Protein Engineering vol.13 no.3 pp.167–177, 2000

Analysis of structural and physico-chemical parameters involved in the specificity of binding between α-amylases and their inhibitors

M.C.M.Da Silva[1], M.F.Grossi de Sá[1], M.J.Chrispeels[3], R.C.Togawa[2] and G.Neshich[2,4]

[1]Laboratório de Biologia Molecular and [2]Laboratório de Bioinformática, EMBRAPA, Recursos Genéticos e Biotecnologia, CP 02372, CEP 70770 Brasília, DF, Brazil and [3]Department of Biology, University of California San Diego, La Jolla, CA 92093-0116, USA

[4]To whom correspondence should be addressed
Email: neshich@cenargen.embrapa.br

Enzyme–inhibitor specificity was studied for α-amylases and their inhibitors. We purified and cloned the cDNAs of two different α-amylase inhibitors from the common bean (*Phaseolus vulgaris*) and have recently cloned the cDNA of an α-amylase of the Mexican bean weevil (*Zabrotes subfasciatus*), which is inhibited by α-amylase inhibitor 2 but not by α-amylase inhibitor 1. The crystal structure of AI-1 complexed with pancreatic porcine α-amylase allowed us to model the structure of AI-2. The structure of *Zabrotes subfasciatus* α-amylase was modeled based on the crystal structure of *Tenebrio molitor* α-amylase. Pairwise AI-1 and AI-2 with PPA and ZSA complexes were modeled. For these complexes we first identified the interface forming residues. In addition, we identified the hydrogen bonds, ionic interactions and loss of hydrophobic surface area resulting from complex formation. The parameters we studied provide insight into the general scheme of binding, but fall short of explaining the specificity of the inhibition. We also introduce three new tools—software packages STING, HORNET and STINGPaint—which efficiently determine the interface forming residues and the ionic interaction data, the hydrogen bond net as well as aid in interpretation of multiple sequence alignment, respectively.
Keywords: amylase inhibitor/charge compatibility/hydrogen bond net/interface forming residues/protein interactions

Introduction

Most biological processes require interactions between proteins and much attention has been directed at understanding what makes protein associations possible and how protein complexes are stabilized. Stabilization is a function of specific interactions such as the formation of hydrogen bonds and salt bridges between amino acid side chains at the interface of two protein molecules in a complex (Xu *et al.*, 1997). In addition, the size of the hydrophobic area buried at the interface contributes significantly to complex stabilization (Chothia, 1974; Jones and Thornton, 1996). Lytic enzymes, such as proteases and amylases and their naturally occurring proteinaceous inhibitors, make excellent model systems to study protein–protein interactions. We are particularly interested in the interactions of insect α-amylases and the inhibitors that occur in their potential food sources, and want to understand more about how the amylases and the amylase inhibitors are co-evolving. Insects gain access to food sources when they evolve amylases that are not affected by the inhibitors present in the food source, and plants become resistant when they evolve inhibitors that inhibit these insect enzymes. Effective inhibitors form stable complexes with the amylase. We wish to understand these interactions at the molecular level.

α-Amylases (α-1,4-glucan-4-glucanohydrolases) are a family of enzymes that hydrolyze α-D-(1,4)-glucan linkages in amylose, amylopectin, glycogen and phytoglycogen and play an important role in the carbohydrate metabolism of many autotrophic and heterotrophic organisms. Heterotrophic organisms use α-amylases primarily to digest starch in their food sources. Insects such as the yellow meal worm (*Tenebrio molitor*) and the Mexican bean weevil (*Zabrotes subfasciatus*), which live in dried stored products (wheat flour and seeds of the common bean, respectively), have evolved mechanisms to overcome the effects of the proteinaceous inhibitors that are present in their food sources. *Tenebrio molitor* highly expresses a single amylase gene and in this manner overcomes the presence of inhibitors in wheat flour. Many insects have several α-amylases that differ in specificity, and successful utilization of a food source is dependent on the presence of an α-amylase for which there is no specific inhibitor. To gain further insight into this problem of enzyme–inhibitor specificity, we purified and cloned the cDNAs of two different α-amylase inhibitors from the common bean (*Phaseolus vulgaris*) (Grossi de Sá and Chrispeels, 1987, 1997; Moreno and Chrispeels, 1989), and recently cloned the cDNA of an α-amylase from the Mexican bean weevil (*Zabrotes subfasciatus*), which is inhibited by α-amylase inhibitor 2 (AI-2), but not by α-amylase inhibitor 1 (AI-1) (Grossi de Sá and Chrispeels, 1997). These two inhibitors, which share 78% amino acid identity, have the converse interaction with pancreatic porcine α-amylase (PPA): AI-1 inhibits PPA, whereas AI-2 does not (Grossi de Sá and Chrispeels, 1997).

The availability of the crystal structure of AI-1 complexed with PPA (Bompard-Gilles *et al.*, 1996) allowed the modeling of the structure of AI-2. The structure of *Zabrotes subfasciatus* α-amylase (ZSA) was modeled based on the crystal structure of *Tenebrio molitor* α-amylase (TMA) (Strobl *et al.*, 1998). Pairwise AI-1 and AI-2 with PPA and ZSA complexes were also modeled. Modeled complexes were analyzed both with and without energy minimization. For this analysis, we first identified the interface forming amino acid residues (IFR) of the four possible complexes. In addition, we identified the hydrogen bonds, ionic interactions and loss of hydrophobic surface area resulting from complex formation.

Although the data we show here fall short of explaining the specificity of the inhibition, they do provide insight into the general scheme of binding. Three new web tools, STING, HORNET and STINGPaint, efficiently determine the IFR and ionic interaction data, the hydrogen bond net, as well as aid in interpretation of the multiple sequence alignment (MSA), respectively.

Materials and methods

Multiple amino acids sequence alignments were carried out using the ClustalW algorithm (Thompson *et al.*, 1994) and

Fig. 1. STINGPaint 2 presentation of CLUSTAL-W multiple sequence alignment for TMA and ZSA. The '#' symbol indicates IFRs (obtained from calculations done on the ZSA–AI-2 complex). Residue color coding is as follows: gray, AVLIMFP; green, STYNQWG; red, DE; dark blue, RK; cyan, H; yellow, C.

WWW Service at the European Bioinformatics Institute, (http://www2.ebi.ac.uk/clustalw; Rodrigo Lopez) using both the Blosum and Pam matrices. No further manual modification was carried out with the alignments.

The program STINGPaint2 (manuscript in preparation), available at http://asparagin.cenargen.embrapa.br/STING/STINGpaint2/ and also at http//:trantor.bioc.columbia.edu/arthrosoft, was used for ease of presentation by means of adequate coloring of sequence identities, hydrophobic patches, charge distribution on critical loops around the active site, as well as for annotation of conserved and calculated structural features on the aligned sequences.

Molecular modeling of AI-2 and ZSA was carried out on a Silicon Graphics ONYX R4400 VTX workstation using the program MODELLER, release 4 (Sali *et al*., 1993). The atomic coordinates of AI-1 (Brookhaven Protein Data code 1DHK.pdb; Bompard-Gilles *et al*., 1996) were used to build the three-dimensional model of AI-2. The C-terminal end of AI-2 (residues 191–213 in the processed sequence) was not modeled because there were no corresponding atomic coordinates in the template structure used for this region. The model of ZSA was built using the atomic coordinates of TMA available from the Protein Data Base (code 1JAE.pdb; Strobl *et al*., 1998). For the building of the inhibitor–enzyme complex, we assumed the same orientation of the inhibitor and the enzyme as in 1DHK.pdb. Adequate superimpositions and substitutions, accompanied by energy minimization, resulted in modeled complexes. The energy minimization procedure used the steepest descent and conjugate gradient, set to run for up to 18 000 steps, or until the maximum derivative of the energy with respect to the atomic positions was less than 0.0002 kcal/mol/Å. To ensure that atom movement is limited for a region far from the interface layer, we previously set fixed coordinates for all atoms in the subset that is far from the interface. With this procedure we gained in CPU time and also in accuracy of modeled complexes. To define subsets of atoms belonging to the interface, we used our program STING (Neshich *et al*., 1998), available at http://asparagin.cenargen.embrapa.br/STING/ and also at http://honiglab.cpmc.columbia.edu/STING and http://www.pdb.bnl.gov/pdb-bin/pdbids. In STING we can define the IFR layer based on a distance set to 8.0 Å and measured between any two atoms belonging to residues in

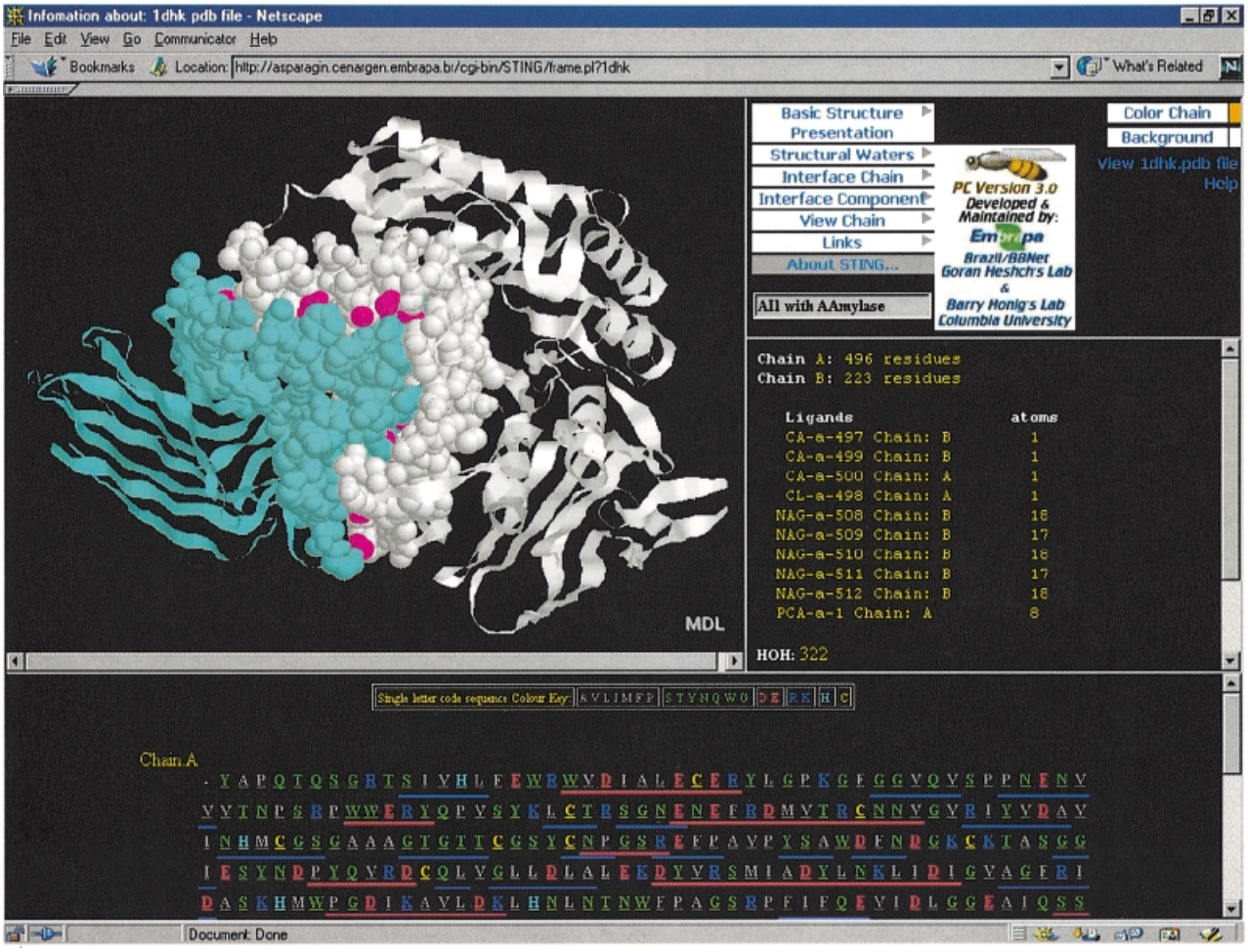

Fig. 2. STING graphical presentation of complex 1DHK.pdb. On the left side, the three-dimensional structure of AI-1 (blue color) complexed with the PPA (white color, right side) is shown. The blue and white CPK presents components of IFRs of both chains and red CPK indicates water molecules at the interface. The upper right window is a pull-down menu showing some of STING's commands. The right middle window contains information about the two chains taken from the PDB [Brookhaven Protein Database] file. The lower window shows the linear sequence of the molecule, color coded according to hydrophobicity and charged groups. Blue and red lines below the sequence present helix and sheet regions.

two different protein chains. Once the IFR ensemble is defined, we can easily use this definition as the subset definition in Discover-3 for fixing all but the IFR atoms. IFR atoms were allowed to adjust freely when docking was performed. In another procedure, all atoms were fixed in the vicinity of their original positions. In this procedure, we did not permit adjustments of IFR atoms to the environment encountered in the active site, except for the simple removal of steric hindrance.

The same procedure was adopted to build the model of three complexes: AI-1–ZSA, AI-2–PPA and AI-2–ZSA. The quality of the models was checked with PROCHECK (Laskowski *et al*., 1993).

Hydrogen bonds were determined using our own package HORNET (available at http://asparagin.cenargen.embrapa.br/HORNET/ and also at http://trantor.bioc.columbia.edu/arthrosoft/HORNET) which specifically shows the hydrogen bond net formed between two protein chains (or protein and DNA). HORNET also shows exact listings of IFR, based on calculated buried surface area. Both graphic and textual descriptions of all inter-chain hydrogen bonds were obtained and analyzed further. We preserved interface buried water molecules. To define subsets of water molecules at the interface, we used STING's feature 'Interface HOH'. The coordinates of 'Interface waters' were then inserted into modeled complexes and relaxed to appropriate positions during Discover minimization.

Charge pairing was analysed using features present in STING that allow for convenient graphic presentation of all charges at the interface of complexes. Using a sequence of STING's commands: 'interface on', 'charges on interface' and adequate hiding of other than IFRs, as well as reading distances between charges, we were able to easily identify charge interactions at the interface.

Electrostatic potential at the interface was calculated and graphically presented using the program GRASP (Nicholls *et al*., 1991). Partial and full charges were used for point and complete electrostatic field descriptions at the protein surfaces.

Results

The sequence identity percentages among the α-amylases are as follows: TMA and ZSA, 61%; PPA and ZSA, 54% and between the α-amylase inhibitors, AI-1 and AI-2, 78%. Such high identity facilitates structural modeling: ZSA was modeled based on the atomic coordinates of TMA (1JAE.pdb) and AI-2 was modeled based on the structure of AI-1 from 1DHK.pdb. The signal peptide was removed from the sequence of AI-1. In addition, we did not model that part of the sequence which is not resolved in the original structure of the inhibitor

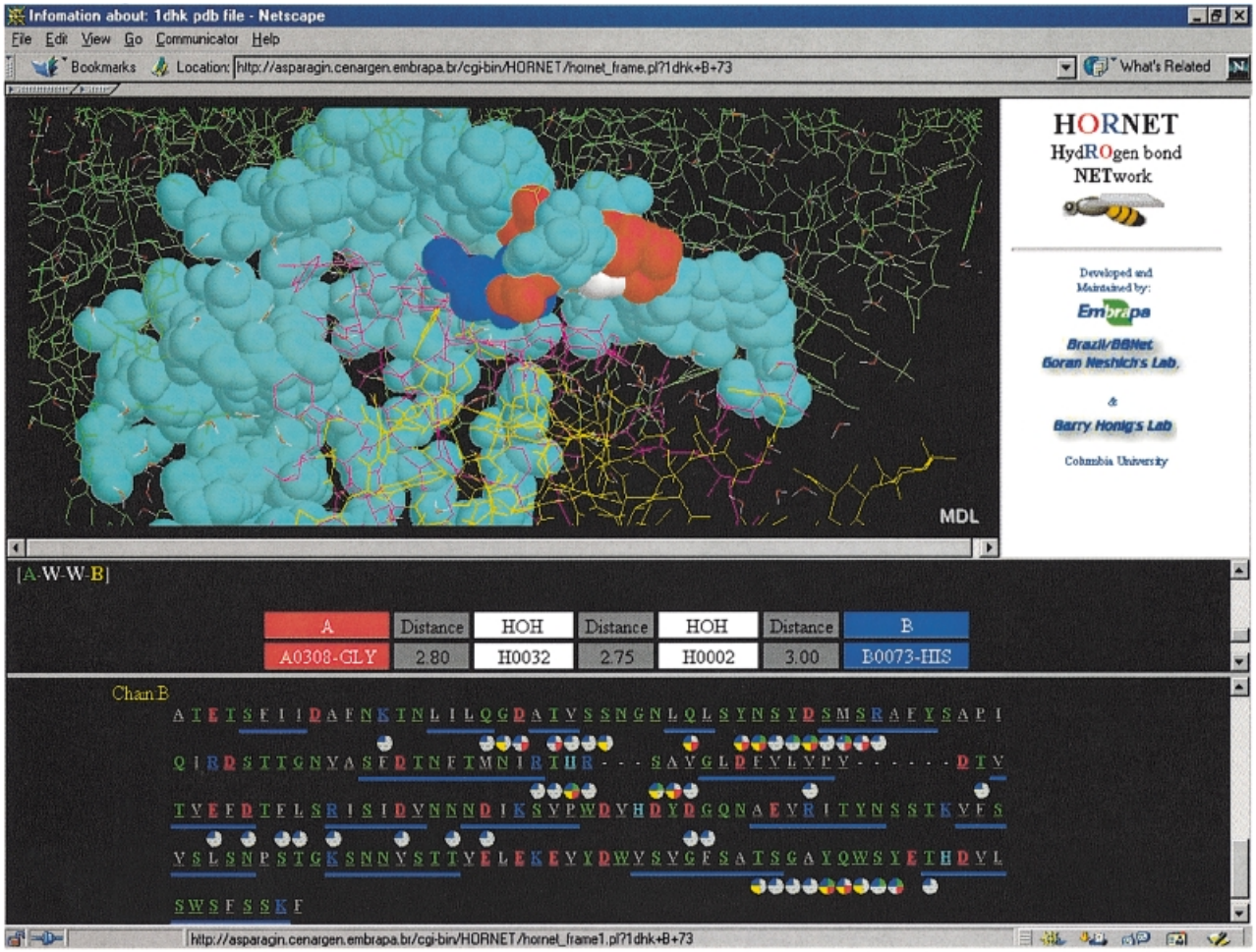

Fig. 3. HORNET graphical presentation; the upper window displays 1DHK.pdb (green wireframe, amylase; yellow wireframe, inhibitor; cyan CPK, IFRs of amylase; red wireframe, IFRs of inhibitor). Dark blue CPK represents residue His73 of the inhibitor chain, chosen for analysis from the sequence window (by clicking with the mouse on the colored four quadrants circled below that residue). Colored quadrants in this case indicate that His73 belongs to IFRs, and creates all three types of H-bond (A–B, A–W–B and A–W–W–B). Red CPK are H-bond paired residues that belong to the A chain (amylase). In the middle window, H-bond type and pairs, together with corresponding distances, are presented.

chain of 1DHK.pdb (QPESKG, 90–95 in AI-1). We also removed proteolytically processed residues (QAN, positions 75–77) in the AI-1 sequence, and the corresponding residues (EAN) in the AI-2 sequence. For modeling of AI-2, we also deleted a two residue insertion (S33 and Y34). At positions where processing and/or lack of coordinates was present in the AI-1 structure, we asked the MODELLER algorithm to introduce appropriate gaps [after residue 74 in AI-1 (corresponding with residue 72 in AI-2) and after residue 86 in AI-1 (corresponding with residue 84 in AI-2)]. The C-terminal end of AI-2 (position 193–215) was not modeled due to lack of template coordinates. The model parameters obtained were well within generously allowed regions of the Ramachandran plot, obtained by PROCHECK. The overall G-factors reported by PROCHECK for each complex were as follows: PPA–AI-2, –0.10; ZSA–AI-1, –0.30; ZSA–AI-2, –0.34. All of the cited G-factors were well above the lower limit (–1.0) considered as non-acceptable. R.m.s.d. values were obtained using CE software (Shindyalov and Bourne, 1998) and were as follows: PPA versus ZSA, 1.04 Å; TMA versus ZSA, 0.36 Å and for AI-2 versus AI-1, 0.46 Å.

The ZSA sequence alignment to TMA (Figure 1) shows the gaps introduced in the ZSA sequence at positions 211 (by removing G210 in TMA), 294 (by removing S294 in TMA), 337 (by removing Q337 and D338 in TMA) and finally at position 347 (by removing 349–356 in TMA). IFRs are indicated by the symbol '#' in Figure 1. The gaps that were introduced are all outside the region where loops form the interface between the proteins. Except for S294 in ZSA, these gaps are not crucially important for binding of the inhibitor at the active site. The structure obtained was checked by the PROCHECK program with extensive parameter analysis, and clearly showed good agreement with expected and allowed values.

The amylase inhibitor complexes were formed in such a way as to allow all IFRs to move slightly to accommodate the best fit at the active site, assuming also that the enzyme chain and the inhibitor chain would have the same orientation in the new complexes as in the PPA–AI-1 [1DHK.pdb] complex. We observed no impeding space clashes either before or after the minimization process. Therefore, we deduced that space fit in all the complex models created did not account for the difference in binding specificity.

Water molecules were introduced at the interfaces of the complexes ZSA–AI1, ZSA–AI2 and PPA–AI2. The STING-L package (command: HOH interface) was used to obtain a list of those water molecules that are found positioned between the IFRs of the two chains in the complexes. The availability

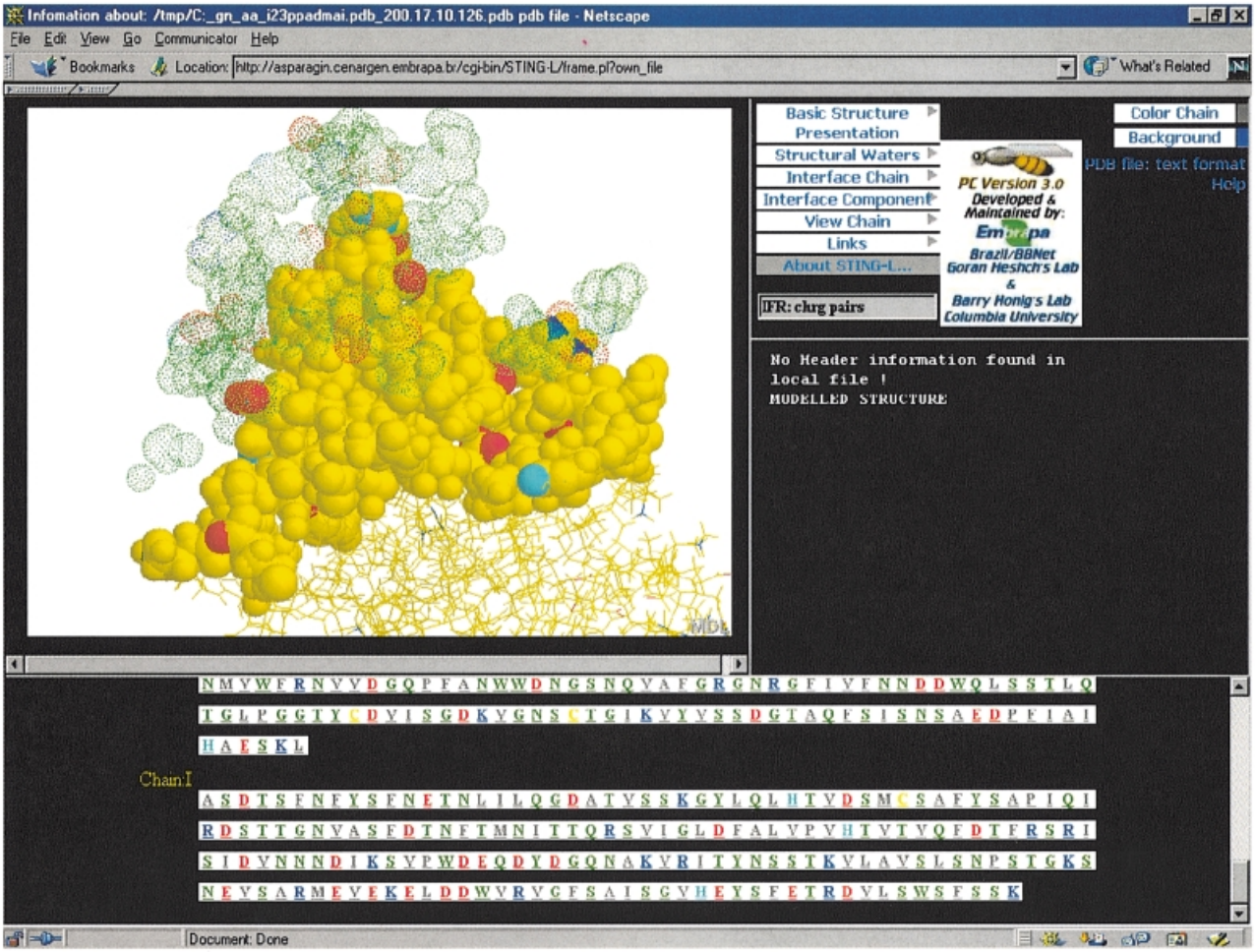

Fig. 4. STING-L graphic representation of charge complementarity. All frames are as in Figure 3. Three-dimensional structure of AI1 with ZSA. The inhibitor chain (wireframe) is shown in yellow and the yellow CPK's are IFRs of the same chain. Red CPK's are residues D and E of the inhibitor IFRs. Blue CPK's are residues R and K, while cyan CPK's are of H in inhibitor's IFR's. Same color code applies to amylase IFRs. Amylase IFRs are green doted spheres of van de Waal's radii. Distances between charged atoms are obtained by using the command 'set picking distance' in the interactive frame (upper right window).

of such a quick identification of water molecules in between IFR's facilitates the determination of water mediated H-bonds formed between two chains. Figure 2 shows an example of the STING-L graphical representation of two IFR areas (blue CPK for AI-1 and white CPK for PPA) with red colored water molecules between them. The water molecules visualized are those that satisfy the geometric condition of being at a 3.3 Å (maximum) distance from both chains. The output of STING-L was then used in INSIGHT to define the subset of water molecules that is inserted in all complexes. These water molecules at the interface were allowed to relax to appropriate positions during the minimization process for all docked complexes (AI-1–ZSA, AI-2–ZSA and AI-2–PPA). In such a way, we were able to get complete information about the hydrogen bond net formed between respective IFRs, including water mediated H-bonds.

The HORNET package produces a complete listing of the hydrogen bonds, including direct inter-chain H-bonds (A–B, where A and B are the protein chains of the α-amylase and α-amylase inhibitor, respectively), H-bond with one mediating water molecule (A–W–B) and H-bond with two mediating water molecules (A–W–W–B). The upper window of Figure 3 shows the three-dimensional constellation of a chosen H-bond donor atom (dark blue CPK: H73) and receptive acceptor (red CPK, A307, G308 and G304), where the water molecules involved are in white CPK(HOH). In the same window, the inhibitor chain (chain B) is presented with a yellow wire-frame, while the amylase chain (chain A) is presented with a green wire-frame. The list of H-bond donors, acceptors and mediating water molecules, as well as the distances between them, is given in the middle window (only part of the list is actually shown). The sequence is color coded with respect to hydrophobicity and is also marked with circles divided in quadrants that indicate participation of each residue in the IFR ensemble as well as in H-bond formation (button window). Structural water molecules at the interface contribute to structural stability of the complexes by forming an extensive hydrogen bond net (Raymer, 1997; Xu *et al*., 1997; Krem and Dicera, 1998). In Table I, we show an extensive listing of all hydrogen bonds of three different types (direct and with one and two mediating waters) encountered at the interface of all complexes. Analysis of the extensive list of H-bonds indicates that the total number found for PPA–AI-1 (52) versus the total number found for PPA–AI-2 (46) corroborates the experimental data, since only AI-1 inhibits PPA, but the hydrogen bonding scheme falls short of explaining the absence of binding in the case of ZSA–AI-1 (total number of H-bonds is 32) with respect to ZSA–AI-2 (total number of H-bonds is 31). Experimental

Table I. Hydrogen bonds net in the AI1–PPA and AI2–PPA, AI1–ZSA and AI2–ZSA complexes

PPA	H-bonds in AI1 R#algn(R#strct)	H-bonds in AI2	ZSA:R#Algn (R#structr)	H-bonds in AI2 R#algn(R#strct)	H-bonds in AI2 R#algn(R#strct)
Q63	Q178(187) 2(w) 2(ww)	E176 (EI)	W59		E176 (EI)
H101	Y177(186) (EI)		Q63	Q178(187) (w)	E176 (w)
I148	–	R95 (EI)	H101		
S145	–	R95 (EI)	D152(135)		
E149	**S77(78)** (EI)	**R97** (ww)	I155(138)	A76(79) (ww)	
		S99 (ww)			
S150		**S73** (ww)	N156(139)	E92(101) (EI)	**S73** (EI)
		S169 (ww)			N104 (EI)
Y151	**S78** (w)	V74 (w) (ww)	N157(140)	**T97** (EI)	
	A76(79) (w)	**D36** (EI) (w) (ww)		**D112** (EI)	
	D38 (EI)	**S172** (ww)			
	T173(182) (w)				
N152		N104 (EI)	Y158(141)	R42 (EI)	**S73** (ww)
		S73 (w)			**D36** (EI)
		S169 (ww)			**D78** (w)
Q156		**R97** (ww)	Q159(142)	R42 (EI)	
		S99 (ww)			
V163	W179(188) (w)		D160(143)		N104 (EI)
	A76 (79) (ww)				
D197	Y177(186) (EI)		R202(185)		
K200	**D38** 2(EI) (ww)	**D36** (EI)	D204(187)	Y177(186) (EI)	H175 (EI) (w)
H201	**D38** (EI)	**D36** (EI)			
E233	Y177(186) (ww)		K207(190)		**D36** (EI)
	Y37 (EI)				
I235	**D38** (w) (ww)		H208(191)	**D38** (EI)	**D36** (EI)
L237	**G20** (w)		E240(223)	Y37 (EI)	H175 (ww)
	M40 (ww)				
	S41 (ww)				
G238		**G20** (EI) (w)	I242(225)	Y37 (ww)	**D36** (w)
				S41 (ww)	
				D38 (ww)	
G239	S41 (ww)		Y244(227)	S41 (ww)	**G20** (w)
	M40 (ww)			**G20** (w) (ww)	**S37** (ww)
D240	S41 (ww)	**S37** (EI)	G245(228)	S41 (ww)	**D21** (ww)
	M40 (EI)				**S37** (ww)
K257			T246(229)	R42 2(EI)	**Q19** (EI)
				Q19 (EI)	S40 2(EI)
					R42 (EI)
K261	**D21** 2(ww)	**G20** (EI)	E247(230)	S41 (w)(ww)	D21 (ww)
		D21 (EI)			**S37** (EI) (ww)
T264	N35 (ww)	**D21** (w) (ww)	Q266(259)	K13 (w)	
		T23 (ww)			
		V24 2(w)			
S270	**S26** (w)	**E13** (EI) (w)	D305(288)	Y181(190)(EI)	**S178** (EI)
		V24 (ww)			
N298	Y37 (EI)		D309(292)	H73 (EI)(w)	H33 2(ww)
H299			R312(295)	N35 (EI) (w) (ww)	**D21** 2(EI)
				D21 2(EI)	
				T23 (w)	
D300	Y177(186) (w)	H175 (EI)			
	S200(189) (EI)	**S178** (EI)			
	Y181(190) (EI)				
N301	N35 (w) (ww)	**D21** (ww)			
		H33 (ww)			
G304	H73 (w)				
H305	S33 (ww)				
	Y34 (w) (ww)				
	Y181(190) (ww)				
A307	H73 (w)				
G308	H73 (ww)				
G309	**T23** (ww)				
	N35 (w) (ww)				
	Q31	(w) (ww)			
S311	N35 (EI)	**D21** (EI) (w) (ww)			
		H33 (w)			
N350		D118 (EI)			
E352	H73 (EI)	T69 (EI)			
D353		Q71 (EI) (w)			
		D179 (ww)			
TOTAL	EI(14) w(16) ww(22)	EI(18) w(11) ww(17)	TOTAL	EI(16) w(7) ww(9)	EI(17) w(5) ww(9)

Bold residues are conserved in both sequences.
Hydrogen bonds: EI, direct bond between enzyme and inhibitor.
W–HB extension over one water molecule; WW–HB extension over two water molecules.
R#algn indicates residue number in sequence alignment; R#strct indicates corresponding residue number in structure after model was created.

Table II. Charge complementary between AI1 and AI2 with PPA and ZSA in pairwise enzyme–inhibitor complexes

PPA:	AI1/PPA Distance[Å] R#algn (R#struct)		AI2/PPA Distance[Å]		ZSA: R#algn (R#struct)	AI1/ZSA R# Distance[Å]		AI2/ZSA R# Distance[Å]	
H101			H175	4.371 –	H101			H175	5.515 –
E149	**R99(108)**	**5.587 +**	**R95**	**4.260 +**	D152(135)			**R95**	**5.767 +**
	D108(117)	**3.864 –**	**R97**	**5.350 +**					
	D103(112)	**5.300 –**	**D106**	**5.010 –**					
	D87(101)	5.500 –							
D197			H175	3.919 +	R202(185)			H175	3.027 –
K200	**D38**	**5.190 +**	**D36**	**5.010 +**	D204(187)			H175	3.027 +
H201	**D38**	**2.780 +**	**D36**	**4.780 +**	K207(190)			**D36**	**5.032 +**
E233	**D38**	**5.488 –**	**D36**	**6.800 –**	H208(191)	**D38**	**4.700 +**	H175	6.383 –
			H175	5.200 +				**D36**	**2.914 +**
D240	**D38**	**6.900 –**			E247(230)			**D36**	**6.511 –**
K257			**D21**	**5.700 +**	D305(288)			H175	6.301 +
K261			**D21**	**2.660 +**	D309(292)	H73	2.847 +		
H299			H175	6.110 –	R312(295)	**D21**	**2.855 +**	**D21**	**2.694 +**
D300			H175	4.480 +					
E352	**D120(129)**	**4.300 –**	**D118**	**5.300 –**					
	H73	4.600 +							
D356	H73	5.600 +	R72	5.370 +					
TOTAL	(+5) + (–6) = (–1)		(+10) + (–5) = (+5)		Total	(+3)		(+6) + (–4) = (+2)	
Total revised	(+1)		(+1)		Total revised	(+2)		(+2)	

Bold residues are conserved in both inhibitors.
'+' indicates attractive force; '–' indicates repulsion.
Total revised indicates sum of attractive and repulsive forces with atoms that are less than 3 Å distant from each other.
R#algn indicates the residue number in the sequence alignment; R#struct indicates the corresponding residue number in the structure.

data show that only AI-2 inhibits ZSA (Grossi de Sá and Chrispeels, 1997).

Electrostatic interactions play a central role in protein–protein interactions (Mathew *et al*., 1985; Gilson *et al*., 1986). In this work, the modeled complexes were further analyzed with respect to charge complementarity at the interface. STING-L software was used for identification of charges at facing surfaces. STING-L is a version of the STING package that is capable of reading and processing not only PDB deposited files, but also modeled structures. Figure 4 demonstrates the graphical interpretation of a STING-L analysis for the complementarity of charges at facing protein surfaces from 1DHK.pdb. Table II shows for all four complexes, the charged atoms identified and their potential pairs from the facing protein chain. In addition, the distance between the charged atoms is also presented. According to Table II, the total number of attracting and repulsing charge interactions does not lead to a complete understanding of why PPA binds to AI-1 and not to AI-2, or why ZSA binds to AI-2 and not to AI-1. Although Table II shows all the data collected, the 'total revised' line indicates the total sum of attractive (+) and repulsive (–) interactions and is more relevant because it considers interactions for atoms which are less than 3 Å apart.

Further analysis of the interface area of the studied complexes was done by calculating the buried surface area for each member of the complex. GRASP software was used for both graphical and numerical presentations. In Table III, we show the surface areas buried by complex formation which are classified with respect to residue type. Figure 5 shows a GRASP-produced image of the facing surfaces (open book view of the complex interface) color coded with respect to hydrophobicity and charge. Analysis of the numerical values in Table III shows that the hydrophobic area buried by complex formation is larger (energetically more favorable) in complex PPA–AI-2 than in PPA–AI-1 (contrary to experimental results

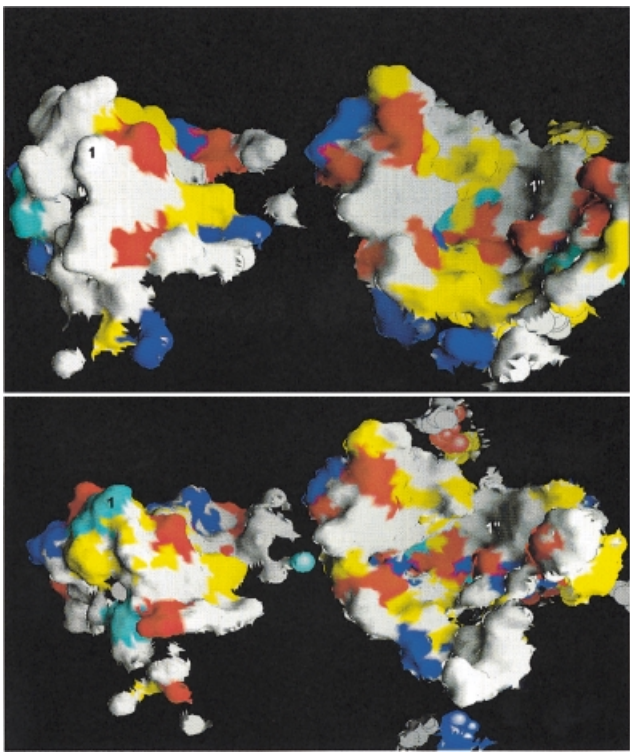

Fig. 5. GRASP-produced image of the two interfaces (open book view): left, the inhibitor; right, the amylase. The upper frame shows AI1–PPA (1DHK.pdb) and the lower frame shows AI2–ZSA. The residues forming the surface are color coded: yellow, hydrophobic residues; red, glutamic and aspartic acids; blue, arginine and lysine; cyan, histidine; white, glycines and polar residues. Numbers 1 [on inhibitor (left) side] and 1′ [on amylase (right) side] are overlapping with each other in the 'closed book' state.

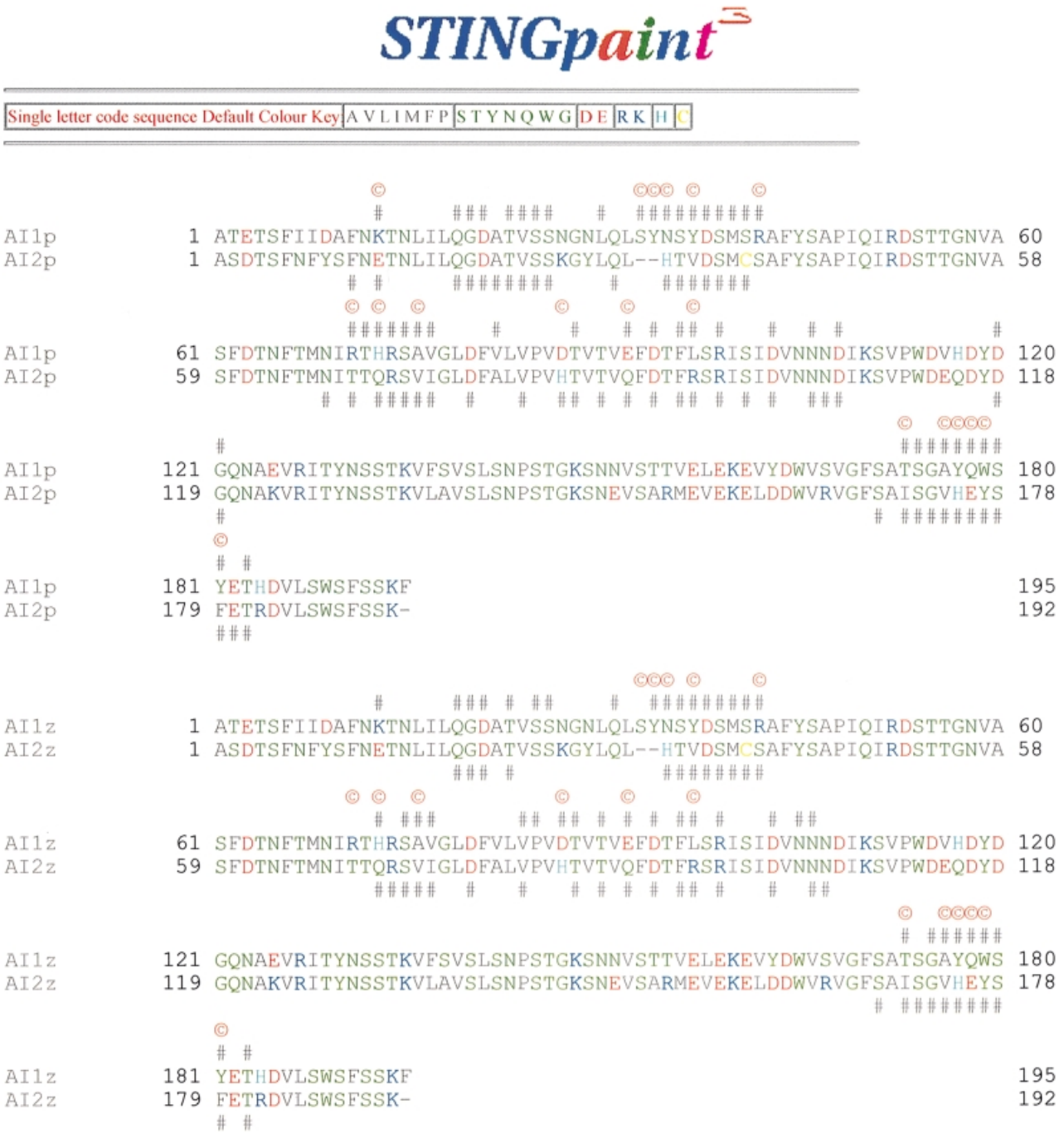

Fig. 6. STINGPaint 2 presentation of CLUSTAL-W MSA for AI-1 and AI-2. '#' symbol indicates IFRs. The upper alignment contains IFR symbols obtained with respective inhibitors in complex with PPA, while the lower alignment shows IFRs of inhibitors complexed with ZSA.'©' symbol marks residues defined by us as those that are likely to be responsible for the specificity differences observed between the two inhibitors.

for the inhibition), and is also larger in complex ZSA–AI-2 than in ZSA–AI-1 (in accordance with the experimental results).

The STINGPaint2 program allowed us to properly annotate the sequence alignment with the calculated and defined structural characteristics. This 'Web' analysis tool allows us to create a single document containing aligned sequences, color coded residues with respect to hydrophobicity and also with annotation that can easily point out important structural features. In Figures 1 and 6, the alignment of two α-amylases and two inhibitors, respectively, is shown together with structural annotations. Figure 6 shows alignment done by ClustalW and further processed with our own STINGPaint 2, where we introduced the symbol '©' for specific differences in sequence/structure, judged by us to be essential for the specificity of the two inhibitors complexed with PPA (see Discussion). We also annotated the participation of the residues in all sequences as being IFR (symbol '#'). We have indicated (symbol '©') the most likely residues to be substituted by site-direct mutagenesis to achieve a change in specificity. However, it is very important to realize that substitution of a single residue or isolated patches will not result in the creation of new specificity. We expect that only concomitant and extensive substitution of critical residues will have the effect of changing the specificity of a given inhibitor.

Table III. Buried surface area in four different inhibitor/amylase complexes

Residue type	E		I		% of Pho Area @ interface wrt total area of (E+I) in isolation	% of Pho Area @ interface wrt total Pho Area of E + I in isolation
	Area in isolation [Å^2 (%)]	Lost area at interface [Å^2 (%)]	Area in isolation [Å^2 (%)]	Lost area at interface [Å^2 (%)]		
PPA/I1						
Glycine	1508.745 (8.5)	208.873 (13.6)	200.783 (2.3)	65.525 (3.9)		
Hydrophobic	3596.28 (20.3)	409.028 (26.9)	1800.313 (20.2)	266.211 (15.8)	2.53	12.51
Charged	5367.337 (30.3)	432.77 (28.5)	2227.505 (25.0)	270.661 (16.0)		
Polar	7189.884 (40.6)	468.709 (30.8)	4692.914 (52.6)	1084.102 (64.3)		
Total	17723.18	1519.38	8921.519	1686.499		
PPA/I2						
Glycine	1547.848 (8.7)	164.21 (9.7)	223.138 (2.4)	90.852 (4.8)		
Hydrophobic	3615.727 (20.3)	461.232 (27.3)	1931.533 (20.8)	522.898 (27.8)	3.62	17.74
Charged	5325.024 (29.8)	426.245 (25.2)	2922.2 (31.4)	623.142 (33.1)		
Polar	7360.603 (41.2)	638.911 (37.8)	4217.819 (45.4)	647.169 (34.3)		
Total	17849.27	1690.598	9294.7	1884.061		
ZSA/I1						
Glycine	1603.864 (9.6)	60.751 (3.7)	192.788 (2.2)	57.554 (3.5)		
Hydrophobic	3045.366 (18.2)	413.218 (25.1)	1875.353 (21.3)	289.736 (17.8)	2.75	14.29
Charged	5102.085 (30.5)	306.18 (18.6)	2176.5 (24.7)	245.727 (15.1)		
Polar	6997.437 (41.8)	867.572 (52.7)	4569.02 (51.8)	1035.362 (63.6)		
Total	16748.8	1647.721	8813.656	1628.379		
ZSA/I2						
Glycine	1602.258 (9.7)	54.588 (3.6)	196.998 (2.2)	62.078 (3.9)		
Hydrophobic	3044.18 (18.5)	386.888 (25.7)	1896.893 (20.8)	554.747 (34.8)	3.68	19.06
Charged	5085.685 (30.8)	328.904 (21.9)	2816.873 (30.9)	417.929 (26.2)		
Polar	6757.147 (41.0)	732.296 (48.7)	4217.35 (46.2)	561.271 (35.2)		
Total	16489.28	1502.676	9128.109	1596.025		

Discussion

Our search to explain how differences in the specificity of α-amylase inhibitors toward α-amylases of different origins are constituted at the atomic level included a number of approaches. An analysis of steric hindrance that would appear upon docking in all complexes was considered (before and after the minimization that allows for IFR occupied volume to relax to its lowest energy conformation). Knowing that structural water can play a significant role in molecular association (Meyer *et al*., 1992; Israelachuli *et al*., 1996), we also considered the extent of the hydrogen bond net (including up to two mediating water molecules). In addition, we analyzed the extent of the total hydrophobic area buried upon complex formation, as well as the charge complementarity at the interface. Finally, we analyzed the electrostatic potential at the interfaces. None of these parameters alone is sufficient to explain the observed specificity in the binding of the two amylase inhibitors to the two amylases.

Our methodology has several specific points; with respect to steric hindrance, we know that, in general, hard body docking approaches have intrinsic limitations due to their lack of consideration of molecular flexibility, so we used a strategy with further minimization. On the other hand, approaches that include the entire molecular structure during minimization can be extremely CPU intensive (Cherfils *et al*., 1991; Shoichet and Kuntz, 1991; Bacon and Moult, 1992; Cherfils and Janin, 1993; Norel *et al*., 1995) as well as producing unexpected changes in parts of the molecule far from the interface. Our methodology defines IFRs and creates a corresponding subset to be the only region allowed to change during minimization. Using our methodology the model complexes contained no steric hindrance of such magnitude that could not be removed by the energy minimization procedure. In other words, in all models built, the inhibitor could penetrate without impeding steric hindrance into the active site. In addition, as reported by Jackson and Stenberg (1995), limiting minimization to only IFRs could avoid computationally expensive minimization (by avoiding 'multiple' problems) as well as a lack of consideration of solvation energy (due to *in vacuo* simulations). As reported by the same authors, surface area burial, as well as electrostatic desolvation, depends weakly on the precise conformation of the docked structure, indicating that the precision in the docking step is not critical as far as these two parameters are concerned.

Specificity of the inhibitor isoforms towards different α-amylases may depend on the strength of the network of hydrogen bonds that forms and stabilizes the inhibitor–enzyme complex. However, this study shows that the hydrogen bond net alone is not a sufficiently clear indicator to explain the known inhibitory specificity of the two amylases.

Horton and Lewis (1992) showed a strong correlation between buried surface area and measured binding strength of many protein complexes. However, in their review, Cherfils *et al*. (1991) noted that when compared with the rest of an enzyme's surfaces, regions involved in interaction with other molecules are neither more hydrophobic nor enriched in groups bearing an electric charge. Our own analysis shows that hydrophobic areas of the enzyme and inhibitor that are buried upon complex formation, with respect to total surface area of two proteins in isolation, are actually larger in PPA–AI-2 than in PPA–AI-1. One would expect that the contribution to the binding energy coming from the thermodynamically favorable process of burying of hydrophobic areas would be more extensive in the case of PPA–AI-1. However, this is not the case. Nevertheless, such a positive contribution to the binding energy, originating from the burial of the hydrophobic area, is observed in the case of ZSA–AI-2 (in accordance with the observed data for inhibition of ZSA by AI-2). The same trend is observed if values are calculated with respect to the total hydrophobic area of the two proteins in isolation rather that

the total surface area of the two proteins in isolation (last column in Table III). It is interesting though, that some other studies have shown that the binding site is not necessarily the most hydrophobic patch among all fitted geometries (Xu *et al*., 1997); this gives an additional degree of complexity to an analysis based on hydrophobicity parameters in binding specificity.

The binding energy, or more generally energetics, of protein–protein association has been reported to be critically dependent on the hydrophobic effect (Chothia and Janin, 1975). Electrostatic interactions, on the other hand, are more involved in specificity. Binding energy results from maximization of surface burial while minimizing penalties arising from desolvation. Although many authors (Cherfils and Janin, 1993; Connolly, 1986) have reported the use of buried surface area and buried hydrophobic surface area as a scoring function for docking, this alone, as confirmed by our results, is not sufficient to determine specificity.

Charge complementary is essential for binding and specificity (Janin, 1995). Proteins utilize residue identity and configuration at the surface as control of the recognition of ligands by electrostatics. The high selectivity found in some complexes may originate, to a large extent, from specific electrostatic interactions governed by polar groups (Warshel, 1979, 1986; Xu *et al*., 1995). Xu *et al*. (1997) has shown that electrostatics can provide not only for specificity but also contributes to binding affinity. In our case, charge compatibility of IFR in the complexes studied indicates that total net charges satisfied with complementary charges, originating from the facing surface, are identical in both complexes: PPA–AI-1 and PPA–AI-2, as well as in case of ZSA–AI-1 and ZSA–AI-2. In the former case, however, the sum of the attractive and repulsive forces between facing IFRs is +1 (in favor of attractive forces), while in the latter case, the sum is +2 (in favor of attractive forces). Thus, this parameter alone would not be sufficient to explain the existing specificity of the inhibitors for the amylases.

The electrostatic potential calculated at the interface of complexes shows some variability around the active site (data not shown). This difference becomes more pronounced if its origin is presented through point charges that are at, or just below, the interactive surface of the inhibitor. Qualitative analysis of charge distribution clearly indicates the importance of site-specific substitutions at the IFR in α-amylase inhibitors, as well as in α-amylases, corroborating the structural basis of the specificity. To make a description of the charge distribution at the interface in a form more readily understandable, we used the sequence annotated by structural parameters, produced by STINGPaint2.

The sequence similarity of both inhibitors and α-amylases is high, so that the general fold and specific structural elements of the modeled (ZSA and AI-2) and known (PPA and AI-1) proteins are alike. Both inhibitors show the same two-chain structure as a product of proteolytic cleavage at position Asn77 and Asn75 in AI-1 and AI-2, respectively. A significant difference between AI-1 and AI-2 is the absence of Ser34 and Tyr35 in AI-2. We judged the lack of these residues to present a serious prejudice to the hydrogen bond net, as well as to the other parameters considered in this study.

After analysis of the hydrogen bonds, electrostatic potential, steric hindrance data and charge complementarity, substitutions in AI-2 (as indicated in Figure 6) are suggested. These are mainly due to a change in hydrophobic character on this specific IFR and/or charge difference between two sequences. Also, these residues, marked with a '©' symbol, make up an ensemble that we judge to be essential for the specificity, by a combination of factors mentioned above, as well as the fact that they are also members of the IFR ensemble.

Conclusions

Specificity of α-amylase inhibitors for α-amylases of different origin has been studied at the molecular level. To address this complex issue, we analyzed separately several structural parameters that participate as components of the total energy of binding. In some cases, we did find satisfactory agreement of these parameters with the known specificity of the protein–protein interaction. However, none of the parameters was able to explain binding across all complexes studied. Obviously, the process of encoding the specificity in an enzyme–inhibitor complex is multifactorial. Qualitative and weighted contributions of each of the elements examined here may enable us to explain the observed specificity. We were nevertheless able to identify single critical elements of the sequence/structure of the inhibitors that together will effectively produce the desired specificity. None of the single sites will be able to change the specificity, but will most likely influence affinity of the inhibitor to the amylase. Suggested IFR substitutions are now being tested experimentally, the result of which will measure the degree of accuracy of the analyzed parameters.

The new tools used in this study–package STING, STINGPaint and HORNET–have been shown to be both didactic tools as well as research tools. They are easy to use and require virtually no training time.

Acknowledgements

We are grateful to Dr Barry Honig (Columbia University) for reading this manuscript and for hospitality in his laboratory during the development of the STING and STINGPaint software. We would also like to thank Dr Daniel Ridgen for critical reading of the manuscript. This work was supported by NFS grant DBI 9001463, CNPq, FAP-DF and EMBRAPA.

References

Bacon,D.J. and Moult,J. (1992) *J. Mol. Biol.*, **225**, 849–858.
Bompard–Gilles,C., Rousseau,P., Rougé,P. and Payan,F. (1996) *Structure*, **4**, 1441–1452.
Cherfils,J., Duquerroy,S. and Janin,J. (1991) *Proteins Struct. Funct. Genet.*, **11**, 271–280.
Cherfils,J. and Janin,J. (1993) *Curr. Opin. Struct. Biol.*, **3**, 265–269.
Connolly,M.L. (1986) *Biopolymers*, **25**, 1229–1247.
Chothia,C. and Janin,J. (1975) *Nature*, **256**, 705–708.
Chothia,C. (1974) *Nature*, **248**, 338–339.
Gilson,M. and Honig,B. (1986) *Biopolymers*, **25**, 2097–2119.
Grossi de Sá,M.F. and Chrispeels,M.J. (1997) *Insect Bioch. Mol. Biol.*, **27**, 271–281.
Horton,N. and Lewis,M. (1992) *Protein Sci.*, **1**, 169–181.
Israelachvili,J. and Wennerstrom,H. (1996) *Nature*, **379**, 219–225.
Jackson,R. and Stenberg,M.J.E. (1995) *J. Mol. Biol.*, **250**, 258–275.
Janin,J. (1995) *Biochemie*, **77**, 497–505.
Jones,S. and Thornton,J. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
Krem,M.M. and Dicera,E. (1998) *Proteins Struct. Funct. Genet.*, **30**, 34–42.
Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.
Matthew,J.B. (1985) *Annu. Rev. Biophys. Chem.*, **14**, 387–401.
Meyer,E. (1992) *Protein Sci.*, **1:12**, 1543–1562.
Mirkov,T.E., Evans,S.V., Wahstrom,J., Gomes,L., Young,N.M. and Chrispeels,J.M. (1995) *Glycobiology*, **5**, 45–50.
Moreno,J. and Chrispeels,M.J. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7885–7889.
Neshich,G., Togawa,R.,Vilella,W. and Honig,B. (1998) *Brookhaven Protein Data Bank Q. Newsl.*, **85**, 6–7.
Nicholls,A., Sharp,K. and Honig,B. (1991) *Protein Struct. Funct. Genet.*, **11**, 281–286.

Norel,R., Lin,S.L., Wolfson,H.J. and Nussinov,R. (1995) *J. Mol. Biol.*, **252**, 263–273.
Raymer,M.L., Sanschagrin,P.C., Punch,W.F., Venkataraman,S., Goodman,E.D. and Kuhn,L.A. (1997) *J. Mol. Biol.*, **265**, 445–464.
Sali,A. and Blundell,T.L. (1993) *J. Mol. Biol.*, **234**, 779–815.
Schoichet,B.K. and Kuntz,I.D. (1991) *J. Mol. Biol.*, **221**, 327–346.
Shindyalov,I.N. and Bourne,P.E. (1998) *Protein Engng*, **11**, 739–747.
Strobl,S., Maskos,K., Betz,M., Wiegand,G., Huber,R., Gomis-Ruth,F.X. and Glockshuber,R (1998) *J. Mol. Biol.*, **278**, 617–628.
Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
Xu,D.,Tsai,C.J. and Nussinov,R. (1997) *Protein Engng*, **10**, 999**–**1012.
Xu,D., Sheves,M. and Schulten,K. (1995) *Biophysics J.*, **69**, 2745–2760.
Warshel,A. (1979) *Photochem. Photobiol.*, **30**, 285–290.
Warshel,A. (1986) *Methods Enzymol.*, **127**, 578–587.

Received May 30, 1999; revised October 26, 1999; accepted November 22, 1999