Structural Determinants of the Bifunctional Corn Hageman Factor Inhibitor: X-ray Crystal Structure at 1.95 Å Resolution[†,‡]

Craig A. Behnke,[§] Vivien C. Yee,[||] Isolde Le Trong,[⊥] Lars C. Pedersen,[#] Ronald E. Stenkamp,[⊥] Seung-Sup Kim,[▽] Gerald R. Reeck,[○] and David C. Teller[*]

The Biomolecular Structure and Design Program, Departments of Biochemistry, and Biological Structure and the Biomolecular Structure Center, University of Washington, Seattle, Washington 98195, and Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506

Received May 26, 1998; Revised Manuscript Received September 4, 1998

ABSTRACT: Corn Hageman factor inhibitor (CHFI) is a bifunctional 127 residue, 13.6 kDa protein isolated from corn seeds. It inhibits mammalian trypsin and Factor XIIa (Hageman Factor) of the contact pathway of coagulation as well as α-amylases from several insect species. Among the plasma proteinases, CHFI specifically inhibits Factor XIIa without affecting the activity of other coagulation proteinases. We have isolated CHFI from corn and determined the crystallographic structure at 1.95 Å resolution. Additionally, we have solved the structure of the recombinant protein produced in *Escherichia coli* at 2.2 Å resolution. The two proteins are essentially identical. The proteinase binding loop is in the canonical conformation for proteinase inhibitors. In an effort to understand α-amylase inhibition by members of the family of 25 cereal trypsin/α-amylase inhibitors, we have made three-dimensional models of several proteins in the family based on the CHFI coordinates and the coordinates determined for wheat α-amylase inhibitor 0.19 [Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T., and Morimoto, T. (1997) *Biochemistry 36*, 13503−13511]. From an analysis of the models and a structure-based sequence analysis, we propose a testable hypothesis for the regions of these proteins which bind α-amylase. In the course of the investigations, we have found that the cereal trypsin/α-amylase inhibitor family is evolutionarily related to the family of nonspecific lipid-transfer proteins of plants. This is a new addition to the group which now consists of the trypsin/α-amylase inhibitors, 2S seed storage albumins, and the lipid-transfer family. Apparently, the four-helix conformation has been a successful vehicle in plant evolution for providing protection from predators, food for the embryo, and lipid transfer.

A proteinaceous inhibitor of human blood coagulation Factor XIIa (a term we will use to denote any activated form of Factor XII or Hageman factor) has been purified from corn seeds (*1*, *2*) and extensively characterized (*3*, *4*). The inhibitor is a useful reagent since, among plasma proteinases, it appears to inhibit only Factor XIIa (*1*, *5*).

Factor XII is part of the "contact system" of plasma, along with prekallikrein, high-molecular-weight kininogen, and Factor XI (*6*). Contact with appropriate surfaces results in an initial production of Factor XIIa, which, acting as a serine proteinase, generates plasma kallikrein and ultimately activates the entire contact system. This system can trigger any of several defense mechanisms, blood coagulation, the complement system, fibrinolysis, or inflammation (*6*). Thus, Factor XIIa occupies a strategic position physiologically, and from this follows the importance of the corn inhibitor as a tool to control and study these systems. As suggested by Hojima et al. (*1*), we will call this protein CHFI,[1] for corn (activated) Hageman factor inhibitor (*1*).

CHFI is notable for its strict selectivity among plasma proteinases for Factor XIIa. In contrast to many inhibitors of Factor XIIa (*7*), CHFI does not inhibit plasma kallikrein

[†] This material is based upon work supported under a National Science Foundation Graduate Fellowship (CAB) and was supported in part by NIH Grant HL 50355. This is a publication 99-150-J from the Agricultural Experiment Station, Kansas State University. We also thank the Murdock Foundation for their support of facilities in the Biomolecular Structure Center.

[‡] The structures determined have been assigned PDB identifiers 1bea and 1bfa.

* Corresponding author. Department of Biochemistry, Box 357350, University of Washington, Seattle, WA 98195.

[§] Biomolecular Structure and Design Program.

[||] Current address: Department of Molecular Cardiology/FF10, The Lerner Research Institute, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland OH 44195.

[⊥] Department of Biological Structure.

[#] National Institute of Environmental Health Sciences, 111 TW Alexander Dr., Research Triangle Park, NC 27709.

[▽] Structural Biology Program, Skirball Institute of Biomolecular Medicine, New York University Medical Center, 540 First Avenue, New York, NY 10016.

[○] Department of Biochemistry, Kansas State University, Manhattan, KS 66506.

[1] Abbreviations: CHFI, corn Hageman factor Inhibitor. 7n-CHFI, recombinant CHFI expressed in *Escherichia coli*; PDB, Protein Data Bank; NMR, nuclear magnetic resonance; ns-LTP, nonspecific lipid transfer protein; PEG, poly(ethylene glycol); MIR, multiple isomorphous replacement; NOE, nuclear Overhauser effect; EPMR, evolutionary programming for molecular replacement; MIRAS, multiple isomorphous replacement with anomalous scattering. PIR, protein identification resource; additional protein name abbreviations and references are presented in Table 1.

10.1021/bi9812266 CCC: $15.00
Published on Web 10/16/1998

Table 1: Protein Name Abbreviations and References

no.	sequence	PDB code	abbrev	ref[a]	description
1	1mzm_maize	1MZM	M-NSLTP	P19656	*Zea mays* (Maize) nonspecific lipid-transfer protein
2	itrf_maize	1BEA	CHFI	P01088	*Zea mays* trypsin/factor XIIa inhibitor
3	iaat_eleco	1BIP	RBI	P01087	*Eleusine coracana* (Millet) bifunctional inhibitor
4	iaae_horvu		BTI	P01086	*Hordeum vulgare* (Barley) trypsin inhibitor Cme
5	itr_rye			S29002	*Secale cereale* (Rye) trypsin inhibitor
6	iaa_horvu			P16969	*Hordeum vulgare* α-amylase/trypsin inhibitor pUP23
7	si5_sorbi			S28202	*Sorghum bicolor* (Sorghum) α-amylase inhibitor SIα5
8	si4_sorbi		SIA4	S28201	*Sorghum bicolor* α-amylase inhibiotr SIα4
9	ia01_wheat			P16850	*Triticum æstivum* (Wheat) α-amylase/trypsin inhibitor CM1
10	ia02_wheat			P16851	*Triticum æstivum* α-amylase/trypsin inhibitor CM2
11	iaaa_horvu		BCMA	P28041	*Hordeum vulgare* α-amylase/trypsin inhibitor Cma
12	ia03_wheat			P17314	*Triticum æstivum* α-amylase/trypsin inhibitor CM3
13	iaad_horvu			P11643	*Hordeum vulgare* α-amylase/trypsin inhibitor CMd
14	ia16_wheat			P16159	*Triticum æstivum* α-amylase/trypsin inhibitor CM16
15	iaab_horvu			P32936	*Hordeum vulgare* α-amylase/trypsin inhibitor CMb
16	pup13_horvu			*57*	*Hordeum vulgare* α-amylase/trypsin inhibitor pUP13
17	iaa1_wheat	1HSS	WAI-19	P01085	*Triticum æstivum* α-amylase inhibitor 0.19
18	iaa5_wheat		WAI-53	P01084	*Triticum æstivum* α-amylase inhibitor 0.53
19	iaa2_wheat		WAI-28	P01083	*Triticum æstivum* α-amylase inhibitor 0.28 (CIII, WMAI-1)
20	wrp27_wheat		WRP27	*56*	*Triticum æstivum* inhibitor of rice weevil α-amylase
21	iaai_horvu			P16969	*Hordeum vulgare* α-amylase inhibitor BDAI-1
22	iaa1_horvu			P16968	*Hordeum vulgare* α-amylase inhibitor BMAI-1
23	ra05_orysa			Q01881	*Oryza sativa* (rice) seed allergenic protein RA5
24	ra17_orysa		RA17	Q01883	*Oryza sativa* seed allergenic protein RA17
25	rag2_orysa			Q01885	*Oryza sativa* seed allergenic protein RAG2
26	ra14_orysa			Q01882	*Oryza sativa* seed allergenic protein RA14
		1HYP		P24337	*Glycine max* (soybean) hydrophobic seed protein
		1PNB		P24565	*Brassica napas* (rape seed) 2S seed storage protein

[a] Entries beginning with Q or P are SwissProt accession numbers. Entries beginning with S are PIR accession numbers. Entries not in SwissProt or PIR are referenced.

(*1*, *4*). Indeed, on the basis of direct (*1*) and indirect (*5*) measurements, it does not appear to inhibit any plasma proteinase other than Factor XIIa. In the terminology introduced by Bode and Huber (*8*), CHFI is a canonical inhibitor and it obeys the "standard mechanism of inhibition" as elaborated by Laskowski (*9*). In the standard mechanism, the proteinase both cleaves and ligates the reactive site scissile bond; the inhibitor at equilibrium exists in both intact and cleaved forms, and each is capable of binding to and thus inhibiting the target proteinase. The peptide bond in CHFI cleaved by Factor XIIa or trypsin (*10*) lies between Arg34 and Leu35 (CHFI sequence numbering) (*3*, *11*).

CHFI is a 127 residue, 13.6 kDa (*11*) member of the "cereal superfamily" (*12*), which includes numerous other proteins of similar length. Among them are proteinase inhibitors, α-amylase inhibitors [at least some of which are major food allergens (*13*)], and bifunctional proteins that inhibit both serine proteinases and α-amylases. CHFI falls in the last category: CHFI inhibits α-amylases from the yellow mealworm, *Tenebrio molitor*, and the red flour beetle, *Tribolium castaneum* (*14*) in addition to its inhibition of the serine proteinases Factor XIIa, mammalian trypsins (*2*), and trypsin from larvae of the yellow mealworm.[2]

Three-dimensional structures have been determined for two other members of this cereal superfamily. The first is that of the ragi (Indian finger millet) bifunctional inhibitor (ref *15*; RBI, See Table 1), obtained using NMR techniques and used in the current work as a molecular replacement model. The second structure in this family is that of the wheat α-amylase inhibitor 0.19 determined crystallographically at 2.06 Å resolution (ref *16*; WAI-19, Table 1). While RBI, like CHFI, is an inhibitor of both amylases and proteinases, WAI-19 inhibits α-amylases only. CHFI shares 67% sequence identity with RBI and 32% identity with WAI-19. Despite the close sequence similarity, there is a surprising degree of structural divergence among the three proteins.

In the search for structures similar to the protein fold of CHFI, we have found that the family of nonspecific lipid transfer proteins (ns-LTP) are homologues to the cereal trypsin/α-amylase inhibitor family. It is interesting to note that the ns-LTP from ragi (Indian finger millet) was once considered to be an α-amylase inhibitor and sequenced under that name but only later found to correspond to ns-LTP (*17*).

Identification of the proteinase inhibitory site in CHFI was straightforward, and the results found in the current work are in line with our expectation for the conformation of this site. In the absence of an atomic resolution structure of Factor XIIa, it is not possible to analyze the CHFI structure in terms of its specificity for this protein.

With the advent of the CHFI structure determination together with the structure of wheat amylase inhibitor 0.19 (*16*), we have attempted to locate the inhibitory site for α-amylases in this family of proteins. To accomplish this goal, we have relied on the extensive sequence information available for the family together with three-dimensional model building of selected family members. The current work has resulted in a structure-based sequence alignment of these proteins. From the models and the sequence alignment, we are able to propose a testable hypothesis concerning the location of the α-amylase inhibition site.

MATERIALS AND METHODS

(A) Protein. Wild-type CHFI was extracted from corn seeds and purified as previously described (*10*, *4*). CHFI

[2] G. R. Reeck and K. J. Kramer, unpublished results.

was also expressed in *Escherichia coli* using the pT7 expression vector and purified as earlier described (*18*); this protein is referred to as 7n-CHFI. The sequence to residue Cys 6 of this protein is −7 MARIPMASAGTSC 6. The Met −7 is removed during expression of the protein.

(B) Crystal Growth. Crystals were grown using conditions similar to those described in Pedersen et al. (*19*). Wild-type CHFI was dissolved in water to a concentration of 25 mg/mL and mixed with an equal volume of reservoir solution (30% PEG-400, 0.2M $MgCl_2$, and 0.1 M Hepes at pH 7.5). Crystals were grown in either a sitting or hanging drop format. The data collection crystal for the wt-CHFI was soaked in 0.1 M sodium acetate, 30% PEG-400, and 0.2 M $MgCl_2$, pH 5.8, prior to mounting. 7n-CHFI crystals were grown in a hanging drop format from a protein solution at 10 mg/mL mixed with 1% 2-propanol, 1% PEG-4000, and 0.1 M sodium citrate, pH 5.4. Crystals for heavy atom derivative searches were soaked using the native conditions or a solution with the 0.1 M Hepes replaced by 0.1 M Tris-maleate at pH 7.0. Both 7n-CHFI and wt-CHFI crystallize with $a = b = 57.12$ Å and $c = 80.24$ Å, space group $P4_22_12$. The ratio of molecular mass to unit cell volume (*20*) is 2.6 Da/Å^3, which converts to 51% solvent content (*19*).

(C) Heavy Atom Derivatization. After many unsuccessful attempts with heavy atom soaking experiments, we placed the crystals in a crystallization solution without $MgCl_2$ (30% PEG-400 and 0.1 M Tris-maleate, pH 7.0) for 24 h. This resulted in visible degradation of the crystals. The crystals were then soaked in 15 mM uranyl nitrate in 0.2 M NaCl or 0.2 M $CdCl_2$ for an additional 48 h.

(D) Diffraction Data Collection. All diffraction data were collected at room temperature with in-house Cu−Kα X-ray sources. Wild-type and 7n-CHFI native data sets were collected on a Rigaku R-AXIS IIC image plate detector equipped with a Rigaku RU200 rotating anode generator. Data from this detector were processed with the R-Axis software, PROCESS (*21*). Derivative data sets were collected on a Siemens multiwire area detector. Data were processed with SADIE and SAINT (*22*). Friedel mates were collected for one uranyl nitrate derivative. Data collection statistics are summarized in Table 2.

(E) Heavy Atom Position Determination and Phasing. Two software packages were used to identify heavy atom sites in the cadmium derivative, xhercules (*23*) and RSPS (*24*). Both packages indicated a single major heavy atom site which was confirmed by examination of the difference Patterson map. Difference Fourier maps indicated a second minor cadmium site. MLPHARE (*24*) was used to refine positions and occupancies of the two sites. Phases from the cadmium derivative were used to prepare cross-difference Fourier maps to determine the positions of uranyl ions in the uranyl nitrate derivatives. All sites were checked against difference Patterson maps calculated with xfft (*23*). All heavy atom positions were combined and refined using MLPHARE, and experimental MIRAS phases were determined. DM (*24*) was used for solvent flattening and histogram matching to improve the experimental map. The protein backbone in the MIR map was traced using xfit (*23*).

(F) Molecular Replacement. Due to the relatively poor phases from multiple isomorphous replacement, we used molecular replacement to complete the phasing of this protein.

Table 2

Data Collection

	maximum resolution (Å)	% completeness		R_{sym}[a]	no. of unique reflections
			last shell		
native					
wt-CHFI	1.95	92.1	81.8	N/A	9033
7n CHFI	2.20	77.8	57.5	0.08	5016
derivative					
cadmium	3.1	97.2	53.4	0.20	2323
uranium 1	3.2	96.8	94.1	0.13	2379
uranium 2	2.5	66.5	25.8	0.09	3282

MIRAS Results

	no. of sites	R_{iso}[b]	phasing power[c]	
			centric	acentric
cadmium	1	0.180	0.61	0.88
uranium 1	4	0.250	0.78	1.12
uranium 2	3	0.340	0.96	1.38

⟨FOM⟩[d] 15−2.5 Å

acentric	centric	all
0.380	0.640	0.430

[a] $R_{sym} = \sum|I_h - \langle I_h\rangle|/\sum I_h$. [b] $R_{iso} = \sum||F_{PH}| - |F_P||/\sum|F_{PH}|$. [c] Phasing power $= \sum|F_{Hcalc}|/\sum|(F_P + F_{Hcalc}) - F_{PH}|$. [d] ⟨FOM⟩ = mean of $\int\{P_{hkl}(\alpha)e^{i\alpha}\}d\alpha$ over all reflections.

A single best model of ragi bifunctional inhibitor, RBI, was calculated from the ensemble of 20 NMR structures deposited in the PDB by Strobl et al. (*15*). This model was produced by Fourier transforming the set of 20 structures (with occupancy set to 0.05) and using the calculated structure factors as "Fobs" in a refinement of a single model of RBI using X-PLOR (*25*). This single model with B factors was reduced to a poly(ala, ser, pro, cys) search model, and parts of the model not well-defined in the NMR structures were removed. Residues were deleted if the root-mean-square deviation of the Cα atoms of the ensemble was greater than 1.0 Å. Additionally, residues 67−78 were removed because the sequence alignment through this loop was uncertain.

The ensemble of 20 structures was also subjected to simulated annealing with constraints from the deposited NOE data and additional restraints imposed on the disulfide torsion angles to bring them into agreement with angles observed in high-resolution X-ray crystal structures. The resulting ensemble of 20 structures was subjected to the procedure above to produce another search model.

The evolutionary programming for molecular replacement (EPMR) (*26)* was used to place the search model in the unit cell. Only one solution was identified. This solution was confirmed by checking against the MIRAS-phased map.

(G) Model Refinement. X-PLOR (*25*) was used for rigid-body refinement of the molecular replacement solution and for subsequent simulated annealing and positional refinement (Table 3). Side chains throughout the molecule, which had been excluded from the search model, had well-defined density in a SIGMAA-weighted $2mF_o - DF_c$ map (*27*). Side chains were rebuilt, and the backbone was traced for residues omitted in the search model.

Following rebuilding and refinement with X-PLOR including individual *B* factor refinement, waters were built into the model. After iterative water building and refinement,

Table 3: Refinement Statistics

	wild-type 7n	recombinant
RMS deviations from ideality		
bonds (Å)	0.01	0.01
angles (deg)	1.75	1.54
dihedrals (deg)	24.34	21.32
no. of protein atoms	889[a]	879
no. of solvent atoms	96	14
Ramachandran Plot,		
no. of residues in regions		
core + allowed + generous	113	113
disallowed	1	1
mean B (RMS B)		
main-chain atoms	33.2 (13.2)	29.1 (12.1)
all protein atoms	37.1 (16.9)	31.9 (15.6)
solvent	59.3 (16.7)	51.8 (10.3)
$R^{b,d}$	0.19	0.20
$R_{free}{}^{c,d}$	0.29	0.27
resolution (Å)	45−1.95	45−2.2

[a] There are 10 fewer atoms in 7n-CHFI due to alternate conformations of Trp10 in wild-type CHFI. [b] $R = \sum||F_{Pobs}| - |F_{Pcalc}|/\sum|F_{Pobs}|$ calculated over the 90% of data used for refinement. [c] $R_{free} = \sum||F_{Pobs}| - |F_{Pcalc}|/\sum|F_{Pobs}|$ calculated over the 10% of data excluded from refinement. [d] Calculated including 2σ cutoff and bulk solvent correction.

96 waters were incorporated. REFMAC (*24*) was employed for positional refinement. Once the water structure was built, X-PLOR was used to calculate a bulk solvent correction and for final *B* factor and positional refinements.

The 7n-CHFI structure was determined by using the wt-CHFI structure without waters to directly phase the isomorphous crystals, followed by simulated annealing and positional and *B* factor refinement with X-PLOR. Following refinement, shaken omit maps were examined using xfit, and side chains were corrected. Only 16 waters were determined due to the lower resolution of the data.

(H) Characterization of CHFI. Electron density was missing for both the N- and C-terminal regions of the wt-CHFI and the 7n-CHFI. Mass spectra of a previous preparation of wt-CHFI yielded a major (60%) and two minor components (20% each). The major component corresponded to residues 1−123, one minor component corresponded to residues 1−122, and the third component fits best to residues 1−126. Edman sequencing with yet another preparation indicated an intact N-terminus on the protein. Hence, there is some possibility for ragged C-termini on the material isolated from corn and used for crystallization in this study.

Mass spectra were also taken on a prior preparation of 7n-CHFI than that used for crystallization and data collection. In this case, the determination yielded the expected molecular mass exactly for residues −6 to 127. Thus, we attribute the missing electron density of both wt-CHFI and 7n-CHFI to flexibility in the terminal regions.

(I) Structure-Based Sequence Alignments. Preliminary sequence alignments were performed by the methods documented in Yee et al. (*28*). To refine the sequence alignments, we used three-dimensional structural models of several proteins. In this way, problematic insertions and deletions could be evaluated from a structural standpoint and revised as necessary. In some cases, the preliminary sequence alignments were significantly modified as a result of the model building.

The basic idea used for model building in this study is to maintain those portions of the three-dimensional structures which are conserved between pairs of proteins and use the geometric restraints and energy minimization capabilities of X-PLOR (version 3.1) together with the visualization capabilities of O to arrive at plausible structures for the unknown portions.

To make the models, the program SOD (*29*) was used to make an O (*30*) macro file. This macro file, when used in O, maps the sequence of the protein being modeled onto the template structure. O was then used to manipulate the modeled residues into appropriate positions. The result of this operation is a model with a reasonable mapping of the sequence to the structure. If the starting sequence alignment is incorrect, it will be seen in a poor fit of the sequence to the template, and appropriate re-alignment can be made.

For two of the models reported here, the coordinates were written out, and X-PLOR (version 3.1) was used to energy minimize the structures with 400 steps of Powell minimization and 100 steps of Verlet dynamics (*25*). Models of BCMA and BTI were constructed in this fashion using CHFI as a coordinate template.

For the WAI-19 model, we wished to build the "flexible loop" between residues 68 and 78 for ease of sequence and structure comparisons and for use as a template for proteins of the same subfamily. This loop was built for convenience only; it is purely hypothetical since no electron density was observed in this region of the experimental structure (*16*). Because we also wished to have a model that accurately represented the known portion of the experimental model, we resorted to a procedure similar to that employed in constructing the molecular replacement model.

Dr. K. Fukuyama kindly made available a prerelease set of the coordinates (1HSS (*16*)]. All four independent molecules of 1HSS were superimposed, and the occupancy of each atom was set to 0.25. The set of four structures was Fourier transformed using SFALL (*24*) and the resulting structure factors subsequently treated as "Fobs". This was done by simply running the program SFTOOLS (*24*) and writing an X-PLOR formatted output file. A plausible loop was built for residues 68−78 using the lego_loop subprogram in O and chain A of the 1HSS coordinates. This constituted the starting model of WAI-19. The *B* factors of the model were set to 20 Å^2 and the model refined using X-PLOR. The refinement steps consisted of rigid body refinement, positional refinement, simulated annealing to 1000 °C, another positional refinement, and individual *B* factor refinement. These steps were carried out with weights as the ideal weights (*25*) except for the simulated annealing which used $^4/_5$ of the ideal weight. The refinement was carried out at a nominal resolution range of 40−2.5 Å. The final *R* factor was 0.10 ($R_{free} = 0.11$ for 10% of the data), and the *B* factors for the constructed loop became very large. However, for the 109 Cα atoms in common between the X-ray derived coordinates and the constructed model, the rms deviation was only 0.29 Å. The majority of this deviation comes from residues 68 and 78 which are adjacent to the modeled insert. These residues have *B* factors near 80 Å^2 in the 1HSS coordinates, so a large deviation (1 Å) of these two residues is not badly out of line.

The same set of Fobs structure factors from 1HSS was used to refine other models in this subfamily. The model of the wheat monomeric inhibitor 0.28 (WAI-28) yielded a final *R* factor of 0.20 (R_{free} was 0.26). The wheat monomeric

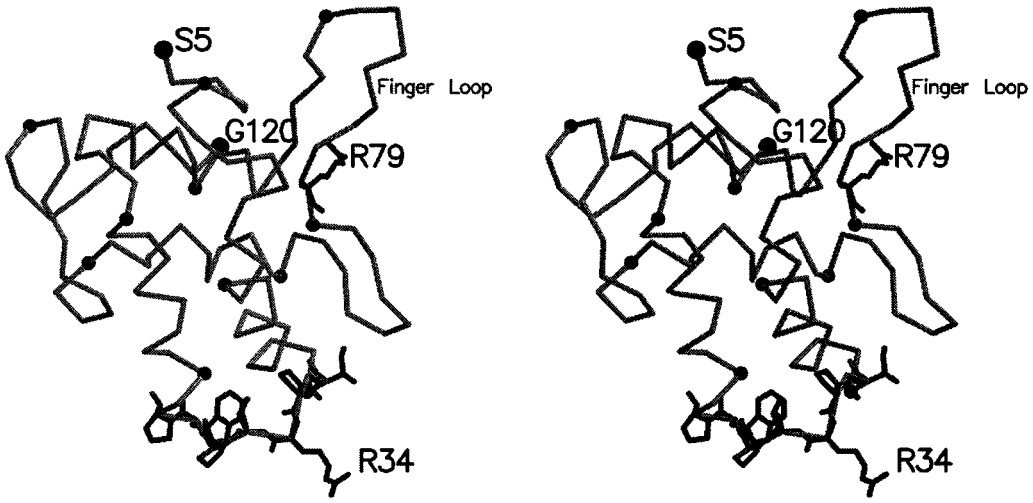


FIGURE 1: Schematic drawing of CHFI residues 5−120. Every 10th residue in the sequence is marked with a small black sphere. The residues of the protease inhibition loop are fully drawn (residues 31−38). Arg 34−Leu 35 contains the scissile bond. Helix A consists of residues 18−29, helix B consists of residues 37−49, helix C consists of residues 56−65, and helix D consists of residues 88−95. Residues 66−79 constitute the finger loop. Arg 79 is labeled and discussed in the text.

α-amylase inhibitor WRP27 gave $R = 0.24$ and $R_{free} = 0.29$. In this case, the resolution range for refinement was 40−3.0 Å. The model of rice allergen 17 (RA17) protein was also refined with these structure factors. The final *R*s were 0.25 and $R_{free} = 0.32$ for the refinement from 40 to 2.5 Å resolution. After refinement, energy minimization with the same protocol as above was used to relieve bad stereochemical contacts for these three proteins. This final energy refinement step did not significantly reduce the fit of the model to the relevant portions of the experimentally determined template.

For the sorghum α-amylase inhibitor protein 4 (SIA4), there are several regions in the sequence which are difficult to align with CHFI, and the Arg in the sequence, IYAVSRLT (CHFI positions 93−99), is particularly problematic. To find the correct alignment, a procedure similar to that used for the WAI-19 model was followed. In this case, CHFI was used as the template and coordinates of CHFI residues surrounding the insertions were deleted. The structure factors computed from this partial CHFI model were then used as Fobs with X-PLOR refinement to constrain most of the SIA4 molecule while allowing the chain to adjust to geometric factors in unknown regions. In the resolution range 40−2.5 Å, the model had $R = 0.24$ and $R_{free} = 0.30$. Once more, energy minimization followed the refinement to relieve bad contacts. In this example, the Arg in the above sequence was nicely accommodated in the modeled structure, was moved from its original alignment position, and forced a gap in the remainder of the proteins. We cite this as a justification for the procedure.

All of the final models had good stereochemistry as indicated by the Ramachandran plots (PROCHECK), and analysis with Prosa II (*31*) indicated that the models were satisfactory. The pairwise comparison of these structures was used to guide the sequence alignment of close relatives. Following modeling, all models were superimposed using locally written least-squares programs with a cutoff for equivalency of Cα atoms of 3.0 Å.

To make the hypothesis concerning the α-amylase inhibitory site on this family of proteins, the superimposed models were colored by residue category and examined with O in a variety of combinations to test conjectures concerning the site. This examination, together with a literature survey summary of inhibitory activity allowed us to postulate the residues involved in α-amylase inhibition.

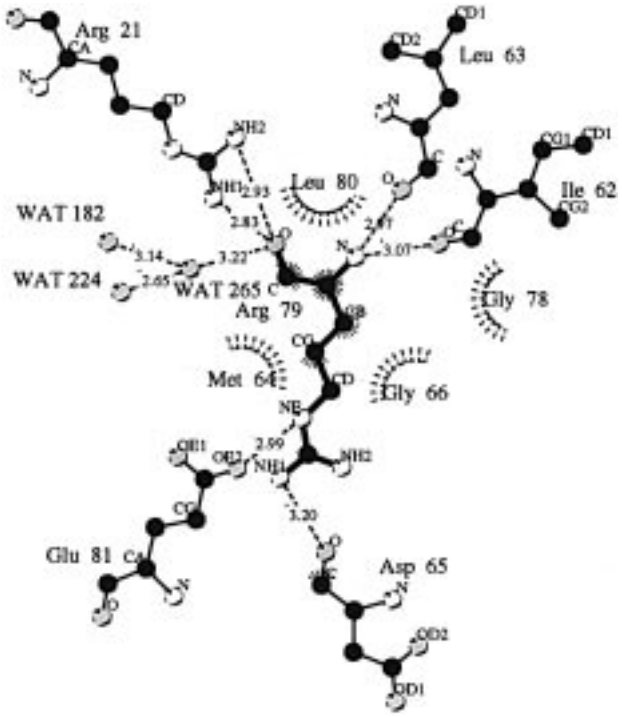


FIGURE 2: Schematic plot of the region surrounding Arg 79. This residue is deviant in the Ramachandran plot (ϕ, $\psi = 84°$, 132°). Hydrogen bonds and their lengths are given by the dashed lines. The rayed arcs denote nonpolar interactions with the residue.

The phylogeny of the sequence alignment was determined using PROTPARS from the PHYLIP package (*32*) to determine the topology of the dendrogram. Then ANCSTR (*33*) was used to construct the branch lengths as a matrix which was reduced using FITCH (*32*).

(J) Figures. Figures 1, 3, 5, and 6 utilized MOLSCRIPT (*34*) and Raster3d (*35*) in their construction. Figure 2 was drawn with LIGPLOT (*36*).

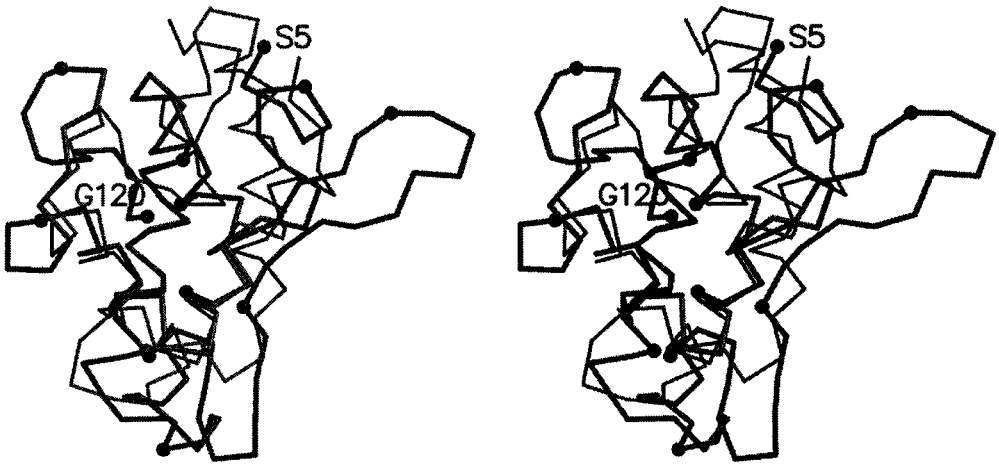


FIGURE 3: Superposition of nonspecific lipid transport protein of maize (M-NSLTP) and CHFI. Three of the four helices superimpose well, but helix-C of M-NSLTP is displaced relative to CHFI in order to form the lipid binding groove between helix-C and helix-D. Every 10th residue of CHFI is labeled as in Figure 1. Thick lines represent CHFI and thin lines mark the Cα atoms of M-NSLTP.

RESULTS

(A) Structure Determination. Difficulties were encountered in heavy atom derivatization of CHFI crystals, despite a reasonable solvent content (51%). We do not observe bound Mg^{2+} in the refined structure, so the difficulty was apparently due to tightly packed protein and large solvent channels, resulting in a small solvent-accessible surface of the molecules in the crystal. It appears that soaking in Mg^{2+}-free solvent was necessary to disrupt the crystal packing enough to make the heavy atom binding sites accessible. All heavy atom binding sites are on large cylindrical solvent channels. The heavy atom phasing was poor, with an MIR figure of merit of only 0.43. We could observe the helices, but the connectivities in the solvent-flattened MIR map were discontinuous and difficult to trace. The attainment of the molecular replacement solution simplified this problem.

Molecular replacement with NMR coordinates is often problematic; in our case, AMoRe (*37*) and X-PLOR were unable to identify the correct solution. Both a high-symmetry space group and a search model with >1.5 Å rms differences for the backbone contributed to these difficulties. Use of the deposited NMR constraints along with restraints on the disulfide bond dihedral angles in simulated annealing produced a search model that was closer to the final refined model than the averaged NMR coordinates. Finally, the six-dimensional search, implemented in an evolutionary algorithm to make the computation time reasonable, was essential for the successful molecular replacement (*26*).

(B) Molecule Description. CHFI is folded into four α-helices. The helices are arranged into two pairs of antiparallel helices, with an angle of approximately 60° between the two antiparallel pairs (Figure 1). These helices are named A−D. There is an extended loop, the "finger loop", between the C and D helices which is absent from the sequence or moved in WAI-19 (*16*) but present in RBI (*15*). The only β-structure in the protein occurs here between residue Ala 67 and Glu 77. This finger loop, extending from residue 66 to 79, forms the only direct protein contact along the *c* axis in our crystal form. Some residues in the finger loop, namely 73−75, were visible but had poor density, with refined *B* factors near 80 Å^2. Portions of the polypeptide not observed in the electron density are residues 1−4 and 121−127.

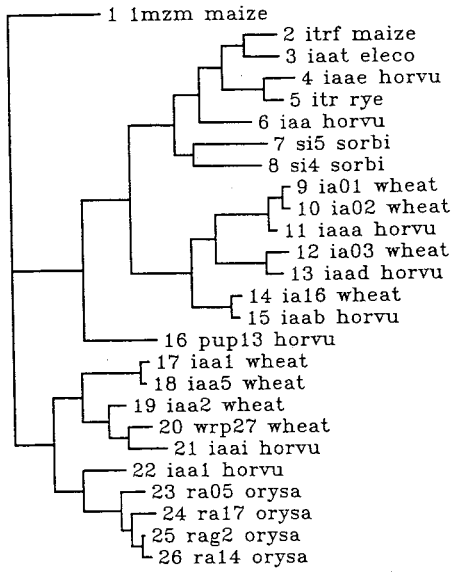
The known proteinase inhibitory site loop from residues Ile 31 to Pro 38 (P4 to P4′) is involved in a crystal-packing contact with the same loop from a symmetry-related molecule. Despite the crystal contacts, the loop backbone is in the canonical conformation for protein inhibitors of serine proteinases. As can be seen in Figure 1, Arg 34 at the scissile bond is fully exposed and projects away from the body of the protein.

In the Ramachandran plot, one residue, Arg 79, is deviant with $\phi,\psi = 84°, 132°$. Figure 2 shows the model in the region surrounding this residue. From the figure, it would appear that the disallowed conformation is generated by the hydrogen bonding. The residue, together with the main-chain atoms of Leu 80, is involved in a network of hydrogen bonds which extends over Arg 79, Arg 21, Glu 81, and Asp 65 as well as the main-chain carbonyl atoms of Ile 62 and Leu 63. In addition to these interactions, several water molecules participate in the network of hydrogen bonds surrounding Arg 21. This arrangement illustrates one of several extended hydrogen bonding networks within the protein and between the protein and major water clusters. Helix C in CHFI ends with Asp 65 as the last helical residue and Gly 66 as the first nonhelical residue. The Cα atom of Arg 79 is only 4.18 Å from Cα of Asp 65. Thus, it appears that Arg 79 "caps" helix C and assists in the formation of the finger loop rather than allowing an extension of helix C such as occurs in WAI-19 (*16*).

The 7n-CHFI structure was found to be essentially identical to that of the wild-type protein (rms deviation of 0.44 Å for all atoms, 0.22 Å for backbone atoms). In wt-CHFI, we find two alternate conformations of residue Trp 10 with relative occupancy of 0.75 and 0.25. In 7n-CHFI, only a single conformation was observed, presumably due to the lower resolution. R_{free} increased when alternate conformations were included in 7n-CHFI at Trp 10. Overall, this 7n-CHFI structure confirms the known biochemical similarity of isolated and expressed forms of CHFI and justifies use of engineered CHFI derivatives in functional and structural studies.

As expected, the CHFI structure is quite similar to the structures determined for RBI (*15*) and WAI-19 (*16*) in overall topology. There are differences in helix packing between CHFI and RBI, with differences of 1.61 and 1.74

A



B

```
__CHFI_No.: ----5...10 ........20 ........30 .----..--- ---------. ....40....
_2_itrf_ma: SAGTSCVPGW AIPHNPLPSC RWYVTSRTCG I----GP--- ---------R LPWPELKRRC
_3_iaat_el: SVGTSCIPGM AIPHNPLDSC RWYVSTRTCG V----GP--- ---------R LATQEMKARC
_8_si4_sor: VDVTACAPGL AIPAPPLPTC RTFARPRTCG LGGPYGP--- ---------V DPSPVLKQRC
11_iaaa_ho: ATGQYCYAGM GLPSNPLEGC REYVAQQTCG V----TI--- ---AGSPVSS EPGDTPKDRC
17_iaa1_wh: SGPWMCYPGQ AFQVPALPAC RPLLRLQCNG SQV------- ---------- --PEAVLRDC
18_iaa5_wh: SGPWMCYPGQ AFQVPALPGC RPLLKLQCNG SQV------- ---------- --PEAVLRDC
19_iaa2_wh: GPWSWCNPAT GYKVSALTGC RAMVKLQCVG SQV------- ---------- --PEAVLRDC
20_wrp27_w: GPWMWCDPAM GHRVRPLMGC RAMVKLQCVG NQV------- ---------- --PEAIQRDC
24_ra17_or: SPGEQCRPGI SYPTYSLPQC RTLVRRQCVG RGASA----- ---------- -ADEQVWQDC
___Conserv: .....*.**. ..8..6*4.* *.4*..8.49 .......... .......... .........*

__CHFI_No.: ....50..-. .....60... .....70... .....8-0-- ---------- ----......
_2_itrf_ma: CRELADIP-A YCRCTALSIL MDGAIPPGPD AQLEGR-L-- ---------- ----EDLPGC
_3_iaat_el: CRQLEAIP-A YCRCEAVRIL MDGVVTPSG- -QHEGRLL-- ---------- ----QDLPGC
_8_si4_sor: CRELAAVP-S RCRCAALGFM MDGV------ ---DAP-L-- ---------- ----QDFRGC
11_iaaa_ho: CQELDEAP-Q HCRCEAVRYF IGRRSH---- -PDWSV-L-- ---------- ----KDLPGC
17_iaa1_wh: CQQLAHIS-E WCRCGALYSM LD-------- -----S-MYK EHGA---QEG QAGTGAFPRC
18_iaa5_wh: CQQLADIS-E WPRCGALYSM LD-------- -----S-MYK EHGV---SEG QAGTGAFPSC
19_iaa2_wh: CQQLADINNE WCRCGDLSSM LR-------- -----S-VYQ ELGV---REG ---KEVLPGC
20_wrp27_w: CQELANITNN WCRCHDLGSM LN-------- -----S-VYQ ELGA---REG ---TEVFPGC
24_ra17_or: CRQLAAVDDG WCRCGALDHM LS-------- -----G-IYR ELGA--TEAG HPMAEVFPGC
___Conserv: *.7*7.4... .***.**... .......... .......... .......... .......*8*

__CHFI_No.: ..90..... .-100..... ..110----- .......120 .......---
_2_itrf_ma: PREVQRGFAA T-LVTEAECN LATIS----- GVAECPWILG GGTMPSK---
_3_iaat_el: PRQVQRAFAP K-LVTEVECN LATIH----- GGPFCLSLLG AGE-------
_8_si4_sor: TREMQRIYAV SRLTRAAECN LPTI------ PGGGCHLSNS PR--------
11_iaaa_ho: PKEPQRDFAK V-LVTPGQCN VLTVH----- NAPYCLGLDI ----------
17_iaa1_wh: RREVVKLTAA S-ITA--VCR LPIVVDASGD GAYVCKDVAA YPDA------
18_iaa5_wh: RREVVKLTAA S-ITA--VCR LPIVVDASGD GAYVCKDVAA YPDA------
19_iaa2_wh: RKEVMKLTAA S-VPE--VCK VPIPNP-SGD RAGVCYGDWA AYPDV-----
20_wrp27_w: RKEVMKLTVA S-VPA--VCK VPIPNL-QRG GEGVCY---- ----------
24_ra17_or: RRGDLERAAA S-LPA--FCN VDIPN----G PGGVCYWLGY PRTPRTGH--
___Conserv: .87.....74 ..9.....** .......... ....*..... ..........
```

FIGURE 4: Sequence analysis of the trypsin/α-amylase inhibitor family. (A) Maximum parsimony phylogenetic tree as determined from the sequence alignment with PROTPARS (*32*) using the maize nonspecific transfer protein as an outgroup. Branch lengths indicate phylogenetic distance. (B) Structure-based sequence alignment of the cereal trypsin/α-amylase family of proteins. In this Figure only selected members of the family are presented. Gaps in all proteins indicate one or more proteins of the family are longer in this region. Conserved residues within the entire family are marked as (*) 9, 8, ..., 4 = 20−25, 19, 18, ..., 14 identities among the 25 cereal trypsin/α-amylase inhibitor proteins in the complete alignment. In panel B, the first line is the CHFI sequence numbering. The next four lines are the sequences of proteins 2, 3, 8, and 11. Five lines of AI-subfamily sequences follow for proteins 17, 18, 19, 20, and 24. A key to these protein numbers and other abbreviations is located in Table 1.

Å between CHFI and the average RBI backbone and all atom coordinates, respectively. While the folds are the same overall, there are differences in the side-chain packing, and major differences in disulfide bond dihedral angles.

(C) Distantly Related Structures. Oda et al. (*16*) have found that WAI-19 possesses the same protein fold as a soybean hydrophobic protein (ref *38*; PDB entry 1HYP), an allergen (*39*) with unknown function. The crystallographic

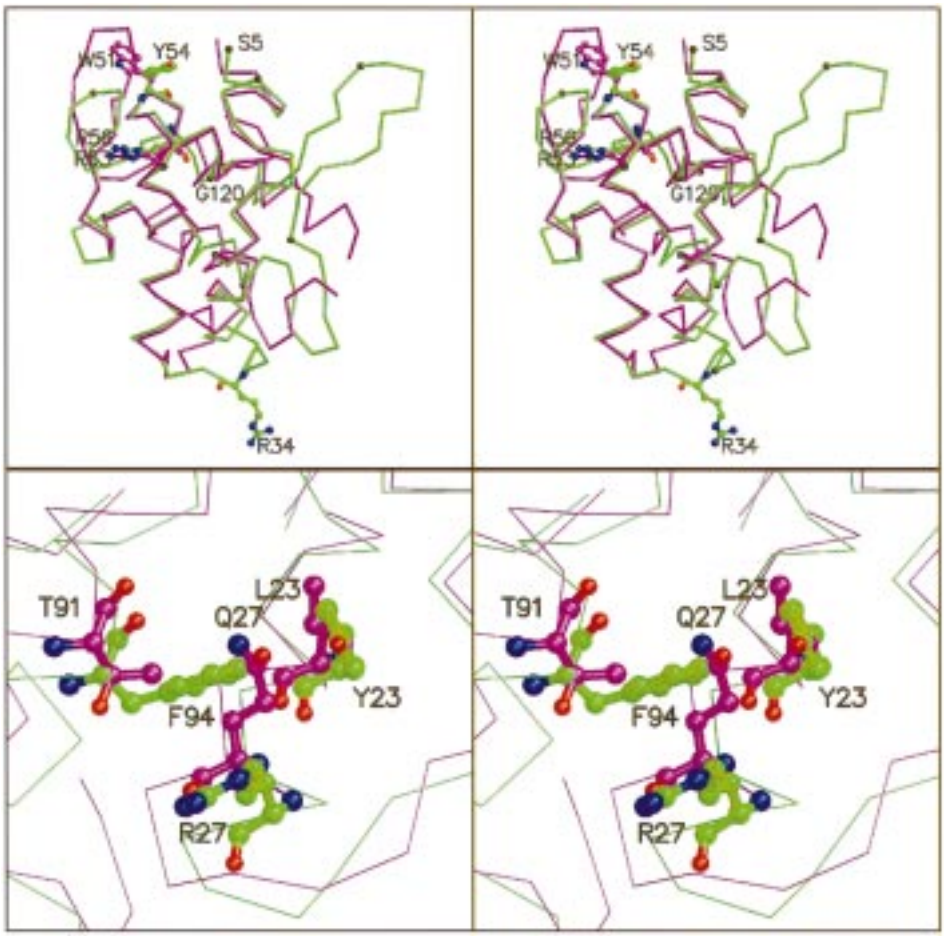


FIGURE 5: Stereo diagrams of the superposition of WAI-19 and CHFI together with noteworthy residues. (A) Superposition of WAI-19 (1HSS, chain A) and CHFI. The specific residues shown include Arg 56 (53 in WAI-19), Tyr 54 (Trp 51 in WAI-19), and Arg 34 for orientation relative to the scissile bond. CHFI is colored green while WAI-19 is magenta. The break in the chain of WAI-19 occurs between residues 68 and 78 of this protein. Every tenth residue of CHFI is marked with a small black sphere. (B) Stereo diagram illustrating the role of Tyr 23 and Phe 94 in serving as a scaffold for the conformation of the protease loop in CHFI. In the figure, Arg 27 of CHFI points out into the solution, while Gln 27 of WAI-19 is directed back toward the hydrophobic core of the protein as a result of smaller residues at CHFI positions 23 and 94. This difference appears to result in the observed difference in conformation of the protease loop between the two proteins.

model of this protein also superimposes well with CHFI (1.48 Å deviation of Cα for 37 equivalenced residues). Additionally, we have found with DALI (*40*) and SCOP (*41*) that the crystallographic model of the maize nonspecific lipid-transfer protein (ref *42*, M-NSLTP) has the same topology as CHFI and fits quite well with three of the four helices (Figure 3). This structural similarity has been noted by the SCOP authors (*41*) because RBI and M-NSLTP are classified in the same superfamily. The disulfide bridge pattern is slightly different from that found with CHFI, apparently an accommodation to allow helix C to translate away from the other three and form the lipid-binding pocket. When the X-ray coordinates of M-NSLTP are superimposed on CHFI, the rms deviation is 1.74 Å for 42 equivalenced Cα atoms. The fit to WAI-19 (1HSS, molecule A) is somewhat closer, 1.43 Å for 44 equivalenced Cα atoms. Sequence alignment based on the structure superposition with CHFI shows that the percent identity of M-NSLTP is marginally closer to WAI-19 (20%) than to CHFI (17%).

The cereal trypsin/α-amylase inhibitor family[3] of proteins is also homologous with the family of 2S seed storage albumins. The similarity is picked up by BLAST searches (*43*) with the CHFI sequence, and the similarity is noted on the PRODOM server (*44*). The structure of one representative of the 2S seed storage family of proteins is available from an NMR study (ref *45*, PDB identifier 1PNB). In this family of proteins, posttranslational proteolysis results in a two-chain mature protein. It is noteworthy that the cleavage into two chains occurs in the proteinase loop of the trypsin/α-amylase inhibitor family. The structural alignment of the energy minimized average of 1PNB with CHFI gives six of the eight Cys residues of 1PNB aligned with the corresponding Cys residues of CHFI. The rms deviation of 37 equivalent Cα atoms is 1.89 Å; however, the NMR structure is only preliminary at this time (*45*).

(D) Sequence Analysis. By the construction of three-dimensional models for sequence alignment of the members

[3] We have termed the family of 25 homologous sequences dealt with in this manuscript as the "cereal trypsin/α-amylase inhibitor" family. This family plus the 2S seed storage proteins was called "cereal superfamily" by Richardson (*12*). The family we consider is termed "trypsin/α-amylase family" in the SwissProt database. Since the close members of this family have been found only in cereals to date and to distinguish the members from other bifunctional inhibitor families, we have attempted to include both Richardson's terminology and the SwissProt terminology in the name "cereal trypsin/α-amylase inhibitor" family.

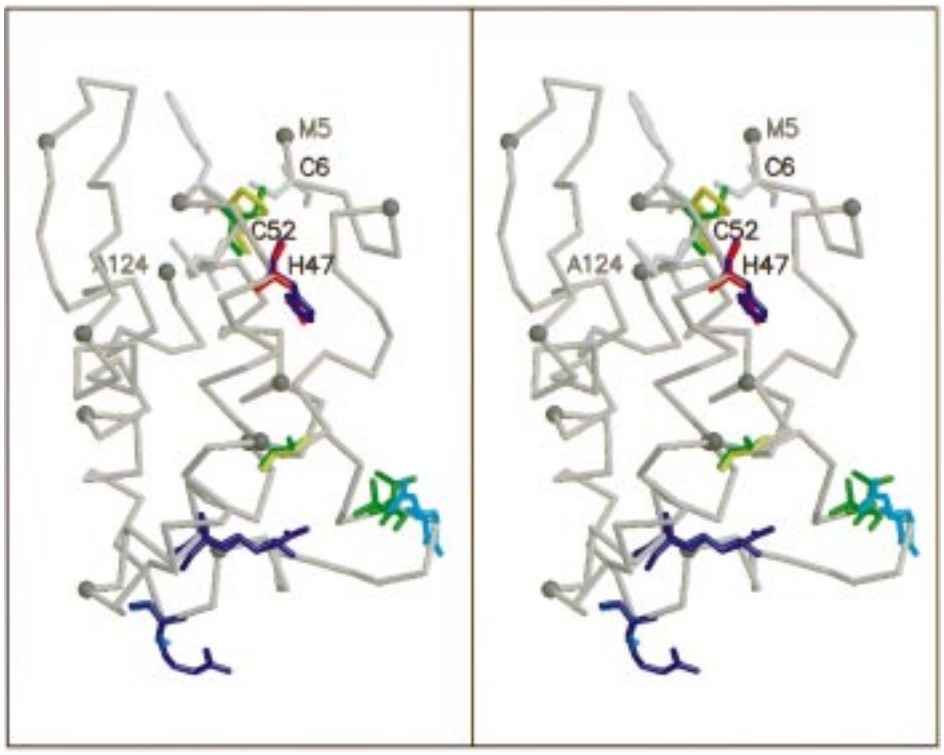


FIGURE 6: Comparison of WAI-19 and WAI-53. Residues differing between these two proteins are shown in color. Note that the change of Cys 52 to Pro (WAI-19 numbering) will make the N-terminal region more flexible in WAI-53; the gray Cys 6 connects to Cys 52 in WAI-19. Additionally, a charge difference results from the conversion of His 47 to Asp. WAI-53 is 50-fold less effective as an inhibitor than WAI-19 against human pancreatic α-amylase (*54*). Coloring is as follows: acidic residues red, basic residues blue, polar residues cyan, nonpolar residues green, Pro and Gly yellow-green. Trp 51 and Arg 53 are gray.

of the cereal trypsin/α-amylase inhibitor family, we hoped to accomplish two goals. First, a structure-based sequence alignment should reflect significant similarities and differences among the family members. This would highlight the importance of particular residues in the architecture of these proteins. Second, we wished to be able to formulate hypotheses concerning the α-amylase inhibition site in this family of proteins. It was felt that the site would be an extensive surface rather than confined to a single contiguous sequence. Consequently, a family of three-dimensional structures might be more useful than a simple sequence alignment to test trial conjectures concerning this site.

Figure 4A gives the phylogeny of the sequences of family members we have found in the published literature. For this analysis, the complete sequence alignment given in the Supporting Information was used. The alignment is condensed in Figure 4B, and the gaps due to alignment with M-NSLTP have been squeezed out. M-NSLTP was aligned with CHFI on a structural basis and used as an outgroup for the family to root the dendrogram.

As has been noted by others, the sequences of the cereal trypsin/α-amylase inhibitor family fall into two major groupings. We have termed these two major subfamilies as AI (α-amylase inhibitors) and TI (trypsin inhibitors). The TI subfamily includes proteins 2−16 of Figure 4A. Actually, only proteins 2−5 are capable of inhibiting trypsin-like proteinases. Some proteins such as 6 and 16 are characterized only by the known sequence. Proteins 9−15 include α-amylase inhibitors, but only some are inhibitory; the sequences have been characterized as subunits of the tetrameric wheat and barley α-amylase inhibitors (*46*−*48*). Proteins 9,10, and 12−15 are not inhibitory as monomers. In the second major grouping, AI, are included both dimeric (proteins 17, 18, and 21) and monomeric (proteins 19, 20, and 27) α-amylase inhibitors. Proteins 22−26 have not been tested for α-amylase inhibition, but are rice or barley allergens associated with Baker's asthma. In Figure 4A, the branch lengths indicate phylogenetic distance.

Figure 4B gives the sequence alignment of a set of proteins limited to a few for which models were constructed. For discussion of positions, the CHFI numbering will be used. Figure 5, top, shows a superposition of CHFI and WAI-19 (1HSS, chain A). As noted by Oda et al. (*16*) significant differences occur between the secondary structure of WAI-19 and the TI subfamily as represented by RBI (*15*). Figure 5, top, is presented as an aid in the following discussion of sequences.

(E) Noteworthy Residues in the Family of Structures. Among the TI family members, Tyr 23 is conserved (Figure 5, bottom). This residue interacts with Phe 94. The pair of residues causes the Arg or Gln of the TI family proteins at position 27 to point out into the solvent and continue helix A as shown in Figure 5, bottom. In the AI group of proteins, Tyr 23 becomes Leu, Met, or Val. These smaller residues and the smaller Thr, Ala, or Leu at 94 allow the residue at 27, generally Gln, to orient toward the hydrophobic core of the protein, ending helix A. The result of these interactions is that Cys 29 of the TI group does not align with Cys 28 of the AI group. Thus, indirectly, the Tyr 23−Phe 94 interaction can be viewed as the scaffold for the proteinase loop configuration among the trypsin inhibitors.

A part of the structure which is conserved among all members of this family is the arrangement surrounding Arg 56. The nonpolar portion of the Arg side chain contributes

significantly to the hydrophobic core of the protein and the guanidinium group ties the inaccessible main-chain carbonyl oxygens of the surrounding residues together with four hydrogen bonds. The main-chain atoms which are bonded in this structure are N$\epsilon\cdots$O−48, N$\epsilon\cdots$O−51, Nη1$\cdots$O−107, and Nη2$\cdots$O−49. We observe similar arrangements in the hydrogen bonding surrounding Arg 56 among all the models constructed for the sequence alignments. This structure appears to be fundamental to forming the proper disulfide bridges and also to set up the change from helix B to helix C in the family of proteins. The arrangement also appears important for exposure of the (usually) aromatic Tyr or Trp at position 54.

DISCUSSION

(A) Structure Determination. The structure determination of CHFI relied on a combination of multiple isomorphous replacement and molecular replacement techniques. We tried many heavy atom compounds in routine soaks without finding an effective derivative. Also, in many surveys of crystallization conditions, we did not find additional crystal forms for this protein. Both of these factors led us to try to distort the crystal by removal of the Mg^{2+} prior to the heavy atom soak. While the electron density of the MIR-phased map was poor, it was nevertheless good enough that we could immediately recognize the correct molecular replacement solution when it was found.

There are several advantages in using a single model rather than the entire ensemble observed in the NMR results in molecular replacement trials. It was done in this case to simplify changes to the model. The idea of using the Fourier transform to average the models was inspired by the work with the "probability map" of Bonvin and Brunger (*49*). In our case, we are simply using the structure factors in the same fashion as Fobs to arrive at a single representative model of the protein using an ordinary crystallographic type of refinement. The *B* factors found upon refinement correlate well with the rms deviations among the models. Parts of the model which have large *B* factors (>80 Å^2) can then be excised for the actual molecular replacement trials.

Initially, upon inspection of the RBI NMR ensemble, we thought the failure in the molecular replacement could be due to improper determination of the dihedral angles in the disulfide bridges. It was for this reason that we decided to try to restrain the disulfide bridge dihedral angles to ±90° as observed in high-resolution crystal structures. Restraining the dihedral angles and also constraining the Cα−Cα distance to 5.5 ± 0.6 Å across the disulfide bridge did not have as much effect on the model as we expected, although it did improve the model slightly as indicated by a drop of 0.1 Å in rms deviation from the final CHFI coordinates.

The six-dimensional search program, EPMR (*26*), was essential for solving this molecular replacement problem. Neither AMoRe nor X-PLOR were able to identify the correct solution in either the rotation or translation functions. Only after partial refinement of the search model with X-PLOR (to an *R* of 33%, R_{free} = 42%) were the molecular replacement methods of X-PLOR able to "recognize" the correct solution.

The conformation of the proteinase loop was pretty much as expected. The molecule can be superimposed on the CMTI-I inhibitor−trypsin complex (*50*) and with some small modifications to the Arg 34 side chain, the CHFI model fits well. In the absence of a structure for Factor XIIa, however, we are unable to answer the question why this inhibitor is so exquisitely selective for this proteinase among all the plasma proteinases.

(B) Sequence Analysis. The first objective of our sequence alignment was to aid in mutagenesis or other functional analysis of family members. Superposition of the CHFI model and the other models constructed in this study indicates that the "finger loop" of the TI subfamily should not be aligned with the "flexible loop" of the AI subfamily. When the two subfamilies are separated and aligned, quite distinct "signature residues" are observed which reflect structural differences in the two subfamilies.

Our second objective in the model construction and alignment was to locate the α-amylase inhibitory site on the surface of CHFI and (presumably) the rest of the family. Taking together the models and experimental observations we can hypothesize the location of the α-amylase inhibition site. First, we assume the site is similar among all the members of the family. Second, because RBI can form a ternary complex of trypsin, RBI, and α-amylase (*51, 52*), the amylase site must be distinct from the proteinase loop. Third, based on the WAI-19 dimerization site, we can exclude the "platform loop" in the C-terminal region. WAI-19 inhibits *Tenebrio molitor* α-amylase as a dimer (*53*), and one face of the "platform loop" is buried in the extensive dimer interface (*16*).

These assumptions and observations can be put together with four experimental results as criteria for the site identification. The observations are as follows. First, study of the complex of RBI with porcine pancreatic α-amylase (*54*) indicated involvement of aromatic residues in the complex. Second, modification of amino groups (Lys or the N-terminus) abrogated α-amylase binding to RBI (*51*). Third, the lack of α-amylase (*Tenebrio*) inhibitory activity of CHFI mutated as 4n-11 (*18*) has further implicated the N-terminal region in the inhibitory site. Fourth, binding of the inhibitors to their cognate α-amylase is tight, in the nanomolar range for K_i (*14*), indicating extensive surface involvement in the inhibitor−enzyme complex.

In the region surrounding the CHFI Arg 56 site, at CHFI position 54, an aromatic tyrosine or tryptophan occurs in almost all of the α-amylase inhibitors. Additionally, the residues in this region superimpose well among all of the models. We think that this is one portion of the site (see Figure 5, top, and 6). A second position of interest is the conserved structure near the N-terminal between residue 6 and 15 (possibly as far as residue 18). The third region which may be part of the site is located in the surroundings of residue Glu/Gln 47 where there are charge differences between the WRP27 and WAI-28 proteins possibly causing the former to lose its *Tenebrio* α-amylase inhibition. The charge pattern of Arg, Glu, and Asp on the surface of helix B (residues 42−44) may be involved in specificity determination of the inhibitors with their cognate α-amylase. A last mention must be made of Asn 105. The residue shows remarkable conservation, and no structural basis for it can be discerned by examination of the RBI or CHFI structures. This portion of the platform loop is close to the exposed aromatic residue at position 54 and so may be involved in

the binding to α-amylase by some of the monomeric inhibitors.

This proposed set of sites has some experimental verification in the following form (Figure 6). Wheat α-amylase inhibitor 0.53 differs by only seven residues from the wheat 0.19 inhibitor. Two positions are Cys 55 (52 in WAI-19) → Pro and nearby His 50 (47 in WAI-19) → Asp. The remaining five changes are near the flexible loop or do not involve alteration of charges. The lack of a Cys 6−Cys 55 disulfide bridge should increase the flexibility of the N-terminal region as well. The profound effects these few changes have on the activity (*55*), taken together with the activity specificity differences in the pair of WAI-28 and WRP-27 (*56*) lend support to our hypothesized inhibition site.

A question which this hypothetical site does not address is why a clustering of aromatic residues occurs in the C−terminal regions of these molecules in this family of proteins. There is quite a bit of variability among the modeled structures in the positioning of these aromatic residues, so we were unable to arrive at a satisfactory hypothesis for their involvement in the inhibition mechanism.

(C) Conclusions. The structures determined in this work indicate that there is no particular, peculiar conformation of the proteinase inhibitory site of CHFI which might determine its specificity for Factor XIIa among all the plasma serine proteinases. While models with trypsin are quite satisfactory, models with thrombin give substantial steric clashes when complexes with CHFI are constructed. We doubt that structural modeling of Factor XIIa would be sufficiently accurate to answer detailed questions regarding interacting residues, so we have not approached that problem in this work.

The finding of structural similarity with the nonspecific transport protein of maize opens further questions on the evolution of this diversified group of proteins. Apparently this four-helix motif and its variants have proven to be of great utility in cereal evolution by providing protection from predators, food for the embryo, and lipid transfer.

We have tried to ask questions about the α-amylase inhibition site on this family of proteins, and have proposed a set of residues for this site, centered around Arg 56 and Tyr 54 of CHFI. It is hoped that this proposal will be tested soon by chemical or molecular biological experiments.

ACKNOWLEDGMENT

We would like to thank Dr. K. Fukuyama for sending us the coordinates of wheat α-amylase inhibitor 0.19 prior to release from the PDB. Without these coordinates the sequence comparisons could not have been made.

SUPPORTING INFORMATION AVAILABLE

Complete alignment of the 25 cereal trypsin/α-amylase inhibitor family of proteins including the alignment of maize nonspecific lipid transfer protein (4 pages). Ordering information is given on any current masthead page.

NOTE ADDED IN PROOF

Since this manuscript was submitted and revised, the following paper has appeared: Strobl, S., Maskos, K., Wiegand, G., Huber, R., Gomis-Ruth, F. X., and Glockshuber, R. (1998) A novel strategy for inhibition of alpha-amylases: yellow meal worm alpha-amylase in complex with the *Ragi* bifunctional inhibitor at 2.5 Å resolution. *Structure 6*, 911−921.

REFERENCES

1. Hojima, Y., Pierce, J. V., and Pisano, J. J. (1980) *Thromb. Res. 20*, 149−162.
2. Swartz, M. J., Mitchell, H. L., Cox, D. J., and Reeck, G. R. (1977) *J. Biol. Chem. 252*, 8105−8107.
3. Mahoney, W. C., Hermodson, M. A., Jones, B., Powers, D. D., Corfman, R. S., and Reeck, G. R. (1984) *J. Biol. Chem. 259*, 8412−8416.
4. Chong, G., and Reeck, G. R. (1987) *Thromb. Res. 48*, 211−221.
5. Ratnoff, O. D., and Moneme, V. (1981) *Proc. Soc. Exp. Biol. Med. 166*, 297−299.
6. Kozin, F., and Cochrane, C. G. (1988) in *Inflammation: Basic Principles and Clinical Correlates* (Gallin, J. I., Goldstein, I. M., and Snyderman, R., Ed.) pp 101−120, Raven Press, New York.
7. Hojima, Y., Pierce, J. V., and Pisano, J. J. (1980) *Thromb. Res. 20*, 163−171.
8. Bode, W., and Huber, R. (1992) *Eur. J. Biochem. 204*, 433−451.
9. Laskowski, M., Jr., and Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593−626.
10. Lei, M. G., and Reeck, G. R. (1986) *J. Chromatogr. 363*, 315−321.
11. Wen, L., Huang, J. K., Zen, K. C., Johnson, B. H., Muthukrishnan, S., MacKay, V., Manney, T. R., Manney, M., and Reeck, G. R. (1992) *Plant Mol. Biol. 18*, 813−814.
12. Richardson, M. (1991) in *Plant Biochemistry* (Rogers, L. J., Ed.) Vol. 5 pp 259−305, Academic Press, New York.
13. Sanchez-Monge, R., Gomez, L., Barber, D., Lopez-Otin, C., Armentia, A., and Salcedo, G. (1992) *Biochem. J. 281*, 401−405.
14. Chen, M. S., Feng, G. H., Richardson, M., Valdez-Rodriguez, S., Reeck, G. R., and Kramer, K. J. (1992) *Insect Biochem. Mol. Biol. 22*, 261−268.
15. Strobl, S., Muhlhahn, P., Bernstein, R., Wiltscheck, R., Maskos, K., Wunderlich, M., Huber, R., Glockshuber, R., and Holak, T. A. (1995) *Biochemistry 34*, 8281−8293.
16. Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T., and Morimoto, T. (1997) *Biochemistry 36*, 13503−13511.
17. Lee, J. Y., Min, K., Cha, H., Shin, D. H., Hwang, K. Y., and Suh, S. W. (1988) *J. Mol. Biol. 276*, 437−448.
18. Hazegh-Azam, M., Kim, S.-S., Masoud, S., Andersson, L., White, F., Johnson, L., Muthukrishnan, S., and Reeck, G. R. (1998) *Protein Expression Purif.* (in press).
19. Pedersen, L. C., Yee, V. C., von Dassow, G., Hazeghazam, M., Reeck, G. R., Stenkamp, R. E., and Teller D. C. (1994) *J. Mol. Biol. 236*, 385−387.
20. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
21. Higashi, T. (1990) PROCESS: A program for indexing and processing R-AXIS II imaging plate data, Tokyo: Rigaku Corp.
22. Siemens Energy and Automation, Inc. (1996) *Saint Version 4 Software Reference Manual*, Siemens Energy & Automation, Inc. Madison WI.
23. McRee, D. E. (1992) *J. Mol. Graph. 10*, 44−46.
24. CCP4, Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
25. Brunger, A. T. (1992) *XPLOR version 3.0: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
26. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1998) American Crystallographic Association Abstract 11.07.04.
27. Read, R. J. (1986) *Acta Crystallogr., Sect. A 42*, 140−149.
28. Yee, V. C., Pratt, K. P., Cote, H. C., Le Trong, I., Chung, D, W., Davie, E. W., Stenkamp, R. E., and Teller, D. C. (1997) *Structure 5*, 125−138.
29. Kleywegt, G. J. (1997) *CCP4/ESF-EACBM Newsletter on Protein Crystallography 34*, 5−8.

30. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
31. Sippl, M. J. (1993) *Proteins 17*, 355−362.
32. Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Washington, Seattle.
33. Fitch, W. M., and Farris, J. S. (1974) *J. Mol. Evol. 3*, 263−278.
34. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
35. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol. 277*, 505−524.
36. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng. 8*, 127−134.
37. Navaza, J. (1994) *Acta Crystallogr., Sect. A 50*, 157−163.
38. Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S., and Lehmann, M. S. (1993) *J. Mol. Biol. 231*, 877−887.
39. Gonzalez, R., Varela, J., Carreira, J., and Polo, F. (1995) *Lancet 346*, 48−49.
40. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
41. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol. 247*, 536−540.
42. Shin, D. H., Lee, J. Y., Hwang, K. Y., Kim, K. K., and Suh, S. W. (1995) *Structure 3*, 189−199.
43. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.
44. Corpet, F., Gouzy, J., and Kahn, D. (1998) *Nucleic Acids Res. 26*, 323−326.
45. Rico, M., Bruix, M., Gonzalez, C., Monsalve, R. I., and Rodriguez, R. (1996) *Biochemistry 35*, 15672−15682.
46. Buonocore, V., De Biasi, M.-G., Giardina, P., Poerio, E., and Silano, V. (1985) *Biochim. Biophys. Acta 831*, 40−48.
47. Gomez, L., Sanchez-Monge, R., Garcia-Olmedo, F., and Salcedo, G. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3242−3246.
48. Barber, D., Sanchez-Monge, R., Mendez, E., Lazaro, A., Garcia-Olmedo, F., and Salcedo, G. (1986) *Biochim. Biophys. Acta 869*, 115−118.
49. Bonvin, A. M., and Brunger, A. T. (1995) *J. Mol. Biol. 250*, 80−93.
50. Bode, W., Greyling, H. J., Huber, R., Otlewski, J., Wilusz, T. (1989) *FEBS Lett. 242*, 285−292.
51. Shivaraj, B., and Pattabiraman, T. N. (1981) *Biochem. J. 193*, 29−36.
52. Maskos, K., Huber-Wunderlich, M., and Glockshuber, R. (1996) *FEBS Lett. 397*, 11−16.
53. Buonocore, V., Gramenzi, F., Pace, W., Petrucci, T., Poerio, E., and Silano, V. (1980) *Biochem J. 187*, 637−645.
54. Alagiri, S., and Singh, T. P. (1993) *Biochim. Biophys. Acta 1203*, 77−84.
55. Maeda, K., Kakabayashi, S., and Matsubara, H. (1985) *Biochim. Biophys. Acta 828*, 213−221.
56. Feng, G. H., Richardson, M., Chen, M. S., Kramer, K. J., Morgan, T. D., Reeck, G. R. (1996) *Insect Biochem. Mol. Biol. 26*, 419−426.
57. Paz-Ares, J., Ponz, F., Lazaro, A., Rodriguez-Palenzuela, P. (1986) *Theor. Appl. Genet. 71*, 842−846.

BI9812266