Implications of the 3-D structure determination of family 18 chitinases
Similarities with FnIII domains, oviductal proteins and narbonins

Anastassis Perrakis, Christos Ouzounis[!],
Keith S. Wilson and Constantin E. Vorgias*
EMBL-Hamburg, DESY, Notkestrasse 85, 22603 Hamburg, Germany
[!] AI Center, SRI International,333 Ravenswood Avenue,Menlo Park, CA 94205, USA
* corresponding author, Email Tasos@Embl-hamburg.de

Abstract

Structure solution of Chitinase A revealed various relationships with other proteins. The fold of the N-termninal domain closely resembles that of Fibronectin type III domains which are present in other chitinases (according to sequence data). The primary structure similarity with oviductal proteins enabled the building of a homology derived model for this class of proteins. The streaking similarities in tertiary structure with narbonin might indicate a possible function for these proteins.

Introduction

Chitinases (E.C.3.2.1.14) hydrolyze the chitin homopolymer of N–acetyl–D–glucosamine into smaller oligomeric saccharides. They compose families 18 and 19 of glycosyl hydrolases [1]. Family 19 is very homogeneous and contains only plant enzymes. Family 18 is evolutionary diverse and contains chitinases from plants, fungi, bacteria and the viral chitinases. In family 18 belong also eukaryotic chitobiases (unlike prokaryotic chitobiases which constitute family 20 together with hexosaminidases) which cleave the dimer of N–acetyl–D–glucosamine. Also proteins secreted in mammalian ovaries (oviductal) belong to this family [2]. These are glycoproteins with unknown function, but they most likely share a common phylogenetic origin with the other family 18 enzymes.

The structures of three family 18 enzymes have been reported almost simultaneously on December 1994. They included:

1. A bacterial chitinase from *Serratia marcescens* [3].

Swissprot	Function	Pos
CHIA_SERMA	*Serratia marcescens*	S
CHIT_ALTSO	*Alteromonas species*	S
U09139*	*Aeromonas caviae*	S
CHIT_NPVAC	*Autographa californica* NPV	S
U12688*	*Bombyx mori* NPV	S
M97906*	*Choristoneura fumiferana* NPV	S
CHI1_BACCI	*Bacillus circulans*	E,E
CHID_BACCI	*Bacillus circulans*	S
CHIT_STRLI	*Streptomyces lividans*	S
CHIT_STRPL	*Streptomyces plicatus*	S
CHIX_STROI	*Streptomyces olivaceoviridis*	S

Table 1: Occurances of CtnI and FnIII domains in chitinases. E,S stand for End and Start respectively. NPV stands for Nuclear Polyhedrosis Virus

2. A hevamine, which combines chitinase and lysozyme activity and was isolated from the plant *Hevea brasiliensis* [4].

3. an endo-β-N-acetyloglucosaminidase, F_1 which is a protein homologous to family 18 chitinases but has a function adapted for a more complex substrate [5] (another endo-β-N-acetyloglucosaminidase, H, was reported a year later, and is highly homologous to F_1 [6]).

Different models have initially been proposed for the enzymatic mechanism, based on the structure of these enzymes. Recent results (unpublished) show that one conserved Glutamate residue must be the proton donor for the reaction and that the breakage of the glycosidic bond leads to retention of configuration of the anomeric conformation of the C_1 atom.

Similarities with FnIII domains

The crystal structure determination of Chitinase A from *Serratia marcescens* [3] revealed that the N-terminus of the mature protein (aminoacids 24-137, after cleavage of the leader peptide) forms a distinct domain. This domain has an immunoglobulin-like fold (seven strands in a two-sheet β sandwich) and is classified in the superfamily of Immunoglobulin

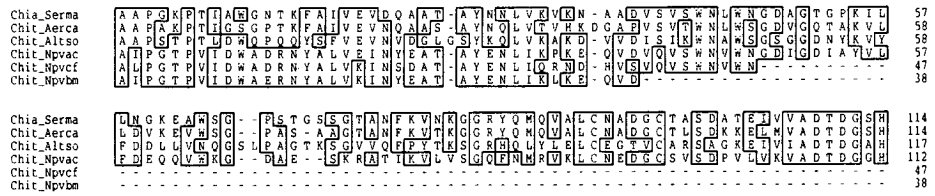

Figure 1: Alignment of the sequences with the ChiN motif. Chia_Serma is for *Serratia marcescens* chitinase ; Chit_Altso for *Alteromonas species* chitinase ; Chit_Aerca for *Aeromonas caviae* chitinase ; Chit_Npvac, Chit_Npvcf and Chit_Npvbm for the probable chitinases from *Autographa californica*, *Bombyx mori* and *Choristoneura fumiferana* nuclear polyhedrosis viruses respectively

domains in the SCOP database. It has similarities with all the other families and superfamilies of this fold, as expected, which include the superfamily of Fibronectin type III (FnIII) domains [7].

Sequence searches against protein and nucleotide databases were performed using the BLASTP suite of programs [8]. The BLOSUM 62 matrix was used for scoring similarities and a non-redundant set of all the major protein databases was searched. The N-terminus domain sequence was identified in four protein sequences (including Chitinase A from *Serratia marcescens*) and two nucleotide sequences, which are listed in Table 1. The sequences from the last two viruses are incomplete and only the N-terminal part of the sequence is present, which is sufficient to show the similarity. The last sequence (M97906) is annotated only as a cathepsin gene. It is very likely that a complete chitinase gene is present in both cases, and could readily be sequenced.

An alignment of these sequences created by CLUSTALW [9] is shown in Figure 1.. A profile constructed from this alignment fails to identify any more sequences that the ones used to construct it, when compared against the protein database.

We propose that the N-terminal domain of chitinases is named ChiN and should marked be accordingly in the corresponding sequence entries.

Serratia marcescens, *Alteromonas species* and *Aeromonas caviae* belong to the superfamily of gracilicutes. All the other bacterial species for which chitinases have been sequenced belong to the superfamily of firmicutes. None of the firmicutes chitinases has the sequence of this domain as gracilicutes enzymes do. Some of the firmicutes chitinases (*Bacillus circulans, Streptomyces lividans, Streptomyces plicatus and Streptomyces*

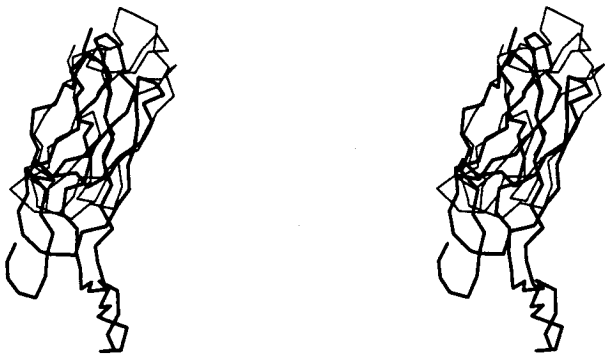

Figure 2: Superimposition of the Cα atoms of the ChiN domain and FnIII domain. The thick line shows the fold of the ChiN domain and the thinner line shows the fold of an FnIII domain (PDB entry 1fna). 55 Cα atoms are superimposed with an *rms* value of 2.258 Å.

olivaceoviridis) have sequences similar to the sequence of FnIII like domains. These domains are located in the N-termini of these proteins, except for one of the two of the *Bacillus circulans* chitinases which has two domains of this type attached to the C-terminus. The FnIII like domains have similar folds to the ChiN domain, as shown in Figure 2. It is likely that the ChiN domain evolved in bacteria shortly after the divergence of the two superfamilies. Possibly it has maintained the same function like the FnIII domain, since at the tertiary structure level they present similar folds.

It has been shown [10] that most likely the FnIII domain did not evolve in bacteria but was acquired from animals later in evolution. The limited spread of the ChiN domain does not allow any clear conclusions for the phylogenetic origin of this domain. However, when a phylogenetic tree is constructed using an alignment only of the catalytic $\alpha\beta$ barrel motif of all bacterial and the one complete viral (Chit_Npvac) chitinase, the four chitinases which have the ChiN domain clearly form a distinct subgroup (Figure 3). The chitinases bearing FnIII like domains do not show any close similarity. This suggests that most likely the ChiN domain evolved in an early ancestor of gracilicutes chitinases, rather than independently acquired by these proteins.

Of interest is the recent identification of chitinases in insect viruses. These viruses infect insects (the species infected are indicated in the virus

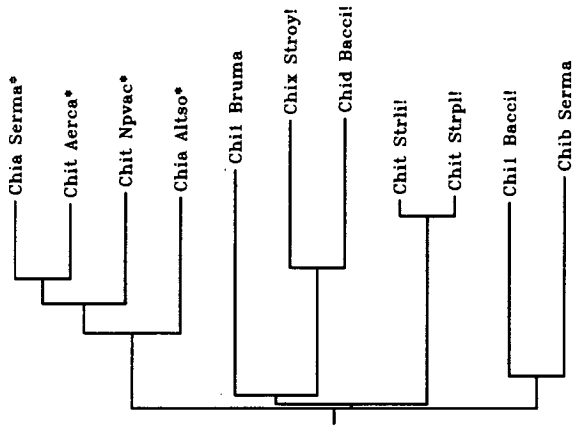

Figure 3: Phylogenetic tree for family 18 bacterial chitinases, constructed using only the sequences forming the $\alpha\beta$ barrel motif of this class of enzymes. The names are the same as for Figure 1. the * show the sequences with the ChiN motif and the ! show the sequences with FnIII motifs. While the former make a distinct cluster the latter are spread over the branches of the rest of the tree, supporting the hypothesis that the FnIII domain was acquired by these proteins late in evolution. table

name) and presumably the ability to secrete a chitinase is crucial for the function of the enzyme, which might help to degrade the insect cuticle, largely composed of chitin. Remarkably, all known viral chitinases have the ChiN domain.

No biochemical data are available for the function of the ChiN domain. However, one of the proteins having a fold similar to the ChiN domain is Cellulase D from *Clostridium thermocellum.* Unfortunately, the coordinates for this domain are not available for comparison. Considering that the role of that domain is to bind cellulose, it is likely that the ChiN domain has a similar role and is responsible for the binding to the filamentous chitin substrate. It must be noted, that FnIII domains in the *Bacillus* chitinase have been shown biochemically not to interact directly with chitin, but their presence affects the hydrolysis of the substrate. Further investigations whould be carried out to clarify the role of the FnIII and ChiN domains in chitinases.

A similar fold for oviductal proteins

Searching the sequence databases with BLASTP and using as a search sequence the barrel motif part of the ChiA sequence, clearly identified similarities with a class of glycosylated proteins secreted by ovaries cells, the oviductal proteins. The similarity observed was significant (similarity scores for the oviductal proteins were better than for some other chitinases of family 18). This similarity was later reported in [2] and the corresponding entries are now annotated for being similar to chitinases. From the initial alignments it was evident that at least one domain of these proteins adopts an $\alpha\beta$-barrel fold. It was decided to build a model for one of these proteins using homology model building techniques. The human oviductal protein (OviH) [11] was chosen because it is the one closest to ChiA and also its apparent importance as a human protein. A homology built model is important because due to the heavy glycosylation of these class of proteins it be very hard to crystallise. Having a three dimensional model and mainly an improved sequence alignment that incorporates the structural information will be helpfull in the future

.

The $\alpha\beta$-barrel motif of oviductal proteins is around 400 residues and is formed by the N-terminal part of the sequence. The rest 200–250 carboxy terminal residues could not be aligned to anything else. Most likely they form an additional domain and also possibly they have the glycosylation sites.

The alignment used for secondary structure prediction was extracted and used for the first step of homology model building based on the ChiA structure. The WHATIF program was used in all steps of homology model building. The homology derived model was checked in the graphics and errors in the alignment were corrected manually. A few cycles of this procedure were used. They proved to be extremely powerful on correcting errors of the alignment that would not have been noticed without the structural information.

The new improved alignment has several insertions and deletions which all can be explained and are structurally 'acceptable' , i.e. they can occur without disrupting the $\alpha\beta$-barrel fold. They are all shown in Figure 4.

The mammalian oviductal proteins (alignment not shown) all have conserved the N-terminal region corresponding to the $\alpha\beta$-barrel . The C-terminal region, 40-50 aa after the end of the barrel starts to be highly

```
AHKLVCYFTNWAHS-RPGPASILPHDLDPFLCTHLIFAFASMN--------------------------NNQIVAKDLQDEK---------------IL
GKVVGSYFVEWGVYGRNFTVDKIP----AQNLTHLLYGFIPICGGNGINDSLKEIEGSFQALQRSCQGREDFKISIHDPFAALQKAQKGVTAWDDPYKGN

YPEFNKLKERNRELKTLLSIGGWNFGTSRFTTMLSTFANREKFIASVISLLRTHD-FDGLDLFFLYPGLR-----GSPMHDRWTFLFLIEELLFAFRKEA
FGQLMALKQAHPDLKILPSIGGWTLSDP--FFFMGDKVKRDRFVGSVKEFLQTWKFFDGVDIDWEFPGGKGANPNLGSPQDGETYVLLMKELRAMLDQLS

LLTMRPRLLLSAAVSGVPHIVQTSYDVRFLGRLLDFINVLSYDLHGSWE-RFTGHNSPLFSLPEDPKSSAYAMNYWRKLG---APSEKLIMGIPTYGRTF
TETGRKYELTSAISAGKDKIDKVAYNVAQN--SMDHIFLMSYDFYGAFDLKNLGHQTALNAPAWKPDTAYTTVNGVNALLAQGVKPGKIVVGTAMYGRGW

RLLKASKNGLQARAIGPASPGKYTKQEGFLAYFEICS-FVWGAKKHWIDY-QYVPYANK--GKEWVGYDNAISFSYKAWFIRREHFGGAMVWTLDMDDVR
TGVNGYQNNIPFTGTATG-PVKGTWENGIVDYRQIAGQFMSGEWQYTYDATAEAPYVFKPSTGDLITFDDARSVQAKGKYVLDKQLGGLFSWEIDADN--
```

Figure 4: Improved alignment of OviH and ChiA after homology model building

variable. The active site residues of chitinases are not conserved, as well as the residues responsible for substrate binding. The overall shape of the binding cleft is obviously conserved in our model, and not filled in with side chains of neighboring residues. It is likely that the oviductal proteins have a role in binding sugar oligomers but not able to have any hydrolizing activity, for any polymeric sugars.

Similarity with narbonins

Narbonins are a class of proteins abundabt in plant seeds whose function remains unidentified. The structure of a narbonin was solved and had an $\alpha\beta$-barrel fold [12]. When this structure was superimposed to the one of ChiA not only the 3-D fit was of good quality (1.91 Å) but also some parts of the sequences were conserved. These parts included the S(IV)GG motif which is overall conserved in family 18 enzymes and the catalytic glutamate residue. BLAST searches using as a probe the ChiA sequence identify narbonins but with low scores, similar however to the ones for eukaryotic chitobiases. It is likely that narbonins share a common phylogenetic origin with family 18 of glycosyl hydrolases.

It has been previously postulated that narbonins might have a chitinolytic action [13]. Despite the fact that the catalytic residue of chitinases is conserved this does not seem very likely. The catalytic groove is 'blocked' by side chains in narbonin and the binding of substrate would not be possible. Biochemical experiments did not show any chitonolytic action for narbonins (Henning, M., personal commmunication).

References

[1] Henrissat, B. & Bairoch, A. (1993). New families in the classification of

glycosyl hydrolases based on amino acid sequence similarities. *Biochem.J.* **293**, 781–788.

[2] DeSouza,M.M. and Murray,M.K. (1995). An estrogen-dependent secretory protein, which shares identity with chitinases, is expressed in a temporally and regionally specific manner in the sheep oviduct at the time of fertilization and embryo development. *Endocrinology.* **136**(6), 2485–2496.

[3] Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A., Chet, I., Wilson, K.S. & Vorgias, C.E. (1994). Crystal structure of a bacterial chitinase at 2.3 Åresolution. *Structure.* **2**, 1169–1180.

[4] Terwischa van Scheltinga, A.C., Kalk, K.H., Beintema, J.J. & Dijkstra, B.W. (1994). Crystal structure of hevamine, a plant defence protein with chitinase and lysozyme activity, and it's complex with an inhibitor. *Structure.* **2**, 1181–1189.

[5] Roey, P.V., Rao, V., Plummer, T.H. & Tarentino, A. (1994). Crystal structure of (endo-β-n-acetyloglucosaminidase f_1, an α/β-barrel enzyme adapted for a complex substrate. *Biochemistry.* **33**, 13989–13996.

[6] Rao, V., Guan, C. & Roey, P.V. (1995). Crystal structure of (Endo-β-N-acetyloglucosaminidase H, at 1.9 Åresolution: active site geometry and substrate recognition. *Structure.* **3**, 449–457.

[7] Campbell, I.D & Spitzfaden, C. (1994). Building proteins with fibronectin type III modules. *Structure.* **2**(5), 333–337.

[8] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipmann P.J. (1990). Basic local alignment search tool. *J.Mol.Biol.* **215**, 403–410.

[9] Higgins, D.G., Bleasby, A.J. & Fuchs, R. (1992). CLUSTALV:improved software for multiple sequnce alignment. *CABIOS.* **8**(2), 189–191.

[10] Bork, P. & Doolittle, R.F. (1992). Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA.* **89**, 8990–8994.

[11] Arias,E.B., Verhage,H.G. and Jaffe,R.C. (1994). Complementary deoxyribonucleic acid cloning and molecular characterization of an estrogen-dependent human oviductal glycoprotein. *Biol. Reprod.* **51**(4), 685–694.

[12] Hennig, M., Schlesier, B., Dauter, Z., Pfefer, S., Betzel, C., Hohne, W.E., & Wilson, K.S. (1992). Crystal structure of narbonin. *FEBS Lett.* **306**, 80–84.

[13] Coulson, A.F.W. (1994). A proposed structure for family 18 chitinases a possible function for narbonin. *FEBS lett.* **354**, 41–44.