Crystal structure of human Charcot–Leyden crystal protein, an eosinophil lysophospholipase, identifies it as a new member of the carbohydrate-binding family of galectins

Demetrios D Leonidas[1], Boris L Elbert[2], Zeqi Zhou[2], Hakon Leffler[3], Steven J Ackerman[2] and K Ravi Acharya[1]*

[1]School of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK, [2]Division of Infectious Diseases, Department of Medicine, Beth Israel Hospital and Harvard Medical School, Boston, MA 02215, USA and [3]Center for Neurobiology and Psychiatry, Departments of Psychiatry and Pharmaceutical Chemistry, University of California, California, CA 94143, USA

Background: The Charcot–Leyden crystal (CLC) protein is a major autocrystallizing constituent of human eosinophils and basophils, comprising ~10% of the total cellular protein in these granulocytes. Identification of the distinctive hexagonal bipyramidal crystals of CLC protein in body fluids and secretions has long been considered a hallmark of eosinophil-associated allergic inflammation. Although CLC protein possesses lysophospholipase activity, its role(s) in eosinophil or basophil function or associated inflammatory responses has remained speculative.

Results: The crystal structure of the CLC protein has been determined at 1.8 Å resolution using X-ray crystallography. The overall structural fold of CLC protein is highly similar to that of galectins -1 and -2, members of an animal lectin family formerly classified as S-type or S-Lac (soluble lactose-binding) lectins. This is the first structure of an eosinophil protein to be determined and the highest resolution structure so far determined for any member of the galectin family.

Conclusions: The CLC protein structure possesses a carbohydrate-recognition domain comprising most, but not all, of the carbohydrate-binding residues that are conserved among the galectins. The protein exhibits specific (albeit weak) carbohydrate-binding activity for simple saccharides including *N*-acetyl-D-glucosamine and lactose. Despite CLC protein having no significant sequence or structural similarities to other lysophospholipases or lipolytic enzymes, a possible lysophospholipase catalytic triad has also been identified within the CLC structure, making it a unique dual-function polypeptide. These structural findings suggest a potential intracellular and/or extracellular role(s) for the galectin-associated activities of CLC protein in eosinophil and basophil function in allergic diseases and inflammation.

Structure 15 December 1995, **3**:1379–1393

Key words: Charcot–Leyden crystal protein, basophil, eosinophil, galectin, lysophospholipase, X-ray crystallography

Introduction

Charcot–Leyden crystals (CLCs) were first described in Paris in 1853 by Charcot and Robin in the post-mortem spleen and heart blood of a leukemia patient [1]. A later report by Leyden described them in the sputum of asthmatics [2], ultimately resulting in their designation as Charcot–Leyden crystals. These distinctive hexagonal, bipyramidal crystals have classically been observed in tissues and secretions from sites of eosinophil-associated inflammatory and related allergic immune reactions in asthma, myeloid leukaemias, allergic, parasitic and other eosinophil-associated diseases [3,4]. Although CLCs are considered a hallmark of the eosinophil, basophils also form CLCs [5], and CLCs have been identified ultrastructurally as intragranular inclusions in the large histamine-containing, particle-filled granule of the human blood basophil and basophils that participate in a variety of inflammatory tissue reactions [6,7].

Eosinophil CLCs are formed by a small (16.5 kDa) markedly hydrophobic polypeptide (CLC protein) [8], comprising 142 amino acids [9,10], that exhibits lysophospholipase (LPLase) activity (lysolecithin acylhydrolase EC 3.1.1.5) [11,12]. CLC protein is indistinguishable from chromatographically purified human eosinophil LPLase and both proteins crystallize to form morphologically identical hexagonal, bipyramidal crystals [12]. The cDNA for CLC protein directs CLC crystallization and the expression of LPLase activity in transiently transfected COS and CHO cells [10,11]. CLC protein is one of the most abundant constituents of the eosinophil [5, 8], comprising an estimated 7–10% of total cellular protein in this granulocyte [12], an amount comparable with the eosinophil's content of secondary granule cationic toxins and enzymes such as the major basic protein and eosinophil peroxidase [13].

The amino-acid sequence of CLC protein shows essentially no significant similarities with available sequences of prokaryotic or eukaryotic LPLases or other lipolytic enzymes [9,10], including LPLases of rat [14] and bovine [15] origin as well as a series of sulphydryl-dependent

*Corresponding author.

LPLases of mouse macrophage [16] and human myeloid cell (HL-60) origins [17] that have been partially sequenced. In contrast, the CLC protein shows limited amino-acid-sequence identity to the family of animal galectins [9,10], previously known as S-type or S-Lac (soluble lactose-binding) lectins [18,19]. So far, eight mammalian galectins, designated galectin-1 to -8, and many non-mammalian galectins, have been reported. The galectins are defined by their affinity for β-galactosides and twelve highly conserved residues in their sequence that comprise their carbohydrate-recognition domain. Most of the conserved residues are involved in interaction with bound carbohydrate, as shown by X-ray crystallography of galectins -1 and -2 [20–22] and site-directed mutagenesis [23–25]. The CLC protein sequence contains seven out of the twelve conserved residues and in addition shares many other residues with the mammalian galectins [9,10]. It is most similar to galectins -3 and -4 (25–30% overall identity) and least similar to galectins -1 and -2 (15–20% overall identity).

In addition to these sequence similarities, CLC protein has some functional similarities to the galectins. Like the galectins, CLC protein is predominantly found as an abundant cytosolic protein and lacks a secretion signal peptide, but may under certain conditions be released extracellularly [26–28]. Regulated non-classical secretion has been demonstrated for galectins [29–32] as well as the CLC protein [33,34]. In addition, like galectins [35], the CLC protein is also found in the nucleus of both eosinophils and basophils [6,11,36,37]. The biological roles of CLC protein and galectins remain unknown, but both CLC protein and galectin-3 have been proposed to be involved in allergic immunological responses. Galectin-3 has been independently identified as the IgE binding protein εBP, the laminin- and IgE-binding macrophage surface antigen Mac-2, as well as the carbohydrate binding proteins CBP35, L-34 and L-29 [18, 19]. Galectin-3 is expressed in eosinophils, neutrophils, macrophages and epithelial cells. It has been suggested that it may mediate the interaction of IgE with neutrophils and eosinophils [38,39] and can act to induce a signalling response in various leukocytes [40–42]. Intracellularly, galectin-3 has been identified as a factor involved in pre-mRNA splicing in HeLa cell nuclear extracts [35]. These findings suggest that CLC protein should possess galectin-like carbohydrate-binding activities and possibly other galectin-like properties that might be extremely pertinent to eosinophil and basophil function in allergic inflammation, especially taking into consideration the major cellular content and subcellular localization of this enigmatic autocrystallizing polypeptide. As very few prokaryotic or eukaryotic proteins undergo autocrystallization intracellularly or extracellularly in a physiological setting, the ability of CLC to form hexagonal bipyramidal crystals is perhaps the most intriguing of its physicochemical properties. Although the physicochemical or structural properties of CLC protein responsible for autocrystallization, and the precise biological role that this plays, have not been determined, the LPLase activity of CLC protein is inactivated by crystallization [10].

In this report, we present a detailed structural analysis of CLC protein to 1.8 Å resolution. Despite minimal overall sequence conservation, the CLC protein structural topology is remarkably similar to that of galectins -1 and -2. In addition, we demonstrate functional, albeit weak, binding of CLC protein to lactose and *N*-acetyl-D-glucosamine, and identify structural differences in the carbohydrate-recognition domain of CLC protein, relative to other galectins, that are likely to be responsible for the low affinity for these saccharides. We also highlight the features of a putative LPLase active site, making CLC a unique protein with a dual functional role.

Results

The crystal structure of CLC protein was determined by the multiple isomorphous replacement method (see the Materials and methods section; Table 1). The experimentally derived phases were used to build a model for the CLC protein structure and refined against the data to 1.8 Å resolution to a final crystallographic R value of 20.0%. In general, the electron-density map is of high quality (Fig. 1). The entire structure is well defined and rigid except for the C-terminal residue, the main chain of Tyr139 and the terminal portions of the side chains Lys23, Arg25, Lys45, Arg60, Gln74, Gln87, Lys134 and Lys141, all of which have high temperature factors and are disordered.

Table 1. Statistics for crystallographic structure determination.

Crystal	Native	*p*-CMBS	K_2PtCl_4
Internal scaling			
Resolution (Å)	40–1.8	45–3.4	45–3.4
Number of crystals	2	1	1
Reflections measured	75363	14557	9526
Unique reflections	18046	3130	3074
Completeness (%)	94.4	99.1	96.6
(outermost shell)	(89.2)	(98.0)	(93.8)
I/σI	9.1	10.0	9.1
R_{sym}* (%)	7.7	6.2	6.5
Derivative scaling			
Resolution (Å)		15–3.4	15–3.4
Mean fractional isomorphous difference (%)		32.4	35.4
Heavy-atom refinement			
Phasing power†		2.42	1.55
Number of sites		2	2
Coordinating residues		Cys29,Cys57	Met44,Met103
R_{cullis}‡ (%)		37.0	52.7
R_{Kraut}§ (%)		15.8	20.6
Figure of merit#		0.65	0.46

*$R_{sym}=\Sigma_h\Sigma_i|I(h)-I_i(h)|/\Sigma h\Sigma_i I_i(h)$, where $I_i(h)$ and I(h) are the ith and the mean measurements of the intensity of reflection h. †Phasing Power is the ratio of the rms calculated heavy-atom structure amplitude to the rms lack of closure. ‡$R_{cullis}=\Sigma||F_{PH}|_{obs}\pm|F_P|_{obs}|-|F_H|_{calc}|/\Sigma||F_{PH}|_{obs}\pm|F_P|_{obs}|$, where F_{PH} and F_P are the observed structure factor amplitudes for the heavy-atom derivative and the native data sets, with the sum taken over all centric reflections, and F_H is the heavy-atom structure factor. §$R_{Kraut}=\Sigma||F_{PH}|_{obs}-|F_{PH}|_{calc}|/\Sigma|F_{PH}|_{obs}$, with the sum taken over all acentric reflections. #The figure of merit after phase combination is 0.60.

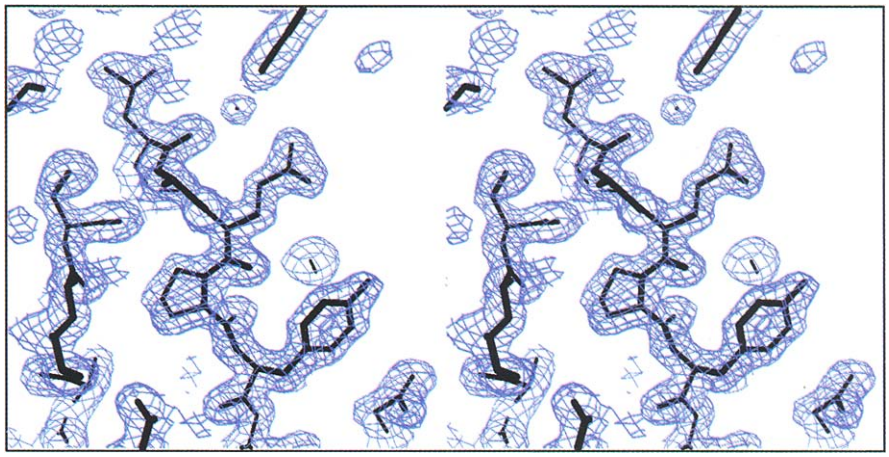

Fig. 1. Stereoview of a portion of the final 1.8 Å $2|F_o|-|F_c|$ electron-density map for CLC protein. The map is contoured at 1.0σ level. (The figure was generated using the program O [68].)

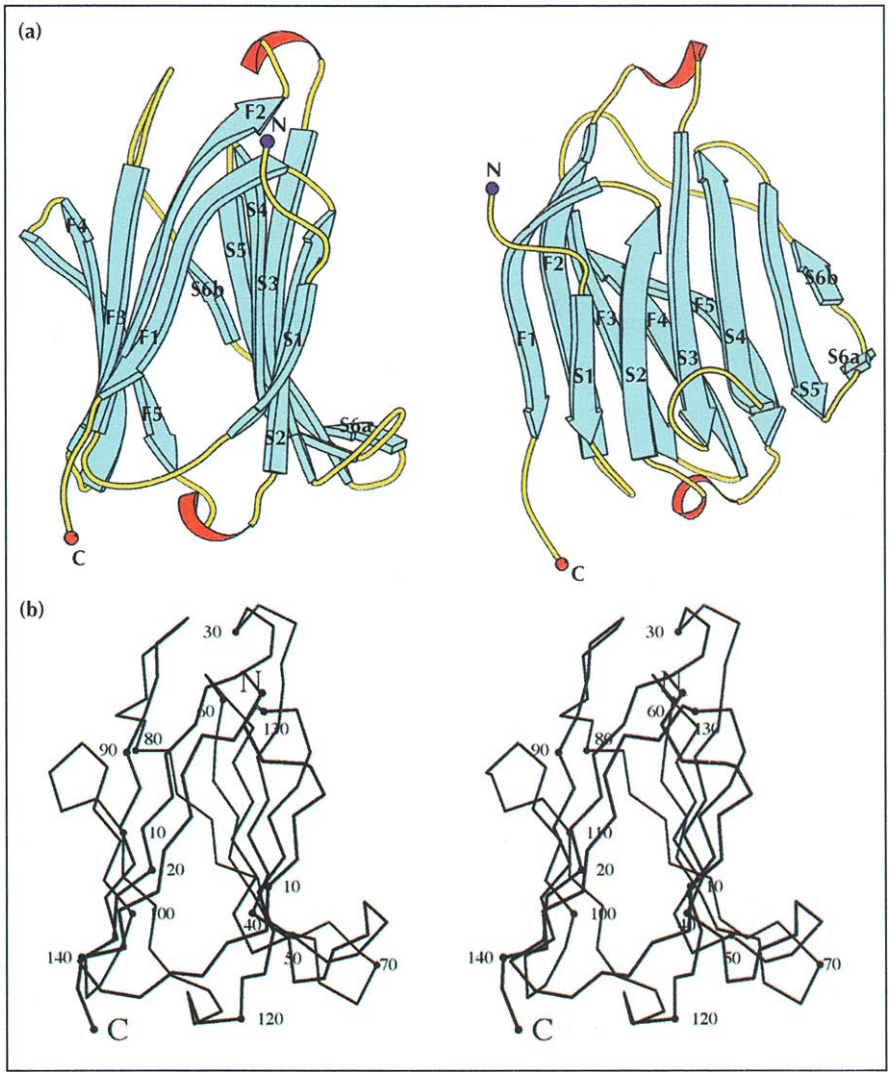

Fig. 2. The three-dimensional structure of human CLC protein. **(a)** Orthogonal views of CLC protein structure. β strands are presented in cyan, the two small 3_{10} helices in red and the loops in yellow. **(b)** Stereoview of the Cα backbone of CLC protein molecule. (The diagrams were drawn using MOLSCRIPT [74].)

Structure description

CLC protein is a compact molecule with approximate monomer dimensions of 43 Å×35 Å×31 Å (Fig. 2). The topological motif consists of a 'jelly-roll', which results from tight association between a five-stranded (F1–F5) and a six-stranded (S1–S6a/S6b) β sheet, as described for galectins -1 and -2 [20–22]. The molecule has very little loop structure, with the majority of residues adopting β conformation. Two short 3_{10} helices (one situated at the top and the other at the bottom in Fig. 2b) connecting β strands F2–S3 and F5–S2, respectively, form part of the decorations of the molecule. The last residue at the C terminus is disordered and is modelled as alanine in the refined structure. One peptide bond connecting residues Val6 and Pro7 adopts a *cis* conformation.

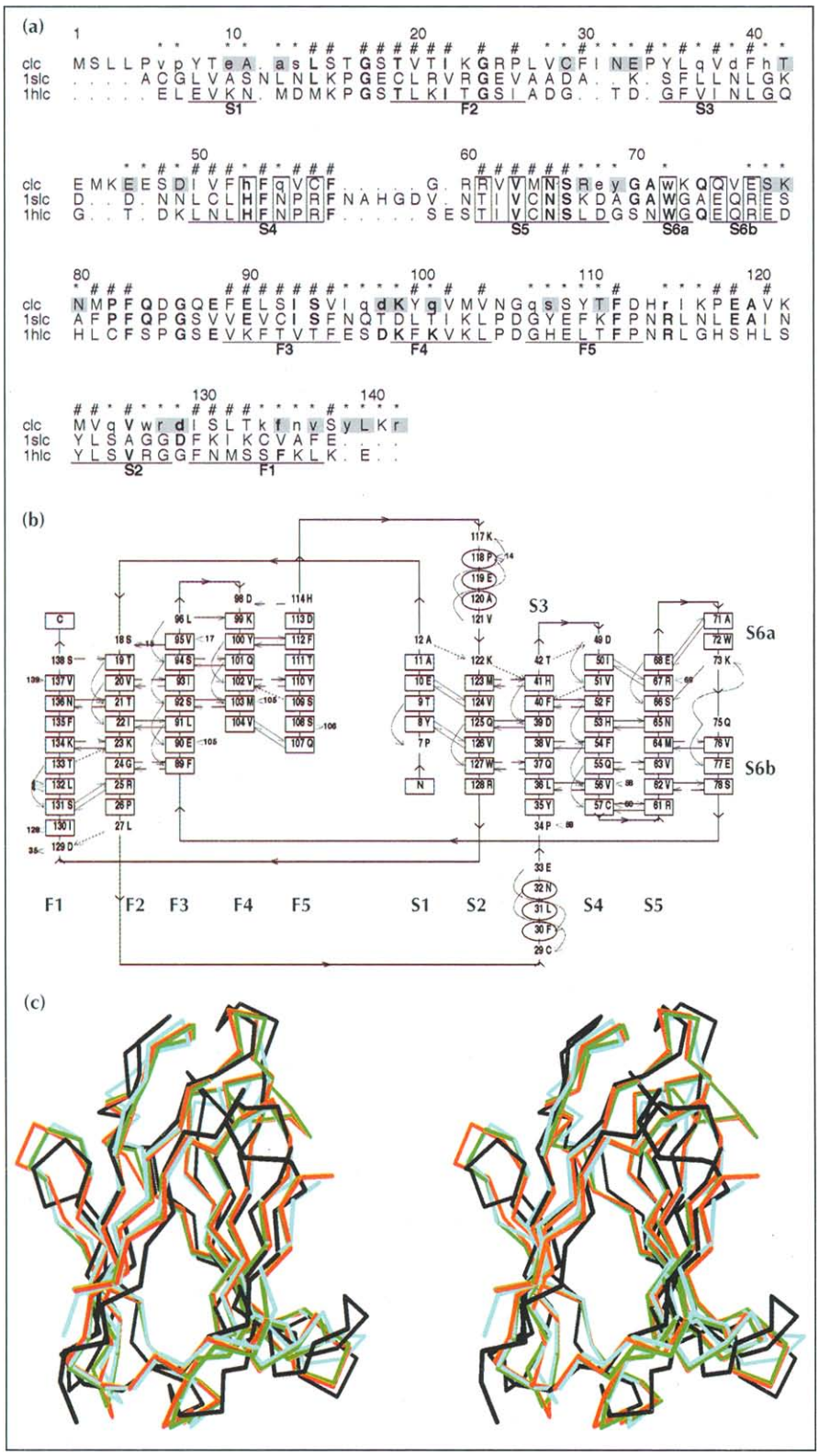

Fig. 3. Structural details of CLC protein. **(a)** Structure-based sequence alignment of CLC protein, bovine heart galectin-1 (1SLC) and human galectin-2 (1HLC) as determined using the program SHP [72]. Every tenth residue in the CLC sequence is numbered. Residues that form the carbohydrate-binding site are boxed. The dots indicate gaps inserted for optimal alignment. In CLC protein the solvent inaccessible residues (less than 20 $Å^2$ of exposed surface as calculated using the program DSSP [71]) are indicated by # and residues involved in crystal packing contacts are marked by *. Polar interactions are indicated by lower case letters and water mediated interactions ≤4.0 Å are highlighted in grey boxes. Secondary structural elements (β strands F1–F5, S1–S6b) as determined by DSSP [71] are also shown. Residues which are identical between CLC protein and galectins are highlighted in bold letters. **(b)** Diagram of hydrogen bonds involving main-chain atoms in the CLC protein structure. Squares indicate residues in β strands and circles indicate residues in 3_{10} turns. Arrows represent hydrogen bonds. (Figure drawn using the program HERA [75].) **(c)** Stereoviews of the structural alignment of CLC protein (black) on bovine heart galectin-1 (1SLC, red), bovine spleen galectin-1 (1SLT, green) and human galectin-2 (1HLC, cyan). The high degree of structural conservation is apparent.

A structure-based sequence alignment of CLC protein with known galectin structures -1 and -2 (Fig. 3a) indicates which CLC protein residues are either solvent inaccessible or involved in crystal contacts. Figure 3b indicates the detailed hydrogen bond network between the strands within the CLC protein molecule. The close structural resemblance between galectins and CLC protein is shown in Figure 3c; the regions that deviate significantly are the loop regions connecting the strands. In a structural comparison of CLC protein and galectin-1, excluding these loops reduces the root mean square (rms) deviation from 1.91 Å to 1.19 Å.

Table 2. Structural superposition of different protein structures that possess similar topology to the CLC protein.

	CLC	1SLT	1SLC	1HLC	WING1	1LTE	1SAC	2AYH	2LTN	WING2	2PHL	1PNG
	1.91											
1SLT	1.91											
	24/127											
	1.91	**0.50**										
1SLC	1.91	0.50										
	24/127	*133/133*										
	2.01	**0.74**	**1.03**									
1HLC	2.01	0.74	1.03									
	24/128	*55/128*	*55/128*									
	2.60	**3.35**	**3.05**	**4.46**								
WING1	2.51	3.01	2.45	3.30								
	11/112	*9/113*	*7/112*	*2/81*								
	3.54	**2.71**	**2.81**	**3.28**	**3.62**							
1LTE	3.25	2.92	3.11	2.97	2.45							
	11/101	*17/124*	*14/123*	*6/119*	*6/157*							
	3.12	**3.18**	**3.16**	**3.32**	**3.27**	**3.37**						
1SAC	3.00	3.18	3.20	3.26	3.47	2.97						
	9/100	*12/115*	*8/93*	*4/95*	*11/134*	*12/142*						
	3.26	**3.10**	**4.55**	**2.76**	**3.14**	**3.15**	**3.41**					
2AYH	2.88	2.85	2.89	2.91	2.55	2.76	2.71					
	7/121	*12/121*	*9/92*	*13/117*	*19/56*	*12/172*	*7/134*					
	3.54	**2.88**	**2.96**	**3.26**	**3.29**	**2.92**	**3.39**	**2.96**				
2LTN	3.25	3.11	2.81	3.12	2.92	1.56	3.36	3.29				
	11/101	*14/127*	*15/128*	*9/121*	*7/121*	*86/178*	*5/97*	*8/128*				
	4.04	**2.90**	**3.14**	**2.90**	**1.53**	**2.84**	**3.70**	**3.32**	**2.83**			
WING2	2.73	2.90	3.97	2.90	1.45	2.71	3.57	3.28	2.83			
	6/90	*9/112*	*7/113*	*11/111*	*33/145*	*9/130*	*10/151*	*13/129*	*5/94*			
	4.25	**3.87**	**5.46**	**4.44**	**4.60**	**4.29**	**4.52**	**5.01**	**4.96**	**5.87**		
2PHL	4.58	4.35	3.92	3.53	4.88	4.94	4.01	4.18	4.22	5.25		
	4/56	*9/63*	*4/57*	*5/58*	*5/77*	*4/80*	*5/77*	*7/77*	*3/70*	*5/72*		
	4.27	**4.68**	**3.09**	**2.89**	**4.53**	**4.73**	**4.17**	**4.96**	**4.55**	**5.36**	**5.00**	
1PNG	3.92	3.26	3.00	4.31	3.30	3.83	3.56	3.69	3.27	4.16	4.20	
	4/55	*3/57*	*4/78*	*9/79*	*5/81*	*6/81*	*4/169*	*7/71*	*2/67*	*4/81*	*7/110*	
	4.14	**4.32**	**3.15**	**4.11**	**4.27**	**3.21**	**3.53**	**2.76**	**3.04**	**4.45**	**3.66**	**4.49**
1CEL	2.97	3.24	3.18	3.44	3.93	3.08	3.75	2.82	3.10	4.53	4.51	3.62
	5/88	*7/91*	*9/118*	*3/100*	*14/145*	*17/167*	*7/92*	*20/181*	*13/127*	*3/89*	*5/84*	*8/100*

Brookhaven Protein Data Bank accession numbering is used for each protein: bovine spleen galectin-1, 1SLT [21]; bovine heart galectin-1, 1SLC [20]; human galectin-2, 1HLC [22]; pea lectin, 2LTN [77]; coral tree lectin (EcorL), 1LTE [76]; serum amyloid P component (SAP), 1SAC [48]; Wing 1 and Wing 2 are the lectin-like domains of the *Vibrio cholerae* neuraminidase [50]; β-glucanase lichenase, 2AYH [78]; cellulase, 1CEL [47]; plant seed storage protein phaseolin, 2PHL [79]; and asparagine amidase, 1PNG [44]. First row (bold) corresponds to the overall rms distance between the proteins, the second row is the rms distance for the part of the structure which has the lectin-like topology, and third row shows the number of identical amino-acid residues over the number of equivalent residues between the two structures. Structural alignments were made using the SHP program [72].

Similarity to other lectin-like structures

A detailed structural comparison of CLC protein with known 'jelly-roll' structures is shown in Table 2 and Figure 4. It is clear from such an extensive comparison that CLC protein and galectins form part of a unique family of soluble calcium-independent carbohydrate-binding animal lectins and differ from cation-dependent carbohydrate-binding plant lectins. The 'jelly-roll' motif in the CLC protein structure is a common, core structural motif found in the diverse group of lectin molecules for which the structures are known so far. But, even if CLC protein and galectin-1 and -2 have the same fold as the legume lectins, they are much more similar in structure to each other than to these lectins. Furthermore, they show no sequence homology to the legume lectins or any of the other proteins with the same type of fold.

More recently, several new molecules possessing the 'jelly-roll' motif have been identified. In particular, the 3D structures of four enzymes have been elucidated: peptide-*N*-glycosidase F (glycosylasparaginase) [43,44], bacterial 1,3-1,4-β-glucanase [45], endo 1,4-β-xylanase II [46] and cellulase cellobiohydrolase I [47]. The carbohydrate-binding mode of these enzymes appears to have similar features to that of the legume lectins, that is, the carbohydrate molecules bind to the concave face of the β sheet with the loops surrounding the ligand.

Interestingly, three new proteins have been identified which possess lectin-like domains (Table 2, Fig. 4) and are relevant to our discussion: human serum amyloid P component protein [48], which has lectin activity and binds to DNA and amyloid fibrils; human C-reactive

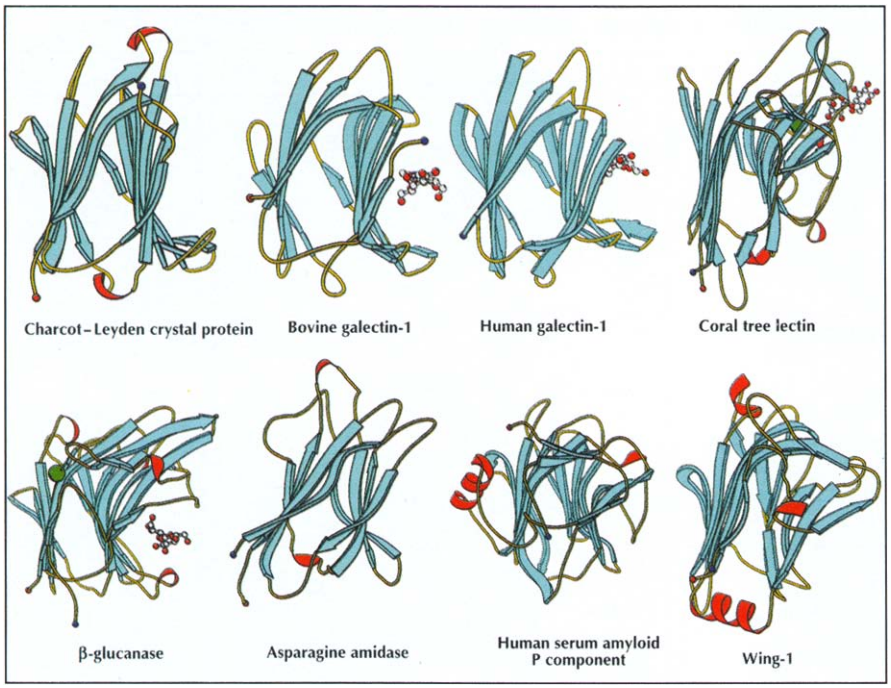

Fig. 4. The lectin domain in CLC protein. This domain is compared with bovine spleen galectin-1 complexed with *N*-acetyl-lactosamine [21]; human galectin-2 complexed lactose [22]; coral tree lectin complexed with lactose [76]; β-D-glucanase complexed with 3,4-epoxy-butyl-β-D-cellobioside [45]; the N-terminal domain of asparagine amidase [44] ; Human serum amyloid P component [48] and Wing-1 from *Vibrio cholerae* neuraminidase [50]. Bound metal ions are shown as green spheres. The N and C termini are shown as blue and red spheres, respectively. The coordinates of the lectin domains of the various proteins have been transformed after superposition onto CLC protein using the program SHP [72].

protein [49], which is an acute phase reactant and has been identified as a galactose-specific particle receptor on liver macrophages, and lectin-like domains (Wings 1 and 2) attached to the catalytic domain in *Vibrio cholerae* neuraminidase [50]. It has not yet been established whether these wings possess any lectin activity, but it has been suggested that they may be involved in promoting adhesion to epithelial cells. Comparison of these domains with CLC protein has revealed considerable similarity with the Wing 1 domain (Table 2). Thus, these proteins appear to possess elaborated forms of the 'jelly-roll' motif and belong to the ConA-like lectin superfamily of proteins, on the basis of structural classification using the SCOP database search [51].

Environment of cysteine residues

The CLC protein molecule contains two cysteine residues (Cys29 and Cys57). Both cysteines appear to be active as they bind mercury atoms, and from the electron-density map it was clear that both cysteines have a reactive sulphydryl group. Cys29 is situated at the beginning of a small 3_{10} helix connecting strands F2 and S3 and Cys57 is located at the end of strand S4. In the crystal structure Cys29 is accessible to solvent, whereas Cys57 is buried, and these cysteines are not involved in disulphide bond formation (they are at a separation of 12 Å). These residues are not involved in any strong crystal packing interactions and the closest atom to Cys29 is Oε1 of Glu46 from a symmetry related molecule (3.9 Å) and to Cys57 is Oη of Tyr35 from the same molecule (4.0 Å). These observations from the molecular structure raise questions as to why only a single reactive sulphydryl can be detected in the denatured protein using either 5,5′-dithiobis(2-nitrobenzoic acid) [8], *p*-chloromercuribenzoic or iodoacetic acids [12].

The carbohydrate-recognition site

On the basis of mutagenesis data for human galectin-1, Hirabayashi and Kasai [24] have suggested that the most conserved residues are located in the central region of the galectin amino-acid sequence. In particular, six hydrophilic residues are well conserved: His44, Asn46, Arg48, Asn61, Glu71 and Arg73. These residues, together with Trp68, form part of the carbohydrate-recognition site in galectins. More detailed pictures of galectin-1 complexed with biantennary saccharides [20] and *N*-acetyl-lactosamine [21], and galectin-2 in complex with lactose [22], have recently been obtained by X-ray crystallography. These studies reveal highly conserved carbohydrate-binding sites in the galectins, which differ from the cation-dependent site in legume lectins.

A high degree of structural resemblance, and amino-acid-sequence conservation between CLC protein and the galectins [9] (at residues which are considered to be important for carbohydrate binding in the galectins) suggested that CLC protein might possess a potential carbohydrate-recognition site. In order to examine the structural basis for these observations, we superimposed the CLC protein and galectin–saccharide complex structures (Fig. 5). The structural alignment (Table 3) highlights the details of this site. Indeed, in the CLC protein structure the conformation of this site is very similar to that observed in the galectins involving strands S4, S5, S6a and S6b on the concave face of one β sheet. In galectins -1 and -2 [20–22] this site consists of seven key residues: His44, Asn46, Arg48, Asn61, Trp68, Glu71 and Arg73, and the structurally equivalent residues in the CLC protein are His53, Gln55, Cys57, Asn65, Trp72, Gln75 and Glu77 (Figs 3,5). Examination of this site in these proteins reveals that there is a high degree of structural conservation on one side of the carbohydrate-binding

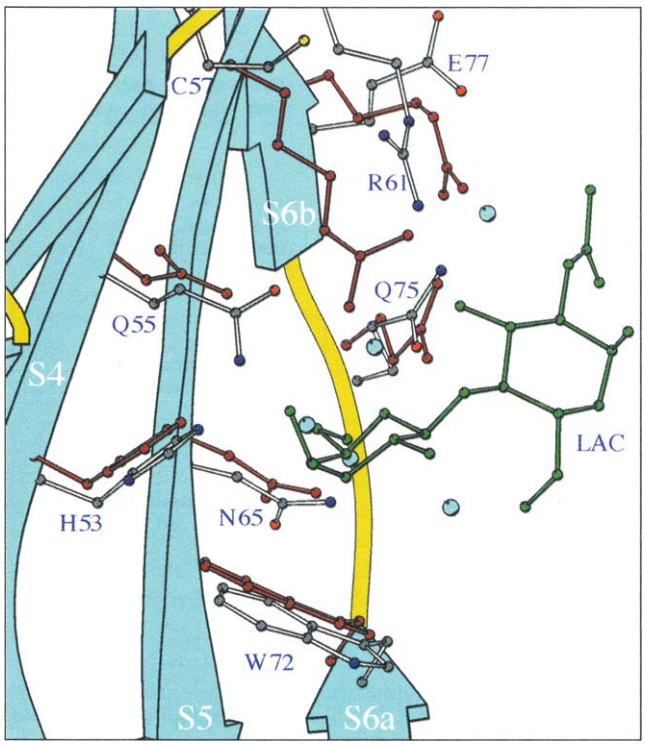

Fig. 5. Superposition of residues in CLC protein and bovine spleen galectin-1 in the vicinity of the carbohydrate-recognition site. The residue labels are from the CLC protein structure. The backbone is shown in cyan. The side chains of CLC protein are shown as coloured atoms and those of galectin-1 in brown. *N*-acetyl-lactosamine (LAC) from the galectin-1 structure is shown in green. Bound water molecules in the CLC structure at the carbohydrate-recognition domain are shown as cyan spheres. (The figure was generated using MOLSCRIPT [74].)

Table 3. Superimposition of the potential carbohydrate-binding site of CLC protein with the carbohydrate-binding site of the galectins.

		rms distance (Å)		
CLC	Galectins*	1SLT	1HLC†	1SLC‡
His53	His44	0.93	1.01	1.01
Gln55	Asn46	0.77	0.88	0.71
Cys57	Arg48	0.78	1.31	0.77
Asn65	Asn61	1.06	1.12	1.14
Trp72	Trp68	1.09	0.92	0.72
Gln75	Glu71	0.99	0.71	1.04
Glu77	Arg73	1.28	1.18	1.17

*The numbering refers to bovine spleen galectin-1(1SLT) [30]. Structural alignments have been performed with the program SHP [72]. †Human galectin-2. ‡Bovine-heart galectin-1.

groove with the crucial amino-acid changes Arg48 to Cys57 and Arg73 to Glu77 occurring on the other side in the CLC protein (Table 4). Moreover, in galectins -1 and -2 these two arginines are involved in important hydrogen-bonding interactions with the bound saccharide [20–22]. These arginines are conserved in all known galectin amino-acid sequences and are responsible for the tight association with carbohydrate ligands. Thus, there is an apparent loss of positive charge in the CLC protein

Table 4. Comparison with CLC of the polar interactions between galectins and various saccharides.

Sugar Atom	Bovine galectin-1	Human galectin-2	CLC
Gal O4	His44 (Nε2) 2.8	His45 (Nε2) 2.7	His53 (Nε2) 2.9
Gal O4	Arg48 (Nη2) 3.0	Arg49 (Nη1) 3.3	–
Gal O4	Asn46 (Nδ2) 3.4	Asn47 (Nδ2) 3.1	Gln55 (Nε2) 2.7
Gal O5	Arg48 (Nη2) 2.9	–	–
Gal O6	Asn61 (Nδ2) 2.7	Asn58 (Nδ2) 2.7	Asn65 (Nδ2) 3.4
Gal O6	Glu71 (Oε2) 2.8	Glu68 (Oε2) 2.6	Gln75 (Oε1) 2.9
Nag/Glc O3	Arg73 (Nη2) 3.3	Arg70 (Nη2) 3.1	–
Nag/Glc O3	Arg48 (Nη1) 2.8	–	Arg61 (Nη2) 2.9
Nag/Glc O3	Glu71 (Oε2) 2.4	Glu68 (Oε2) 2.8	Gln75 (Oε1) 2.8
Nag/Glc O3	Arg48 (Nη2) 3.2	–	–

Numbers are the distances in Å. Gal, Glc and Nag are the galactose, the glucose and the *N*-acetyl-glucosamine moieties, respectively. The possible CLC polar interactions with *N*-acetyl-lactosamine which are shown were deduced after structural superposition of bovine galectin-1 with CLC protein using the program SHP [72].

which to some extent is compensated for by the presence of the Arg61 side chain, that mimics the role played by the two crucial arginines in the galectins. This arginine could make a direct hydrogen-bond interaction with the carbohydrate ligand (Table 4). Two other changes in the CLC protein structure, Asn46→Gln55 and Glu71→Gln75, do not appear to cause any major disruption in possible saccharide–protein interactions. Thus, there is a 'core binding site' in the CLC protein structure which could bind potential carbohydrate ligands, and in our unliganded structure the putative carbohydrate position is occupied by five water molecules which are involved in a network of hydrogen bonds stabilizing the conformation of this site (Fig. 5).

Carbohydrate-binding activity

In order to determine whether CLC protein possesses carbohydrate-binding activity, we have performed a series of experiments using solid-phased saccharides for affinity chromatography (Figs 6,7). Native CLC protein in whole cell lysates of AML14.3D10 cells [52], recombinant CLC protein in lysates of COS cells transfected with wild-type or mutant CLC protein expression vectors [11,53], antibody affinity-purified native CLC protein, and CLC protein resolubilized from fresh, or lyophilized and frozen crystals [8], was analysed for its ability to bind to solid-phased sugars including lactose, *N*-acetyl-D-glucosamine, and mannose. Unlike most galectins, CLC protein did not show strong binding to any of the solid-phased saccharides tested, regardless of the source of the protein. In contrast, CLC protein in lysates of AML14.3D10 cells showed a highly reproducible retarded (delayed) elution on affinity columns of *N*-acetyl-D-glucosamine (Fig. 6a), which could be inhibited by inclusion of the free sugar in the cell lysis and column running buffer (Fig. 6b). This weak binding to *N*-acetyl-D-glucosamine was not inhibited by mannose (Fig. 6c) and CLC protein in cell lysates did not show a similar weak binding (delayed elution) to solid-phased mannose (Fig. 6d). CLC protein was also found to have a weak affinity for lactose, provided the cells were first lysed in the presence of lactose and then extensively

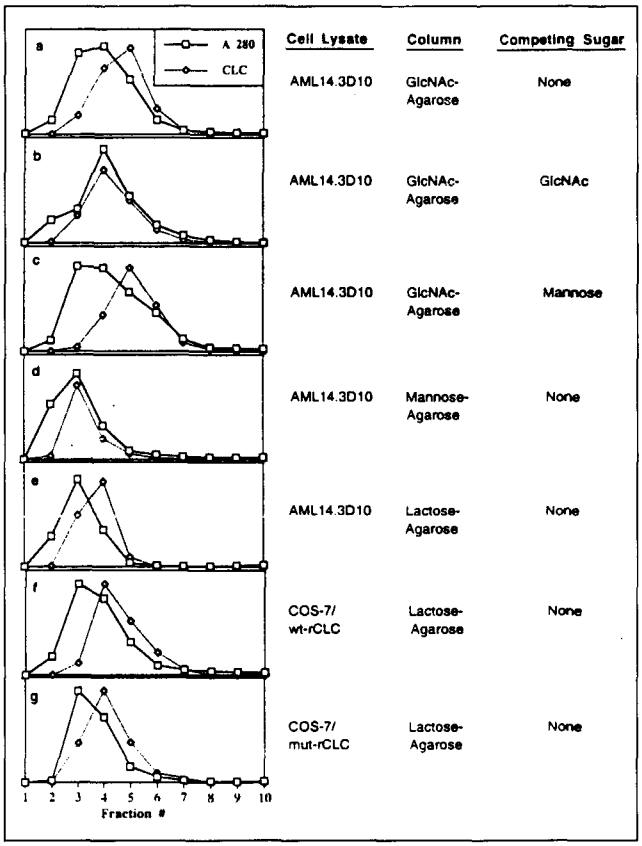

Fig. 6. Low-affinity binding of CLC protein to *N*-acetyl-D-glucosamine and lactose. CLC protein in lysates of AML14.3D10 cells showed a one to two fraction delay in elution, compared with the major protein peak, when the lysate was run on an **(a)** *N*-acetyl-D-glucosamine–agarose or **(e)** lactose–agarose column. **(b)** This weak binding could be inhibited by adding 150 mM *N*-acetyl-D-glucosamine to the lysis and column running buffer. **(d)** The presence of a non-competing sugar (D-mannose) in the lysis and running buffer did not affect binding **(c)**. CLC did not bind to D-mannose agarose. **(f)** Wild-type recombinant CLC (wt-rCLC) from transiently transfected COS-7 cells showed the same one fraction delay in elution. **(g)** Mutation of the Cys29 and Cys57 residues to Gly (the double mutant) did not block the weak binding of rCLC to the lactose–agarose column.

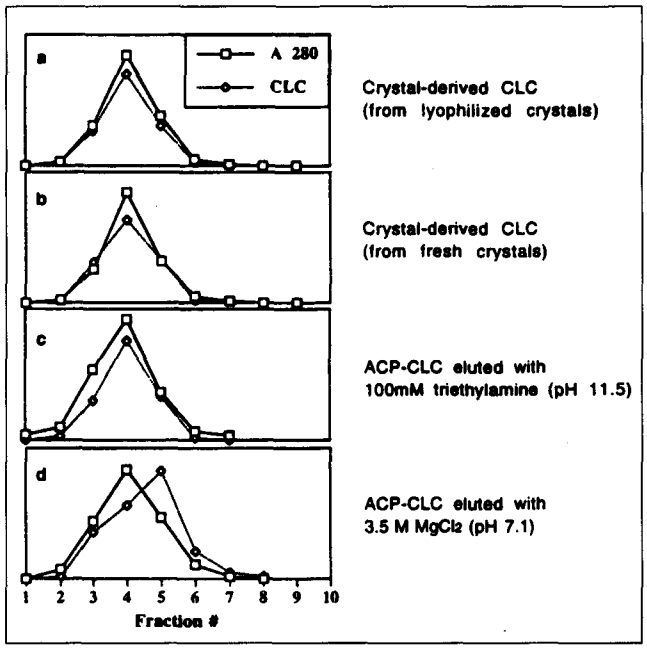

Fig. 7. Carbohydrate-binding activity of crystal-derived and affinity-purified CLC protein. Crystallized and resolubilized CLC from either **(a)** lyophilized or **(b)** freshly prepared crystals showed no binding activity to *N*-acetyl-D-glucosamine-agarose. Either BSA or horse myoglobin (1 mg) was mixed with the CLC protein as an elution marker before loading onto the column. Functionally active carbohydrate-binding CLC protein could be affinity chromatography purified (ACP) using 3.5 M $MgCl_2$ for elution from the antibody column **(d)**, but elution at high pH (triethylamine pH 11.5) completely inactivated the carbohydrate-binding activity of the purified protein **(c)**, a lability shared by other galectins.

dialysed to remove the free sugar and putative endogenous oligosaccharide competitors before affinity chromatography (Fig. 6e). Hirabayashi and Kasai [23,24] have utilized a similar chromatographic retardation analysis on immobilized asialofetuin to assess the carbohydrate-binding activities of site-specific mutants of human galectin-1 that do not bind strongly to the oligosaccharide ligand. In addition to the native eosinophil cell line derived protein, recombinantly expressed wild type CLC protein (Fig. 6f) showed a similarly weak affinity for the lactose–agarose resin. As mutations at cysteine residues to block sulphydryl-dependent oxidative inactivation of carbohydrate-binding activity have been shown to stabilize the β-galactoside-binding activity of the 14 kDa human galectin-1 [23], we tested recombinant CLC protein containing both a Cys29→Gly and Cys57→Gly mutation (Fig. 6g); however, these amino-acid substitutions neither inhibited nor enhanced the weak affinity of the recombinant protein for lactose–agarose (Fig. 6g).

Crystal-derived CLC protein, whether resolubilized from lyophilized and frozen or from freshly prepared crystals, did not show any affinity whatsoever for *N*-acetyl-D-glucosamine (Fig. 7a,b) or lactose (data not shown). In this regard, attempts to co-crystallize CLC protein (resolubilized from fresh or from frozen and lyophilized CLC crystals) with oligosaccharide ligands such as lactose, galactose, *N*-acetyl-D-glucosamine and mannose have yielded excellent CLC crystals for X-ray diffraction, but have been uniformly unsuccessful for demonstrating bound saccharides in the resulting crystal structures. In contrast, native eosinophil cell line derived CLC protein, isolated by affinity chromatography on an anti-CLC protein antibody column, retains its low affinity binding to *N*-acetyl-D-glucosamine, an activity that was inactivated when CLC protein was first purified using alkaline pH elution of the protein from the antibody column (Fig. 7c,d). Taken together, these findings demonstrate low affinity, but specific, binding of CLC protein to carbohydrate ligands normally bound with high affinity by most other members of the galectin superfamily. Moreover, it is evident from the crystal structure of the CLC protein that it possesses extensive crystal contacts (Fig. 3a), but these should not compete with the carbohydrate ligand for space.

Based on the CLC structure, it is possible that the charge distribution and the absence of the full complement of

Table 5. Comparative lysophospholipase activity of mutants vs wild type CLC protein before and after crystallization and resolubilization.

Mutation	Mutant 1	Mutant 2	Mutant 3	Wild Type
Codon	TGT→GGT	TGC→GGC	TGT/TGC→GGT/GGC	None
Amino acid change	Cys29→Gly	Cys57→Gly	Cys29/57→Gly	None
Lysophospholipase activity*				
Before crystallization	3.44 (2.41–4.77)	6.46 (5.19–8.77)	2.45 (2.26–2.62)	1.84 (0.92–2.57)
After crystallization	0.09 (0.05–0.13)	1.91 (0.28–3.82)	0.13 (0.05–0.28)	0.02 (0.0–0.04)

*One unit of lysophospholipase activity represents 1 μmol of fatty acid released per hour per mg of CLC protein at 37°C and pH 7.5. (Numbers in parentheses indicate the range of values for 3–4 experiments.)

hydrogen-bond interactions (Table 4) may account for the weak binding of lactose; there is enough room for lactose to fit into the carbohydrate-binding site by displacing the cluster of water molecules (Fig. 5). Similarly, the *N*-acetyl-D-glucosamine could take the place of the *N*-acetyl-D-glucosamine moiety of *N*-acetyl-D-lactosamine. Alternatively, it might bind in an extended site at the non-reducing side of galactose [19,22]. This may be a better proposition, as Arg61 could make a direct hydrogen-bond interaction in CLC protein with the sugar moiety (or the non-reducing side of galactose, in which a *N*-acetyl-D-glucosamine would be attached at position-3), involving some conformational rearrangement in the structure.

The location of the lysophospholipase active site

CLC protein directs the expression of LPLase activity and spontaneous crystallization in transiently transfected COS cells [11,53]; however, when purified crystals of native or recombinant CLC protein are resolubilized in Tris buffer (pH 8.9), the enzyme is essentially inactive (Table 5). Interestingly, mutation of the Cys57 residue to glycine stabilized the LPLase activity such that the recombinant mutant retained enzyme activity after crystallization (Table 5). The 3D structure of CLC protein reported here is in the absence of bound ligand/substrate and does not provide direct information on the location of the LPLase active site; however, a potential active site is located (Fig. 8) in a hydrophobic channel topologically situated at one end of the molecule, some 15 Å away from the carbohydrate-recognition domain. Two residues, Tyr100 and His114 (located on strands F4 and F5, respectively), pointing towards this hydrophobic cavity (the only cavity in the CLC protein structure as identified using the program GRASP [54]), appear to be part of this putative active site and make hydrogen-bond interactions with a bound water molecule (B factor 14 $Å^2$), as shown in Figure 8.

The potential site described here is reminiscent of those identified in the crystal structures of extracellular phospholipase A_2. The most prominent feature of this catalytic site (the catalytic triad), which is formed by a histidine, an aspartic-acid and a tyrosine residue (e.g. His47, Asp91 and Tyr51 in human secretory phospholipase A_2 [55]), is that it binds to the substrate through hydrogen-bond interactions. It is known that both histidine and aspartic-acid residues contribute the catalytic machinery and the tyrosine residue lends structural support during catalysis, the activation of which is dependent on Ca^{2+} ions. On the basis of our comparative analysis, it is tempting to speculate that, in CLC protein, the bound water molecule might play the role of the aspartic-acid residue (in phospholipase A_2). It is thus possible that in the CLC protein structure, the histidine residue might be a proton donor, the water molecule may act like a nucleophile, and the tyrosine residue might be involved in the catalytic network, thus assisting the binding of the substrate (i.e. a lysophospholipid); however we can not rule out the existence of an alternative LPLase active site, a possibility that requires further investigation.

Discussion

The structural similarity of CLC to galectins -1 and -2 and sequence similarities to the lectin domain of galectin-3 (the collective term for the 31–35 kDa IgE- and

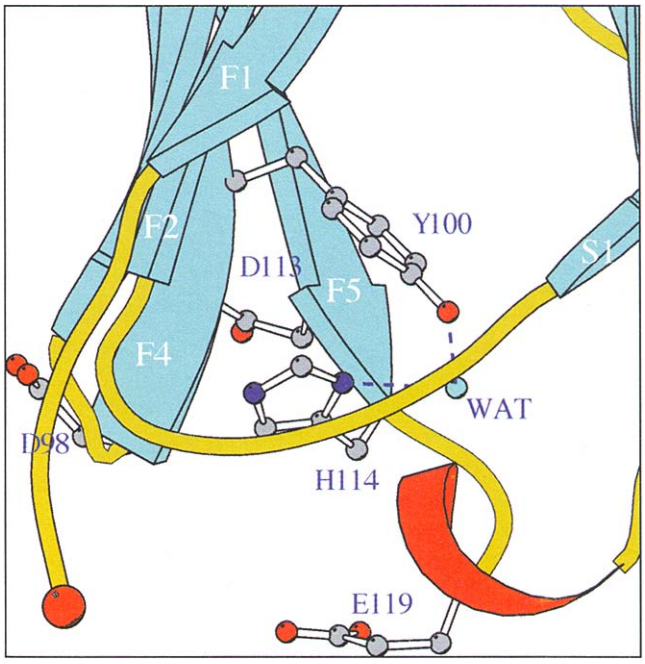

Fig. 8. A view of the residues in the potential LPLase active site in the CLC protein structure showing residues Tyr100, His114 and the water molecule form the potential catalytic triad, as observed in the native CLC protein structure. Amongst residues Asp98, Asp113 and Glu119, one residue may replace the water molecule in the formation of the catalytic triad in the active form of the structure.

laminin-binding members of the galectin family) makes it possible to speculate on its functional role(s) in eosinophils and basophils. Members of the galectin superfamily have been implicated in cell–matrix interactions and inflammation (for a review, see [19]). The amino-acid sequence of CLC protein has similarity to the C-terminal lectin domain of galectin-3, which includes the major non-integrin laminin-binding protein of macrophages (Mac-2), suggesting that CLC protein might be involved in eosinophil interactions with the polylactosamine oligosaccharides of extracellular matrix substrates such as laminin. Galectin-3 binds avidly to laminin [56], shows differential recognition of certain IgE glycoforms [57] and induces mast cell histamine secretion by binding the glycosylated high affinity IgE receptor (FcεRI) [41]. In contrast, the nuclear localization of galectin-3 [58] and its recent identification as an important factor in pre-mRNA splicing in the nucleus of HeLa cells [35] emphasizes the potential multifunctional nature of many members of this superfamily. CLC protein has also been localized, by immunofluorescence and ultrastructural immunocytochemistry, to two major sites in the eosinophil: the nucleus and a granule-free space in the cytoplasm immediately beneath the plasma membrane [5,11]. This is in addition to its localization in other minor subcellular compartments, including a minor population of crystalloid-free 'primary' granules [36] and the cytosol and vesicles of both activated tissue eosinophils and *in vitro* differentiated eosinophils derived from IL-5-induced umbilical cord blood progenitors [34]. In basophils, CLC protein is present in the large, histamine-containing granules (as intragranular crystals or crystalloids in activated basophils [6]) and in the nucleus [37]. Significant amounts of CLC protein are not secreted during receptor-mediated basophil activation, and instead the protein undergoes a marked cellular rearrangement with increased nuclear localization [37] (AM Dvorak and SJ Ackerman, unpublished data). The possibility that CLC protein could function intracellularly in pre-mRNA splicing is intriguing and requires further investigation.

Eosinophilic promyelocytes and myelocytes secrete CLC protein *in vitro* during eosinophil differentiation and colony formation in soft agar, possibly by exocytosis of primary granules [33], and crystals of CLC protein have been identified in developing eosinophilic promyelocytes in the bone marrow [7]. Ultrastructural analyses suggest a process of degranulation/exocytosis of primary granules in the bone marrow during eosinophil development (reviewed in [33]) and CLC protein levels are significantly elevated in the sinusoidal cavity of the bone marrow in normal individuals [33]. Observations that the expression and localization of the galectins is developmentally regulated suggest that secreted CLC protein in the bone marrow might play a role in eosinophil or basophil development, perhaps by functioning extracellularly in the organization of their hematopoietic microenvironment. By analogy, externalization of mouse muscle galectin-1 (L-14) accompanies myoblast differentiation and this lectin co-distributes with laminin in the myotube extracellular matrix [29]. Galectin-1 has been shown to inhibit myoblast adhesion to laminin, suggesting it plays a role in regulating myoblast detachment from laminin during differentiation and fusion into tubular myofibers [59].

Despite no significant sequence or structural similarities to other lipolytic enzymes including LPLases, phospholipases, and lipases, a possible LPLase catalytic triad has been identified in the CLC structure, thus making CLC protein a unique dual function polypeptide. As one of the most abundant eosinophil constituents, it has been hypothesized that CLC protein must play a critical role in eosinophil effector function, either in the pathogenesis of inflammation and tissue damage in allergic diseases or in host–parasite interactions. A number of, as yet unsubstantiated, hypotheses have been proposed. For example, the released fatty acid may have membrane perturbing effects on the eosinophil's target cell. Alternatively, this LPLase may protect eosinophils from lysophospholipids and other toxic parasite products. Another possibility is a direct anti-parasitic effect by hydrolyzing lysophospholipids in the cuticle of parasite targets [60,61]. The fact that eosinophils (and basophils) contain such large quantities of this enzyme suggests a role for these cells and CLC protein in the metabolism of lysophosphoglycerides that may be generated in inflammatory foci by neutrophils, in anaphylactic reactions by mast cells, directly by pathogenic microorganisms, or by helminth parasites such as schistosomula of *Schistosoma mansoni* [61].

Schistosomula secrete considerable quantities of monopalmitoyl-lysophosphatidylcholine, a potent, biologically active lysophosphatide that may mediate membrane fusion and lysis of cytotoxic effector cells by the parasite [62]; as eosinophils, but not neutrophils, can effectively kill schistosomula *in vitro*, they may be resistant to the lytic and membrane-perturbing effects of parasite lysophosphatides by virtue of their considerable content of this enzyme. Enzymatically active CLC protein in or on the plasma membrane could inactivate these lysophosphatides, potentially protecting the eosinophil. An intriguing possibility is whether schistosomula also express a functional carbohydrate ligand for CLC protein that might serve to target this LPLase to the parasite surface. By altering the stoichiometric balance of phospholipid and free fatty acids in membranes, CLC protein might act synergistically to enhance the cytotoxicity or helminthotoxicity of eosinophil granule cationic proteins such as major basic protein (MBP) and eosinophil cationic protein (ECP), both of which have been shown to alter membrane structure [13,63]. By analogy, bee venom is particularly rich in both phospholipases (PLA_2) and highly cationic polypeptides such as melittin that interact with membranes and cause cell lysis. The potential for CLC protein's lipolytic activity to alter membrane fluidity and integrity would suggest the need for careful intracellular regulation of its LPLase activity by eosinophils and basophils. The question of whether endogenous carbohydrate ligands are capable of regulating the LPLase activity of this dual function galectin is currently under investigation.

The biological significance of both the LPLase activity and crystallization of CLC protein in general remains unclear. In this regard, the crystallization of CLC protein enzymatically inactivates both its LPLase activity and weak carbohydrate binding. A detailed structure–function analysis, probing the dual activities of CLC protein using protein engineering experiments, should enable characterisation of the carbohydrate-recognition domain and enzymatic active site. It should also provide insight into the regulatory mechanisms for intracellular and extracellular expression of LPLase and galectin activities by these granulocytes, as well as address the unique propensity and biological role of autocrystallization.

Biological implications

For more than a hundred years, identification of the distinctive hexagonal bipyramidal crystals of Charcot–Leyden crystal (CLC) at sites of eosinophil infiltration in tissues, and in body fluids and secretions, has been a hallmark of eosinophil involvement in allergic, parasitic and inflammatory disorders. Eosinophils, through secretion of granule-associated toxic cationic proteins and enzymes, and lipid mediators, function as proinflammatory and cytoxic-effector cells in asthma and eosinophil-associated allergic and parasitic diseases. The CLC protein comprises ~10% of total eosinophil protein and possesses weak lysophospholipase (LPLase) activity. Several potential consequences of CLC protein's LPLase activity in the pathogenesis of asthma and eosinophil-associated inflammation have been suggested, but the physiological role of this enigmatic autocrystallizing protein has remained highly speculative.

The propensity of CLC protein to crystallize, a process which inactivates its LPLase activity, is perhaps the most unique and intriguing of its physicochemical properties. Although it has been shown that crystals of the CLC protein are uniquely formed by primate but not non-primate eosinophils, which nevertheless express LPLase activity, the physiological relevance of this crystal formation to human eosinophil function is unclear. The amino-acid sequence of CLC protein shows no similarities to LPLases, but does share minimal (20–30%) overall sequence identity with members of the animal galectin family formerly classified as S-type or S-Lac (soluble lactose-binding) lectins.

We have determined the three-dimensional structure of CLC protein at 1.8 Å resolution. Its molecular fold is topologically identical to human galectins -1 and -2, members of the aforementioned galectin family. The structure reveals a carbohydrate-recognition domain with most of the carbohydrate-binding residues conserved among the galectins. From detailed structural comparisons of CLC protein with galectins and other 'jelly-roll' proteins, it is evident that the stable core β-sandwich motif has been conserved during evolution; however, each of these lectins have different functions and carbohydrate-ligand specificities, reflecting multiple modes of carbohydrate binding. Although CLC protein shows specific (albeit weak) affinity for simple saccharides such as *N*-acetyl-D-glucosamine and lactose, a biologically relevant carbohydrate ligand remains to be identified.

Galectins may function in cell–cell and cell–matrix interactions in development and inflammation. Both CLC protein and galectin-3 have been proposed to be involved in allergic responses. Galectin-3, the major non-integrin laminin and IgE-binding protein of macrophages (also known as Mac-2), is expressed in eosinophils, neutrophils and epithelial cells. Galectin-3 can activate mast cell mediator release through binding to the high-affinity IgE receptor. Intracellularly, galectin-3 functions as a nuclear factor in pre-mRNA splicing. Considering CLC protein's cellular content, cytosolic and nuclear localization, and the potential multifunctional nature of many members of the galectin superfamily, our structural and functional findings suggest that it will possess galectin-associated activities that may be extremely important to eosinophil development, function and inflammatory activities, both in allergic diseases and host–parasite interactions.

Materials and methods

Preparation of affinity-purified and crystal-derived CLC protein

Rabbit anti-CLC antibodies were affinity purified, as described previously [6,36], using a solid-phase CLC-Sepharose column. Antibodies were bound directly to protein A–Sepharose (Pharmacia, Piscataway, NJ) and chemically cross-linked using dimethylpimelimidate. CLC protein was affinity purified from AML14.3D10 cell lysates using the anti-CLC protein column. 100 mM triethylamine or 3.5 M $MgCl_2$ in PBS pH 7.1 was used to elute CLC protein bound to the antibody affinity column. Following elution, the CLC protein was dialysed for 24 h against pH 7.1 PBS and concentrated 10–50 fold using a stirred ultrafiltration cell (Amicon, Beverly, MA) equipped with a YM10 membrane (Amicon). Purity of the affinity-isolated CLC protein was assessed by silver-stained SDS-PAGE gels. CLC crystals were made as previously described [8]. CLC protein from either freshly prepared, or lyophilized and frozen CLC, was resolubilized in PBS pH 7.4 by stirring at room temperature for 3 h. Solutions of CLC protein containing approximately 150–300 μg ml^{-1} were obtained using this method.

Crystallization

Details of the crystallization of CLC protein during preliminary X-ray diffraction analysis have been described elsewhere [64]. Purified crystal-derived CLC protein was recrystallized using the hanging drop vapour diffusion method. Drops containing protein (5 mg ml^{-1}) at pH 10.5 in 0.1 M Tris were equilibrated against reservoirs of 0.2 M Tris-acetate buffer (pH 7.0). Hexagonal bipyramidal crystals (approx. 2.0×0.25×0.25 mm) were

grown after 3–4 days at 16°C; these crystals possess the same morphology as those found *in vivo*. CLC protein crystals diffract to minimum Bragg spacings of 1.8 Å using a synchrotron radiation source. The systematic absences and symmetry were consistent with space group $P6_122$ or $P6_522$ (later confirmed to be $P6_522$) with unit cell dimensions of a=b=49.6 Å, c=262.2 Å. There is one molecule of CLC protein per crystallographic asymmetric unit and approximately 52% of the crystal volume is occupied by solvent.

Diffraction measurements

Diffraction data were collected on stations PX 9.5 (derivatives) and PX 9.6 (native) of the Synchrotron Radiation Source, Daresbury, UK, using a 30 cm diameter MAR-research image plate. Native data to 1.8 Å resolution were collected from two crystals at 16°C. A 0.2 mm collimator was used. A typical exposure time was 120 s for 0.5° oscillation. The crystal to image plate distance was kept at 287 mm during data collection for the native crystals.

Heavy atom derivatives obtained using *p*-chloromercuribenzenesulfonic acid (p-CMBS) (SIGMA), and potassium tetrachloro platinate (K_2PtCl_4) (SIGMA) were used to determine the CLC structure. The native crystals were soaked for 22 h in solutions of the mother liquor containing 10 mM of K_2PtCl_4 while the crystals were equilibrated in 3% PEG 4000 to prevent the dissolution of the crystal prior to the soaking experiments. Crystals cocrystallized with 2 mM *p*-CMBS were also used. Data were collected to 3.4 Å resolution with an X-ray beam (0.2 mm collimator) of wavelength 0.92 Å. One crystal per data set was used (at 16°C) with an oscillation range of 1.5° per image. The exposure time was 90 s and the crystal to image plate distance was set at 325 mm.

The bipyramidal crystals were generally mounted with the long axis of the crystal lying parallel to the capillary axis, but were generally offset by a few degrees relative to the oscillation axis to achieve a more efficient coverage of measured reflections. For the native crystals , 50° of data were collected to minimize the 'missing cone'of data.

Raw data images were indexed, integrated and data reduction were performed using the program DENZO [65]. The details of data processing statistics for the three datasets used in the structure determination are presented in Table 1.

Phasing

The structure of CLC protein was determined by the method of multiple isomorphous replacement (MIR). Scaling of native and derivative data, Patterson map calculations, heavy-atom refinement, solvent flattening, combination of model and MIR phases, along with electron-density map calculations, were performed using the programs available in the PHASES package [66]. The initial MIR phases were calculated using reflections with $F/\sigma F>2.0$ to a resolution limit of 3.4 Å. For the conventional heavy-atom parameter refinement only reflections with phases with figure of merit (FOM) better than 0.60 were used (Table 1). During refinement of these parameters of one derivative the contribution from both derivatives to the phasing was used. The final FOM was 0.60 for 3014 phased reflections between 15 and 3.4 Å. The protein phases were improved by solvent flattening the electron-density map [67] (solvent content 52%), using the protocol implemented in the PHASES package and further heavy-atom refinement against the protein phases calculated from the solvent-flattened MIR electron-density map.

Map interpretation

Electron-density maps were displayed and model coordinates were fitted on an Indigo 2 Silicon Graphics workstation using the interactive computer graphics program O, version 5.10.2 [68]. First, Cα atoms were assigned to the skeleton positions and then a polyalanine model was built by performing a database search to obtain the best-fitting peptide with respect to the Cα position. Using the amino-acid sequence of CLC protein [9] two models, containing 138 residues with all the side chains, were built (at 3.4 Å), one for each enantiomer ($P6_122$ and $P6_522$).

Refinement

All refinement was carried out using the program X-PLOR [69]. In order to resolve the ambiguity regarding the space group assignment for CLC crystals, refinement was performed using molecular dynamics and energy minimization for the two enantiomers. After one round of refinement, for the model in space group $P6_122$, the crystallographic R factor (R_{cryst}) converged to 33.5% but the R_{free}, calculated with 5% of data which was omitted from the refinement, was 57.4%. In contrast, the model in space group $P6_522$ had an R_{cryst} value of 28.4% and the R_{free} was 48.3%, using the same dataset as in space group $P6_122$, indicating $P6_522$ was the correct space group. Further calculations were carried out in this space group.

Calculated phases obtained from the refined structure were combined with the original (that is, not the solvent flattened) MIR phases and used to determine a new electron-density map. Careful examination of the maps allowed corrections to be incorporated into the model. Alternating cycles of manual building, conventional positional refinement and simulated annealing method, as implemented in X-PLOR, improved the model. Rebuilding was initially performed at 3.4 Å resolution, using electron-density maps calculated from combined model/MIR phases. Extension of refinement from 3.4 Å to 1.8 Å resolution was performed in 0.1 Å resolution steps and the quality of the model was monitored using $2|F_o|-|F_c|\phi_{calc}$ maps. The behaviour of the R_{free} value was also monitored all through refinement. After several rounds of refinement, model building and individual B-factor refinement (using all data from 8–1.8 Å resolution, 17687 reflections) gave a final model with R_{cryst} of 20.0% and last recorded R_{free} of 25.8% for a randomly selected 5% subset of reflections not used in the refinement [69]. Rms deviations from ideality of bond lengths and bond angles were 0.010 Å and 1.61°, respectively. During the final stages of refinement, water molecules were inserted into the model only if there were peaks in the $|F_o|-|F_c|$ electron-density maps with heights greater than 3σ and they were at hydrogen-bond forming distances from appropriate atoms. $2|F_o|-|F_c|\phi_{calc}$ maps were also used to check the consistency in peaks. Water molecules with a temperature factor higher than 60 $Å^2$ were excluded from subsequent refinement steps. A portion of the $2|F_o|-|F_c|\phi_{calc}$ final electron-density map (at 1.8 Å resolution), contoured at 1.0σ, is shown in Figure 1.

Validation of the model

The final refined model at 1.8 Å resolution comprised 141 residues (1145 protein atoms) and 77 water molecules. The program PROCHECK [70] was used to assess the quality of the final structure. Analysis of the Ramachandran (φ-ψ) plot showed that all residues lie in the allowed regions except Arg128 which is in a generously allowed region. Secondary structural alignments and solvent accessibilities were calculated utilising the DSSP program [71]. Structural superpositions were performed using the program SHP [72]. Atomic

coordinates for structural comparisons were obtained from the Brookhaven Protein Data Bank.

Cell culture, plasmids and transfections

AML14.3D10 is a fully differentiated eosinophilic myelocyte subline of the AML14 and AML14. eos cell lines [52] that continues to proliferate and maintain a differentiated phenotype without cytokine supplementation. AML14.3D10 cells were maintained in RPMI 1640 medium supplemented with L-glutamine (2 mM), 2-mercaptoethanol (0.05 mM), sodium pyruvate (1 mM) and 8% fetal bovine serum (GIBCO, Gaithersburg, MD). COS-7 cells were obtained from the American Type Culture Collection. COS-7 cells were maintained as monolayers at 37°C in a 5% CO_2 atmosphere in Dulbecco's modification of Eagle's medium (Mediatech, Herndon, VA) supplemented with L-glutamine (2 mM), penicillin (100 U ml^{-1}), streptomycin (0.1 mg ml^{-1}) and 10% fetal bovine serum (HyClone, Logan, UT). Site-directed mutagenesis of Cys29→Gly (TGT→GGT) and Cys57→Gly (TGC→GGC) was accomplished using primer 1 (5′-GCCTgaattcGGGCAA-TTC-3′), containing an EcoRI cutting site (indicated in lower case letters), primer 2 (5′-CACctgcagATATGGTTCATTCA-AGAAACCGAC-3′, antisense), containing a Pst1 site and a mismatched base (underlined) used to mutagenize Cys29, and primer 3 (5′GTtcatgaCCACACGACGACCAAAGCCCAC-3′), containing a mismatched base (used to mutagenize Cys57) and a BspH 1 site. Primers 1 and 2 were designed to create a PCR fragment containing the Cys29→Gly mutation and primers 1 and 3 were used to create a PCR fragment containing the Cys 57→Gly mutation. These PCR fragments, used to substitute for the corresponding segments in the wild-type CLC cDNA, were ligated into an EcoRI-digested, phosphatase-treated pMT2 eukaryotic expression vector [73] to create the desired mutant constructs [53]. The double mutation (Cys29→Gly and Cys57→Gly), derived from a combination of the Cys29 and Cys57 mutants, was created by substitution of the Cys29→Gly mutant PCR fragment at the corresponding position in the Cys57→Gly mutant. Both mutations were confirmed by DNA sequencing of the PCR generated constructs. The wild type CLC cDNA and the Cys29/Cys57→Gly mutant cDNA in the pMT2 expression vector were transfected into COS-7 cells using the DEAE dextran technique as previously described for the production of recombinant CLC protein [11]. After 72 h, cells were harvested and lysed in pH 7.4 PBS, containing 1% NP-40, centrifuged at 35000 *g* for 20 min, and the supernatant was applied immediately to the glycoconjugate affinity columns. AML14.3D10 cells were lysed under the same conditions.

Assessment of carbohydrate-binding activity

The carbohydrate-binding activity of native or recombinant CLC protein was determined as a delay in the elution [24] of CLC protein (measured by RIA in μg ml^{-1}, range 0.01–100 μg ml^{-1}, as described previously [8]), compared with the major protein peak (measured as absorbance at 280 nm, range 0–15 units). The bed volume of the *N*-acetyl-D-glucosamine-agarose (SIGMA, St Louis, MO), lactose–agarose (SIGMA) and D-mannose–agarose (SIGMA) columns was 5 ml. Flow rate was 6 ml h^{-1}. After loading the sample, the columns were washed with pH 7.4 PBS running buffer in the absence or presence of competing sugar until the A_{280} returned to a zero baseline. All columns were then eluted with 150 mM competing sugar. In the case of pure CLC protein, 1 mg of bovine serum albumin or horse myoglobin was mixed with the CLC protein prior to loading on the column to provide a protein marker detectable at A_{280}.

Lysophospholipase activity assay

The effects of the Cys to Gly mutations on LPLase activity were determined using transiently transfected COS cells as previously described [11,53]. COS cells were collected by cell scraping at 72 h after transfection and washed with PBS. The cells were sonicated (4×15 s sonication bursts with 1 min intervals for cooling) on ice using a Model 450 sonifier (Branson Ultrasonic Corp., Danbury, CT) with a microtip at a constant duty cycle and power setting of 3.5. The crude sonicates were centrifuged at 10000 *g* at 4°C for 20 min and the supernatant was used for measurement of LPLase activity [11,12]. One unit of LPLase activity represents 1 μmol of released fatty acid h^{-1} mg^{-1} CLC protein (as measured by RIA) at 37°C and pH 7.5. In addition, the mutant and wild-type recombinant (wt-rCLC) proteins were also used for determination of LPLase activity after crystallization and resolubilization. The concentrations of the resolubilized crystal-derived mutants and wild type rCLC protein were measured by either RIA [8] or Bradford protein assay as described by the manufacturer (Bio-Rad, Richmond, CA).

Atomic coordinates for human CLC protein are being deposited with the Brookhaven Protein Data Bank.

Acknowledgements: We are grateful to the staff at the Synchrotron Radiation Source, CLRC Laboratory at Daresbury, Nikos Oikonomakos for help with X-ray data collection, Garry Taylor and Susan Crennell for providing the atomic coordinates of Wing 1 and Wing 2 components of *Vibrio cholerae* neuraminidase, Cassandra Paul and Michael Baumann (Wright State University, Dayton, OH) for developing and kindly making available the AML14.3D10 cell line and Vasanta Subramanian for helpful discussion. We also thank Daniel Tenen for frequent scientific input and helpful discussions on the expression of recombinant CLC protein and site-directed mutagenesis and Ron Stenkamp and Larry Sieker for their initial contributions to the project. This work was supported by the Medical Research Council and the Nuffield Foundation, UK, through grants to KRA, and by the National Institutes of Health, grants AI25230 and AI34953, to SJA.

References

1. Charcot, J.M., & Robin, C. (1853). Observation de Leukocythemia. *C. R. Mem. Soc. Biol.* **5**, 44.
2. Leyden, E. (1872). Zur Kenntniss des bronchial asthma. *Arch. Pathol. Anat.* **54**, 324–344.
3. Beaver, P.C. (1962). Observations on the nature of Charcot–Leyden crystals. *Bull. Soc. Pathol. Exot. Filiales* **55**, 471–476.
4. Beeson, P.B. & Bass, D.A. (1977). The eosinophil. In *Problems in Internal Medicine.* (Smith, L.H., ed), vol. **15**, pp. 39–43, W.B. Saunders Co., Philadelphia.
5. Ackerman, S.J., Weil, G.J. & Gleich, G.J. (1982). Formation of Charcot–Leyden crystal by human basophils. *J. Exp. Med.* **155**, 1597–1609.
6. Dvorak, A.M & Ackerman, S.J. (1989). Ultrastructural localization of the Charcot–Leyden crystal protein (lysophospholipase) to granules and intragranular crystals in mature human basophils. *Lab. Invest.* **60**, 557–567.
7. Dvorak, A.M., Ishizaka, T., Weller, P.F. & Ackerman, S.J. (1993). Ultrastructural contributions to the understanding of the cell biology of human eosinophils: Mechanisms of growth factor-induced development, secretion, and resolution of released constituents from the microenvironment. In *Eosinophils: Biological and Clinical Aspects.* (Makino, S. & Fukuda, T., eds.), pp. 13-32, CRC Press, Boca Raton, FL.
8. Ackerman, S.J., Loegering, D.A. & Gleich, G.J. (1980). The human eosinophil Charcot–Leyden crystal protein: biochemical characteristics and measurement by radioimmunoassay. *J. Immunol.* **125**, 2118–2126.
9. Ackerman, S.J., *et al.*, & Tenen, D.G. (1993). Molecular cloning & characterization of human eosinophil Charcot–Leyden crystal protein (lysophospholipase). Similarities to IgE binding proteins and the S-type animal lectin superfamily. *J. Immunol.* **150**, 456–468.

10. Ackerman, S.J., Zhou, Z.-Q., Tenen, D.G., Clark, M.A., Tu, Y.-P. & Irvin, C.G. (1994). Human eosinophil lysophospholipase (Charcot–Leyden crystal protein): Molecular cloning, expression, and potential functions in asthma. In *Eosinophils In Allergy and Inflammation.* (Gleich, G.J. & Kay, A.B., eds), pp. 21–54, Marcel Dekker, New York.
11. Zhou, Z., Tenen, D.G., Dvorak, A.M., & Ackerman, S.J. (1992). The gene for human eosinophil Charcot–Leyden crystal protein directs expression of lysophospholipase activity and spontaneous crystallization in transiently transfected COS cells. *J. Leukoc. Biol.* **52**, 588–595.
12. Weller, P.F., Bach, D.S., & Austen, K.F. (1984). Biochemical characterization of human eosinophil Charcot–Leyden crystal protein (lysophospholipase). *J. Biol. Chem.* **259**, 15100–15105.
13. Abu-Ghazaleh, R. I, *et al.*, & Gleich, G.J. (1992). Eosinophil granule proteins in peripheral blood granulocytes. *J. Leukoc. Biol.* **52**, 611–618.
14. Han, J.H., Stratowa, C. & Rutter, W.J. (1987). Isolation of full-length putative rat lysophospholipase cDNA using improved methods for mRNA isolation and cDNA cloning. *Biochemistry* **26**, 1617–1625.
15. Kyger, E.M., Wiegand, R.C. & Lange, L.G. (1989). Cloning of the bovine pancreatic cholesterol esterase/lysophospholipase. *Biochem. Biophys. Res. Commun.* **164**, 1302–1309.
16. Garsetti, D.E., Ozgur, L.E., Steiner, M.R., Egan, R.W. & Clark, M.A. (1992). Isolation and characterization of three lysophospholipases from the murine macrophage cell line WEHI 265.1. *Biochim. Biophys. Acta* **1165**, 229–238.
17. Garsetti, D.E., Steiner, M.R., Holtsberg, F., Ozgur, L.E., Egan, R.W. & Clark, M.A. (1993). Comparison of six mammalian lysophospholipases. *J. Lipid Mediat.* **6**, 223–232.
18. Barondes, S.H., *et al.*, & Wang, J.L. (1994). Galectins: a family of animal beta-galactoside-binding lectins. *Cell* **76**, 597–598.
19. Barondes, S.H., Cooper, D.N., Gitt, M.A., Leffler, H. (1994). Galectins. Structure and function of a large family of animal lectins. *J. Biol. Chem.* **269**, 20807–20810.
20. Bourne, Y., *et al.*, & Cambillau, C. (1994). Crosslinking of mammalian lectin (galectin-1) by complex bintennary saccharides. *Nat. Struct. Biol.* **1**, 863–869.
21. Liao, D.I., Kapadia, G., Ahmed, H., Vasta, G.R. & Herzberg, O. (1994). Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein. *Proc. Natl. Acad. Sci. USA* **91**, 1428–1432.
22. Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H., & Rini, J.M. (1993). X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-Å resolution. *J. Biol. Chem.* **268**, 27034–27038.
23. Hirabayashi, J. & Kasai, K. (1991). Effect of amino-acid substitution by site directed mutagenesis on the carbohydrate recognition and stability of human 14 kDa β-galactoside binding lectin. *J. Biol. Chem.* **266**, 23648–23653.
24. Hirabayashi, J. & Kasai, K. (1994). Further evidence by site-directed mutagenesis that conserved hydrophilic residues form a carbohydrate-binding site of human galectin-1. *Glyconjugate J.* **11**, 437–442.
25. Abbot, W.M. & Feizi, T. (1991). Soluble 14-kDa β-galactoside-specific bovine lectin. Evidence from mutagenesis and proteolysis that almost the complete polypeptide chain is necessary for integrity of the carbohydrate recognition domain. *J. Biol. Chem.* **266**, 5552–5557.
26. Ackerman, S.J., Gleich, G.J., Weller, P.F., & Ottesen, E.A. (1981). Eosinophilia and elevated serum levels of eosinophil major basic protein and Charcot–Leyden crystal protein (lysophospholipase) following treatment of patients with Bancroft's filariasis. *J. Immunol.* **127**, 1093–1098.
27. Fukuda, T., Ackerman, S.J., Reed, C.E., Peters, M.S., Dunnette, S.L & Gleich, G.J. (1985). Calcium ionophore A23187 calcium-dependent degranulation in human eosinophils. *J. Immunol.* **135**, 1349–1356.
28. Dor, P.J., Ackerman, S.J., & Gleich, G.J. (1984). Charcot–Leyden crystal protein and eosinophil granule major basic protein sputum of patients with respiratory diseases. *Am. Rev. Respir. Dis.* **130**, 1072–1077.
29. Cooper, D.N. & Barondes, S.H. (1990). Evidence for export of a muscle lectin from cytosol to extracellular matrix and for a novel secretory mechanism. *J. Cell. Biol.* **110**, 1681–1691.
30. Lindstedt, R., Apodaca, G., Barondes, S.H., Mostov, K.E. & Leffler, H. (1993). Apical secretion of a cytosolic protein by madin-darby canine kidney cells — Evidence for polarized release of an endogenous lectin by a nonclassical secretory pathway. *J. Biol. Chem.* **268**, 11750–11757.
31. Sato, S., Burdett, I.D.J. & Hughes, R.C. (1993). Secretion of the baby hamster kidney carbohydrate-binding protein 30 kD (CBP30) from polarized and non-polarized cells. *Glycoconjugate J.* **10**, 271–272.
32. Sato, S. & Hughes, R.C. (1994). Regulation of secretion and surface expression of Mac-2, a galactoside-binding protein of macrophages. *J. Biol. Chem.* **269**, 4424–4430.
33. Butterfield, J.H., Ackerman, S.J., Scott, R.E., Pierre, R.V. & Gleich, G.J. (1984). Evidence for secretion of human eosinophil granule major basic protein and Charcot–Leyden crystal protein during eosinophil maturation. *Exp. Hematol.* **12**, 163–170.
34. Dvorak, A.M., Furitsu, T., Letourneau, L., Ishizaka, T. & Ackerman, S.J. (1991). Mature eosinophils stimulated to develop in human cord blood mononuclear cell cultures supplemented with recombinant human interleukin-5. Part 1. Piecemeal degranulation of specific granules and distribution of Charcot–Leyden Crystal protein. *Amer. J. Pathol.* **138**, 69–82.
35. Dagher, S.F., Wang, J.L., Patterson, R.J. (1995). Identification of galectin-3 as a factor in pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA* **92**, 1213–1217.
36. Dvorak, A.M., Letourneau, L., Login, G.R., Weller, P.F. & Ackerman, S.J. (1988). Ultrastructural localization of the Charcot–Leyden crystal protein (lysophospholipase) to a distinct crystalloid-free granule population in mature human eosinophils. *Blood* **72**, 152–158.
37. Golightly, L.M., Thomas, L.L., Dvorak, A.M. & Ackerman, S.J. (1992). Charcot–Leyden crystal protein in the degranulation & recovery of activated basophils. *J. Leukoc. Biol.* **51**, 386–392.
38. Truong, M.-J., *et al.*, & Capron, M. (1993). Human neutrophils express immunoglobulin-E (IgE)-binding proteins (Mac-2/ε-BP) of the S-type lectin family — role in IgE-dependent activation. *J. Exp. Med.* **177**, 243–248.
39. Truong, M.-J., Gruart, V., Liu, F.T., Prin, L., Capron, A. & Capron, M. (1993). IgE-binding molecules(Mac-2/ε-BP) expressed by human eosinophils — Implication in IgE-dependent eosinophil cytotoxicity. *Eur. J. Immunol.* **23**, 3230–3235.
40. Yamaoka, A., Kuwabara, I., Frigeri, L.G. & Liu, F.T. (1995). A human lectin, galectin-3 (ε-BP/Mac-2), stimulates superoxide production by neutrophils. *J. Immunol.* **154**, 3479–3487.
41. Frigeri, L.G., Zuberi, R.I. & Liu, F.T. (1993). Epsilon-BP, a β-galactoside-binding animal lectin, recognizes IgE receptor (FC-epsilon-RI) and activates mast cells. *Biochemistry* **32**, 7644–7649.
42. Jeng, K.C., Frigeri, L.G. & Liu, F.T. (1994). An endogenous lectin, galectin-3 (εBP/Mac-2), potentiates IL-1 production by human monocytes. *Immun. Lett.* **42**, 113–116.
43. Norris, G.E., Stillman, T.J., Anderson, B.F. & Baker, E.N. (1994). The three-dimensional structure of PNGase F, a glycosyl-asparaginase from *Flavobacterium meningosepticum. Structure* **2**, 1049–1059.
44. Kuhn, P., Tarentino, A.L., Plummer, T.H. & van Roey, P. (1994). Crystal structure of peptide-N(4)-(N-acetyl-β-D-glucosaminyl) asparagine amidase at 2.2 Å resolution. *Biochemistry* **33**, 11699–11706.
45. Keitel, U., Simon, O., Borriss, R. & Heinemann, U. (1993). Molecular & active site structure of a Bacillus 1,3-1,4-β-glucanase. *Proc. Natl. Acad. Sci. USA* **90**, 5287–5291.
46. Torronen, A., Harkki, A. & Rouvinen, J. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesi. Science* **265**, 524–527.
47. Divne, C., *et al.*, & Jones, T.A. (1994). The three dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesi. Science* **265**, 524–528.
48. Emsley, J., *et al.*, & Wood, S.P. (1994). The structure of pentameric human serum amyloid P component. *Nature* **367**, 338–345.
49. Srinivasan, N., White, H.E., Emsley, J., Wood, S.P., Pepys, M.B. & Blundell, T.L. (1994). Comparative analyses of pentraxins — implications for protomer assembly and ligand binding. *Structure* **2**, 1017–1027.
50. Crennell, S., Garman, E., Laver, G., Vimr, E. & Taylor, G. (1994). Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* **2**, 535–544.
51. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
52. Paul, C.C., *et al.*, & Baumann, M.A. (1995). Changing the differentiation program of hematopoietic cells: Retinoic acid-induced shift of eosinophil-committed cells to neutrophils. *Blood* **86**, 3737–3744.
53. Zhou, Z., Liu, J.-N., Chen, H.-M., Tenen, D.G. & Ackerman, S.J. (1993). Site-directed mutagenesis at Cys29 and Cys57 enhances the enzyme activity of human eosinophil lysophospholipase. Advances in Gene Technology; Protein Engineering and Beyond. In *Miami Short Reports vol. 3.* (Brew, K. & Petsko, G.A., eds)], p. 54, IRL Oxford University Press, Oxford.
54. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.
55. Scott, D.L., White, S.P., Browning, J.L., Rosa, S.J., Gelb, M.H. & Sigler, P.B. (1991). Structures of free and inhibited human secretory phospholipase A_2 from inflammatory exudate. *Science* **254**, 1007–1010.
56. Cherayil, B.J., Weiner, S.J. & Pillai, S. (1989). The Mac-2 antigen is a galactose-specific lectin that binds IgE. *J. Exp. Med.* **170**, 1959–1972.

57. Robertson, M.W., Albrandt, K., Keller, D., & Liu, F.T. (1990). Human IgE-binding protein: a soluble lectin exhibiting a highly conserved interspecies sequence and differential recognition of IgE glycoforms. *Biochemistry* **29**, 8093–8100.
58. Gritzmacher, C.A., Robertson, M.W., & Liu, F.T. (1988). IgE-binding protein. Subcellular location and gene expression in many murine tissues and cells. *J. Immunol.* **141**, 2801–2806.
59. Cooper, D.N., Massa, S.M. & Barondes, S.H. (1991). Endogenous muscle lectin inhibits myoblast adhesion to laminin. *J. Cell. Biol.* **115**, 1437–1448.
60. Adewusi, K. & Goven, A.J. (1987). Effect of anti-thymocyte serum on the eosinophil and lysophospholipase responses in mice infected with *Trichinella spiralis*. *Parasitology* **94**, 115–122.
61. Furlong, S.T. & Caulfield, J.P. (1989). *Schistosoma mansoni*: synthesis and release of phospholipids, lysophospholipids and neutral lipids by schistosomula. *Exp. Parasitol.* **69**, 65–77.
62. Golan, D.E., Furlong, S.T., Brown, C.S. & Caulfield, J.P. (1988). Monopalmitoyl-phosphatidylcholine incorporation into human erythrocyte ghost membranes causes protein and lipid immobilization and cholesterol depletion. *Biochemistry* **27**, 2661–2667.
63. Young, J.D., Peterson, C.G., Venge, P. & Cohn, Z.A. (1986). Mechanism of membrane damage mediated by human eosinophil cationic protein. *Nature* **321**, 613–616.
64. Sieker, L.C., Turley, S., Le Trong, I., Stenkamp, R.E., Weller, P.F. & Ackerman, S.J. (1988). Crystallographic characterization of human eosinophil Charcot–Leyden crystals. *J. Mol. Biol.* **204**, 489–491.
65. Otwinowski, Z. (1993). Oscillation data reduction program. In *Data Collection and Processing* (Sawyer, L., Isaacs, N. & Bailey, S.S., eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
66. Furey, W. & Swaminathan, S. (1990). PHASES — A Program Package for the Processing and Analysis of Diffraction Data from Macromolecules. *American Crystallographic Association Meeting Abstracts* **18**, 73.
67. Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112.
68. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
69. Brunger, A.T. (1992). *X-PLOR Version 3.1 Manual: A system for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
70. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK — A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
71. Kabsch, W. & Sander, C. (1983). Dictionary of the protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
72. Stuart, D.I., Levine, M., Muirhead, H. & Stammers, D.K. (1979). The catalytic structure of cat pyruvate kinase at a resolution of 2.6 Å. *J. Mol. Biol.* **134**, 109–142.
73. Kaufman, R.J., Davies, M.V., Pathak, V.K. & Hershey, J.W. (1989). The phosphorylation state of eukaryotic initiation factor 2 alters translational efficiency of specific mRNAs. *Mol. Cell Biol.* **9**, 946–958.
74. Kraulis, P.J. (1991). MOLSCRIPT — A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
75. Hutchinson, E.G. & Thornton, J.M. (1990). HERA — A program to draw schematic diagrams of protein secondary structures. *Proteins* **8**, 203–212.
76. Shaanan, B., Lis, H. & Sharon, N. (1991). Structure of a legume lectin with an ordered *N*-linked carbohydrate in complex with lactose. *Science* **254**, 862–866.
77. Einspahr, H., Parks, E.H., Suguna, K. & Subramanian, E. (1986). The crystal structure of pea lectin at 3.0 Å resolution. *J. Biol. Chem.* **261**, 6518–6527.
78. Hahn, M., Olsen, O., Politz, O., Borriss, R., & Heinemann, U. (1995). Crystal structure and site-directed mutagenesis of *Bacillus macerans* endo-1,3-1,4-β-D-glucanase. *J. Biol. Chem.* **270**, 3081–3088.
79. Lawrence, M.C., Izard, T., Beuchat, M., Blagrove, R.J. & Colman, P.M. (1994). The structure of phaseolin at 2.2 Å resolution: implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. *J. Mol. Biol.* **238**, 748–776.

Received: **26 Sep 1995**; revisions requested: **16 Oct 1995**; revisions received: **27 Oct 1995**. Accepted: **30 Oct 1995**.