielectric environment provided by the active site relative to that in aqueous solutions. It might, therefore, be possible that the influence of DMSO on the solvation of the A1 structure might lead to a preferential loop stabilization in the closed conformation without strong direct binding to the lid loop region. In addition, slight, but distinct DMSO-induced rearrangements of the molecules in the A1 ASU as compared to A2 crystals could increase the specific incorporation of the minor closed apo TIM species into the growing crystal.

3.5 Evidence of structural heterogeneity of TIM in solution

Although the possible influence of DMSO on crystal growth may be involved in the B chain loop closure in the A1 structure, this effect is not general. TIM from Trypanosoma brucei brucei (PDB entry 1TPF⁵⁴), whose structure was also determined in the presence of DMSO, has the lid loop in the open conformation despite the clearly defined DMSO binding sites observed in the structure. Thus, DMSO might be necessary, but not sufficient for the acquisition of a closed active site loop conformation in the present apo A1 structure. Along the same line, the trypanosomal TIM holo structure, PDB entry 1TSI 48 , clearly shows the lid loop in the open conformation together with a ligand bound to the active site. Another paradoxical fact is that complexes of yeast TIM have been crystallized with both loops in the closed conformation (PDB entry 1YPI⁹), while the crystal of a human TIM holo structure (PDB entry 1HTI¹⁵) shows dimers in which the loop is observed in either the open or closed conformation in different subunits, in spite of the fact that in both yeast and human TIM experiments the inhibitor was co-crystallized with the protein. It is important to note that not only the structure of the active site loop but also residues important for catalysis are highly conserved across the various TIM species.

One possible answer to these questions is the existence of structural differences in the TIM dimers in solution, e. g., leading to different ligand binding affinities in solution, which could be specifically selected during the crystallization process. In this regard, it is known that the crystallization process can be seen as a molecular separation technique. The existence of heterogeneous conformational populations in chemically homogeneous protein samples has been reported for a number of proteins^{57,58,59,60,61,62,63,64}. In the particular case of rabbit muscle TIM, a persistent conformational/thermodynamic heterogeneity has been observed in solution studies using hydrostatic pressure and chaotropic agents as denaturants^{62,63}. The existence of a conformationally heterogeneous ensemble of TIM dimers in solution was postulated on the basis of these results^{62,63}. Remarkably, the energy barrier for interconversion between different conformers was found to be 99 kJ/mol⁶³. Such an energy barrier will effectively "freeze" the conformational species, setting the time of conformational exchange between them to hours or days. In this regard, it is interesting to note that an early solution NMR study of the dynamics of loop opening/closing of TIM indicated that the rate of interconversion between the two conformations was indeed quite slow, with a corresponding activation energy of 144 kJ/mol and a characteristic time of 11 hours at 40 °C⁶⁵. This persistent conformational heterogeneity is distinct from the conformational variability described by McDermot and collaborators⁵³ in the sense that, in the latter study, the energy barrier involved in the active site loop opening and closing was found to be significantly lower (about 50 kJ/mol), corresponding to a significantly higher loop motion rate (in the kHz range). In a way, such fast loop motion could be considered as a source of dynamic conformational heterogeneity. Thus, an interesting question is whether the dynamic or persistent structural heterogeneity of TIM molecules in solution, or a combination of the two, can provide an explanation for the closed loop conformation observed in the present study. On the basis of the present crystallographic structures alone, we cannot distinguish between these possibilities. The observed conformational differences between the open and closed loop conformations of apo rabbit muscle TIM could be related to the persistent heterogeneity of the rabbit muscle TIM dimers in solution 62,63 , but, given the magnitude of the energy barrier involved, this might not be the only reason. Conceivably, it could also result from selection and stabilization of the closed conformation of the flexible lid loop by the crystallization process. More experiments aiming to establish the role of ligands in stabilization of the conformation of the active site of TIM and the time scales involved in a variety of conformational changes of TIM molecules in solution are needed in order to answer this question.

3.6 Concluding remarks

In the present work, we describe three rabbit muscle apo TIM structures determined in two different crystal forms. Surprisingly, one of the structures, crystallized in the presence of DMSO and refined to 1.5 Å resolution, shows, simultaneously, the active site loop in open and closed conformations in different subunits. In the apo structure closed conformation, the lid loop residues participate in stabilizing hydrogen bonds characteristic of holo TIM structures. Distinct binding sites for the additives used in crystallization (DMSO and TRIS molecules and Mg atoms) were observed and modeled in the electron density maps. However, none of them is observed near the lid loop or at the active site. Removal of DMSO from the crystallization conditions resulted in a crystal in which the active site loop is in an open conformation in all the subunits of the asymmetric unit of the crystal. Careful analysis of all three structures confirmed that the observed closed conformation cannot be explained by crystal contacts. Instead, the observed closed loop conformation of the chemically homogeneous apo TIM population might be directly related to conformational heterogeneity of TIM in solution. More structural and biophysical studies of the protein in solution are needed to clearly determine whether this heterogeneity is of dynamic or persistent/quasi-static nature.

4 Materials and Methods

4.1 Crystallization, data collection and structure solution

Crystallization, data collection and structure solution of rabbit muscle TIM in two crystalline forms has been described earlier⁶⁶. Data were collected at room temperature for crystal form A and at a cryogenic temperature for crystal form B (following the new nomenclature adopted in the present work). The resulting structures were named A2 and B, respectively. Better diffracting crystals of the form A were obtained by adding dimethyl sulfoxide (DMSO, $C_2H_6O_1S_1$), to the original crystallization conditions. Briefly, a solution containing 24% PEG 4000, 0.2 M MgCl2, 0.1 M Tris–HCl, pH 8.5, was prepared. A final solution containing 1000 μ l of this solution plus 100 μ l of DMSO (final concentration $\simeq 1.1$ M) was used as reservoir solution in the hanging drop method. The reservoir solution was mixed with protein solution (10 mg/ml in water) in equal amounts and equilibrated against the reservoir solution (final protein concentration $\simeq 0.2$ mM). Crystals were grown within 1-2 weeks at room temperature. To prevent radiation damage, crystals were briefly soaked in a cryoprotectant solution containing 10% (v/v) ethylene glycol and rapidly cooled to 100 K⁶⁷. Diffraction patterns were assessed first by using an in-house diffractometer. Data were collected at ESRF, beamline ID14-2, equipped with an ADSC Q4 CCD detector. Images were processed using Denzo and Scalepack packages⁶⁸. The new structure, name A1, was determined by molecular replacement using a partially refined (to 2.25 Å resolution) dimer of form B as a search model. Calculations using 10–4 Å resolution limits and default parameters in Amore^{69,70,71} led to clear solutions for 2 dimers in the ASU. After rigid body refinement, a correlation coefficient of 73.5% and R_{factor} of 32.4% were obtained. A summary of crystal parameters and data collection statistics for all three data sets is presented in Table 2.

4.2 Refinement and quality of the structure

The structures were refined using Refmac^{70,71,72}. In the three structures, TLS parameters^{73,74} were established before maximum likelihood atomic refinement. Tight, medium and loose NCS restraints for main and side chains were set up taking into account different conformations and disordered regions. The *weight_matrix* parameter, which controls the relative weights of experimental X-ray data and biochemical restraints, was adjusted according to the high resolution limit of each data set. Default bulk solvent correction was employed. Initially, solvent molecules were located with Arp_Warp^{70,71}. In all cases, further removal and addition of new water molecules was done manually using the *water_pekpik* utility implemented in the program O^{75,76}. Model refinement and quality statistics for the three structures are summarized in Table 2. No residues were found in disallowed regions of the Ramachandran plot. The only non-glycine residue observed in generously allowed regions was Lys13. This residue probably adopts an unfavorable conformation to stabilize the intermediate in the reaction pathway⁷⁷.

The N-terminal residue, in addition to the last but one residue in some chains, was not visible in any of the structures. The C-terminal residue was modeled in all structures except in chains C and D of structure A2. Some regions contained disordered side chains, mainly lysine residues in the A2 structure for which data were collected at room temperature to a modest resolution. On

the other hand, a number of side chains, mainly serine, leucine and isoleucine residues were modeled in double conformations, mostly in the A1 structure, refined to 1.5 Å. Specifically, the N_{ζ} of Lys13 and O_{γ} of Ser96 in the A1 structure B chain were modeled in double conformations.

4.3 Modelling of ligands

A number of ligands, including Mg ions with coordinations 2 and 6 and DMSO and TRIS molecules, were located by visual inspection in the A1 structure and, in some cases, using the programs WASP⁷⁸ and XPAND⁷⁵ (Figure 5). A detailed description of all the interactions between protein molecules and heteroatoms, as well as their modes of binding, is beyond the scope of this paper. Thus, we restrict ourselves to report the most important features. An interesting feature of the A1 structure is the contact between residues of the active site (His95, Lys13 and Glu165) and water molecules. At an early stage of refinement, in an attempt to explain the stabilization of the closed conformation of the active site loop (see *Results*), the electron density at this location was putatively modeled as a DMSO molecule. However, as the refinement proceeded, clear indications of a wrong solution were found (no acceptable bond distances and unsatisfactory difference maps). There was also no indication of the presence of the DMSO sulfur atom in the anomalous diffraction maps. Furthermore, Fourier synthesis and difference electron density maps did not indicate any significant occupancy for a DMSO molecule in this location. Therefore, the DMSO molecule was removed and water molecules were modeled instead. The calculation and graphical analysis of crystal contacts with the ligands (DMSO, TRIS and Mg ions) show that a number of ligands related by non-crystallographic symmetry seem to stabilize mobile regions at the surface of the two dimers contained in the ASU. Also, ligands (including those related by NCS) are responsible for a more compact packing within the unit cell.

Electron density maps in all the loop regions were very well defined, especially in the case of the A1 structure. With the exception of a few double conformations for side chains, no ambiguities were found in these regions. The overall profile of the B-factor plot for the four chains in the A1 structure (Supplementary Material) is the same as observed in previously reported TIM structures, with higher B-factors corresponding to loop regions. In the A1 structure, the active site loop regions of chains A and C show higher B-factors when comparing to the same regions in chains B and D. Thus, comparing chains A and B, the active site loop region shows higher B-factors in chain A, and, at the same time, chain C exhibits higher B-factors than chain D. In the A2 structure, chains A and B forms a dimer which has significantly higher overall B-factors than the dimer formed by chains C and D (Supplementary Material). This difference in the overall B-factors of dimers AB and CD in structure A2 can be explained by a looser packing of the former. This feature is not observed in the A1 structure, which has a more compact packing, probably as a result of the inclusion of DMSO in the crystallization conditions (*Results*). In the case of crystal form B, the active site loop region in subunit A is stabilized by crystal contacts and has much lower B-factors than in chain B.

Most of the ligands modeled in the A1 structure have B-factor values in the same range as those observed for loop regions in the protein. In two DMSO molecules, with low B-factors, the sulfur atoms were also clearly confirmed by a peak search in an anomalous map, despite the fact that the anomalous signal of S atoms is weak for the wavelength used for data collection. Conversely, there were cases of ligands for which the B-factors are high, with values around 60-75 $Å^2$. High values could be expected in view of the rotational freedom and disorder of these molecules. We opted for keeping these ligands in the final model based on the fact that electron density and difference Fourier maps were clear in the corresponding regions. All structural superpositions were carried out using Lsqman, which gives globally optimal structural alignments ^{75,79,80}. Molecular contacts were calculated using Contact^{70,71} with a default distance cutoff of 5 Å.

Table 1

RMS distance per residue in the active site loop region. The first column refers to the amino acid sequence of rabbit muscle TIM. A single mutation (Thr175 \rightarrow Val177) is observed in the sequence of trypanosomal TIM, for which the corresponding residues in the loop region are Pro168–Ala178. PDB entries 6TIM¹⁰ and 1TRD⁴⁵ correspond to trypanosomal TIM complexes with G3P and PGH, respectively. Both G3P and PGH complexes present the loop in the closed conformation. Superpositions have been carried out using subunit 2 of each structure. The second column shows the RMS distance per residue obtained when superposing the current A1 structure onto 6TIM structure. The third column presents, for comparison, the RMS distance per residue resulting from the superposition of the 6TIM structure onto 1TRD.

	RMS distance per residue (Å)			
Aminoacid sequence	A1 vs. 6TIM	6TIM $vs. 1$ TRD		
Pro166	0.39	0.50		
Val167	0.33	0.45		
Trp168	0.05	0.16		
Ala169	0.19	0.31		
Ile170	0.17	0.33		
Gly171	0.76	0.69		
Thr 172	0.33	0.87		
Gly 173	0.64	0.69		
Lys174	0.54	0.41		
Thr 175	0.58	0.37		
Ala176	0.36	0.17		
Average	0.39	0.45		

Table 2 Data collection and refinement statistics a

	Structure			
	A1	A2	В	
Data collection statistics:				
Aditive	DMSO	_	—	
Temperature	$100 \mathrm{K}$	$298~{\rm K}$	$85~{ m K}$	
Number of images	260	50	75	
Space group	$P2_{1}2_{1}2_{1}$	$P2_{1}2_{1}2_{1}$	$P2_{1}2_{1}2_{1}$	
ASU contents	2 dimers	2 dimers	1 dimer	
Solvent content (%)	44	48	41	
Cell dimensions (Å):				
a	74.68	73.02	65.14	
b	76.14	79.80	72.45	
с	166.70	172.75	93.24	
Resolution range (Å)	32.79 - 1.50	14.00 - 2.85	13.5 - 2.25	
Last resolution shell $(Å)$	1.52 - 1.50	2.92 - 2.85	2.3 - 2.25	
Rsym $(\%)$	6.7(32.7)	12.4(26.4)	10.7 (36.6)	
$< I/\sigma(I) >$	18.34(2.25)	5.51(1.99)	8.39(2.14)	
Refinement Statistics:				
No. of reflections in refinement	127887	18929	19784	
Completeness $(\%)$	89.10	81.63	97.53	
No. of reflections used for Rfree (5%)	6747	1026	1032	
R-factor - all reflections $(\%)$	15.73	17.40	18.41	
R-factor - excluding test set $(\%)$	15.56	17.17	18.22	
R-free $(\%)$	18.95	21.19	22.00	
protein atoms	7657	7486	3755	
water molecules	1118	135	337	
atoms in DMSO molecules	68	0	0	
atoms in TRIS molecules	42	0	0	
Mg ions	16	0	0	
Average B-factor for protein atoms $(Å^2)$	20.48	52.09	35.03	
RMS deviations from ideality:				
Bond lengths $(Å)$	0.014	0.012	0.011	
Bond angles (degrees)	1.574	1.309	1.358	
Ramachandran analysis:				
most favoured regions $(\%)$	93.3	91.9	93.6	
additional allowed regions $(\%)$	6.7	7.7	6.2	
generously allowed regions $(\%)$	0.0	0.5	0.2	
disallowed regions $(\%)$	0.0	0.0	0.0	

 $^{a}\mathit{values}$ in parentheses refer to the last resolution shell











Figure 3.





(b)



Figure 3—continued.

(c)



Figure 4.









Figure Captions.

Figure 1. Cartoon representation of the superposition of the subunits present in the assimetric unit of the A1 structure. The active site loop (Pro166–Ala176) is colored in red in subunit 2 (chain B) and in different hues of blue in the other three subunits. The conformation of the subunit 2 lid loop is clearly distinct from the open conformation of the active site loops of the other three subunits. Figures were prepared using PyMOL (DeLano Scientific, San Carlos, CA, http://www.pymol.org) under Linux.

Figure 2. Stereo drawings of the electron density map and structural alignement of the active site loop region. Water molecules are omitted for clarity. Labels refer to the A1 structure, subunit 2 (chain B). A single conformation of Lys13 N_{ζ} (Materials and Methods) was adopted to prepare the Figure. (a) The active site loop comprises Pro166–Val167–Trp168–Ala169– Ile170-Gly171-Thr172-Gly173-Lys174-Thr175-Ala176. The residues responsible for catalysis are Lys13, His95 and Glu165. The electron density map $(2mF_{obs} - DF_{calc})$ was plotted at 1.25σ . Three hydrogen bonds (dashed lines in green) characteristic of the closed conformation are shown: Trp168 $N_{\epsilon 1} - O_{\epsilon 2}$ Glu129 (distance of 2.88 Å); Ala176 N–OH Tyr208 (distance of 2.96 Å); Gly173 $N-O_{\gamma}$ Ser211 (distance of 3.06 Å). (b) Superposition onto the A1 structure of the same region from subunit 2 of the trypanosomal TIM structure complexed with G3P (PDB entry 6TIM¹⁰, carbon atoms colored in gray), in which the loop assumes the closed conformation. The G3P molecule bound to the active site of the TIM-G3P complex is shown with the phosphorous atom colored in brown. Except for the side-chain of Lys174 of rabbit TIM, which points to a solvent region, the residues occupy nearly identical positions in both structures. Drawings were prepared using PyMOL (DeLano Scientific, San Carlos, CA, http://www.pymol.org) and edited using GIMP (http://www.gimp.org) under Linux.

Figure 3. Stereo views of the active site region in subunit 2 (chain B) of the A1 structure. A single conformation of Lys13 N_{ζ} and Ser96 O_{γ} (Materials and Methods) was adopted to prepare the Figure. (a) Electron density map $(2mF_{obs} - DF_{calc})$ countored at 1.25σ . The positions of water molecules interacting with residues in the active site region are also shown. The Figure illustrates in detail the well-defined conformation of main- and side-chain residues derived from experimental data. (b) The same region of trypanosomal TIM structure (carbons atoms colored in gray) complexed with G3P (PDB entry 6TIM¹⁰, subunit 2) superposed onto the A1 structure. For clarity, water molecules are not labeled. The residues shown are strictly conserved and superpose very well, including the catalytic residue Glu165 side-chain. Note that the position of the G3P molecule is occupied by water molecules in the A1 structure. Three water molecules, W65, W98 and W488, labeled in (a), have nearly identical positions in both structures. Two additional water molecules, W108 and W377, are very close to the carbon hydroxyl groups of G3P (distances of 0.62 Å and 0.88 Å, respectively) and a third molecule, W171, is at a distance of 0.87 Å from the phosphorous atom in G3P. The juxtaposition of the residues shown in this panel and in the Figure 2b unequivocaly indicates that the active site loop is in the closed conformation in the A1 structure. (c) Schematic diagram of the hydrogen bonding network in the active site region of the A1 structure. Hydrogen bonds are represented by dashed lines and bond lengths are given in Å. Water oxygens are colored in green. Panels (a) and (b) were prepared using PyMOL (DeLano Scientific, San Carlos, CA, http://www.pymol.org) and edited using GIMP (http://www.gimp.org) under Linux. Panel (c) was produced using LIGPLOT⁸¹ and edited using GIMP (http://www.gimp.org) under Linux.

Figure 4. Stereo diagram of crystal packing in the A1 structure. Symmetryrelated molecules (colored in slate blue) are drawn around the active site loop region (carbon atoms shown in yellow) of subunit 2 (chain B; colored in gray). Water molecules were omitted for clarity. (a) The molecular arrangement as obtained from experimental data. The contact distances of the nearest subunit to the active site loop of subunit 2 are larger than 4.3 Å (Supplementary Material). Note the gap separating subunit 2 and the symmetry-related molecules. (b) Hypothetical structure in which subunit 1 (chain A) with the active site loop in the open conformation was superposed onto subunit 2. Neither steric clashes nor unacceptable close contacts are observed (Supplementary Material). The Figure demonstrates that the lid loop could potentially adopt an open conformation in subunit 2 of the A1 structure. Figures were generated with PyMOL (DeLano Scientific, San Carlos, CA, http://www.pymol.org) un-
der Linux.

Figure 5. Stereo drawings of some examples of ligand molecules modelled in the electron density maps of the A1 structure. Ligands have been named (see below) according to the nearest interacting subunit. Both the ligand subunit and the neighbouring subunits of the catalytic dimers present in the assimetric unit have the carbon atoms colored in yellow. Symmetry-related molecules are shown in slate blue. All of the ligands modelled in the A1 structure lie at the interface between different subunits. Experimental electron density maps $(2mF_{obs} - DF_{calc})$ were contoured at 1σ around the ligands. Water molecules, except in (b), are not shown. (a) a DMSO molecule (B253). (b) a Mg ion (colored in khaki) coordinated by six water molecules (C257). Water oxygens are colored in red. (c) a TRIS molecule (C254). Drawings were prepared using PyMOL (DeLano Scientific, San Carlos, CA, http://www.pymol.org) and edited using GIMP (http://www.gimp.org) under Linux.

Acknowledgements

We gratefully acknowledge the staff of the York Structural Biology Laboratory, in particular, Drs. G. N. Murshudov, E. J. Dodson and A. M. Brzozowski, for help in crystallization and refinement; ESRF (Grenoble/France) and LNLS (Campinas/Brazil) for use of the Protein Crystallography beamlines. This work has been supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), via grants 98/06761-1 and 99/03387-4; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Howard Hughes Medical Institute (STF).

References

 Yüksel, K. U. & Gracy, R. W. (1991). Triose- and hexose-phosphate isomerases. In *A study of enzymes*, edited by S. A. Kuby, volume II, chapter 19, pages 457–484. CRC Press.

- [2] Rieder, S. V. & Rose, I. A. (1959). The mechanism of the triosephosphate isomerase reaction. J. Biol. Chem., 234, 1007–1010.
- [3] Rose, I. A. (1962). Mechanisms of C-H bond cleavage in aldolase and isomerase reactions. *Brookhaven Symp. Biol.*, 15, 293–309.
- [4] Noble, M. E. M., Zeelen, J. P., Wierenga, R. K., Mainfroid, V., Goraj, K., Gohimont, A. C. & Martial, J. A. (1993). Structure of triosephosphate isomerase from *Escherichia coli* determined at 2.6 Å resolution. *Acta Cryst. D*, 49, 403–417.
- [5] Delboni, L. F., Mande, S. C., Rentier-Delrue, F., Mainfroid, V., Turley, S., Vellieux, F. M. D., Martial, J. A. & Hol, W. G. J. (1995). Crystal structure of recombinant triosephosphate isomerase from *Bacillus* stearothermophilus. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. *Protein Sci.*, 4, 2594–2604.
- [6] Alvarez, M., Zeelen, J. P., Mainfroid, V., Rentier-Delrue, F., Martial, J. A., Wyns, L., Wierenga, R. K. & Maes, D. (1998). Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. J. Biol. Chem., 273, 2199–2206.
- [7] Maes, D., Zeelen, J. P., Thanki, N., Beaucamp, N., Alvarez, M., Dao Thi, M. H., Backmann, J., Martial, J. A., Wyns, L., Jaenicke, R. & Wierenga, R. K. (1999). The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: a comparative thermostability structural analysis of ten different TIM structures. *Proteins*, **37**, 441–453.
- [8] Walden, H., Bell, G. S., Russell, R. J. M., Siebers, B., Hensel, R. & Taylor, G. L. (2001). Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase. J. Mol. Biol., 306, 745–757.
- [9] Lolis, E., Alber, T., Davenport, R. C., Rose, D., Hartman, F. C. & Petsko,
 G. A. (1990). Structure of yeast triosephosphate isomerase at 1.9 Å resolution. *Biochemistry*, 29, 6609–6618.
- [10] Wierenga, R. K., Noble, M. E., Vriend, G., Nauche, S. & Hol, W. G. (1991). Refined 1.83 Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M-ammonium sulphate. A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. J. Mol. Biol., 220, 995–1015.
- [11] Velanker, S. S., Ray, S. S., Gokhale, R. S., Suma, S., Balaram, H.,

Balaram, P. & Murthy, M. R. N. (1997). Triosephosphate isomerase from *Plasmodium falciparum*: the crystal structure provides insights into antimalarial drug design. *Structure*, **5**, 751–761.

- [12] Maldonado, E., Soriano-García, M., Moreno, A., Cabrera, N., Garza-Ramos, G., Tuena de Gómez-Puyou, M., Gómez-Puyou, A. & Pérez-Montfort, R. (1998). Differences in the intersubunit contacts in triosephosphate isomerase from two closely related pathogenic trypanosomes. J. Mol. Biol., 283, 193–203.
- [13] Williams, J. C., Zeelen, J. P., Neubauer, G., Vried, G., Backmann, J., Michels, P. A. M., Lambeir, A. M. & Wierenga, R. K. (1999). Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. *Protein Eng.*, **12**, 243–250.
- [14] Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. (1975). Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature*, 255, 609– 14.
- [15] Mande, S. C., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A. & Hol, W. G. (1994). Crystal structure of recombinant human triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase-related human genetic disorders and comparison with the trypanosomal enzyme. *Protein Sci.*, **3**, 810–821.
- [16] Kohlhoff, M., Dahm, A. & Hensel, R. (1996). Tetrameric triosephosphate isomerase from hyperthermophilic archaea. *FEBS Lett.*, 383, 245–250.
- [17] Borchert, T. V., Pratt, K., Zeelen, J. P., Callens, M., Noble, M. E., Opperdoes, F. R., Michels, P. A. & Wierenga, R. K. (1993). Overexpression of trypanosomal triosephosphate isomerase in *Escherichia coli* and characterisation of a dimer-interface mutant. *Eur. J. Biochem.*, **211**, 703–710.
- [18] Mainfroid, V., Terpstra, P., Beauregard, M., Frere, J. M., Mande, S. C., Hol, W. G., Martial, J. A. & Goraj, K. (1996). Three hTIM mutants that provide new insights on why TIM is a dimer. J. Mol. Biol., 257, 441–456.
- [19] Schliebs, W., Thanki, N., Eritja, R. & Wierenga, R. (1996). Active site properties of monomeric triosephosphate isomerase (monoTIM) as de-

duced from mutational and structural studies. Protein Sci., 5, 229–239.

- [20] Schliebs, W., Thanki, N., Jaenicke, R. & Wierenga, R. K. (1997). A double mutation at the tip of the dimer interface loop of triosephosphate isomerase generates active monomers with reduced stability. *Biochemistry*, **36**, 9655–9662.
- [21] Albery, W. J. & Knowles, J. R. (1976). Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry*, 15, 5631–5640.
- [22] Alber, T., Banner, D. W., Bloomer, A. C., Petsko, G. A., Philips, D. C., Rivers, P. S. & Wilson, I. A. (1981). On the three-dimensional structure and catalytic mechanism of triosephosphate isomerase. *Phil. Trans. Roy. Soc. ser. B*, **293**, 159–171.
- [23] Nickbarg, E. B. & Knowles, J. R. (1988). Triosephosphate isomerase: energetics of the reaction catalyzed by the yeast enzyme expressed in *Escherichia coli. Biochemistry*, 27, 5939–5947.
- [24] Knowles, J. R. (1991). Enzyme catalysis: not different, just better. Nature, 350, 121–124.
- [25] Knowles, J. R. & Albery, W. J. (1977). Perfection in enzyme catalysis: the energetics of triosephosphate isomerase. Acc. Chem. Res., 10, 105–111.
- [26] Blacklow, S. C., Raines, R. T., Lim, W. A., Zamore, P. D. & Knowles, J. R. (1988). Triosephosphate isomerase catalysis is diffusion controlled. *Biochemistry*, 27, 1158–1165.
- [27] Branden, C. & Tooze, J. (1991). Introduction to protein structure. Garland Publishing, Inc., New York, USA.
- [28] Farber, G. K. (1993). An α/β-barrel full of evolutionary trouble. Curr. Op. Struc. Biol., 3, 409–412.
- [29] Doenhoff, M. J. (1998). A vaccine for schistosomiasis: alternative approaches. *Parasitol. Today*, 14, 105–109.
- [30] Wierenga, R. K., Noble, M. E. & Davenport, R. C. (1992). Comparison of the refined crystal structures of liganded and unliganded chicken, yeast and trypanosomal triosephosphate isomerase. J. Mol. Biol., 224, 1115– 26.
- [31] Daar, I. O., Artymiuk, P. J., Phillips, D. C. & Maquat, L. E. (1986). Human triose-phosphate isomerase deficiency: a single amino acid substitution results in a thermolabile enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, 83, 7903–7907.

- [32] Watanabe, M., Zingg, B. C. & Mohrenweiser, H. W. (1996). Molecular analysis of a series of alleles in humans with reduced activity at the triosephosphate isomerase locus. Am. J. Hum. Genet., 58, 308–316.
- [33] Joseph, D., Petsko, G. A. & Karplus, M. (1990). Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science*, 249, 1425–8.
- [34] Davenport, R. C., Bash, P. A., Seaton, B. A., Karplus, M., Petsko, G. A. & Ringe, D. (1991). Structure of the triosephosphate isomerasephosphoglycolohydroxamate complex: an analog of the intermediate on the reaction pathway. *Biochemistry*, **30**, 5821–5826.
- [35] Lolis, E. & Petsko, G. A. (1990). Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5 Å resolution. Implications for catalysis. *Biochemistry*, 29, 6619–6625.
- [36] Wierenga, R. K., Noble, M. E., Postma, J. P., Groendijk, H., Kalk, K. H., Hol, W. G. J. & Opperdoes, F. R. (1991). The crystal structure of the "open" and the "closed" conformation of the flexible loop of trypanosomal triosephosphate isomerase. *Proteins*, **10**, 33–49.
- [37] Wierenga, R., Borchert, T. V. & Noble, M. E. (1992). Crystallographic binding studies with triosephosphate isomerases: conformational changes induced by substrate and substrate-analogues. *FEBS Lett.*, **307**, 34–9.
- [38] McMillan, F. M., Cahoon, M., White, A., Hedstrom, L., Petsko, G. A. & Ringe, D. (2000). Crystal structure at 2.4 Å resolution of *Borrelia burgdorferi* inosine 5'-monophosphate dehydrogenase: evidence of a substrate-induced hinged-lid motion by loop 6. *Biochemistry*, **39**, 4533– 4542.
- [39] Brzovic, P. S., Hyde, C. C., Miles, E. W. & Dunn, M. F. (1993). Characterization of the functional role of a flexible loop in the α-subunit of tryptophan synthase from *Salmonella typhimurium* by rapid-scanning, stopped-flow spectroscopy and site-directed mutagenesis. *Biochemistry*, **32**, 10404–10413.
- [40] Lundqvist, T. & Schneider, G. (1991). Crystal structure of activated ribulose-1,5-bisphosphate carboxylase complexed with its substrate, ribulose-1,5-bisphosphate. J. Biol. Chem., 266, 12604–12611.
- [41] Krietsch, W. K., Pentchev, P. G., Klingenburg, H., Hofstatter, T. & Bucher, T. (1970). The isolation and crystallization of yeast and rab-

bit liver triose phosphate isomerase and a comparative characterization with the rabbit muscle enzyme. *Eur. J. Biochem.*, **14**, 289–300.

- [42] Scopes, R. K. (1977). Purification of glycolytic enzymes by using affinityelution chromatography. *Biochem. J.*, 161, 253–63.
- [43] Scopes, R. K. (1977). Multiple enzyme purifications from muscle extracts by using affinity-elution-chromatographic procedures. *Biochem. J.*, 161, 265–277.
- [44] Xiang, J., Sun, J. & Sampson, N. S. (2001). The importance of hinge sequence for loop function and catalytic activity in the reaction catalyzed by triosephosphate isomerase. J. Mol. Biol., 307, 1103–1112.
- [45] Noble, M. E., Zeelen, J. P. & Wierenga, R. K. (1993). Structures of the "open" and "closed" state of trypanosomal triosephosphate isomerase, as observed in a new crystal form: implications for the reaction mechanism. *Proteins*, 16, 311–26.
- [46] Kursula, I., Partanen, S., Lambeir, A., Antonov, D. M., Augustyns, K. & Wierenga, R. K. (2001). Structural determinants for ligand binding and catalysis of triosephosphate isomerase. *Eur. J. Biochem.*, 268, 5189–5196.
- [47] Zhang, Z., Komives, E. A., Sugio, S., Blacklow, S. C., Narayana, N., Xuong, N. H., Stock, A. M., Petsko, G. A. & Ringe, D. (1999). The role of water in the catalytic efficiency of triosephosphate isomerase. *Biochemistry*, **38**, 4389–4397.
- [48] Verlinde, C. L. M. J., Witmans, C. J., Pijning, T., Kalk, K. H., Hol, W. G. J., Callens, M. & Opperdoes, F. R. (1992). Structure of the complex between trypanosomal triosephosphate isomerase and N-hydroxy-4-phosphono-butanamide: binding at the active site despite an "open" flexible loop conformation. *Protein Sci.*, 1, 1578–1584.
- [49] Rozovsky, S., Jog, G., Tong, L. & McDermott, A. E. (2001). Solutionstate NMR investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics. J. Mol. Biol., 310, 271–280.
- [50] Plomp, M., McPherson, A., Larson, S. B. & Malkin, A. J. (2001). Growth mechanisms and kinetics of trypsin crystallization. J. Phys. Chem. B, 105, 542–551.
- [51] Wiechmann, M., Enders, O., Zeilinger, C. & Kolb, H.-A. (2001). Analysis of protein crystal growth at molecular resolution by atomic force

microscopy. Ultramicroscopy, 86, 159–166.

- [52] Yoshizaki, I., Sato, T., Igarashi, N., Natsuisaka, M., Tanaka, N., Komatsu, H. & Yoda, S. (2001). Systematic analysis of supersaturation and lysozyme crystal quality. *Acta Cryst. D*, 57, 1621–1629.
- [53] Williams, J. C. & McDermott, A. E. (1995). Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated. *Biochemistry*, 34, 8309–8319.
- [54] Kishan, K. V., Zeelen, J. P., Noble, M. E., Borchert, T. V. & Wierenga, R. K. (1994). Comparison of the structures and the crystal contacts of trypanosomal triosephosphate isomerase in four different crystal forms. *Protein Sci.*, 3, 779–87.
- [55] Lehmann, M. S. & Stansfieldt, R. F. D. (1989). Binding of dimethyl sulfoxide to lysozyme in crystals, studied with neutron diffraction. *Biochemistry*, 28, 7028–7033.
- [56] Shan, S. O. & Herschlag, D. (1999). Hydrogen bonding in enzymatic catalysis: analysis of energetic contributions. *Meth. Enzymol.*, **308**, 246– 276.
- [57] Erijman, L. & Weber, G. (1991). Oligomeric protein associations: transition from stochastic to deterministic equilibrium. *Biochemistry*, **30**, 1595–1599.
- [58] Erijman, L., Paladini, A. A., Lorimer, G. H. & Weber, G. (1993). Plurality of protein conformations of ribulose-1,5-bisphosphate carboxylase/oxygenase monomers probed by high pressure electrophoresis. J. Biol. Chem., 268, 25914–25919.
- [59] Ruan, K. & Weber, G. (1989). Hysteresis and conformational drift of pressure-dissociated glyceraldehydephosphate dehydrogenase. *Biochemistry*, 28, 2144–2153.
- [60] Ruan, K. & Weber, G. (1993). Physical heterogeneity of muscle glycogen phosphorylase revealed by hydrostatic pressure dissociation. *Biochemistry*, **32**, 6295–6301.
- [61] Pedrosa, C. & Ferreira, S. T. (1994). Deterministic pressure-induced dissociation of vicilin, the 7S storage globulin from pea seeds: effects of pH and cosolvents on oligomer stability. *Biochemistry*, 33, 4046–4055.
- [62] Rietveld, A. W. M. & Ferreira, S. T. (1996). Deterministic pressure dissociation and unfolding of triose phosphate isomerase: persistent het-

erogeneity of a protein dimer. Biochemistry, 35, 7743–7751.

- [63] Rietveld, A. W. M. & Ferreira, S. T. (1998). Kinetics and energetics of subunit dissociation/unfolding of TIM: the importance of oligomerization for conformational persistence and chemical stability of proteins. *Biochemistry*, 37, 933–937.
- [64] Panda, M. & Horowitz, P. M. (2002). Conformational heterogeneity is revealed in the dissociation of the oligomeric chaperonin GroEL by high hydrostatic pressure. *Biochemistry*, **41**, 1869–1876.
- [65] Yüksel, K. U., Sun, A. Q., Gracy, R. W. & Schnackerz, K. D. (1994). The hinged lid of yeast triose-phosphate isomerase. Determination of the energy barrier between the two conformations. J. Biol. Chem., 269, 5005–5008.
- [66] Aparicio, R., Ferreira, S. T., Leite, N. R. & Polikarpov, I. (2000). Preliminary X-ray diffraction studies of rabbit muscle triose phosphate isomerase (TIM). Acta Cryst. D, 56, 1492–1494.
- [67] Garman, E. F. & Schneider, T. R. (1997). Macromolecular cryocrystallography. J. Appl. Cryst., 30, 211–237.
- [68] Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307–326.
- [69] Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Cryst. A, 50, 157–163.
- [70] Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. Acta Cryst. D, 50, 760–763.
- [71] Winn, M., Dodson, E. J. & Ralph, A. (1997). Collaborative Computational Project, Number 4: providing programs for Protein Crystallography. *Meth. Enzymol.*, 277, 620–633.
- [72] Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst. D, 53, 240–255.
- [73] Schomaker, V. & Trueblood, K. N. (1968). On the rigid-body motion of molecules in crystals. Acta Cryst. B, 24, 63–76.
- [74] Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Cryst. D, 57, 122–133.
- [75] Kleywegt, G., Zou, J., Kjeldgaard, M. & Jones, T. (2001). Around O. In

International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules, edited by M. G. Rossmann & E. Arnold, chapter 17.1, pages 353–356, 366–367. Dordrecht: Kluwer Academic Publishers.

- [76] Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst. A, 47, 110–119.
- [77] Zhang, Z., Sugio, S., Komives, E. A., Liu, K. D., Knowles, J. R., Petsko, G. A. & Ringe, D. (1994). Crystal structure of recombinant chicken triosephosphate isomerase-phosphoglycolohydroxamate complex at 1.8-Å resolution. *Biochemistry*, **33**, 2830–2837.
- [78] Nayal, M. & Di Cera, E. (1996). Valence screening of water in protein crystals reveals potential Na⁺ binding sites. J. Mol. Biol., 256, 228–234.
- [79] Cohen, G. H. (1997). Align: a program to superimpose protein coordinates, accounting for insertions and deletions. J. Appl. Cryst., 30, 1160–1161.
- [80] Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. J. Mol. Biol., 190, 593–604.
- [81] Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). Ligplot: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, 8, 127–134.

6.1 Supporting information



Figura 6.1: Estrutura A1 do artigo em anexo: fatores de temperatura por resíduos para as cadeias A, B, C e D.



Figura 6.2: Estrutura A2 do artigo em anexo: fatores de temperatura por resíduos para as cadeias A, B, C e D.



Figura 6.3: Estrutura B do artigo em anexo: fatores de temperatura por resíduos para as cadeias A e B.

Capítulo 7

Conclusão

A presente tese descreve os resultados do projeto de doutorado do autor, com a realização de estudos estruturais da β -manosidase do fungo *Trichoderma reesei* e da Triose fosfato isomerase (TIM) de músculo de coelho. Durante o período, houve também participações em outros trabalhos, descritos sucintamente ao final desta conclusão. Além disso, o programa de doutorado incluiu o desenvolvimento de atividades acadêmicas e participação em cursos e congressos, apresentados no Apêndice B.

Embora não relacionados pela função enzimática das proteínas em questão, os estudos propostos tinham em comum o objetivo inicial, que era a determinação de suas estruturas cristalográficas. No entanto, uma diferença fundamental foi o método proposto para resolução do problema das fases em cada caso, o que levou a uma complementaridade no que diz respeito aos conhecimentos adquiridos. Se, por um lado, houve amplo uso de técnicas experimentais para obtenção de fases (β -manosidase), por outro, foi adquirida extensa experiência em reconstrução e refinamento de modelos cristalográficos (TIM).

O estágio realizado no York Structural Biology Laboratory (YSBL), em York, na Inglaterra, foi de extrema importância para o sucesso de ambos os projetos. No caso da β -manosidase, antes do estágio, uma vasta quantidade de dados cristalográficos havia sido coletada no LNLS e grande esforço para faseamento da proteína havia sido feito sem sucesso (Capítulo 3). Problemas nos dados haviam sido identificados mas não havia certeza se eram suficiente razão para o insucesso na obtenção de fases: por que uma metologia tão bem estabelecida não levava ao resultado esperado?

Durante o estágio, conclui-se que parte dos problemas têm origem nos próprios cristais de β -manosidase, o que resulta em dados cristalográficos de qualidade insuficiente para utilização do método proposto. Como alternativa, o problema estrutural foi atacado pela técnica de espalhamento de Raios X a baixo ângulo (Small-Angle X-ray Scattering ou SAXS), com dados coletados a partir da proteína em solução. Dois aspectos merecem destaque nestes estudos, que foram descritos em artigo publicado em conceituada revista (Capítulo 4). O primeiro foi a obtenção de informação estrutural biologicamente relevante, levando a um modelo de baixa resolução e ao provável enovelamento da enzima. O segundo foi o uso conjunto de SAXS e Cristalografia, que proporcionou um melhoramento do modelo de SAXS e permitiu formular hipóteses que apontam as causas, em nível molecular, dos problemas apresentados pelos cristais, notadamente as variações nos parâmetros de rede e grupo de espaço, espalhamento anisotrópico, baixa resolução e baixa relação sinal/ruído. Além disso, o método estabelecido para busca de um modelo provável para a proteína¹ pode ter aplicabilidade em outros casos, principalmente de Projetos Genoma, onde técnicas aplicáveis em larga escala são necessárias.

Também durante o estágio, um novo conjunto de dados cristalográficos da TIM foi coletado no ESRF. Este conjunto levou à estrutura de TIM de maior resolução até hoje. Mais do que isso. Pela primeira vez, desde que

 $^{^1}$ A comparação de funções de distribuição de distâncias foi sugerida pelo amigo e mestre Hannes Fischer.

a primeira estrutura de TIM foi resolvida, em 1975 (Banner *et al.*, 1975), o *loop* do sítio ativo foi observado na conformação fechada sem a presença de ligante no sítio ativo. Os resultados e sua análise foram descritos num artigo a ser submetido em breve para publicação (Capítulo 6).

Em suma, não foi possível determinar a estrutura cristalográfica da β manosidase, fato justificável em vista de todos os problemas já mencionados. Em contra-partida, uma estrutura de baixa resolução e o provável enovelamento da proteína foram determinados por espalhamento a baixo ângulo, resultando num trabalho publicado que, além da importância biológica intrínseca, descreve uma nova metodologia que pode ser aplicada a outros casos. Foram resolvidas três estruturas de TIM, em duas formas cristalinas e a partir de dados coletados a diferentes temperaturas. A conformação fechada do *loop* do sítio ativo em uma das subunidades da estrutura de maior resolução (1.5 Å) tem importância biológica inequívoca. Três artigos foram publicados e um quarto está em preparação (incluídos na tese). Além destes, houve participação em dois outros trabalhos, já publicados, e num outro, também em preparação, conforme descrito abaixo.

7.1 Outras publicações

Os estudos estruturais descritos nesta tese foram o principal objetivo do projeto de doutorado. Além destes, houve participações menores em outras pesquisas que resultaram em trabalhos não incluídos aqui mas que devem ser mencionados.

Uma primeira participação, a mais recente, compreende estudos cristalográficos e de baixa resolução de uma glicosidase para a qual já havia dados coletados e parte do trabalho feito. O artigo que descreve este trabalho e encontra-se em preparação, é o seguinte

Aparicio, R., Fischer, H., Golubev, A. M., Forato, L. A., Colnago, L. A. and

Polikarpov, I.

Molecular shape of α -Galactosidase from *Trichoderma reesei*: Structure-Function Similarity with α -N-acetyl galactosaminidase

Em duas outras participações, a colaboração limitou-se ao manuseio de cristais e coleta de dados a temperaturas criogênicas. Uma delas foi na determinação da estrutura cristalográfica de um fragmento Fab da KAU, um anti-corpo IgM humano monoclonal, com atividade anti-I, e que é uma aglutinina de baixa temperatura envolvida na anemia hemolítica. O trabalho foi publicado sob a referência

Cauerhff, A., Braden, B.C., Carvalho, J. G., Aparicio, R., Polikarpov, I., Leoni, J. and Goldbaum, F.A. (2000)

Three-dimensional structure of the Fab from a human IgM cold agglutinin *J. Immunol.*, **165**, 6422–6428.

A outra trata de um estudo cristalográfico da Catalase humana (óxidoredutase, EC 1.1.1.6), enzima que desempenha papel importante na defesa celular contra espécies ativas para oxigênio e está relacionada ao processo de envelhecimento, que resultou no artigo cuja referência é

Nagem, R. A. P., Martins, E. A. L., Gonçalves, V. M., Aparicio, R. and Polikarpov, I. (1999)

Crystallization and preliminary X-ray diffraction studies of human catalase Acta Cryst. D, 55, 1614–1615.

Apêndice A

Teoria: definições elementares

 R_{merge} (ou R_{sym})

$$R_{merge} = \frac{\sum_{\boldsymbol{h},i} |I_i(\boldsymbol{h}) - \langle I(\boldsymbol{h}) \rangle|}{\sum_{\boldsymbol{h},i} I_i(\boldsymbol{h})}, \qquad (A.1)$$

onde o somatório em *i* estende-se sobre todas as observações independentes (por simetria de grupo e medidas repetidas em várias imagens) de cada reflexão única h. A estatística, também chamada R_{sym} , reflete a dispersão em torno da intensidade média de um conjunto equivalente de reflexões e é uma medida da coerência interna de um conjunto de dados. Além do R_{merge} , outros indicadores de qualidade, mais eficientes, já foram descritos (Diederichs & Karplus, 1997).

Pesos m e D nos mapas de densidade eletrônica

Os pesos nos mapas $2mF_{obs} - DF_{calc}$ e $mF_{obs} - DF_{calc}$ (Murshudov *et al.*, 1997) são dados por

$$D_j(s) = \langle exp(-\Delta B_j |s|^2/4) \cos(2\pi s \Delta x_j) \rangle, \qquad (A.2)$$

onde $|s| = 2sin\theta/\lambda$ é o módulo do vetor de posição no espaço recíproco e Δx_j e ΔB_j são os erros na posição e fator de temperatura, respectivamente, dos átomos na *j*th estrutura parcial, e

$$m = (\langle \cos \varphi \rangle^2 + \langle \sin \varphi \rangle^2)^{1/2}, \tag{A.3}$$

onde φ é a fase calculada. Pode-se mostrar que $m = \langle \cos \Delta \varphi \rangle$, onde $\Delta \varphi$ é o erro nas fases do modelo. Esta é a figura de mérito calculada quando já se tem informação advinda de um modelo atômico.

Figura de Mérito Experimental

As fórmulas abaixa são utilizadas durante a determinação de fases através de métodos de átomos pesados:

$$\boldsymbol{F}_{best} = \frac{\int |F_p| exp(i\alpha) P(\alpha) d\alpha}{\int P(\alpha) d\alpha} = m |F_p| exp(i\alpha_{best}), \tag{A.4}$$

onde m é a figura de mérito, dada por

$$m = \frac{\int P(\alpha) exp(i\alpha) d\alpha}{\int P(\alpha) d\alpha},$$
 (A.5)

e $P(\alpha)$ é a função densidade de probabilidade das fases α .

Fator de estrutura

Embora os fatores de estrutura sejam números complexos, normalmente eles são designados por símbolos em negrito (como se usa para vetores em R^3). Assim, temos

$$\boldsymbol{F}(\boldsymbol{h}) \simeq \sum_{1}^{N} n_k f_k(\boldsymbol{h}) T_k(\boldsymbol{h}) exp(2\pi i \boldsymbol{h} \cdot \boldsymbol{r}_{k_0}), \qquad (A.6)$$

onde

$$f_k(\boldsymbol{h}) = \int \rho_k(\boldsymbol{v}) exp(2\pi i \boldsymbol{h} \cdot \boldsymbol{v}) d^3 v, \qquad (A.7)$$

é o fator de forma (ou de espalhamento) atômica, que, assumindo-se uma distribuição esférica de carga, fica

$$f_k(|\boldsymbol{h}|) = 4\pi \int_0^\infty \rho_k(u) \frac{\sin(2\pi|\boldsymbol{h}|u)}{(2\pi|\boldsymbol{h}|u)} u^2 du, \qquad (A.8)$$

е

$$T_k(\boldsymbol{h}) = \int p_k(\boldsymbol{u}) exp(2\pi i \boldsymbol{h} \cdot \boldsymbol{u}) d^3 \boldsymbol{u}, \qquad (A.9)$$

é o fator de temperatura ou fator atômico de Debye-Waller.

Para $p_k(\boldsymbol{u})$ gaussiana,

$$T(\boldsymbol{h}) = exp[-8\pi^2 \langle u_{\boldsymbol{h}^2} \rangle \frac{\sin^2}{\lambda^2}], \qquad (A.10)$$

é o fator de temperatura anisotrópico.

Na ausência de espalhamento anômalo, vale a chamada "Lei de Friedel"

$$|\boldsymbol{F}(\boldsymbol{h})| = |\boldsymbol{F}(-\boldsymbol{h})|, \qquad (A.11)$$

Pares de Bijvoet

Na presença de espalhamento anômalo, o fator de espalhamento atômico total, que é a somatória para todos os átomos no cristal, adquire uma componente complexa (espalhamento ressonante fora de fase)

$$f = f' + if'', \tag{A.12}$$

e a Lei de Friedel não mais é válida.

No caso de um derivado de átomo pesado¹ com espalhamento anômalo, isso significa que

$$|\boldsymbol{F}_{PH}(\boldsymbol{h})| \neq |\boldsymbol{F}_{PH}(-\boldsymbol{h})| \tag{A.13}$$

A partir das chamadas diferenças de Bijvoet,

$$|\boldsymbol{F}_{PH}(\boldsymbol{h})| - |\boldsymbol{F}_{PH}(-\boldsymbol{h})|, \qquad (A.14)$$

é possível obter fases para os fatores de estrutura da proteína, $\boldsymbol{F}_{P}(\boldsymbol{h})$.

 $F_{PH}(h) \in F_{PH}(-h)$ também são chamados de $F_{PH}^+ \in F_{PH}^-$, respectivamente.

¹ indicado pelo sub-índice *PH*, de Protein plus Heavy atom.

Apêndice B

Cursos, congressos e atividades acadêmicas

Além do trabalho de pesquisa propriamente dito, outras atividades foram desenvolvidas em paralelo e estão expostas abaixo.

B.1 Atividades acadêmicas

Em adição à elaboração da presente tese, as disciplinas da Tabela abaixo foram cursadas para cumprimento do programa de doutorado.

Período	Código	Descrição	$Conceito^a$	$\mathbf{C}\mathbf{H}^{b}$	\mathbf{Crd}^c
1S/1998	FI001	Mecânica Quântica I	В	180	12
	FI008	Eletrodinâmica I	В	180	12
2S/1998	FI500	Tese de Doutoramento	Р		
	NB161	Química de Proteínas	А	180	12
	NG243	Cristalografia de Proteínas	А	135	9
1S/1999	FI010	Mecânica Clássica	А	180	12
	QP222	Métodos Físicos em Química Orgânica	А	180	12
2S/1999	FI002	Mecânica Quântica II	А	180	12
	FI204	Tópicos da Física da Matéria Condensada I	А	180	12
	FI250	Tópicos em Física de Aceleração Anéis de			
		Estocagem	S	90	6
	FI500	Tese de Doutoramento	Р		
1S/2000	FI500	Tese de Doutoramento	Р		
	FI600	Tópicos em Ensino de Física I	S	90	6
2S/2000	FI500	Tese de Doutoramento	Р		
1S/2001	FI500	Tese de Doutoramento	Р		
2S/2001	FI500	Tese de Doutoramento	Р		
1S/2002	FI500	Tese de Doutoramento	Р		

Tabela: Histórico escolar.

 a Conceitos:

A - Excelente - Aprovado;

B - Bom - Aprovado;

P - Pendente - Aprovado;

S - Suficiente - Aprovado

 $^b\mathrm{Carga}$ horária (h).

 $^c\mathrm{N}\mathrm{\acute{u}mero}$ de créditos

B.2 Cursos e congressos

Outra parte importante do doutorado foi a participação em cursos e congressos, com a apresentação de trabalhos, enumerados abaixo:

 American Crystallographic Association (ACA) Annual Meeting
 San Antonio, Texas, USA
 25 – 30 de May 2002

Oral presentation (Lecture):

Low Resolution SAXS Model of β-manosidase from Trichoderma reesei Enhanced by X-Ray Crystallography Aparicio, R.; Fischer, H.; Craievich, A. F.; Golubev, A. M.; Neustroev, K. N.; Eneiskaya, E. V.; Kulminskaya, A. A.; Savel'ev, A. N. and Polikarpov, I. Program and Abstract Book, 10.01.18, pp. 65.

- Euroconference on Phasing Biological Macromolecules (PhaBio) Martina Franca, Italy
 23th – 27th June 2001
- Protein Crystallography Data Collection Workshop Daresbury Laboratory, Daresbury, UK February 5th – 9th 2001,
- CCP4 Study Weekend University of York, York, UK

January 5 – 6th 2001

- EU/CCP4 Refinement Workshop University of York, York, UK January 3rd – 4th and 7 – 9th 2001
- XV Encontro da Sociedade Brasileira de Cristalografia LNLS, Campinas - SP 2000

Crystallofraphic structure determination of β -mannosidase from Trichoderma reesei by the MIRAS method.

R. Aparicio, E. V. Eneiskaya, A. A. Kulminskaya, A. N. Savel'ev, A.M. Golubev, K. N. Neustroev, J. Kobarg & I. Polikarpov.

IV Congresso de Biofísica do Cone Sul LNLS, Campinas - SP 2000

X-ray structure determination of T. reesei β -mannosidase using synchrotron radiation.

R. Aparicio, E. V. Eneiskaya, A. A. Kulminskaya, A. N. Savel'ev, A.M. Golubev, K. N. Neustroev, J. Kobarg & I. Polikarpov.

Livro de Resumos do IV Congresso de Biofísica do Cone Sul, E10: 75.

• XXIX Reunião Anual da SBBq (Soc. Bras. de Bioquímica e Biologia Molecular) Caxambu - MG

Maio de 2000

Crystallographic structure determination of β -mannosidase from Trichoderma reesei by the MIRAS method.

R. Aparicio, E. V. Eneiskaya, A. A. Kulminskaya, A. N. Savel'ev, A.M. Golubev, K. N. Neustroev, J. Kobarg & I. Polikarpov.Programa e Resumos da XXIX Reunião Annual da SBBq, M22: 126

Initial X-ray studies of two crystal forms of rabbit muscle triosephosphate isomerase (TIM).

R. Aparicio, Leite, N. R., Ferreira. S. T. & I. Polikarpov.Programa e Resumos da XXIX Reunião Annual da SBBq, M18: 125

Crystallization and structure determination of human catalase.R. A. P. Nagem, G. J. A. de Souza Jr., E. A. L. Martins, V. M. Goncalves, R. Aparicio & I. Polikarpov.Programa e Resumos da XXIX Reunião Annual da SBBq, M12: 124

• X Reunião Anual de Usuários do LNLS LNLS, Campinas - SP Fevereiro de 1999

Determinação da estrutura cristalográfica da β-manosidase de Trichoderma reesei por MIRAS. R. Aparicio, A. M. Golubev & I. Polikarpov Livro de Resumos da X Reunião Anual de Usuários do LNLS, 1.24: 50. Editor: R. Medeiros. Crystallization and preliminary X-ray Diffraction studies of human catalase.

R. A. P. Nagem, E. A. L. Martins, V. M. Goncalves, R. Aparicio & I. Polikarpov.

Livro de Resumos da X Reunião Anual de Usuários do LNLS, 1.18: 44. Editor: R. Medeiros.

• Minicurso de Biologia Molecular Estrutural LNLS - 27 a 30 de Julho de 1999.

Apresentação de seminário sobre cristalização de proteínas.

XXVIII Reunião Anual da SBBq (Soc. Bras. de Bioquímica e Biologia Molecular) Caxambu - MG
22 a 25 de Maio de 1999

Crystallization and preliminary X-ray diffraction studies of βmannosidase from Trichoderma ressei
R. Aparicio, A. M. Golubev and I. Polikarpov
Programa e Resumos da XXVIII Reunião Annual da SBBq, M23: 124

• IX Reunião Anual de Usuários do LNLS LNLS, Campinas - SP 22 a 24 de Fevereiro de 1999

Cristalização e coleta de dados da β -manosidase de Trichoderma ressei

R. Aparicio, A. M. Golubev and I. Polikarpov

Livro de Resumos da IX Reunião Anual de Usuários do LNLS, 6.3: 139. Editores R. Rodrigues e R. Medeiros

Cristalização e coleta de dados de duas enzimas da via glicolítica: triosefosfato isomerase (TIM) e hexoquinase PII R. Aparicio, Ferreira ST and I. Polikarpov Livro de Resumos da IX Reunião Anual de Usuários do LNLS, 6.4: 140. Editores R. Rodrigues e R. Medeiros

Instrumentação para a linha de Cristalografia de Proteínas do LNLS: implementação de equipamentos e métodos para coleta de dados a temperaturas criogênicas

R. Aparicio, J. R. B. Neto e I. Polikarpov

Livro de Resumos da IX Reunião Anual de Usuários do LNLS, 6.5: 141. Editores R. Rodrigues e R. Medeiros

Referências Bibliográficas

Adams, M. R., Richtmyer, N. K. & Hudson, C. S. (1943). Some enzymes present in highly purified invertase preparations: a contribution to the study of fructofuranosidases, galactosidases, glucosidases and mannosidases. J. Am. Chem. Soc., 65, 1369–1380. 1

Akino, T., Nakamura, N. & Horikoshi, K. (1988). Characterization of betamannosidase of an alkalophilic *Bacillus* sp. *Agr. Biol. Chem. Tokyo*, **52**, 1459–1464. 1, 6

Alber, T., Banner, D. W., Bloomer, A. C., Petsko, G. A., Philips, D. C., Rivers, P. S. & Wilson, I. A. (1981). On the three-dimensional structure and catalytic mechanism of triosephosphate isomerase. *Phil. Trans. Roy. Soc. ser. B*, **293**, 159–171. **11**, 12

Albery, W. J. & Knowles, J. R. (1976). Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry*, **15**, 5631–5640. **11**

Alkhayat, A. H., Kraemer, S. A., Leipprandt, J. R., Macek, M., Kleijer, W. J. & Friderici, K. H. (1998). Human beta-mannosidase cDNA characterization and first identification of a mutation associated with human beta-mannosidosis. *Hum Mol Genet.*, **7**, 75–83. **5**

Alvarez, M., Zeelen, J. P., Mainfroid, V., Rentier-Delrue, F., Martial, J. A., Wyns, L., Wierenga, R. K. & Maes, D. (1998). Triose-phosphate isomerase (TIM) of the psychrophilic bacterium Vibrio marinus. J. Biol. Chem., 273, 2199–2206. 9

Aparicio, R., Neto, J. R. B. & Polikarpov, I. (1999). Instrumenta o para a linha de Cristalografia de Prote nas do LNLS: implementa o de equipamentos e m todos para coleta de dados a temperaturas criog nicas. In *Livro de Resumos da IX Reuni o Anual de Usu rios do LNLS*, edited by R. Rodrigues & R. Medeiros, pages 145–145. LNLS, Campinas, SP. **31**, **32**

Azàroff, L. V. (1968). *Elements of X-ray Crystallography*. McGraw-Hill, New York, USA. **30**

Baker, D., Bystroff, C., Fletterick, R. J. & Agard, D. A. (1994). PRISM: topologically constrained phased refinement for macromolecular crystallography. *Acta Cryst. D*, **49**, 429–439. **48**

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson,
C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord,
R. E., Priddle, J. D. & Waley, S. G. (1975). Structure of chicken muscle
triose phosphate isomerase determined crystallographically at 2.5 angstrom
resolution using amino acid sequence data. *Nature*, 255, 609–14. 9, 11, 175

Bauer, M. W., Bylina, E. J., Ronald, V. S. & Kelly, R. M. (1996). Comparison of a beta-glucosidase and a beta-mannosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Purification, characterization, gene cloning, and sequence analysis. *J Biol Chem.*, **271**, 23749–23755. 7

Bernard, M., Brochet, C. & Percheron, F. (1985). Decreased serum beta-dmannosidase activity in diabetic patients, in comparison with other glycosidases. *Clin Chim Acta*, **152**, 171–174. 7 Black, G. W., Hazlewood, G. P., Millward-Sadler, S. J., Laurie, J. I. & Gilbert, H. J. (1995). A modular xylanase containing a novel non-catalytic xylan-specific binding domain. *Biochem. J.*, **307**, 191–195. **8**

Blundell, T. L. & Johnson, L. N. (1976). Protein Crystallography. Academic Press, New York, USA. 18, 20

Borchert, T. V., Pratt, K., Zeelen, J. P., Callens, M., Noble, M. E., Opperdoes, F. R., Michels, P. A. & Wierenga, R. K. (1993). Overexpression of trypanosomal triosephosphate isomerase in *Escherichia coli* and characterisation of a dimer-interface mutant. *Eur. J. Biochem.*, **211**, 703–710. 11

Bouquelet, S., Spik, G. & Montreuil, J. (1978). Properties of a β -D-mannosidase from Aspergillus niger. Biochim. Biophys. Acta., **522**, 521–530. 1

Branden, C. & Tooze, J. (1991). *Introduction to protein structure*. Garland Publishing, Inc., New York, USA. 11, 18

Branden, C. I. & Jones, T. A. (1990). Between objectivity and subjectivity. *Nature*, **343**, 687–689. **55**

Bricogne, G. (1974). Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Cryst. A*, **30**, 395–405. **4**9

Brünger, A. T. (1992). The Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474. **54**

Brünger, A. T. (1997). Free R value: cross-validation in Crystallography. In Carter Jr. & Sweet (1997b), pages 366–396. 54

Brünger, A. T., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N., Read, R., Rice, L., Simonson, T. & Warren, G. (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta. Cryst. D*, **54**, 905–921. **46**, 53, 74, 77

Bryan, L., Schmutz, S., Hodges, S. D. & Snyder, F. F. (1990). Bovine betamannosidase deficiency. *Biochem Biophys Res Commun*, **173**, 491–495. **5**

Brzovic, P. S., Hyde, C. C., Miles, E. W. & Dunn, M. F. (1993). Characterization of the functional role of a flexible loop in the α -subunit of tryptophan synthase from *Salmonella typhimurium* by rapid-scanning, stoppedflow spectroscopy and site-directed mutagenesis. *Biochemistry*, **32**, 10404– 10413. **12**

Carter Jr., C. W. & Sweet, R. M. (eds.) (1997a). *Macromolecular Crystal*lography, Part A, volume 276 of Methods in Enzymology. Academic Press, San Diego. 18, 194, 207

Carter Jr., C. W. & Sweet, R. M. (eds.) (1997b). *Macromolecular Crystallography, Part B*, volume 277 of *Methods in Enzymology*. Academic Press, San Diego. 18, 191, 197, 208

Castresana, J., Muga, A. & Arrondo, J. L. (1988). The structure of proteins in aqueous solutions: an assessment of triose phosphate isomerase structure by Fourier-transform infrared spectroscopy. *Biochem Bioph Res Co*, **152**, 69–75. **13**

Cavanagh, K. T., Fisher, R. A., Legler, G., Herrchen, M., Jones, M. Z., Julich, E., Sewell-Alger, R. P., Sinnott, M. L. & Wilkinson, F. E. (1985). Goat liver beta-mannosidases: molecular properties, inhibition and inactivation of the lysosomal and nonlysosomal forms. *Enzyme*, **34**, 75–82. 1 CCP4 (1994). Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Cryst. D, 50, 760–763.
18, 28, 46, 50, 51, 53, 56, 73, 81, 82, 88, 92, 98, 118

Chen, H., Leipprandt, J. R., Traviss, C. E., Sopher, B. L., Jones, M. Z., Cavanagh, K. T. & Friderici, K. H. (1995). Molecular cloning and characterization of bovine beta-mannosidase. *J. Biol. Chem.*, **270**, 3841–3848. **5**, 9

Cooper, A. & Sardharwalla, I. B. (1986). Human beta-mannosidase deficiency. *N Engl J Med*, **315**, 1231. **5**

Corran, P. H. & Ley, S. G. (1975). The amino acid sequence of rabbit muscle triose phosphate isomerase. *Biochem J*, **145**, 335–344. **13**

Cullity, B. D. (2001). *Elements of X-ray Diffraction*. Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition. 30

Daar, I. O., Artymiuk, P. J., Phillips, D. C. & Maquat, L. E. (1986). Human triose-phosphate isomerase deficiency: a single amino acid substitution results in a thermolabile enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 7903– 7907. **13**

Dauter, Z. (1999). Data-collection strategies. Acta Cryst. D, 55, 1703–1717. 70

Dauter, Z., Dauter, M. & Dodson, E. (2002). Jolly sad. Acta Cryst. D, 58, 494–506. 45

Davenport, R. C., Bash, P. A., Seaton, B. A., Karplus, M., Petsko, G. A. & Ringe, D. (1991). Structure of the triosephosphate isomerasephosphoglycolohydroxamate complex: an analog of the intermediate on the reaction pathway. *Biochemistry*, **30**, 5821–5826. **1**2 de La Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In Carter Jr. & Sweet (1997a), pages 472–494. 46, 74, 77, 90, 92

Delboni, L. F., Mande, S. C., Rentier-Delrue, F., Mainfroid, V., Turley, S., Vellieux, F. M. D., Martial, J. A. & Hol, W. G. J. (1995). Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. *Protein Sci.*, 4, 2594–2604. 9

Dey, P. M. & del Campillo, E. (1984). Biochemistry of the multiple forms of glycosidases in plants. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **56**, 141–249. 6

Diederichs, K. & Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Struc. Biol.*, **4**, 269–275. **177**

Dodson, E., Kleywegt, G. J. & Wilson, K. (1996). Report of a workshop on the use of statistical validators in Protein X-ray Crystallography. *Acta Cryst. D*, **52**, 228–234. **54**

Doenhoff, M. J. (1998). A vaccine for schistosomiasis: alternative approaches. *Parasitol. Today*, **14**, 105–109. **13**

Duffaud, G. D., McCutchen, C. M., Leduc, P., Parker, K. N. & Kelly, R. M. (1997). Purification and characterization of extremely thermostable beta-mannanase, beta-mannosidase, and alpha-galactosidase from the hyperthermophilic eubacterium *Thermotoga neapolitana* 5068. *Appl. Environ. Microbiol.*, **63**, 169–177. 7
Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B. & Mornon, J. P. (1997). Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology*, **7**, 277–284. 9

EU 3-D Validation Network (1998). Who check the checkers? Four validation tools applied to eight atomic resolution structures. J. Mol. Biol., **276**, 417–436. 54, 57

Farber, G. K. (1993). An α/β -barrel full of evolutionary trouble. *Curr. Op.* Struc. Biol., **3**, 409–412. 12

Fernandez-Velasco, D. A., Sepulveda-Becerra, M., Galina, A., Darszon, A., Tuena de Gómez-Puyou, M. & Gómez-Puyou, A. (1995). Water requirements in monomer folding and dimerization of triosephosphate isomerase in reverse micelles. Intrinsic fluorescence of conformers related to reactivation. *Biochemistry*, 34, 361–369. 13

Garman, E. F. & Schneider, T. R. (1997). Macromolecular Cryocrystallography. J. Appl. Cryst., **30**, 211–237. **23**

Garza-Ramos, G., Tuena de Gómez-Puyou, M., Gómez-Puyou, A. & Gracy,
R. W. (1992). Dimerization and reactivation of triosephosphate isomerase
in reverse micelles. *Eur J Biochem.*, 208, 389–395. 13

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller Jr., R. C. & Warren, R.
A. J. (1991). Domains in microbial b-1,4-glycanases: sequence conservation, function, and enzyme families. *Microb. R.*, 55, 303–315.

Gilmore, C. J. (2000). Direct methods and protein crystallography at low resolution. *Acta Cryst. D*, **56**, 1205–1214. **38**

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc. Roy. Soc. A*, **225**, 287–307. **4**3

Hao, Q. (2001). Phasing from an envelope. *Acta Cryst. D*, **57**, 1410–1414. 109

Houston, C. W., Latimer, S. B. & Mitchell, E. D. (1974). Purification and properties of beta-mannosidase from malted barley. *Biochim. Biophys. Acta.*, **370**, 276–282. 1

Iwasaki, Y., Tsuji, A., Omura, K. & Suzuki, Y. (1989). Purification and characterization of beta-mannosidase from human placenta. *J Biochem* (*Tokyo*), **106**, 331–335. 1

Jackson, J. D. (1975). Classical electrodynamics. John Wiley & Sons Inc, New York., 2nd edition. 30

Jancarik, J. & Kim, S.-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Cryst*, **24**, 409–411. **21**

Jones, M. Z. & Dawson, G. (1981). Caprine beta-mannosidosis. Inherited deficiency of beta-D-mannosidase. *J Biol Chem*, **256**, 5185–5188. **5**

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A*, **47**, 110–119. **52**

Joseph, D., Petsko, G. A. & Karplus, M. (1990). Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science*, **249**, 1425–8. **12** Karle, J. & Hauptman, H. A. (1956). A theory of phase determination for the four types of non-centrosymmetric space groups 1P222, 2P22, $3P_12$, $3P_22$. Acta Cryst., **9**, 635–651. **38**

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. Acta Cryst. D, 55, 484–491. 81

Kleywegt, G., Zou, J., Kjeldgaard, M. & Jones, T. (2001). Around O. In Rossmann & Arnold (2001), chapter 17.1, pages 353–356, 366–367. 78

Kleywegt, G. J. (2000). Validation of protein crystal structures. Acta Cryst. D, 56, 249–265. 55, 57

Kleywegt, G. J. & Brünger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure*, **4**, 897–904. **54**, **55**

Kleywegt, G. J. & Jones, T. A. (1995). Where freedom is given, liberties are taken. *Structure*, **3**, 535–540. **54**

Kleywegt, G. J. & Jones, T. A. (1996). Phi/psi-chology: Ramachandran revisited. *Structure*, **4**, 1395–1400. **57**

Kleywegt, G. J. & Jones, T. A. (1997). Model building and refinement practice. In Carter Jr. & Sweet (1997b), pages 208–230. 54

Kleywegt, G. J. & Jones, T. A. (2002). Homo crystallographicus - quo vadis? *Structure*, **10**, 465–472. **15**

Knowles, J. R. (1991). Enzyme catalysis: not different, just better. *Nature*, **350**, 121–124. **11**

Knowles, J. R. & Albery, W. J. (1977). Perfection in enzyme catalysis: the energetics of triosephosphate isomerase. *Acc. Chem. Res.*, **10**, 105–111. **11**

Kohlhoff, M., Dahm, A. & Hensel, R. (1996). Tetrameric triosephosphate isomerase from hyperthermophilic archaea. *FEBS Lett.*, **383**, 245–250. **11**

Kornfeld, R. & Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Ann. Rev. Biochem., 54, 631–664. 2

Krietsch, W. K., Pentchev, P. G., Klingenburg, H., Hofstatter, T. & Bucher, T. (1970). The isolation and crystallization of yeast and rabbit liver triose phosphate isomerase and a comparative characterization with the rabbit muscle enzyme. *Eur J Biochem.*, **14**, 289–300. **13**

Kuhad, R. C., Singh, A. & Eriksson, K. L. (1997). Microorganisms and enzymes involved in the degradation of plant fiber cell walls. *Biochem. Eng. Biotechnol.*, **57**, 45–125. **6**, **8**

Kulminskaya, A. A., Eneiskaya, E. V., Isaeva-Ivanova, L. S., Savel'ev, A. N.,
Sidorenko, I. A., Shabalin, K. A., Golubev, A. M. & Neustroev, K. N. (1999). Enzymatic activity and betagalactomannan binding property of beta-mannosidase from *Trichoderma reesei*. *Enzyme Microb. Technol.*, 25, 372–377. 1, 2, 7, 9

Lambeir, A. M., Opperdoes, F. R. & Wierenga, R. K. (1987). Kinetic properties of triose-phosphate isomerase from *Trypanosoma brucei brucei*. A comparison with the rabbit muscle and yeast enzymes. *Eur J Biochem*, **168**, 69–74. **13**

Leipprandt, J. R., Kraemer, S. A., Haithcock, B. E., Chen, H., Dyme, J. L., Cavanagh, K. T., Friderici, K. H. & Jones, M. Z. (1996). Caprine beta-mannosidase: sequencing and characterization of the cDNA and identification of the molecular defect of caprine beta-mannosidosis. *Genomics*, **37**, 51–56. 5

Lesk, A. M., Br nd n, C. & Chothia, C. (1989). Structural principles of α/β barrel proteins: the packing of the interior of the sheet. *Prot. Struc. Func.* Gen., 5, 139–148. 11

Leslie, A. G. W. (1988). A reciprocal space algorithm for calculating molecular envelope using the algorithm of B. C. Wang. In *Improving Protein Phases*, edited by S. Bailey, E. Dodson & S. Phillips, CCP4 Daresbury Study Weekend, pages 25–31. Daresbury Laboratory. 48

Li, Y. & Lee, Y. C. (1972). Pineapple α and β -d-mannopyranosidases and their action on core glycopeptides. *J. Biol. Chem.*, **247**, 3677–3683. 1

Lindley, P. F. (1999). Macromolecular crystallography with a thirdgeneration synchrotron source. *Acta Cryst. D*, **55**, 1654–1662. **31**

Lolis, E. & Petsko, G. A. (1990). Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5 Å resolution. Implications for catalysis. *Biochemistry*, **29**, 6619–6625. **1**2

Lolis, E., Alber, T., Davenport, R. C., Rose, D., Hartman, F. C. & Petsko,
G. A. (1990). Structure of Yeast Triosephosphate Isomerase at 1.9 Å Resolution. *Biochemistry*, 29, 6609–6618. 9, 12

Lundqvist, T. & Schneider, G. (1991). Crystal structure of activated ribulose-1,5-bisphosphate carboxylase complexed with its substrate, ribulose-1,5-bisphosphate. J. Biol. Chem., **266**, 12604–12611. 12

Maes, D., Zeelen, J. P., Thanki, N., Beaucamp, N., Alvarez, M., Dao Thi,
M. H., Backmann, J., Martial, J. A., Wyns, L., Jaenicke, R. & Wierenga,
R. K. (1999). The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: a comparative thermostability structural analysis of ten different TIM structures. *Proteins*, **37**, 441–453.

Mainfroid, V., Terpstra, P., Beauregard, M., Frere, J. M., Mande, S. C., Hol, W. G., Martial, J. A. & Goraj, K. (1996). Three hTIM mutants that provide new insights on why TIM is a dimer. *J. Mol. Biol.*, **257**, 441–456. 11

Maldonado, E., Soriano-García, M., Moreno, A., Cabrera, N., Garza-Ramos, G., Tuena de Gómez-Puyou, M., Gómez-Puyou, A. & Pérez-Montfort, R. (1998). Differences in the intersubunit contacts in triosephosphate isomerase from two closely related pathogenic trypanosomes. *J. Mol. Biol.*, **283**, 193–203. **9**, **13**

Mande, S. C., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A. & Hol, W. G. (1994). Crystal structure of recombinant human triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase-related human genetic disorders and comparison with the trypanosomal enzyme. *Protein Sci*, **3**, 810–821. **9**, **12**, **13**

Marti-Renom, M. A., Stuart, A., Fiser, A., S nchez, R., Melo, F. & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct., **29**, 291–325. **118**

Mason, R. W. (1986). Observations on the effects of management stress on beta-mannosidase activity in goats. *Aust Vet J.*, **63**, 339–340. **5**

McMillan, F. M., Cahoon, M., White, A., Hedstrom, L., Petsko, G. A. & Ringe, D. (2000). Crystal structure at 2.4 Å resolution of *Borrelia burg-dorferi* inosine 5'-monophosphate dehydrogenase: evidence of a substrate-induced hinged-lid motion by loop 6. *Biochemistry*, **39**, 4533–4542. **12**

McPherson, A. (1982). *Preparation and analysis of protein crystals*. John Wiley & Sons Inc, Malabar, Florida. 15, 16, 18

McRee, D. (1999). A versatile program for manipulating atomic coordinates and electron density. *Journal Structural Biology*, **125**, 156–165. **52**

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst. D*, **53**, 240–255. **56**, 177

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst. A*, **50**, 157–163. **81**, 118

Neustroev, K. N., Krylov, A. S., Firsov, L. M., Abroskina, O. N. & Khorlin, A. Y. (1992). Isolation and properties of beta-mannosidase from *Aspergillus awamori*. *Biochemistry-Moscow, part 1*, **56**, 982–986. 1, 2

Neustroev, K. N., Valter, S. N., Timchenko, N. V., Firsov, L. M., Golubev, A. M. & Khokhlov, S. E. (1993). Adsorption of glucoamylase from *Aspergillus awamori* X-100/D27 on cell walls. *Biochem. Mol. Biol. Int.*, **30**, 115–120. **8**

Nickbarg, E. B. & Knowles, J. R. (1988). Triosephosphate isomerase: energetics of the reaction catalyzed by the yeast enzyme expressed in *Escherichia coli*. *Biochemistry*, **27**, 5939–5947. **11**

Noble, M. E., Zeelen, J. P. & Wierenga, R. K. (1993a). Structures of the "open" and "closed" state of trypanosomal triosephosphate isomerase, as observed in a new crystal form: implications for the reaction mechanism. *Proteins*, **16**, 311–26. **12**

Noble, M. E. M., Wierenga, R. K., Lambeir, A.-M., Opperdoes, F. R., Thunnissen, A.-M. W. H., Kalk, K. H., Groendijk, H. & Hol, W. G. J. (1991). The adaptability of the active site of trypanosomal triosephosphate isomerase as observed in the crystal structures of three different complexes. *Proteins*, **10**, 50–69. **12** Noble, M. E. M., Zeelen, J. P., Wierenga, R. K., Mainfroid, V., Goraj, K., Gohimont, A. C. & Martial, J. A. (1993b). Structure of triosephosphate isomerase from *Escherichia coli* determined at 2.6 Å resolution. *Acta Cryst. D*, 49, 403–417.

Oda, Y., Komaki, T. & Tonomura, K. (1993). Production of betamannanase and beta-manosidase by *Enterococcus casseliflavus* FL2121 isolated from decayed konjac. *Food Microbiol.*, **10**, 353–358. **1**

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*, **276**, 307–326. 70, 72, 73

Percheron, F., Foglietti, M. J., Bernard, M. & Ricard, B. (1992). Mammalian beta-d-mannosidase and beta-mannosidosis. *Biochimie*, **74**, 5–11. 1, 2, 5

Perrakis, A., Morris, R. M. & Lamzin, V. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.*, 6, 458–463. 51

Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R. C., Arruda, P., Leite,
A. & Craievich, A. (1998a). Protein crystallography station at LNLS, The
Brazilian National Synchrotron Light Source. Nucl Instrum Methods A,
405, 159–164. 30

Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998b). Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. J Synchrotron Radiat, 5, 72–76. 31

Read, R. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst. D*, **57**, 1373–82. **81**, **118**

Rieder, S. V. & Rose, I. A. (1959). The mechanism of the triosephosphate phosphate reaction. J. Biol. Chem., 234, 1007–1010. 9

Rietveld, A. W. M. & Ferreira, S. T. (1996). Deterministic pressure dissociation and unfolding of triose phosphate isomerase: persistent heterogeneity of a protein dimer. *Biochemistry*, **35**, 7743–7751. 15

Rietveld, A. W. M. & Ferreira, S. T. (1998). Kinetics and energetics of subunit dissociation/unfolding of TIM: the importance of oligomerization for conformational persistence and chemical stability of proteins. *Biochemistry*, 37, 933–937. 15

Rodriguez-Serna, M., Botella-Estrada, R. M. D., Chabas, A., Coll, M., Oliver, V., Febrer, M. & Aliaga, A. (1996). Angiokeratoma corporis diffusum associated with beta-mannosidase deficiency. *Arch. Dermatol.*, **132**, 1219– 1222. 5

Rose, I. A. (1962). Mechanisms of C-H bond cleavage in aldolase and isomerase reactions. *Brookhaven Symp. Biol.*, **15**, 293–309. 9

Rossmann, M. G. & Arnold, E. (eds.) (2001). Dordrecht: Kluwer Academic Publishers. 18, 197

Rossmann, M. G. and van Beek, C. G. (1999). Data processing. *Acta Cryst.* D, 55, 1631–1640. 36

Sayre, D. (1974). Least-squares phase refinement. II. High-resolution phasing of a small protein. Acta Cryst. A, **30**, 180–184. **4**9

Schliebs, W., Thanki, N., Eritja, R. & Wierenga, R. (1996). Active site properties of monomeric triosephosphate isomerase (monoTIM) as deduced from mutational and structural studies. *Protein Sci.*, **5**, 229–239. **11** Schliebs, W., Thanki, N., Jaenicke, R. & Wierenga, R. K. (1997). A double mutation at the tip of the dimer interface loop of triosephosphate isomerase generates active monomers with reduced stability. *Biochemistry*, **36**, 9655–9662. **11**

Schuller, D. J. (1996). MAGICSQUASH: more versatile noncrystallographic averaging with multiple constraints. *Acta Cryst. D*, **52**, 425–434. **4**9

Scopes, R. K. (1977a). Multiple enzyme purifications from muscle extracts by using affinity-elution-chromatographic procedures. *Biochem J.*, 161, 265–277. 13

Scopes, R. K. (1977b). Purification of glycolytic enzymes by using affinityelution chromatography. *Biochem J.*, **161**, 253–63. **13**

Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (1989). The metabolic basis of inherited disease, volume II. McGraw-Hill, Inc., USA, 6th edition.

Sheldrick, G. M. (1998). Direct methods for solving macromolecular structures. pages 401–411. Dordrecht: Kluwer Academic Publishers. 38, 46

Sheldrick, G. M. & Schneider, T. R. (2001). Direct methods for macromolecules. In *Methods in Macromolecular Crystallography*, edited by L. Johnson & D. Turk, volume 325, pages 72–81. NATO Science Series: Life Sciences, IOS PRESS. 46

Spraggon, G. (1999). Envelope skeletonization as a means to determine monomer masks and non-crystallographic symmetry relationships: application in the solution of the structure of fibrinogen fragment D. Acta Cryst. D, 55, 458–463. 78 Suurnakki, A., Tenkanen, M., Buchert, J. & Viikari, L. (1997). Hemicellulases in the bleaching of chemical pulps. Adv. Biochem. Eng. Biotechnol., 57, 261–287. 6, 8

Swanson, S. M. (1994). Core tracing: depicting connections between features in electron density. Acta Cryst. D, 50, 695–708. 48

Tanaka, T., Kimura, H., Hayashi, M., Fujioshi, Y., Fukuhara, K. & Nakamura, H. (1994). Characteristics of a *de novo* designed protein. *Protein Sci*, 3, 419–427. 15

Taylor, C. A. (1981). A Non-Mathematical Introduction to X-ray Crystallography. In *IUCR Teaching Pamphlets*, Series 1. International Union of Crystallography and University College Cardiff Press, Cardiff, Wales. 37, 38

Teramoto, Y., Kira, I. & Hayashida, S. (1989). Multiplicity and preferential inactivation by proteolysis as to raw starch-digestibility of bacterial α -amylases. *Agric. Biol. Chem.*, **53**, 601–605. **8**

Terwilliger, T. (2000). Maximum-likelihood density modification. Acta. Cryst. D, 56, 965–972. 50, 51, 92

Terwilliger, T. & Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Cryst. D*, **55**, 849–861. **46**, **74**, **90**

Velanker, S. S., Ray, S. S., Gokhale, R. S., Suma, S., Balaram, H., Balaram, P. & Murthy, M. R. N. (1997). Triosephosphate isomerase from *Plasmodium falciparum*: the crystal structure provides insights into antimalarial drug design. *Structure*, 5, 751–761. 9

Verlinde, C. L. M. J., Noble, M. E. M., Kalk, K. H., Groendijk, H., Wierenga, R. K. & Hol, W. G. J. (1991). Anion binding at the active site

of trypanosomal triosephosphate isomerase. Monohydrogen phosphate does not mimic sulphate. *Eur. J. Biochem.*, **198**, 53–57. **1**2

Voet, D. & Voet, J. (1995). *Biochemistry*. John Wiley & Sons, Inc., New York, USA, 2nd edition. 2, 8, 18

Walden, H., Bell, G. S., Russell, R. J. M., Siebers, B., Hensel, R. & Taylor, G. L. (2001). Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase. *J. Mol. Biol.*, **306**, 745–757. **9**

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. In *Diffraction Methods for Biological Macromolecules*, edited by H. Wyckoff, C. H. W. Hirs & S. N. Timasheff, volume 115 of *Methods in Enzymology*, pages 90–113. Academic Press. 48

Warren, B. E. (1969). X-ray diffraction. Addison-Wesley series in metallurgy and materials. Reading, Mass., Addison-Wesley Pub. Co., Ontario, USA. 38

Watanabe, M., Zingg, B. C. & Mohrenweiser, H. W. (1996). Molecular analysis of a series of alleles in humans with reduced activity at the triosephosphate isomerase locus. *Am. J. Hum. Genet.*, **58**, 308–316. **13**

Watts, R. W. E. & Gibbs, D. A. (1986). Lysossomal storage deseases: biochemical and clinical aspects. Taylor and Francis, Inc., Philadelphia, USA. 3

Weeks, C. M. & Miller, R. (1999). The design and implementation of SnB version 2.0. J. Appl. Cryst., **32**, 120–124. **38**, 46, 77

Wenger, D. A., Sujanski, E., Fennessey, P. V. & Thompson, J. N. (1986). Human beta-mannosidase deficiency. *N Engl J Med*, **315**, 1201–1205. **5** Westbrook, E. M. & Naday, I. (1997). Charge-coupled device-based area detectors. In Carter Jr. & Sweet (1997a), pages 244–268. **31**

Wierenga, R., Borchert, T. V. & Noble, M. E. (1992a). Crystallographic binding studies with triosephosphate isomerases: conformational changes induced by substrate and substrate-analogues. *FEBS Lett*, **307**, 34–9. **12**

Wierenga, R. K., Noble, M. E., Vriend, G., Nauche, S. & Hol, W. G. (1991a). Refined 1.83 Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M-ammonium sulphate. A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. J. Mol. Biol., **220**, 995–1015. 9, **12**, **13**, **14**

Wierenga, R. K., Noble, M. E., Postma, J. P., Groendijk, H., Kalk, K. H., Hol, W. G. J. & Opperdoes, F. R. (1991b). The crystal structure of the "open" and the "closed" conformation of the flexible loop of trypanosomal triosephosphate isomerase. *Proteins*, **10**, 33–49. **12**

Wierenga, R. K., Noble, M. E. & Davenport, R. C. (1992b). Comparison of the refined crystal structures of liganded and unliganded chicken, yeast and trypanosomal triosephosphate isomerase. J. Mol. Biol., 224, 1115–26. 13

Williams, J. C., Zeelen, J. P., Neubauer, G., Vried, G., Backmann, J., Michels, P. A. M., Lambeir, A. M. & Wierenga, R. K. (1999). Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. *Protein Eng.*, **12**, 243–250. **9**

Winn, M., Dodson, E. J. & Ralph, A. (1997). Collaborative Computational Project, Number 4: Providing Programs for Protein Crystallography. In Carter Jr. & Sweet (1997b), pages 620–633. 18, 28, 46, 50, 51, 53, 56, 73, 81, 82, 88, 92, 98, 118

Xiang, J., Sun, J. & Sampson, N. S. (2001). The importance of hinge sequence for loop function and catalytic activity in the reaction catalyzed by triosephosphate isomerase. *J. Mol. Biol.*, **307**, 1103–1112. **12**

Yüksel, K. U. & Gracy, R. W. (1991). Triose- and hexose-phosphate isomerases. In *A study of enzymes*, edited by S. A. Kuby, volume II, chapter 19, pages 457–484. CRC Press. 9

Zabori, S., Rudolph, R. & Jaenick, R. (1980). Folding and association of triose phosphate isomerase from rabbit muscle. Z. Naturforsch C, **35**, 999–1004. **15**

Zhang, K. Y. J. & Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Cryst A*, **46**, 377–381. **48**, 49