Exploring the subsite specificity of *Schistosoma mansoni* aspartyl hemoglobinase through comparative molecular modelling

F.P. Silva Jr.[a], F. Ribeiro[b], N. Katz[b], S. Giovanni-De-Simone[a,c,*]

[a] *Laboratório de Bioquímica de Proteínas e Peptídeos, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro-RJ, Brazil*
[b] *Centro de Pesquisas René Rachou, FIOCRUZ, Belo Horizonte-MG, Brazil*
[c] *Departamento de Biologia Celular e Molecular, Instituto de Biologia, UFF, Niterói-RJ, Brazil*

Received 4 October 2001; revised 21 December 2001; accepted 21 December 2001

First published online 15 February 2002

Edited by Gunnar von Heijne

Abstract Blood flukes of the genus *Schistosoma* currently infect millions of people in tropical and subtropical countries. An enzyme playing a major role in hemoglobin (Hb) degradation by *Schistosoma mansoni* has been cloned and shown to be highly similar to the human cathepsin D aspartyl proteinase, although presenting a distinct substrate specificity from the latter. Investigating the structural features responsible for this difference has a major application in the design of selective anti-schistosomal drugs. In order to achieve this goal a homology model for the *S. mansoni* aspartyl hemoglobinase was constructed and then used to simulate the complexes formed with two transition state analogues of Hb-derived octapeptide substrates. Comparison with human cathepsin D showed that different pocket volumes and surface electrostatic potentials arise from substitutions in residues comprising the S4, S3, S2 and S3′ subsites. Since the primary specificity of the *S. mansoni* enzyme resembles that of HIV-1 protease, we have discussed the applicability of current retroviral protease inhibitors as leads for the design of new anti-schistosomal drugs.

Key words: Schistosomiasis; Aspartyl hemoglobinase; Specificity; Cathepsin D; *Schistosoma mansoni*

1. Introduction

Schistosomiasis is a major human parasitic disease. An estimative from the World Health Organization in 1993 [1] indicates that the blood flukes of the genus *Schistosoma* may infect over 200 million people in more than 74 countries throughout the world. Schistosome infestation is currently both treatable and preventable. The former is being achieved by oral administration of praziquantel or oxamniquine and the later by health educational programs. Nevertheless, there is no apparent sign of improvement in schistosomiasis morbidity indexes in most developing countries, possibly due to a premature increased tolerance to praziquantel [2]. Thus, seeking novel strategies for schistosomiasis chemotherapy is of major importance.

Schistosomes as any other blood-feeding parasite depend on hemoglobin (Hb) catabolism for survival. Several already characterized or putative Hb degrading enzymes have been isolated from protozoans and helminths, including: *Plasmodium* sp. [3–6], *Ancylostoma caninum* [7], *Strongyloides stercoralis* [8] and *Haemonchus contortus* [9]. These proteolytic enzymes have been shown to belong either to the cysteinyl or to the aspartyl families of peptidases [10]. In schistosomes, a cathepsin D-like proteinase from both *Schistosoma mansoni* and *Schistosoma japonicum* species has been cloned [11,12]. Human cathepsin D is an aspartyl protease that has been assumed to be involved in the generation of peptides for antigen presentation [13], cancer metastasis [14] as well as in the formation of the plaques encountered in the Alzheimer's disease [15]. Today, a lot of knowledge has been accumulated from structural studies of aspartyl proteinases involved in the pathogeny of important human diseases [16]. Nevertheless, where no experimental information on the structure of the studied protease is available, molecular modelling plays a crucial role on the prediction of structure–function relationships [17–20]. The value of these computational techniques has been proved by the subsequent confirmation of numerous predictions through comparison with the structures resolved by nuclear magnetic resonance techniques or, in a greater extent, by X-ray crystallography [21–23].

Brindley et al. [24] have shown that the *S. mansoni* enzyme presents a discrete substrate preference from its orthologous host enzyme, cleaving human Hb in several distinct sites. Therefore, exploring this difference has a potential application on the design of selective inhibitors of *S. mansoni* cathepsin D-like hemoglobinase, which can represent a novel strategy for the combat of schistosomiasis. Noticeably, Brinkworth et al. [25] have recently focused the structural features responsible for the α-chain Phe36–Pro37 cleavage site on human Hb. This is a very unusual site for an eukaryotic aspartyl proteinase since it was essentially only observed in the aspartyl proteases of retroviral origin [16]. Pursuing the Phe–Pro cleavage site is a very interesting approach for substrate-based inhibitor design because it can furnish selectivity over eukaryotic enzymes. Nevertheless, addressing the general topological, electrostatic and steric features responsible for the discrete substrate preference shown by the schistosome and human enzymes should be of greater interest for broad-spectrum projects in structure-based drug design. Following this approach, a homology model for the *S. mansoni* aspartyl hemoglobinase has been constructed and used to simulate the complexes formed with two transition state analogues of

*Corresponding author. Fax: (55)-21-25903495.
E-mail address: dsimone@ioc.fiocruz.br (S. Giovanni-De-Simone).

Abbreviations: 3D, three dimensional

0014-5793/02/$22.00
PII: S0014-5793(02)02270-6

Hb-derived octapeptide substrates (αL105–L106/SHCL* LVTL and βL31–L32/LGRL*LVVY). The modelled complexes were used to pinpoint the structural discrepancies found in the subsites of the schistosome and human cathepsin D enzymes that could be correlated to the enzymatic activities and binding affinities reported in the literature for these proteases. In addition, in view of the similarity at the substrate preference level of the aspartyl proteases of schistosoma and HIV-1, we have discussed the feasibility of current HIV-1 protease inhibitors as leads for the design of novel anti-schistosomal drugs.

2. Materials and methods

2.1. Sequence analysis

The sequence of the mature *S. mansoni* cathepsin D-like protease (CatD_Sman), comprising 377 residues, was extracted from the zymogen sequence deposited in TrEMBL protein databank under accession number P91802. Aspartyl proteases found on the complexes of cathepsin D-pepstatin (PDB ID 1LYB, 2.50 Å resolution) [21], mouse renin–CH-66 (PDB ID 1SMR, 2.00 Å resolution) [22] and pepsin 3A-phosphonate inhibitor (PDB ID 1QRP, 1.96 Å resolution) [26] were used as templates in the modelling procedure. A reliable alignment between CatD_Sman and its templates was achieved by adjusting the multiple sequence alignments obtained either from the pepsin A1 family in MEROPS database [27] or from ClustalW v. 1.72 [28], in order to maximize the placement of insertions and deletions in loops and to align features conserved in the pepsin family of proteases, such as the three pairs of cysteine residues. This alignment was also used as an input to the program MEGA v. 2.1 [29] to calculate the pairwise distances (using the Poisson correction model and a pairwise deletion of gap containing segments), which were used to generate a phenogram by the neighbor-joining method. The reliability of each internal branch was assessed by the bootstrap method implemented in MEGA.

2.2. Comparative molecular modelling

The sequence of CatD_Sman was folded onto the superposed structures of its templates according to the optimal alignment, obtained as explained above, using the SwissPDB Viewer v. 3.6 program and then submitted in optimize mode to the Swiss-Model automated modelling server [30]. This modelling method has been rigorously assessed by a large-scale initiative called the 3DCrunch project (http://www.expasy.ch/swissmod/SM_3DCrunch.html). Briefly, models whose sequences shared around 50% identity to templates deviate from experimental structures by approximately 1 Å relative mean square deviation in the common core. Suitability of the modelled CatD_Sman three dimensional (3D) structure for this study was assessed with the analytical tools available in SwissPDB Viewer v. 3.6 and with the programs in the PROCHECK suite [31]. Residues corresponding to the carboxy-terminal extension found in *S. mansoni* aspartyl protease were not considered in the modelling because this segment showed to have no counterpart on protein structures determined to date. As no function has been assigned to this region yet and as schistosomes are believed to express both forms of the enzyme, with and without this segment [11], we decided not to address this question here.

2.3. Modelling of active site residues

The geometry of the residues composing the catalytic triad in the active site is highly conserved among the aspartyl proteases [16]. Following this principle, we have applied a manual torsion on the χ^2 angle of D33 and D219 side chains, taking as a guide the catalytic aspartate side chains from human cathepsin D structure [21], in order to guarantee that both groups belong to the plane observed in aspartyl proteinases. Finally, 20 cycles of steepest descent and 20 cycles of conjugate gradient energy minimization with the Gromos96 force field [32] implemented in SwissPDB Viewer v. 3.6 were applied to release any problems with stereochemistry that would have been imposed by the manual torsion procedure.

2.4. Substrate modelling

The 3D structure of mouse renin complexed with the decapeptide inhibitor CH-66 was used as the archetype to model the transition state isoster of the pseudo-substrates LGRLψ[CH(OH)CH_2]LVVY and SHCLψ[CH(OH)CH_2]LVYL. CH-66 was chosen because: (i) it almost perfectly resembles the transition state intermediate formed during the cleavage of a peptide bond and (ii) it would be large enough to furnish the extended backbone necessary to model an octapeptide. Substrate modelling was achieved by superimposing the CatD_Sman model on the enzyme–inhibitor complex of mouse renin (0.70 Å in 1264 backbone atoms) and then transferring the coordinates of the inhibitor onto the model structure. Finally, collisions and improper stereochemistry of the mutated side chains in the inhibitor–enzyme complex were released by 20 cycles of steepest descent and 20 cycles of conjugate gradient energy minimization with the Gromos96 force field. Hydrogen bonds and non-bonded contacts between the modelled transition state isosteres of Hb-derived substrates and CatD_Sman model were calculated with HBPLUS [33].

3. Results and discussion

3.1. Overall model structure and validation

The structure of CatD_Sman, as expected, resembles that of most eukaryotic aspartyl proteinases (Fig. 2A). A six-stranded

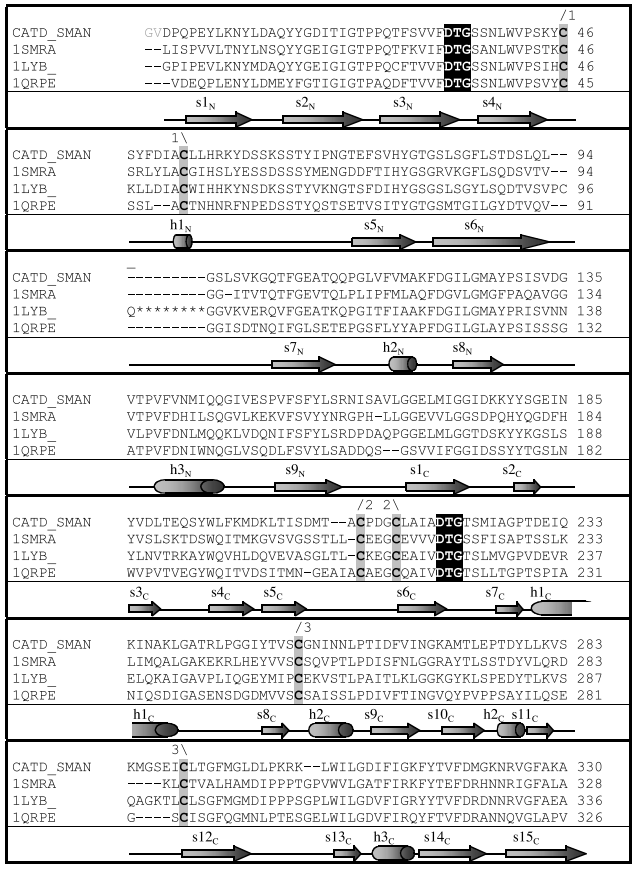

Fig. 1. Multiple sequence alignment used on the construction of *S. mansoni* aspartyl hemoglobinase model. CatD_Sman – *S. mansoni* aspartyl hemoglobinase; 1SMRA – mouse renin; 1LYB – human cathepsin D and 1QRPE – human pepsin 3A. The black arrowhead indicates the end of cathepsin D light chain. Asterisks in 1LYB sequence correspond to residues excised during post-translational processing of human cathepsin D. The three conserved cysteine pairs are highlighted in gray while the two catalytic triads are highlighted in black. In the bottom of each alignment block is represented the secondary structure of CatD_Sman derived from its model (coded to facilitate discussion in the text). Secondary structure elements are represented as cylinders (α-helices) and arrows (β-strands). The first two residues in the CatD_Sman sequence (shown in grayish) were not modelled neither considered in the numeration.

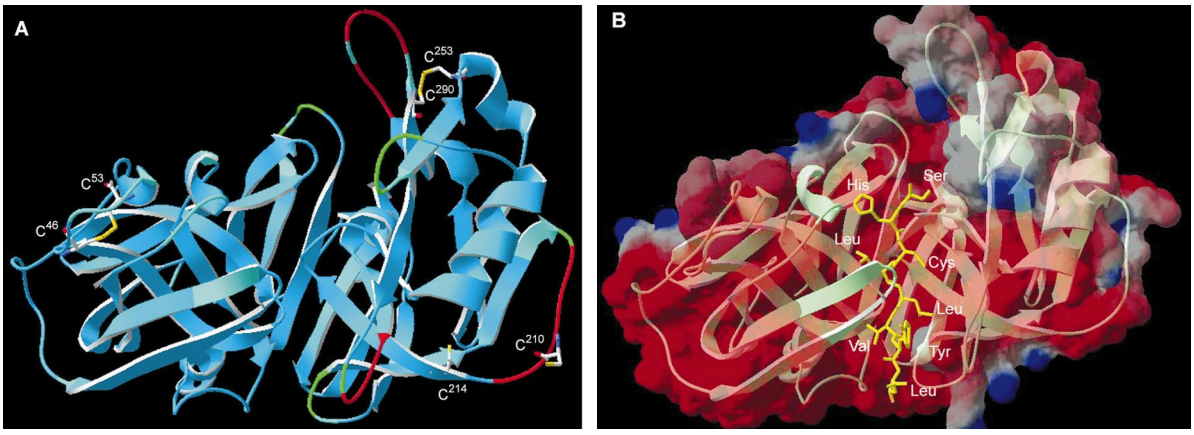

Fig. 2. A: *S. mansoni* aspartyl hemoglobinase structure colored by model B-factor. Cold residues are shown in blue while hot residues are in red. Labelled cysteine residues are shown as CPK-colored sticks. Disulfide linkages are shown in yellow. B: Modelled complex between the Hb-derived peptide αL105–L106 and CatD_Sman model. Top view of the active site binding groove where the surface was made transparent to allow the perception of the internal protease architecture, depicted as greenish ribbons. The molecular surfaces in this work were calculated in SwissPDB Viewer v. 1.6 [30] and colored according to the coulombic electrostatic potential: red – negative, blue – positive and white – neutral.

β-sheet connects the two lobes of the protein and constitutes the active site floor. The catalytic apparatus is assembled in the middle of the protease by two loops localized on the tips of structures known as the 'greek keys' for their similarity to the *psi* (ψ) [16]. Each of these loops contributes with three residues Asp, Thr and Gly (D33, T34, G35 and D219, T220, G221 according to the numeration used in Fig. 1) that are highly conserved among the aspartyl proteases [16]. The β-turn positioned over the active site cleft, known as the 'flap', is composed of residues H77, Y78, G79 and T80, in the proposed model. The CatD_Sman sequence does not present the insertion between residues L94 and G95, where would be encountered the surface loop that is excised from human cathepsin D (residues S98–L105, shown as asterisks in Fig. 1) resulting in the formation of the light and heavy chains of this enzyme. In the CatD_Sman model this loop was predicted as a short β-turn between the $s6_N$ and $s7_N$ β-strands based on the structures of mouse renin and human pepsin 3A.

Most uncertainties related to the CatD_Sman model were identified in structurally variable regions (SVRs), as indicated by the red-colored loops in Fig. 2A. Surprisingly, only two disulfide linkages could be detected in the model (represented by the yellow lines connecting C46 to C53 and C253 to C290 in Fig. 3). The CatD_Sman model cannot assure the absence of the linkage that would be formed by residues C210 and C214 since the former is located in a SVR. The PROCHECK suite of programs was used to assess the stereochemical quality of the model. Around 86% of the residues were plotted in the most favored regions (A, B, L) of the Ramachandran plot. Similar values were encountered on the analysis of the templates used in CatD_Sman modelling. Six residues (2.2%) felt in generously allowed regions of the plot (S96, S98, V99, I163, A164 and T305) and only four residues (1.4%) were found to be in disallowed regions (N161, S163, S205 and K284). None of these is found in CatD_Sman subsite pockets. On the main-chain parameters check, the ω angle standard deviation showed to be 1.9 times above the typical value for high-resolution structures. This angle, in general terms, represents a measure of the peptide bond planarity. Residues with irregular ω torsions that are near the substrate binding groove are F32, L113 and M224. Remaining stereochemical parameters

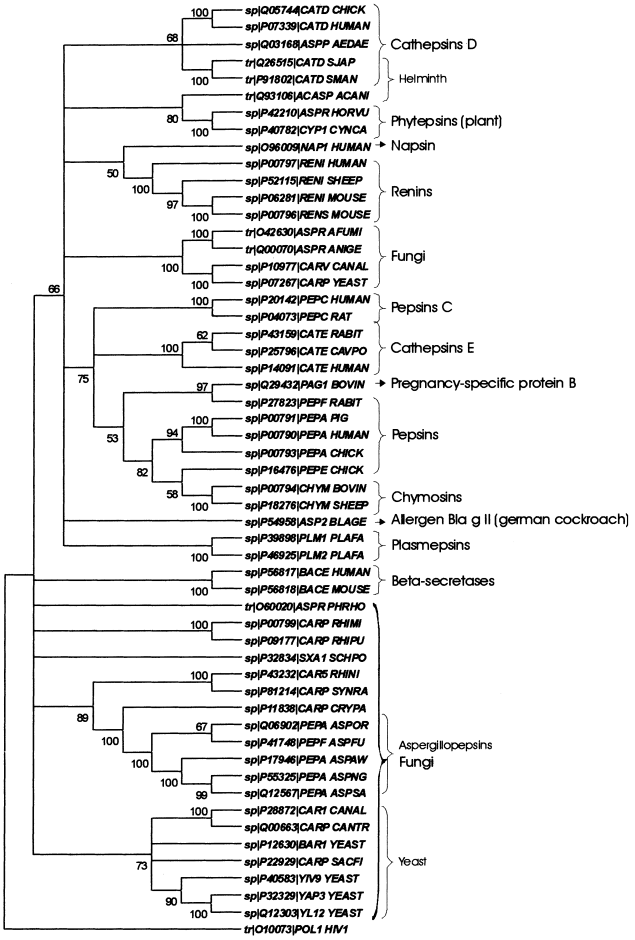

Fig. 3. Phenogram of 54 mature sequences of representative aspartyl proteinases (pepsin A1 family [27]) rooted with the HIV-1 proteinase monomer. The sequences were aligned using the ClustalW v. 1.81 program [28]. The tree was constructed according to the neighbor-joining method from the calculated pairwise distances (MEGA v. 2.1 [29]). Numbers on the branches represent percent agreement values obtained by the bootstrap procedure. Sequences are denoted with database accession numbers (tr: translated EMBL; sp: SwissProt) followed by codes in SwissProt format.

Table 1
Hydrogen bonds between residues mapped in the S4–S4′ subsites of the CatD_Sman model and the atoms of the substrates transition state analogues

Enzyme residue and atom		Substrate analogue	Substrate analogue moiety	Distance (Å)
S223	Oγ	αL105–L106	His (P3) N	2.99
		βL31–L32	Gly (P3) N	2.92
	N	αL105–L106	His (P3) O	2.90
		βL31–L32	Gly (P3) O	2.80
T80	$O\gamma^1$	αL105–L106	Cys (P2) N	2.62
		βL31–L32	Arg (P2) N	2.49
	N	αL105–L106	Cys (P2) O	3.31
		βL31–L32	Arg (P2) O	2.90
G79	N	αL105–L106	Cys (P2) O	3.08
		βL31–L32	Arg (P2) O	3.13
		αL105–L106	Leu (P1′) O	3.31
		βL31–L32	Leu (P1″) O	3.31
D33	$O\delta^1$	αL105–L106	Leu (P1) O	2.85
	$O\delta^2$	αL105–L106	Leu (P1) O	2.94
D219	$O\delta^2$	βL31–L32	Leu (P1) O	2.38
G221	O	αL105–L106	Leu (P1) N	3.10
		βL31–L32	Leu (P1) N	3.07
G35	O	αL105–L106	Val (P2′) N	2.68
		βL31–L32	Val (P2′) N	2.73
H77	O	βL31–L32	Val (P3′) N	3.27
K301	O	βL31–L32	Tyr (P4′) OH	2.75

for the whole CatD_Sman model were inside the expected values.

3.2. Subsite specificity and comparison with human cathepsin D active site

Brindley et al. [24] have observed that among the 17 Hb sites cleaved by the schistosome aspartyl proteases, 12 P1 residues and six P1′ residues were either Leu or Phe. In addition, two out of the three initial cleavage sites had a Leu residue either on P1 or P1′ positions. Considering that this markedly preference must be a result of a higher number of favorable contacts between substrate moieties and the active site residues atoms we have modelled two complexes between the CatD_Sman model and the transition state analogues of the peptides representing positions P4–P4′ in the cleavage sites αL105–L106 (SHCL*LVTL; Fig. 2B) and βL31–L32 (LGRL*LVVY) from α- and β-chains, respectively, of human Hb. These substrates could be appropriately modelled from the decapeptide renin inhibitor CH-66, which contains a Leu-ψ[$CHOHCH_2$]-Leu moiety, since a very tight superposition of the CatD_Sman model and mouse renin could be achieved. This procedure allowed us to define the enzyme residues corresponding to the subsite pockets S4–S4′ (Schechter and Berger's nomenclature [34]). For all of the subsites studied, with exception of S2, several hydrogen bonds could be detected between substrate backbone atoms and enzyme moieties (Table 1). Nevertheless, not all hydrogen bonds predicted between the $O\delta^1$ and $O\delta^2$ atoms from the catalytic aspartates and the P1-Leu carbonyl oxygen (as detected in the mouse renin–CH-66 complex [22]) were detected in CatD_Sman and its modelled substrates. This observation is probably explained by the existence of a slight difference in the rotameric position of the aspartate side chains in the crystallized mouse renin structure in comparison to the CatD_Sman model.

Table 2
Comparison between subsite composition of *S. mansoni* and human cathepsin D proteases[a]

Subsite	*S. mansoni*	Human	Subsite	*S. mansoni*	Human
S4	S223	S235	S4′	K301	P314
	M224	L236		R302	P313
	M295	M307			
	L291	L303			
S3	Q14	Q14	S3′	**H77**	H77
	T80	S80		R302	P313
	F120	F131		*L299*	I311
	F115	F126			
S2	T222	T234	S2′	G35	G35
	A226	V238		S36	S36
	L297	M309		H77	H77
	I306	I320		I131	I142
	I217	I229		*Y194*	Y205
S1	V31	V31	S1′	Y78	Y78
	D33	D33		G79	G79
	Y78	Y78		Y194	Y205
	G79	G79		I217	I229
	I123	I134		D219	D231
	G221	G233		L299	*I317*

[a]Highlighted in bold are residues mutated in *S. mansoni* protease in relation to cathepsin D. Detached in italic are the residues from CatD_Sman not detected in the stipulated contact range of 4.0 Å.

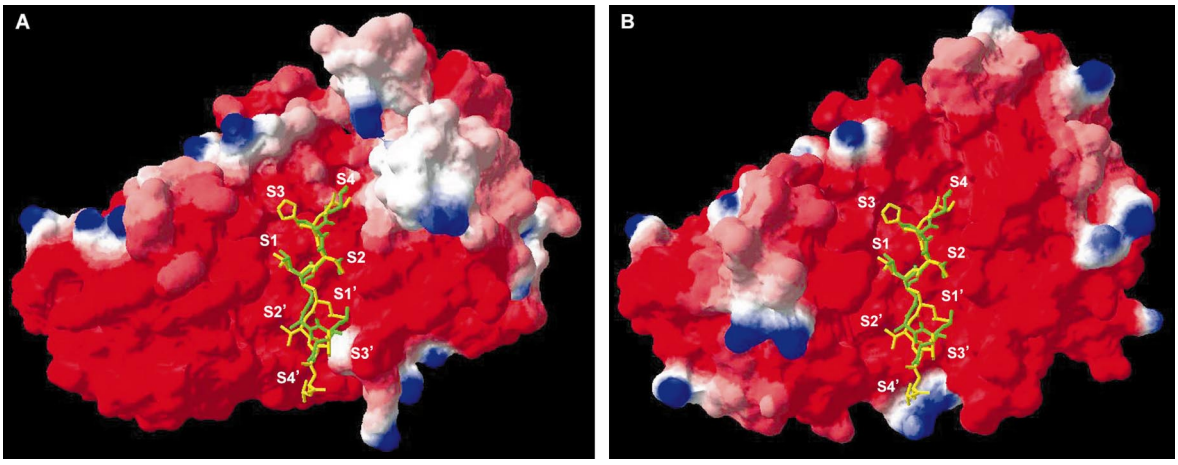

Fig. 4. Comparison between the electrostatic potential at the active site cleft surfaces of the CatD_Sman model (A) and the human cathepsin D structure (B). Subsite pockets are indicated in each structure. The surface corresponding to the flap and the second N-terminal helix are not shown for a better visualization of the substrate binding groove.

The first step in validating a parasitic protease as a target for chemotherapy should be the study on how to selectively block its activity over the host orthologous enzyme. This can be effectively achieved by correlating the differences found in substrate preference of the parasitic and human proteases with the structures in the active site of these enzymes. *S. mansoni* aspartyl hemoglobinase shows only moderated similarity to most human aspartyl proteases (including cathepsin E, renin and pepsin). Nevertheless, it is significantly similar to the lysosomal human enzyme cathepsin D (71% of similarity in 240 residues from the mature enzymes). Indeed it can be assumed from the phenogram shown in Fig. 3 that CatD_Sman is clearly closer to human cathepsin D than to any other host protease. The tree shown in Fig. 3 was rooted with HIV-1 proteinase monomer sequence, since it would better represent the ancestral gene that has given rise to the known eukaryotic aspartyl proteases [35]. This close relationship between the schistosome protease and human cathepsin D led us to focus on the comparison between the residue composition and the electrostatic properties on the surfaces of the subsite pockets from these enzymes. In Table 2 are listed the residues occupying homologous positions in the S4–S4′ subsites, according to a structural superimposition of the human and schistosome protease. Differences were encountered in the residue composition as well as in the electrostatic properties on the surfaces of the S4, S3, S2 and S3′ pockets.

The S2 and S4 subsites are fundamentally formed by the same molecular framework, where a clear distinction between the individual pockets was difficult to define. These pockets are formed by the β-turn between the $s6_C$ and $s7_C$ β-strands (spanning residues S223–G227, as defined in Fig. 1) and by the $s12_C$ strand (comprising residues C290–L297). Although a similar analysis by Brinkworth et al. [25] has classified this pocket as essentially hydrophobic, we found that this subsite architecture results in a shallow and amphipathic S4 pocket, which is similar to that described for human cathepsin D [21]. Mutation of L236 to M224 in the parasite enzyme provides a hydrophobic environment which can result in a more favorably interaction with a short hydrophobic side chain in opposition to the human enzyme. This could explain why a serine can be found among the residues occupying the P4 position in the sites cleaved in bovine Hb by the human cathepsin D [36] while this amino acid never occurred in the same position of human Hb sites cleaved by the schistosome enzyme. Important discrepancies between the parasitic and human enzymes could be encountered in the S2 subsite. In the human enzyme, A226 is substituted for V238 in the loop between $s6_C$ and $s7_C$ strands and L297 is substituted for M309 in the tip of the $s12_C$ β-strand. The large diversity in the kind of side chain favored in the large and hydrophobic S2 subsite of human cathepsin D poses further difficulty in interpreting the structure–activity relationships of the same pocket in the schistosome enzyme. Gulnick et al. [37] have observed that the human enzyme more effectively cleaves substrates having sulfur-containing residues such as methionine or cysteine in the P2 position. However, Scaborough and Dunn [38] have shown that the wild-type human enzyme bearing M287 (M309 according to the numeration in Fig. 1) displays a preference for hydrophobic residues such as Leu while moderately accommodating polar side chains such as Glu in the P2 position of

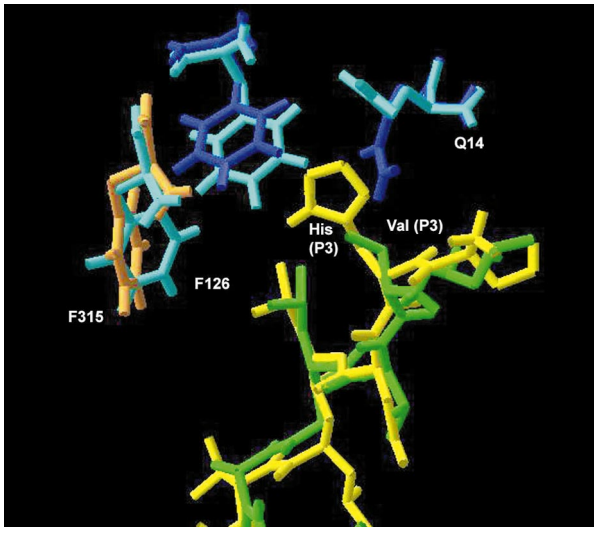

Fig. 5. Superimposition of S3 residues of CatD_Sman (cyan) and human cathepsin D (blue). Q14 assumes different conformations in both enzymes. F115 (orange) in the parasitic enzyme is in a different orientation from that of the correspondent F126 residue in the human protease.

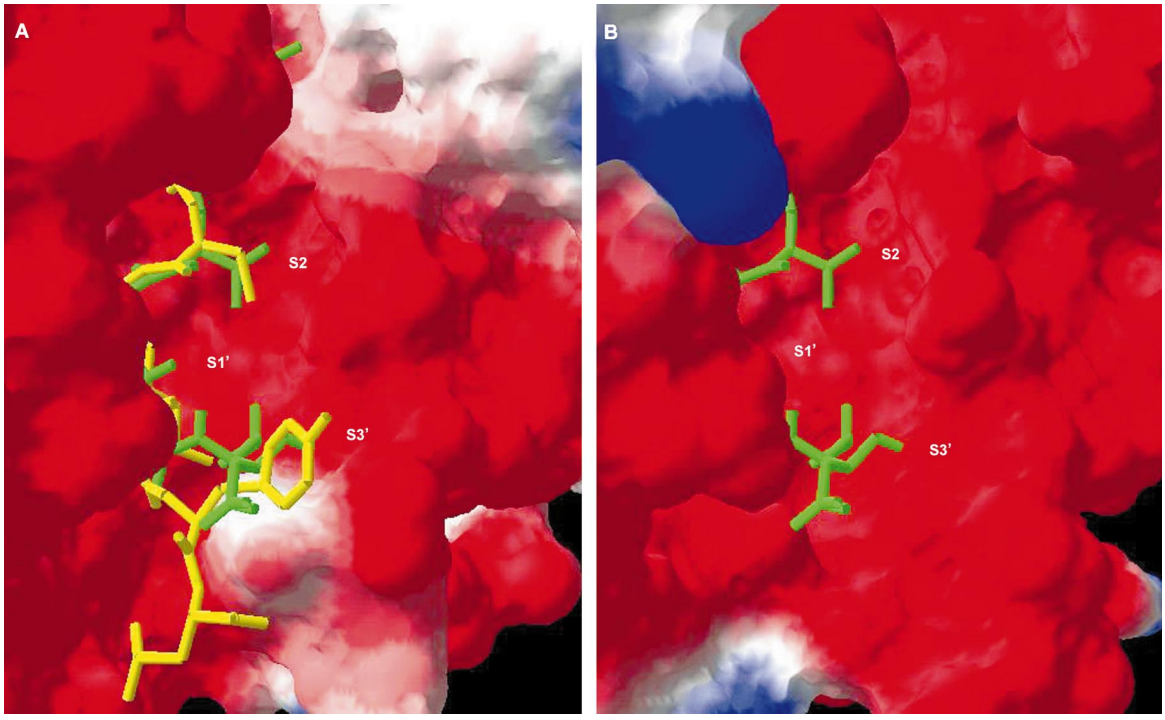

Fig. 6. Comparison of the surfaces at the active site C-terminal boundaries in CatD_Sman model (A) and human cathepsin D structure (B). A: The substrate analogue comprising αL105–L106 cleavage site of human Hb (SHCLψ[CH(OH)CH_2]LVTL; yellow) is shown superimposed on the statin-based inhibitor of human cathepsin D (Iva-VV-Sta-ψ[CH(OH)CH_2OH]GA-Sta; green [21]). B: Only the statin-based inhibitor is shown docked at the human cathepsin D active site cleft. Iva = isovaline and Sta = statin.

substrates. Thus, large hydrophobic (Met or Leu), medium polar (cysteine) as well as large positively charged (Glu) residues can fit the S2 pocket of human cathepsin D. According to the mutations observed in the S2 pocket of the parasitic enzyme we can only propose that this subsite being even larger would favor the binding of large hydrophobic groups. In addition, examining the surface of both enzymes, two major differences that can affect S4 and S2 subsites are encountered (Fig. 4). First, a protuberance in the C-terminal border is clearly differentiable in the human enzyme, which can reflect in less conformational freedom in the binding of the P4 and P2 residues of the substrate. This difference is caused by the mutation of G247 for Q258 in the human enzyme. Second, there is a considerable hydrophobic patch near the S4 pocket surface in CatD_Sman that is absent from the human enzyme, which may also favor the binding of hydrophobic residues in this subsite. In the flap position above the S3 subsite region, S80 is mutated for T80 in the parasite enzyme. This change seemed to have no major effect on the hydrogen-bonding pattern observed between the flap and the substrate backbone.

Although much information can be obtained directly by comparison of the primary sequence of the two enzymes, certain observations can only be made using 3D models. For instance, as depicted in Fig. 5, residue Q14 in the S3 subsite, in spite of being conserved, assumes different conformations in both enzymes. According to our model, the rotamer for Q14 in the parasitic protease has its side chain pointed outside the pocket, while in the human enzyme it is directed toward the binding cleft. This would enable bulkier side chains, such as that of histidine in the αL105–L106 substrate, occupy the S3 pocket (yellow in Fig. 5) where only a short side chain, such as that of valine in the P3 position of the statin-based inhibitor of human cathepsin D [21], would fit (green in Fig. 5). Indeed, carefully analysis of the cleavage sites in human Hb suggests that residues with longer side chains, such as arginine or glutamine, are preferred by the parasitic enzyme in the P3 position. On the other hand, it has recently been shown that the human enzyme attacks with greater efficiency fluorogenic substrates bearing the residues leucine, valine or glutamic acid [39] in this position. We also detected that F115 (orange in Fig. 5) in the parasitic enzyme is in a different orientation from that of the homologous F126 residue in the human protease. This possibly accounts for the fact that F115 was not initially detected in the set of residues within 4.0 Å from substrate atoms. Other residues have been substituted between the two enzymes in the loop connecting $s7_N$ and $h2_N$ (F115–I123) making up the framework of the S3 pocket, as noted before by Wong et al. [11], but none of these should affect interactions with P3 side chains of substrates.

None of the residues that are near to substrate atoms differs in S2′ subsite of both schistosome and human proteases. However, mutations in residues surrounding I131 in the loop connecting the $s8_N$ strand and the $h2_N$ helix (S130 for R141, D134 for N145 and G135 for N146) would turn the pocket in the mammalian enzyme more hydrophilic than in the parasitic homologue. It has been proposed before that the hydrophobic S2′ pocket in human enzyme can accommodate polar or charged residues by enabling the polar group to point out of the cleft and be stabilized by the solvent [40]. Indeed, inspection of the residues favored in P2′ position of protein substrates cleaved by human cathepsin D characterizes a preference for short and polar residues.

A special feature of the structures of renin and cathepsin D is the presence of a proline-rich loop (also referred to as the

'rigid loop') connecting the $s11_C$ and $s12_C$ β-strands forming the border of the S3′ and S4′ subsites. The equivalent structure in the CatD_Sman model is formed by a sequence of three basic residues – K301, R302 and K303, for this reason being hereafter referred to by us as the 'basic loop'. The proline-rich loop in human cathepsin D is displaced by 2.7 Å away from the center of the binding groove in relation to the basic loop in CatD_Sman. An immediate consequence of this difference can be noted as a protuberance in the surface of the CatD_Sman model, which would leave much less space for accommodating large side chains in these pockets (Fig. 6A). A good example of how this can affect the binding of substrates/inhibitors can be found in the backbone torsion within the statin-based inhibitor of human cathepsin D responsible for making P3′ pointing toward S1′ in the complex [21]. Comparison of the surfaces of the complexes on the S3′ binding pockets of both enzymes (Fig. 6A,B) shows that it would be impossible for an inhibitor of *S. mansoni* aspartyl protease to suffer this backbone torsion because of the steric blockage that is imposed by the basic loop. The complex between CatD_Sman and the modelled substrate (αL105–L106 cleavage site of Hb) shows that the bulk Tyr side chain in P3′ would have to be torsioned out of the binding pocket in order to be accommodated in S3′ (Fig. 6A). Based on the precedent discussion we propose that although larger side chains can occupy P3′ position in CatD_Sman substrates, a favorable fit would only be achieved by short and polar residues. Again, our conclusions disagree with the observations made by Brinkworth et al. [25] that considered the S3′ pocket as markedly hydrophobic. Analysis of interactions within the S4′ pocket of CatD_Sman was made difficult by the fact that: (i) P4′ in modelled substrates was positioned in the limit of the active site N-terminal half or prime region (adopting the nomenclature of Schechter and Berger [34]) and (ii) a direct comparison with the human cathepsin D complexed inhibitor was impossible in view of its particular binding mode as discussed before.

S1 and S1′ subsite structures of both schistosome and human enzymes were very similar. They showed an almost perfect superimposition of the strands forming the ψ-like structure in the active site of several aspartyl proteinases and all of the residues within 4.0 Å from the atoms of the modelled substrates were conserved between the two proteases. Nevertheless, several sites in human Hb α-chain are cleaved by the schistosome enzyme but are not recognized by the human protease, based on the orthologous sites in bovine Hb. Among these sites, αPhe36–Pro37 deserves special attention because it represents a very unusual primary specificity for an eukaryotic aspartyl proteinase. Brinkworth et al. [25] have proposed that a rationale for this would be that when Pro is at P1′ the backbone atom locations in P2, P1 and P1′ are changed, which in human cathepsin D would produce a destabilization because of the lack of important hydrophobic contacts between Phe at P1 and residues in the S1 pocket. By careful inspection of the models generated we have observed that the substitution of Ala preceding D219 in CatD_Sman by Val in the human protease would make the S1′ pocket of the later smaller and more hydrophobic. The larger S1′ pocket in CatD_Sman would collaborate to a better accommodation of the rigid proline cycle at this position and thus help Phe at P1 to find a more stable conformation in S1.

Cleavage sites bearing Phe–Pro at P1–P1′ positions have only been efficiently hydrolyzed by retroviral aspartyl proteinases [41] and it has been suggested that the selectivity evidenced by HIV-1 protease is the consequence of an extensive network of interactions, not only within the S4–S3′ subsites, but also extending beyond this segment [41]. This pattern would indicate that only a more rigorous and systematical study involving substitutions in P4–P4′ positions of synthetic substrates will be able to identify any subtle differences existent in S1–S1′ subsites from the schistosome and human proteases. The exclusive specificity of retroviral aspartyl proteinases has been extensively explored, as it became the first approach to reach selectivity within the anti-HIV-1 drugs [40]. Although successful in finding selective peptidic leads, this approach has fallen short in designing effective drugs mainly because of the low bioavailability of the designed peptidomimetics compounds [41]. The two compounds designed following this concept that have reached the market are the Hoffmann-La Roche RO318959 (Saquinavir) and the Merck L-735,524 (Indinavir). Most recent approaches that have taken full advantage on the 3D structure of the HIV-1 protease have been more productive in finding anti-viral drugs [41,42]. At this time three drugs were successfully designed in this way: the Vertex VX478 (Amprenavir), the Agouron AG-1343 (Nelfinavir) and the Pharmacia-Upjohn PNU-140690 (Tipranavir). The approach followed in this work, in which the general topological and chemical features responsible for substrate recognition by the schistosome and human orthologous enzyme were compared, would facilitate this type of drug design strategy.

The abnormal similarity with the HIV-1 protease can be explored in the design of inhibitors of the schistosome aspartyl hemoglobinase through the in vitro or in silico screening of known HIV-1 protease inhibitors. Nevertheless, besides the affinity for substrates bearing a hydrophobic Pro motif as P1–P1′ residues, a number of studies have pointed out the essentially hydrophobic S2 and S2′ subsites as the most critical in defining effective substrate interactions with the HIV-1 protease [43]. Thus, many substrate-based HIV-1 protease inhibitors reflect these requirements [44]. For example, the *cis*-decahydroisoqinoline moiety found in Saquinavir and Nelfinavir mimics the proline at P1′ as well as it provides some interactions at the S3′ binding site, while the bulky *t*-butylamide fills S2′. The general hydrophilic character of its S2′ binding pocket would lower the binding affinity of such inhibitors to the schistosome enzyme but it would also enhance the selectivity for the parasitic enzyme over human cathepsin D, since this subsite in the human orthologue tends to be even more hydrophilic. As discussed above, S2 subsites on both the parasitic and human enzymes are hydrophobic. However, this pocket seems to be narrower in the schistosome protease, which would favor HIV-1 protease inhibitors with shorter P2 groups.

Acknowledgements: This work has been partially supported by the CNPq, Faperj Fapemig and Fiocruz (Papes). F.P.S.Jr. was a PIBIC CNPq/Fiocruz fellowship recipient.

References

[1] World Health Organization (1993) WHO Technical Report 830.
[2] Doenhoff, M.J., Kimani, G. and Cioli, D. (2000) Parasitol. Today 16, 364–366.
[3] Berry, C., Humphreys, M.J., Matharu, P., Granger, R., Hor-

rocks, P., Moon, P.R., Certa, U., Ridley, R.G., Bur, D. and Kay, J. (1999) FEBS Lett. 447, 149–154.
[4] Moon, R.P., Tyas, L., Certa, U., Rupp, K., Bur, D., Jacquet, C., Matile, H., Loetscher, H., Grueninger-Leitch, F., Kay, J., Dunn, B.M., Berry, C. and Ridley, R.G. (1997) Eur. J. Biochem. 244, 552–560.
[5] Silva, A.M., Lee, A.Y., Gulnik, S.V., Majer, P., Collins, J., Bhat, T.N., Collins, P.J., Cachau, R.E., Luker, K.E., Gluzman, I.Y., Francis, S.E., Oksman, A., Goldberg, D.E. and Erickson, J.W. (1996) Proc. Natl. Acad. Sci. USA 93, 10034–10039.
[6] Francis, S.E., Gluzman, I.Y., Oksman, A., Banerjee, D. and Goldberg, D.E. (1996) Mol. Biochem. Parasitol. 83, 189–200.
[7] Harrop, S.A., Prociv, P. and Brindley, P.J. (1996) Biochem. Biophys. Res. Commun. 227, 294–302.
[8] Gallego, S.G., Slade, R.W. and Brindley, P.J. (1998) Acta Trop. 71, 17–26.
[9] Longbottom, D., Redmond, D.L., Russell, M., Liddell, S., Smith, W.D. and Knox, D.P. (1997) Mol. Biochem. Parasitol. 88, 63–72.
[10] Dalton, J.P., Smith, A.M., Clough, K.A. and Brindley, P.J. (1995) Parasitol. Today 11, 299–303.
[11] Wong, J.Y.M., Harrop, S.A., Day, S.R. and Brindley, P.J. (1997) Biochem. Biophys. Acta 1338, 156–160.
[12] Becker, M.M., Harrop, S.A., Dalton, J.P., Kalinna, B.H., McManus, D.P. and Brindley, P.J. (1995) J. Biol. Chem. 270, 24496–24501.
[13] Hewitt, E.W., Treuman, A., Morrice, N., Tatnell, P.J., Kay, J. and Watts, C. (1997) J. Immunol. 159, 4693–4699.
[14] Rochefort, H. and Liaudet-Coopman, E. (1999) APMIS 107, 86–95.
[15] Chevallier, N., Vizzavona, J., Marambaud, P., Baur, C.P., Spillantini, M., Fulcrand, P., Martinez, J., Goedert, M., Vincent, J.P. and Checler, F. (1997) Brain Res. 750, 11–19.
[16] Davies, D.R. (1990) Annu. Ver. Biophys. Chem. 19, 189–215.
[17] Chou, K.-C., Taomasselli, A.G. and Heinrikson, R.L. (2000) FEBS Lett. 470, 249–256.
[18] Sauder, J.M., Arthur, J.W. and Dunbrack Jr., R.L. (2000) J. Mol. Biol. 300, 241–248.
[19] Guruprasad, K., Törmäkangas, K., Kervinen, J. and Blundell, T.L. (1994) FEBS Lett. 352, 131–136.
[20] Azim, M.K. and Zaidi, Z.H. (1999) Biochem. Biophys. Res. Commun. 264, 825–832.
[21] Baldwin, E.T., Bhat, T.N., Gulnick, S., Hosur, M.V., Sowder, R.C., Cachau, R.E., Collins, J., Silva, A.M. and Erickson, J.W. (1993) Proc. Natl. Acad. Sci. USA 90, 6796–6800.
[22] Dhanaraj, V., Dealwis, C.G., Frazao, C., Badasso, M., Sibanda, B.L., Tickle, I.J., Cooper, J.B., Driessen, H.P.C., Newman, M., Aguilar, C., Wood, S.P., Blundell, T.L., Hobart, P.M., Geoghgan, K.F., Ammirati, M.J., Danley, D.E., O'Connor, B.A. and Hoover, D.J. (1992) Nature 357, 466–472.
[23] Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. and Kent, S.B.H. (1989) Science 245, 616–621.
[24] Brindley, P.J., Kalinna, B.H., Wong, J.Y.M., Bogitsh, B.J., King, L.T., Smyth, D.J., Verity, C.K., Abbenante, G., Brinkworth, R.I., Fairlie, D.P., Smythe, M.L., Milburn, P.J., Bielefeldt-Ohmann, H., Zheng, Y. and McManus, D.P. (2001) Mol. Biochem. Parasitol. 112, 103–112.
[25] Brinkworth, R.I., Prociv, P., Loukas, A. and Brindley, P.J. (2001) J. Biol. Chem. 276, 38844–38851.
[26] Fujinaga, M., Cherney, M.M., Tarasova, N.I., Bartlett, P.A., Hanson, J.E. and James, M.N.G. (2000) Acta Crystallogr. Sect. D 56, 272–280.
[27] Rawlings, N.D. and Barrett, A.J. (2000) Nucleic Acids Res. 28, 323–325.
[28] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[29] Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001) Bioinformatics (submitted).
[30] Guex, N. and Peitsch, M.C. (1997) Electrophoresis 18, 2714–2723.
[31] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Crystallogr. 26, 283–291.
[32] Van Gunsteren, W.F. and Mark, A.E. (1992) Eur. J. Biochem. 204, 947–961.
[33] McDonald, I.K. and Thornton, J.M. (1994) J. Mol. Biol. 238, 777–793.
[34] Schechter, I. and Berger, A. (1967) Biochem. Byophys. Res. Commun. 27, 157–162.
[35] Tang, J., James, M.N.G., Hsu, I.N., Jenkins, J.A. and Blundell, T.L. (1978) Nature 271, 618–621.
[36] Fruitier, I., Garreau, I. and Piot, J.M. (1998) Biochem. Biophys. Res. Commun. 246, 719–724.
[37] Gulnick, S.V., Suvorov, L.I., Majer, P., Collins, J., Kane, B.P., Johnson, D.G. and Erickson, J.W. (1997) FEBS Lett. 413, 379–384.
[38] Scarborough, P.E. and Dunn, B.M. (1994) Protein Eng. 7, 495–502.
[39] Pimenta, D.C., Oliveira, A., Juliano, M.A. and Juliano, L. (2001) Biochim. Biophys. Acta 1544, 113–122.
[40] Beyer, B.M. and Dunn, B.M. (1997) Protein Sci. 7, 88–95.
[41] Darke, P.L., Nutt, R.F., Brady, S.F., Garsky, V.M., Ciccarone, T.M., Leu, C.T., Lumma, P.K., Freidinger, R.M., Veber, D.F. and Sigal, I.S. (1988) Biochem. Biophys. Res. Commun. 156, 297–303.
[42] Tomasselli, A.G. and Heinrikson, R.L. (2000) Biochim. Biophys. Acta 1477, 189–214.
[43] Kay, J. and Dunn, B.M. (1992) Scand. J. Clin. Lab. Ivest. 52 (Suppl. 210), 23–30.
[44] Vacca, J.P. (1994) Methods Enzymol. 241, 311–334.