ELSEVIER

FEBS Letters 340 (1994) 129–132

FEBS 13713

FEBS LETTERS

Structural similarity of plant chitinase and lysozymes from animals and phage

An evolutionary connection

Liisa Holm*, Chris Sander

European Molecular Biology Laboratory, D-69012 Heidelberg, Germany

Received 28 December 1993; revised version received 20 January 1994

Abstract

A search in the database of known three-dimensional protein structures with the structure of a plant endochitinase revealed a subtle but unambiguous similarity to lysozymes from animals and phages. An evolutionary connection between plant endochitinases and lysozymes is supported by similar overall topology of fold, overlapping substrate specificities and remarkable conservation of some sequence and architectural detail around the active site. Much of the knowledge about lysozyme can now be extended by analogy to endochitinase. New insights into the mechanism of endochitinase are expected to stimulate genetic engineering studies into plant defense mechanisms against pests and pathogens.

Key words: Lysozyme; Chitinase; Homology; Evolution

1. Introduction

Lysozyme from chicken egg-white was the first enzyme structurally determined by crystallography. The catalytic mechanism by which lysozyme breaks down bacterial cell walls is well known to every biochemist [1]. Lysozyme from bacteriophage T4 is perhaps the most thoroughly studied protein by means of protein engineering [2,3]. The crystal structures of hundreds of point mutants are already known. Much less is understood about the molecular mechanism of chitinases. This class of enzymes is thought to play an important role in plant defense reactions against fungi and insects which contain chitin in their cell walls and exoskeleton [4,5].

Both endochitinases and lysozymes catalyze the hydrolysis of β-1,4 glycosidic bonds. The main substrate of lysozyme are polysaccharide chains made up of alternating β-1,4-linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramate (MurNAc) residues, but some lysozymes are also active on chitin, a homopolymer of β-1,4-linked *N*-acetylglucosamine. Similarly, lysozyme activity has been demonstrated for some chitinases [6–8]. The difference between the two sugars is that MurNAc is an ether of GlcNAc with lactic acid.

From the sequence point of view (Table 1), chitinases are classified either into family number 18 or 19 of glycosyl hydrolases [9]. Family 19 contains both vacuolar and secreted plant enzymes. Family 18 includes extracellular chitinases from higher plants as well as bacteria, yeast and fungi. Lysozymes fall into several distinct sequence families. Because of structural similarities between g-, c- and phage-type lysozymes (families no. 22–24 [9]) it has been proposed that these enzymes are related by divergent protein evolution [10].

Here, we report a striking structural resemblance between the endochitinase from barley seeds [11] and lysozymes. The similarity places the endochitinases of family 19 in a common superfamily with animal and phage lysozymes, and reveals the active site of the endochitinase in a way not seen from the crystal structure alone.

2. Methods and databases

Pairwise structural alignments were generated automatically using the program Dali [12]. The method finds an optimal set of structurally equivalent residues by maximizing the similarities of intramolecular C^{α}–C^{α} distances within the common core. The structure of barley endochitinase was compared against a database of 330 representative protein structures with lower than 30% pairwise sequence identity [13]. Atomic co-ordinates were obtained from the Protein Data Bank (October 1993) [14], except for goose lysozyme [10]. Multiple sequence alignments were generated using the program MaxHom [15] and the Swiss-Prot 27 (October 1993) protein sequence database [16]. For example, the family alignment of barley-type endochitinases consisted of 27 sequences (Table 1).

*Corresponding author. Fax: (49) (6221) 387 517.

SSDI 0014-5793(94)00097-F

3. Results and discussion

T4 and human lysozyme were identified as top scoring hits in the structure database search against barley endochitinase (Fig. 1). Not only is the general $\alpha + \beta$ architecture of the endochitinase fold similar to that of lysozymes but four α-helices and a 3-stranded β-sheet are topologically identical and closely superimposable in 3D (Fig. 2). These elements make up the common structural core of g-type, c-type and T4 lysozymes (Fig. 3).

Unexpectedly, the structural similarity leads to the identification of the catalytic residue in endochitinase: the side chain of Glu^{67} coincides exactly in 3D with the essential catalytic glutamic acid of lysozymes (Glu^{35} in human lysozyme). Glu^{67} is invariant among relatives of barley endochitinase (with one exception, a deduced amino acid sequence from stinging nettle [17]). Further invariant residues cluster around the proposed active site of endochitinase. For example, Thr^{68} and Gln^{118} make hydrogen bonds to the protein backbone. They appear to play a similar role in maintaining the geometry of the active site as do strongly conserved Ser and Gln residues, respectively, in g- and c-type lysozymes (Ser^{36} and Gln^{58} in human lysozyme). Apart from the catalytic glutamic acid, only one glycine residue is invariant in all lysozymes (Fig. 3). This glycine (Gly^{55} in human lysozyme), located in a β-turn near the active site, is structurally equivalent to the invariant Gly^{115} in endochitinase and has a positive φ angle in the known structures.

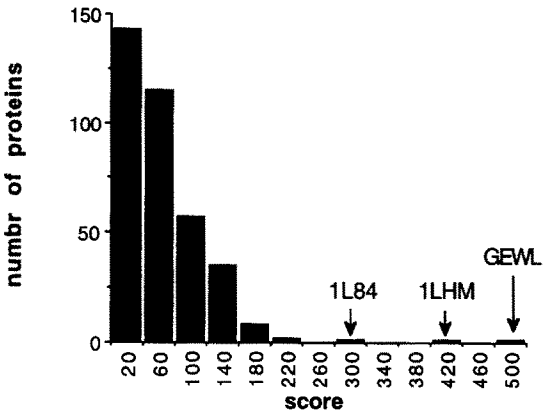

Fig. 1. Histogram of structural similarity scores for the pairwise comparisons of barley endochitinase against a representative structure database. Sequential alignment of structurally equivalent segments was required. Lysozymes from animal and phage score highest (GEWL, goose lysozyme; 1LHM, human lysozyme; 1L84, T4 lysozyme). The next-best scoring proteins, with different overall fold than in endochitinase, are colicin A (1COL) and thermolysin (4TLN).

Table 1
Classification of lysozyme and chitinase sequences. Sequences annotated as lysozymes or chitinases in the Swissprot 27 database are grouped according to sequence similarity, conserved sequence patterns, and structure similarity (part A).

(A) Lysosyme/chitinase superfamily defined by structural similarity

No. of known 3D structures	Sequence family	Family number**	No. of homologous sequences	Representative***
1	g-type (goose) lysozyme	23	4	lyg_ansan
6 + 1	c-type (chicken) lysozyme + α-lactal-bumin	22	39# + 15	lyc_chick, lca_papcy
1	phage (T4) lysozymes	24	8	lycv_bpph2
1	plant endochitinases (class I + II)	19	27	chi2_horvu

(B) Other lysozyme/chitinase families

No. of known 3D structures*	Sequence family	Family number**	No. of homologous sequences	Representative***
1	lysozyme M1	25	2	lysm_strgl
–	phage SF6 lysozyme	–	2	ly_bpsf6
–	*B. subtilis* lysozyme	–	1	lyb_bacsu
–	T3 lysozyme	–	1	naaa_bpt3
–	phage T4 tail lysozyme	–	1	vg25_bpt4
–	plant, yeast and bacterial endochitinases (class III)	18	12§	chly_hevbr
			10§	chit_strpl
			1§	chit_sacer
			1§	ktxa_klula

* Excluding mutants
** According to the classification of glycosyl hydrolases [9]
*** Swissprot identifier. The homologous sequences have significant sequence identity to the representative as defined by the significance threshold of Sander and Schneider [15]
# Excluding isozymes. The total number of known lysozyme c sequences is close to 60.
§ The bacterial sequences have less than 25% sequence identity with the eukaryotic sequences. The family is identified by a conserved sequence pattern [23].

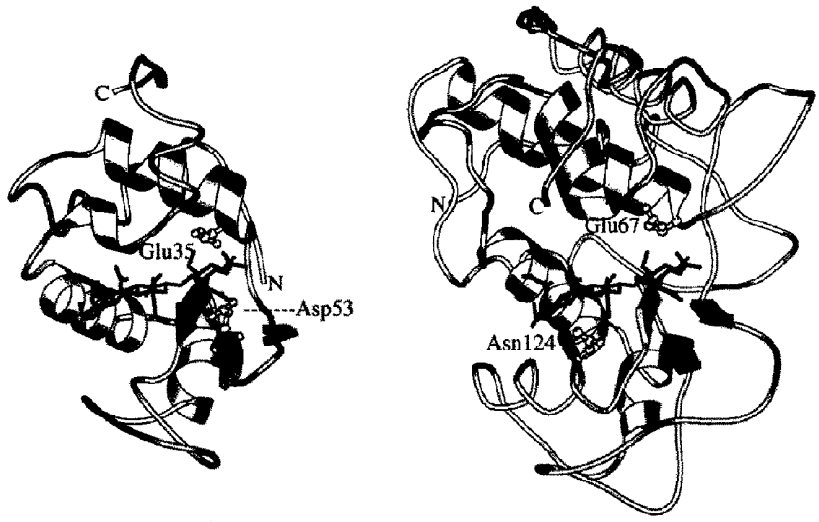

Fig. 2. MolScript [20] drawings of human lysozyme (left, Protein Data Bank entry 1LHM) and barley endochitinase (right, entry 1BAA). The structurally equivalent residues are (1LHM/1BAA): 4–7/10–13, 12–16/34–38, 19–22/43–46, 24–36/56–68, 37–40/83–86, 42–51/87–96, 52–65/112–125, 75–78/126–129, 80–86/139–145, 88–102/146–160, 103–106/165–168, 107–110/196–199, 113–117/212–216, 121–129/219–227. A trisaccharide inhibitor of lysozyme (black) was transferred from the cocrystal with human lysozyme (entry 9LYZ) to an equivalent position in the structure of endochitinase. Ribbons highlight helices and strands in the common core of c-type (human), g-type [10] and T4 lysozymes (1L84), and the structurally equivalent elements in endochitinase. *Human lysozyme:* Glu^{35} is the catalytic residue. Asp^{53} is also implicated in the catalytic mechanism of human lysozyme but is not conserved in the other types of lysozymes. *Barley endochitinase:* Glu^{67} is the proposed catalytic residue. Asn^{124} is invariant and could hydrogen bond to substrate. The other invariant residues in the barley-type endochitinase family are buried or apolar.

Surprisingly, endochitinase is structurally more closely related to g-type lysozyme than are either c-type or T4 lysozymes (Fig. 4). The combined conservation of fold, function and even some sequence and architectural detail strongly suggests that all four families descend from a common ancestor. Convergent evolution appears

```
RTDCYGNVNRIDTTGASCKTAKPEGLSYCGVSASKKIAGRDLQAMD.....EYKTLIK.KVGEKL.CV.....  61 GEWL
.....
.....
.....434.....4554455.315204.23.....
KVF.....ERa.....ELARTLKRLGMDGYRGI.....  23 1LHM
.....405444.....0323422.332344.1.....
SVSSIVSRAQFD.....RMLLRRNDGAaQAKGFYTYDAFV.AAAAAF.PGFGTTGSADA  52 1BAA
                                  *   **      *       *

...EPAVIAGIISRES.....HAGKVLKNGWGD...R....G.....NGFGLMQVDK  98 GEWL
....022443454554.....3433...3....1.....6553.53666
....MNIFEMLRIDEG.....LRL.....KIYKDTEG....Y.....YTIG.IGHLL  33 1L84
...4455555566666.....666....14653...5....0.....6765666666
...SLANWMbLAKWES.....GYNT...RATNY...NAGDRS.....TDYGIFQINS  61 1LHM
...2233355455666.....4642....3555...3.....4555546655
QKREVAAFLAQTSHETTGGWATAPDGAFAWGYaF....KQER...G....ASSDYDTPSAQWPbAPGKRYYGRGPIQLSH 121 1BAA
   * **       **          *                          *  *** *

RSHK.....PQGTWN..GEVHITQGTTILINEI.TIQKKF.....PSWTKD.....QQLKGGIS 144 GEWL
6.....011..6667666766665663.43222.....1343322.....21233345
TKSPSLNAAKSELDKAIGRNCNGVIT..KDEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAV.....RRCALINM 102 1L84
6545.....30413...6456435533533514.....234
RYWcNDGKTPGAVNAdHLS.cSALLQDNIADAVAdAKRVVRDPQ.....GIR 107 1LHM
3365.....0423...54545545555445.....2023.....32145333
NYNYGPAGRAIGVDLLANP..DLVATD.ATVGFKTAIWFWMTAQ.PPKPSSHAVIA.GQWSPSGADRAAGRVPGFGVITN 195 1BAA
  **          *   *       *

AYNAGAGNVR.SYARMDIGT.....TH....DDYANDVVAR.AQYYKQHGY..... 184 GEWL
6656655543.1-13.....06....211.14002.....
VFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQT....PNRAKRVITT.FRTGTWDAYK..... 162 1L84
334.....143542...24643523.....
AWVA.....WRNRbQNRDVRQYVQGaGV..... 130 1LHM
5531.....35.....-....2322455456.45654543.....
IING.....GI.....EcGHGQDSRVADRIGE.YKRYCDILGVGYGNNLDcYSQRPFA 243 1BAA
                                                *
```

Fig. 3. Structural alignments. The goose enzyme (GEWL) was used as a reference structure against which human lysozyme (1LHM), T4 lysozyme (1L84) and barley endochitinase (1BAA) were aligned. The numbers refer to the strength of similarity in structural environments in the pairwise alignment with GEWL, on a scale from 0 (least similar) to 9 (most similar). Dots denote gaps and trailing ends. Asterisks below the barley endochitinase sequence mark invariant residues in the barley-type endochitinase family. The secondary structure [21] is shown as helix and β-strands. Lowercase characters are disulphides (a with a, b with b, and so forth).

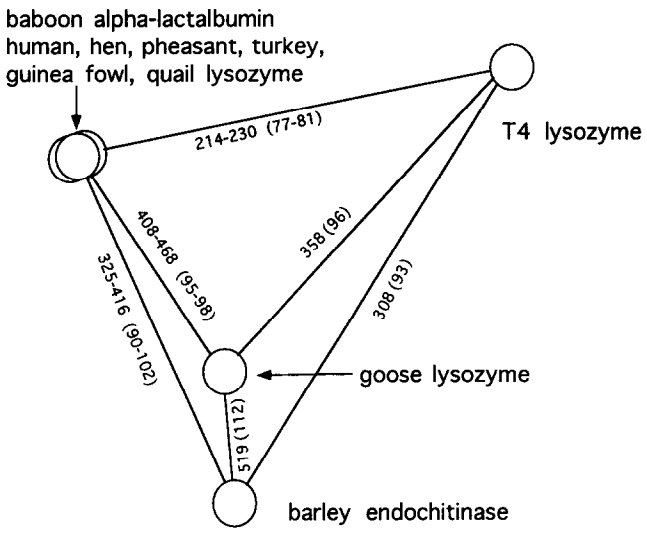

Fig. 4. Overall structural relations between lysozymes and endochitinase. The similarity score and the number of structurally equivalent residues (in parentheses) are given for each pair (smallest – largest value for cluster at top left). The planar projection was generated from the table of pairwise similarity scores by a multivariate analysis method [22]. C-type lysozymes and α-lactalbumin have essentially identical structures (>30% sequence identity) and cluster together at the top left (1LHM, 1GHL, 1HHL, 1HEL, 2IHL and 1ALC). Similarity scores within the c-type family are 1090–1464. Pairwise sequence identities between the families of c-type, g-type or T4 lysozyme and barley endochitinase are below 15%.

improbable. The structural elements making up the active site are colinearly arranged in the amino acid sequences. These elements come together in 3D in spite of massive insertions/deletions in the intervening sequences. As at least one alternative structural solution for building the same functionality is known to exist (the lysozyme from *Streptomyces erythraeus* [18], family 25 in Table 1), it is difficult to envisage why one particular fold should arise independently as many as four times. The wide taxonomic distribution and vanishing sequence similarity indicate an ancient origin of the lysozyme/endochitinase superfamily, which could account for the genomic reorganization of c- and g-type genes in chicken [19].

Endochitinases are of interest in the engineering of crops with improved resistance to fungi, and the present analysis may be useful in directing mutagenesis experiments. When more high-resolution 3D structures become available, it will be interesting to see whether additional protein families bear any similarity to the present lysozyme/endochitinase superfamily.

References

[1] Stryer, L. (1988) Biochemistry, Freeman, New York.
[2] Matsumura, M., Wozniak, J.A., Sun, D.P. and Matthews, B.W. (1989) J. Biol. Chem. 264, 16–59–16066.
[3] Rennell, D., Bouvier, S.E., Hardy, L.W. and Poteete, A. (1991) J. Mol. Biol. 222, 67–88.
[4] Schlumbaum, A., Mauch, F., Vögeli, U. and Boller, T. (1986) Nature 324, 365–367.
[5] Eyal, Y. and Fluhr, R. (1991) Oxford Surveys of Plant Molecular and Cell Biology 7, 223–254.
[6] Boller, T. Gehri, A., Mauch, F. and Vögeli, U. (1983) Planta 157, 22–31.
[7] Jekel, P.A., Hartmann, B.H. and Beintema, J.J. (1991) Eur. J. Biochem. 200, 123–130.
[8] Chang, C.T., Lo, H.F., Wu, C.J. and Sung, H.Y. (1992) Biochem. Int. 28, 707–715.
[9] Henrissat, B. (1991) Biochem. J. 280, 309–316.
[10] Weaver, L.H., Grütter, M.G., Remington, S.J., Gray, T.M., Isaacs, N.W. and Matthews, B.W. (1985) J. Mol. Evol. 21, 97–111.
[11] Hart, P.J., Monzingo, A.F., Ready, M.P., Ernst, S.R. and Robertus, J.D. (1993) J. Mol. Biol. 229, 189–193.
[12] Holm, L. and Sander, C. (1993) J. Mol. Biol. 233, 123–138.
[13] Hobohm, U., Scharf, M., Schneider, R. and Sander, C. (1992) Protein Sci. 1, 409–417.
[14] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[15] Sander, C. and Schneider, R. (1991) Proteins 9, 56–68.
[16] Bairoch, A. and Boeckmann, B. (1992) Nucleic Acids Res. 20, 2019–2022.
[17] Lerner, D.R. and Raikhel, N.V. (1992) J. Biol. Chem. 267, 11085–11091; correction 22694.
[18] Harada, S., Sarma, R., Kakudo, M., Hara, S. and Ikenaka, T. (1981) J. Biol. Chem. 256, 11600–11602.
[19] Nakano, T. and Graf, T. (1991) Biochim. Biophys. Acta 1090, 273–276.
[20] Kraulis, P. (1991) J. Appl. Cryst. 24, 946–950.
[21] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[22] Hill, M.O. (1973) J. Ecol. 61, 237–251.
[23] Henrissat, B. (1990) Protein Seq. Data Anal. 3, 523–526.