Ancestral βγ-crystallin precursor structure in a yeast killer toxin

The structure of the killer toxin from the yeast *Williopsis mrakii* (WmKT) corresponds to a so far hypothetical ancestral single-domain form of the βγ-crystallin structure type, although there is neither significant sequence homology nor coincidence of physiological functions.

Sir—Here we describe a close three-dimensional (3D) structure homology between the newly determined nuclear magnetic resonance (NMR) solution structure of the killer toxin from *Williopsis mrakii* (WmKT) (Fig. 1*a*)[1] and the 3D structure of γB-crystallin, determined by X-ray crystallography (Fig. 1*b*,*c*)[2,3]. γB-crystallin is present in high concentration in the vertebrate eye lens. All other known members of the βγ-crystallin protein superfamily are found to consist of two closely similar domains: the WmKT structure provides the first stable single-domain representative of this structure type and is an important addition to our knowledge on the βγ-crystallin superfamily, which has already become a textbook example for evolution by gene duplication and fusion (Fig. 2*a*)[4,5].

Killer toxins are proteinaceous compounds secreted into the extracellular medium by some yeast strains in order to prevent simultaneous growth of other, foreign yeast strains. WmKT is a monomeric 88-residue protein with a molecular mass of 9,500 M_r[6] which affects cell-wall construction in sensitive yeasts by inhibiting β-glucan synthesis. It is stable over a wide range of pH (2.0 to 11.0) and temperature (100 °C, 10 minutes) and has a basic isoelectric point (pI 9.1), which contrasts with the mostly acidic nature of most other reported killer toxins[7].

The topology of the NMR structure of WmKT (Fig. 1*a*) is dominated by two Greek key motifs, each formed by four β-strands. The polypeptide segments comprising residues 4–9 (A), 17–19 (B), 28–30 (C) and 35–40 (D) form the first motif and the segments 48–54 (a), 61–67 (b), 71–75 (c) and 82–87 (d) the second one. The major secondary structure elements of the toxin are two twisted four-stranded antiparallel β-sheets. The strands A, B and D from motif 1 and strand c from motif 2 form the first β-sheet (blue in Fig. 1*a*), the strands a, b and d from motif 2 and strand C from motif 1 the second β-sheet (red in Fig. 1*a*). The space bounded by the two β-sheets is packed mostly with hydrophobic side chains, which form a well defined hydrophobic protein core[1].

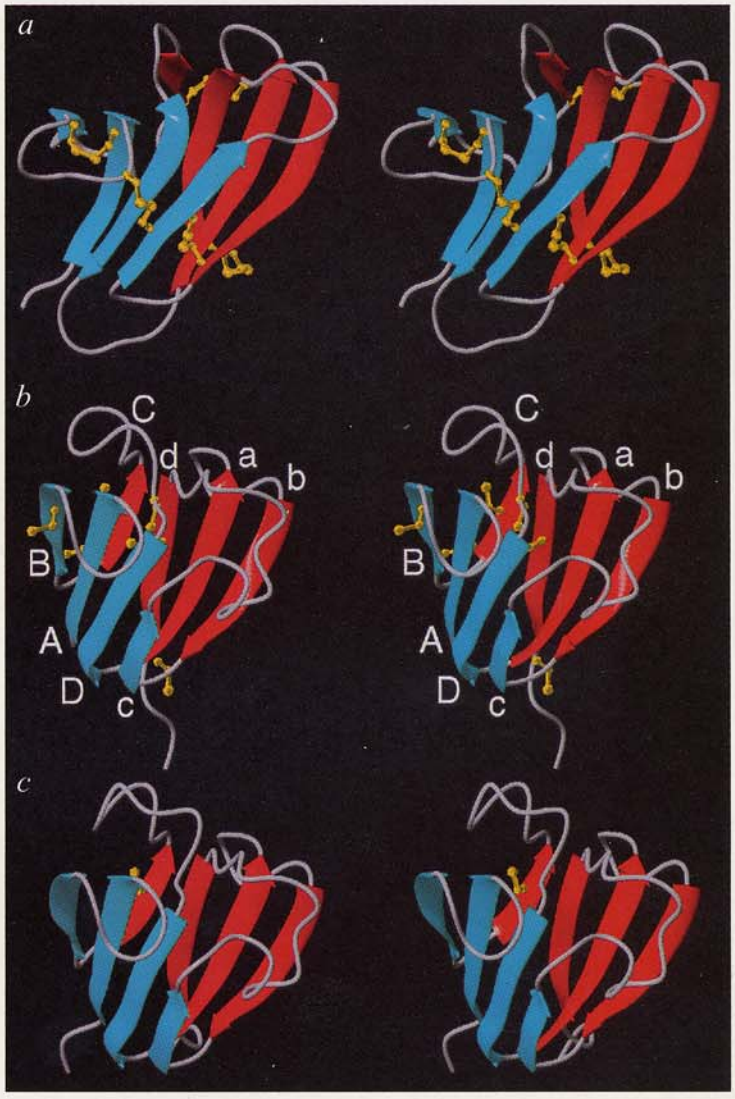

Fig. 1 Ribbon drawings of the protein domains of WmKT and γB-crystallin. *a*, NMR solution structure of the *Williopsis mrakii* killer toxin (WmKT). *b*, and *c*, First (γBC-I) and second (γBC-II) domain, respectively, of the X-ray crystal structure of γB-crystallin (Brookhaven Protein Data Bank entry 4GCR)[2,3]. All three structures are shown in identical orientations. The four strands of the first β-sheet are indicated by blue arrows, those of the second sheet by red arrows. In (*b*) the β-strands in the first greek key motif are further labelled A–D, and those in the second greek key motif a–d (see text). The polypeptide segments that link the regular β-structures are shown as a grey rope, and cysteinyl side chains are drawn as yellow ball and stick models. Drawings produced with the program MOLMOL[24].

γB-crystallin is an outstandingly stable monomeric protein of 174 amino acids, which was originally isolated from bovine eye lenses. Very similar proteins are present in high concentration in the eye lenses of other mammals, where they play an important role in maintaining tissue transparency[8]. High stability of γB-crystallin is essential for its function, since the nucleus of the lens is formed at the fetal stage and there is almost no protein turnover in the cells which form the central part of the lens[4,8]. In the crystal structure, each of two domains (γBC-I and γBC-II) is composed of two Greek key motifs (Fig. 1*b*,*c*) and the two domains are linked by a short connecting peptide[2,3]. Similar two-domain structures have been found for βB2-crystallin[9], and for a development-specific protein from the bacterium *Myxococcus xanthus*, protein S[10]. Sequence comparisons led to the identification of spherulin 3a from the slime mold *Physarum polycephalum*[11] as a possible candidate for the anticipated corresponding one-domain structure (Fig. 2*a*), but the 3D structure of spherulin 3a is not yet available.

The βγ-crystallin fold in the WmKT structure was revealed by a search with the program DALI[12] for 3D structures with similar folds in the Brookhaven Protein Data Bank. DALI compares protein structures by aligning distance matrices for the Cα atoms so that exclusively 3D structure information is used. The similarity of the folds of WmKT and the two domains of γB-crystallin is readily apparent (Fig. 1), although there are noticeable differences in the length of corresponding loops that link regular secondary structure elements. The topology of the two β-sheets is the same in all three domains, with root-mean-square deviations (r.m.s.d.s) calculated between WmKT and either γBC-I or γBC-II for the backbone atoms N, Cα and C' of the coinciding segments of the β-strands of 2.4 Å for the first β-sheet, and 1.9 Å for the second sheet.

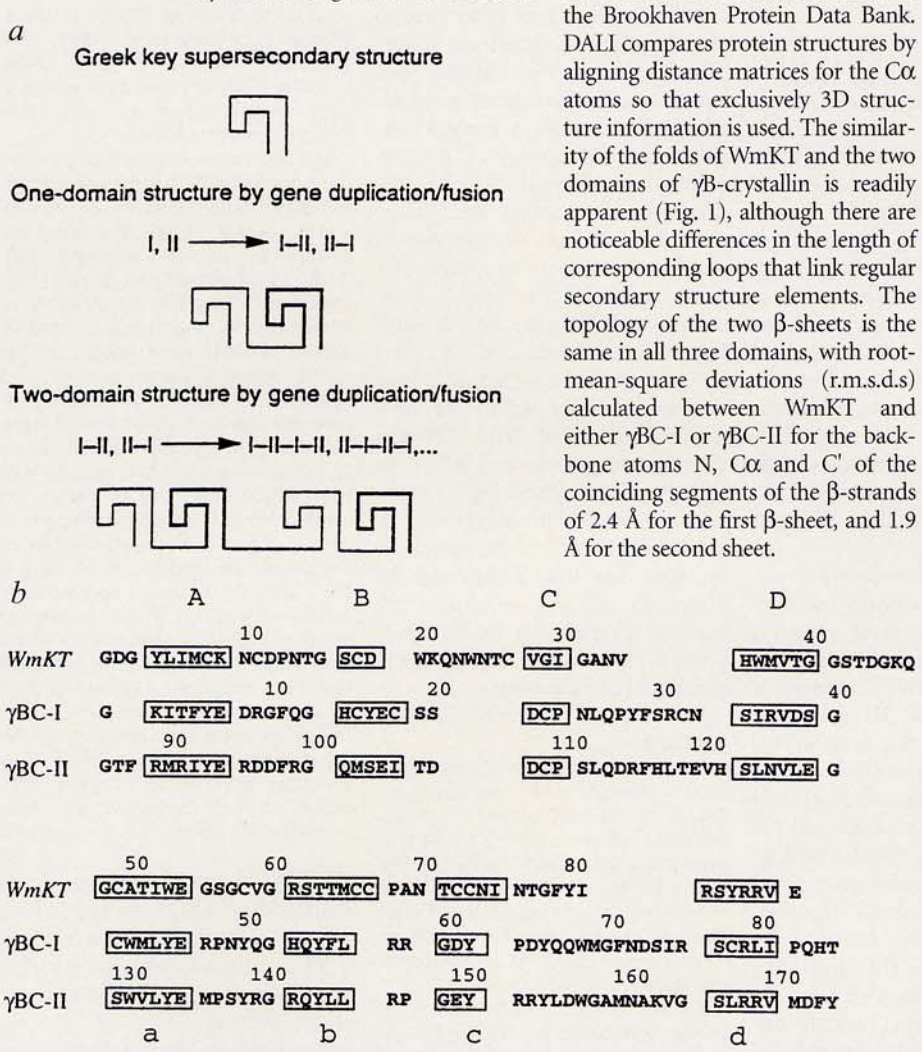

Fig. 2 *a*, Proposed gene duplication and gene fusion events in the evolution of βγ-crystallins[4,5,8]. These proteins form two-domain structures, with each domain composed of two similar Greek key motifs. *b*, 3D structure-based alignment of the *Williopsis mrakii* killer toxin, WmKT, and the two domains of γB-crystallin, γBC-I and γBC-II,[2,3] without consideration of sequence homology. β-strands are indicated by boxes and labelled A–D and a–d for the first and second Greek key motif, respectively (see Fig. 1). Corresponding β-strands are aligned for best structural coincidence, that is, such that the side chains of matched residues have the same orientation with respect to the planes of the β-sheets[12]. Residues outside of the β-strands have not been aligned.

WmKT shows no significant sequence homology to any protein of known 3D structure. Attempts at classification of the amino acid sequences for 3D structure motifs with the program PROSITE[13] even failed to assign WmKT to a specific structural class, since the consensus sequence for identification of greek key motifs did not match those in WmKT sufficiently closely (Fig. 2*b*). The alignment of WmKT and the γB-crystallin domains in Fig. 2*b* was based entirely on 3D structure similarities. Among the 37 corresponding residues within the β-sheets of the three domains, 18 (49%) are conserved between the two γB-crystallin domains, but only 6 and 7 (16% and 19%) are identical in WmKT and γBC-I, and WmKT and γBC-ll respectively. The first Greek key motif shows significantly higher divergence among the three sequences (39, 11 and 6% identity between γBC-I and γBC-II, WmKT and γBC-I, and WmKT and γBC-II respectively) than the second one (corresponding figures: 58, 21 and 32%).

The β-hairpin loops between strands A and B, or a and b, respectively, provide further evidence for closer similarity of the second Greek key motif between WmKT and γBC-I or γBC-II. The loop is one residue longer in the first Greek key motif of WmKT, and the conserved Gly residue (residue 16 in WmKT) at the end of the loop does not adapt a ϕ-value >0°, which is characteristic of the second Greek key motif in WmKT and in both domains of γB-crystallin[2–4]. In addition, the otherwise conserved, completely buried Ser (residues 83 in WmKt, 34 and 77 in γBC-I, 123 and 126 in γBC-II) at the beginning of β-strands D, and d respectively, is replaced in the first Greek key motif of WmKT by the partially solvent-exposed His 35. The concomitant removal of the characteristic hydrogen bonds between the Ser side chain and the aforementioned β-hairpin loop is compensated for by stacking interactions and hydrogen bonding with the sequentially neighbouring Trp 36 side chain. The high conservation of the Gly residue preceding the second β-strands, B and b, and of the Ser residue in the first position of the fourth β-strands, D and d, previous-

Table 1 Structural statistics for the WmKT NMR structure[1]

DIANA target function (Å^2)	3.0 ± 0.6
AMBER energies (kcal mol^{-1}):	
Physical	−2953 ± 47
Van der Waals	−296 ± 11
Electrostatic	−3029 ± 49
NOE constraint violations	
Number ≥ 0.1 Å	1.0 ± 0.9
Sum (Å)	9.5 ± 0.6
Maximum (Å)	0.1 ± 0.0
Dihedral angle constraint violations:	
Number ≥ 2.0°	0.8 ± 0.3
Sum (Å)	30.0 ± 4.6
Maximum (°)	2.1 ± 0.3
R.m.s.d. to the mean for N, Cα and C' of residues 1–88 (Å):	1.08 ± 0.17

[1]Average values and standard deviations are listed for the group of 20 energy-minimized DIANA conformers. The target function value is given for the 20 DIANA conformers before energy minimization. All other data are for the same group of conformers after energy-refinement with the program OPAL.

ly formed the basis for predicting that protein S and spherulin 3a contain the βγ-crystallin fold[11]. Clearly, the alignment of Fig. 2*b* could not have been obtained without reference to the 3D structures, since corresponding loops have different lengths in the three domains and the residues used to assemble the hydrophobic core are not conserved, which results in different packing arrangements of the core and slightly different relative orientations of the two β-sheets in the three domains (Fig. 1).

Examination of hydrophobic contacts and disulphide bonds indicates that different structural elements have been used in the evolution of γB-crystallin and WmKT to arrive at outstandingly stable 3D structures. On the one hand the X-ray crystal structure of γB-crystallin[2,3] shows that the hydrophobic residues Met 43 and Phe 56 in γBC-I, and Val 132 and Leu 136 in γBC-II mediate hydrophobic interdomain contacts[14], whereas in WmKT the corresponding positions contain the hydrophilic residues Thr 51 and Thr 64 (Fig. 2*b*). On the other hand, all six Cys residues in γBC-I and the single Cys in γBC-II are in the reduced state under physiological conditions[2,3], whereas a network of disulphide bonds appears to be necessary for the stability of the single domain WmKT structure. The disulphide bonds cross-link a hairpin in each of the two β-sheets (Cys 8–Cys 18 and Cys 49–Cys 66), anchor the two β-sheets against each other (Cys 67–Cys 73), and tie two loops to the regular secondary structures (Cys 11–Cys 72 and Cys 27–Cys 58)[1].

In conclusion, WmKT provides a first example of a single-domain protein structure that corresponds to the predicted ancestral fold from which the two-domain proteins in the βγ-crystallin superfamily are believed to derive[4,5,8]. The WmKT structure shows that this fold can exist as a stable single-domain structure and thus provides direct support for the proposed model of evolution for the βγ-crystallins (Fig. 2*a*)[4]. The two domains of γB-crystallin are more closely related to each other than to WmKT (Fig. 1), which is readily rationalized if one assumes that WmKT has evolved from a similar ancestral one-domain form as the two-domain structures of the βγ-crystallin superfamily (Fig. 2*a*): the first gene duplication and fusion event that lead from a single primordial Greek key motif to a stable single-domain protein must have occurred much earlier than the subsequent gene duplication and gene fusion steps that resulted in formation of the βγ-crystallin-type two-domain structure. The different evolutionary pressure exerted by the widely different physiological functions of yeast killer toxins and of structural proteins of the vertebrate eye lens may have accelerated the divergence of the sequences, for example with respect to the disulphide bridges, while maintaining the global fold and topology.

Methods

Cells of the *Williopsis mrakii* strain IFO 0895 (Institute of Fermentation, Osaka, Japan) were grown in a bioreactor (9 l culture medium) for 21 h at pH 6.0 and 28 °C with uniformly ^{15}N-labelled $(NH_4)_2SO_4$ (Isotec) as the sole nitrogen source. 44 mg of WmKT were purified from the centrifuged medium by two cation exchange chromatography steps and ultrafiltration[1,15].

The NMR structure determination of WmKT was based on NMR measurements with Varian Unity-*plus* 750 and Bruker AMX 600 spectrometers, with samples of 4 mM unlabelled WmKT and 3 mM uniformly ^{15}N-labelled WmKT in aqueous solution at pH 5.0 and 36 °C. Complete sequence-specific ^{1}H and ^{15}N resonance assignments for the backbone and ^{1}H assignments for the aliphatic side chains were obtained with the NOE-based sequential assignment method[16–18], using the program XEASY[19] for the spectral analysis. 1053 upper distance bounds were obtained from two NOESY spectra recorded with a mixing time of 40 ms in H_2O and ^2H_2O solution, and 206 dihedral angle restraints resulted from combined analysis of the local NOEs and the scalar coupling constants with the program HABAS[20]. For the final structure calculation we used the program DIANA[21] with the standard REDAC protocol[22], starting from a set of 100 randomized conformers. The 20 best DIANA conformers were energy-minimized with the AMBER force field[23] in the program OPAL (P. Luginbühl, P. Güntert, M. Billeter, K. Wüthrich, personal communication), and the refined conformers are used to represent the solution structure of WmKT (Table 1, Protein Data Bank entry 1WKT). The disulphide bridges 8–18, 11–72, 27–58, 49–66 and 67–73 could be unambiguously established during the NMR structure determination.

We attempted sequence alignments using the PROSITE[13] consensus pattern for identification of greek key motifs, which consists of the following 18-residue polypeptide segment: x(2)-[LIVMFYWA]-x(1)-[DEHRKSTP]-[FY]-[DEQHKY]-x(3)-[FY]-x(1)-G-x(4)-[LIVMFCST], where one of the residues listed in square brackets must appear at the corresponding sequence position, and x(n) denotes a stretch of n residues of arbitrary type. In the first WmKT greek key motif, the aromatic residue Phe or Tyr at position 6 in the consensus pattern is substituted by Cys, there is an insertion in the loop connecting the β-strands A and B that causes a shift of the conserved residue Gly 13 (Figs 1, 2*b*), and there is a non-aromatic residue at position 11. In the second motif of WmKT the positions 5 and 6 are occupied by Ile and Trp, and position 11 is occupied by Cys rather than by an aromatic residue. These deviations were apparently sufficient for the otherwise highly regular greek key structures in WmKT to escape identification, so that the above consensus sequence may need to be re-examined.

Walfrido Antuch[1,2], Peter Güntert[1] and Kurt Wüthrich[1]
[1]Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland
[2]Present address: Department of Biochemistry, University of Connecticut Health Center, Farmington, Connecticut 06032, USA

Correspondence should be addressed to K.W.

Received 25 March; accepted 9 May 1996.

Acknowledgements We thank Y. Furuichi and H. Senn for initiating the NMR structure determination of WmKT, W. Sidler for help in the preparation of WmKT, R. Marani for the careful processing of the typescript, and R. Glockshuber and one of the anonymous referees for helpful comments.

1. Antuch, W. Ph.D. thesis 11401, ETH Zürich, Switzerland (1995).
2. Blundell, T. *et al. Nature* **289**, 771–777 (1981).
3. Najmudin, S. *et al. Acta Cryst.* D**49**, 223–233 (1993).
4. Wistow, G. *Trends Biochem. Sci.* **18**, 301–306 (1993).
5. Branden, C. & Tooze, J. *Introduction to Protein Structure*, pp. 67 ff. (Garland, New York, NY, 1991).
6. Yamamoto, T., Hiratani, T., Hirata, H., Imai, M. & Yamaguchi, H. *FEBS Lett.* **197**, 50–54 (1986).
7. Yamamoto, T., Imai, M., Tachibara, K. & Mayumi, M. *FEBS Lett.* **195**, 253–257 (1986).
8. Wistow, G. & Piatigorsky, J. *Ann. Rev. Biochem.* **57**, 479–504 (1988).
9. Bax, B. *et al. Nature* **347**, 776–780 (1990).
10. Bagby, S., Harvey, T.S., Eagle, S.G., Inouye, S. & Ikura, M. *Structure* **2**, 107–122 (1994).
11. Wistow, G. *J. molec. Evol.* **30**, 140–145 (1990).
12. Holm, L. & Sander, C. *J. Mol. Biol.* **233**, 123–138 (1993).
13. Bairoch, A. & Bucher, P. *Nucleic Acids Res.* **22**, 3583–3589 (1994).
14. Mayr, E.M., Jaenicke, R. & Glockshuber, R. *J. Mol. Biol.* **235**, 84–88 (1994).
15. Kasahara, S. *et al. FEBS Lett.* 348, 27–32 (1994).
16. Billeter, M., Braun, W. & Wüthrich, K. *J. Mol. Biol.* **155**, 321–346 (1982).
17. Wagner, G. & Wüthrich, K. *J. Mol. Biol.* **155**, 347–366 (1982).
18. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, New York, NY, 1986).
19. Bartels, C., Xia, T.H., Billeter, M., Güntert, P., Wüthrich, K. *J. biomol. NMR* **5**, 1–10 (1995).
20. Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. *J. Am. Chem. Soc.* **111**, 3997–4004 (1989).
21. Güntert, P., Braun, W. & Wüthrich, K. *J. Mol. Biol.* 217, 517–530 (1991).
22. Güntert, P. & Wüthrich, K. *J. Biomol. NMR* 1, 447–456 (1991).
23. Singh, U.C., Weiner, P.K., Caldwell, J.W., Kollman, P.A. *Amber 3.0* (University of California, San Francisco, CA) (1986).
24. Koradi, R., Billeter, M. & Wüthrich, K. *J. Mol. Graphics* **14**, 51–55 (1996).