The EMBO Journal vol.15 no.24 pp.6810–6821, 1996

Giant protein kinases: domain interactions and structural basis of autoregulation

Bostjan Kobe[1,2], Jörg Heierhorst[1,3], Susanne C.Feil[1], Michael W.Parker[1], Guy M.Benian[4], Klaudiusz R.Weiss[3] and Bruce E.Kemp[1,5]

[1]St Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia, [3]Department of Physiology and Biophysics, Mount Sinai School of Medicine, 1 Gustave L.Levy Place, New York, NY 10029 and [4]Department of Pathology, Emory University, Atlanta, GA 30322, USA

[2]On leave from: Jozef Stefan Institute, Department of Biochemistry and Molecular Biology, Jamova 39, 61111 Ljubljana, Slovenia

[5]Corresponding author

B.Kobe and J.Heierhorst contributed equally to this paper

The myosin-associated giant protein kinases twitchin and titin are composed predominantly of fibronectin- and immunoglobulin-like modules. We report the crystal structures of two autoinhibited twitchin kinase fragments, one from *Aplysia* and a larger fragment from *Caenorhabditis elegans* containing an additional C-terminal immunoglobulin-like domain. The structure of the longer fragment shows that the immunoglobulin domain contacts the protein kinase domain on the opposite side from the catalytic cleft, laterally exposing potential myosin binding residues. Together, the structures reveal the cooperative interactions between the autoregulatory region and the residues from the catalytic domain involved in protein substrate binding, ATP binding, catalysis and the activation loop, and explain the differences between the observed autoinhibitory mechanism and the one found in the structure of calmodulin-dependent kinase I.

Keywords: crystal structure/immunoglobulin/intrasteric regulation/protein kinase/twitchin

Introduction

Giant protein kinases twitchin, projectin and titin have remarkable molecular masses ranging from 700 to >3000 kDa. The majority of the polypeptide is built of multiple repeats of immunoglobulin (Ig)-like and fibronectin type III (Fn3)-like domains ranging from 61 in nematode twitchin (Benian *et al.*, 1989, 1993) to 297 in skeletal muscle titin (Labeit and Kolmerer, 1995). A single autoinhibited protein kinase domain is located near the C-terminus (Benian *et al.*, 1989; Hu *et al.*, 1994; Ayme-Southgate *et al.*, 1995; Labeit and Kolmerer, 1995) (Figure 1). Twitchin's role in the regulation of contraction–relaxation cycles and in myofilament assembly was first apparent in *Caenorhabditis elegans* where homozygous mutants of the *unc-22* gene (*unc*; uncoordinated) were characterized by impaired movement, a constant twitch of their body wall muscles and an abnormal sarcomere structure (Waterston *et al.*, 1980; Dibb *et al.*, 1985). The largest of the giant protein kinases, titin, an ~1 μm long molecule, spans the entire half-sarcomere (from the Z-line to the M-line) of striated muscle and acts as an elastic cytoskeletal ruler for myosin thick filament assembly (Trinick, 1994). The numerous Fn3 and Ig domains bind myosin and myosin-associated proteins. Recently, the C-terminal tail of the myosin heavy chain was shown to contain a titin binding site (Houmeida *et al.*, 1995).

The kinase domains of the giant protein kinases are related to vertebrate smooth muscle myosin light chain kinases (smMLCKs) and share an identical pattern of the surrounding Fn3 and Ig domains (Olson *et al.*, 1990) (Figure 1). The Ig domain C-terminal to the protein kinase domain in smMLCK can be transcribed independently from the myosin light chain kinase gene and expressed as the protein telokin (Shirinsky *et al.*, 1993). Actin (Kanoh *et al.*, 1993) and myosin (Shirinsky *et al.*, 1993) binding sites are present at the N- and C-termini of smMLCK, respectively. Vertebrate MLCKs have masses ranging from 68 to 208 kDa and phosphorylate the regulatory light chains of myosin II, initiating contractions in smooth muscle and non-muscle cells, or enhancing contractions in striated muscle (Kamm and Stull, 1985). *Aplysia* twitchin also phosphorylates the MLCKs and may serve dual cytoskeletal and regulatory roles in molluscan muscle (Heierhorst *et al.*, 1995).

Giant protein kinases, MLCKs and other protein kinases such as calmodulin (CaM)-dependent protein kinases (CaMK-I–IV) are regulated by an intrasteric mechanism (Kemp *et al.*, 1994). An autoregulatory sequence maintains an inactive state; activation occurs when this sequence is sequestered by an activator protein [Ca^{2+}/CaM in MLCKs and CaMKs (Kemp *et al.*, 1994) and Ca^{2+}/S100A1$_2$ in twitchin (Heierhorst *et al.*, 1996)]. The crystal structure of a fragment of the twitchin kinase (5890–6262) containing the catalytic core and the autoregulatory sequence showed that the autoinhibition resulted from the autoregulatory sequence binding to the active site. The autoregulatory sequence contains a prominent α-helix in the substrate binding groove that makes substrate-like contacts with the catalytic core, while the rest of the sequence occupies part of the ATP binding pocket and makes contacts with critical catalytic residues (Hu *et al.*, 1994). Recently, the structure of a CaM-regulated protein kinase, a fragment of CaMK-I, has been reported (Goldberg *et al.*, 1996). The autoregulatory sequence in CaMK-I also contains a strikingly similar α-helix that makes substrate-like contacts, but the structure of the rest of the sequence differs dramatically from twitchin. The twitchin autoregulatory sequence exits the active site along the substrate binding groove, passing over the enzyme's activation loop, in marked contrast to the CaMK-I

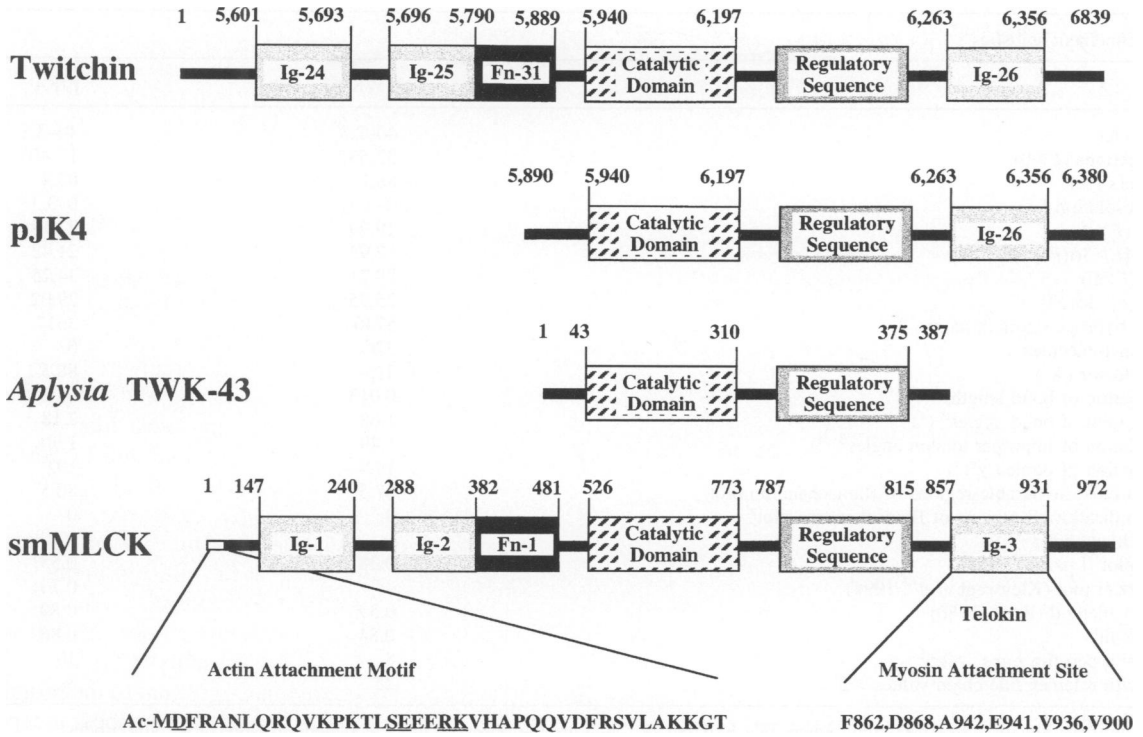

Fig. 1. Schematic diagrams of domain arrangements. Domain arrangements in the nematode twitchin (Benian *et al.*, 1993), the crystallized fragments of nematode (pJK4) and *Aplysia* (TWK-43) twitchins and chicken gizzard MLCK (Olson *et al.*, 1990) (smMLCK) are shown. Actin and myosin attachment sites on smMLCK are also shown.

sequence that turns in the opposite direction and avoids the activation loop entirely.

Here we report two new crystal structures of autoinhibited twitchin fragments: the 387 residue protein kinase from the marine mollusc *Aplysia* (TWK-43), and a longer, 490 residue protein kinase from *C.elegans* (pJK4) that includes the C-terminal Ig domain [number 26 in the twitchin sequence (Benian *et al.*, 1993); Ig-26 (Figure 1)]. These structures provide the first insight into the interdomain relationships of the modular giant protein kinase and smMLCK subfamilies, and advance our understanding of the regulatory mechanisms in intrasterically regulated protein kinases. The Ig-26 domain contacts the kinase domain opposite the active site and suggests a relatively rigid interdomain association. A model of the analogous domains in smMLCK reveals the juxtaposition of the protein kinase active site and the myosin binding site within the contractile apparatus. The structures highlight the importance of the contacts between the autoregulatory sequence and the activation loop, and provide an explanation for the structure of the autoregulatory sequence seen in the crystal structure of the CaMK-I fragment (1–320) (Goldberg *et al.*, 1996) that is missing the residues required to contact the activation loop.

Results and discussion

Overall structures of the twitchin kinase fragments

The *Aplysia* (TWK-43) and the *C.elegans* (pJK4) fragments comprise an autoinhibited protein Ser/Thr kinase domain plus the C-terminal Ig domain in pJK4 (Figure 1). The structures of TWK-43 and pJK4 were refined at 2.3 and 3.3 Å resolution, respectively; despite the low resolution limit of pJK4 crystals, the positions of the protein kinase and Ig-26 domains could be assigned unequivocally and the model refined using R_{free} (Brünger, 1992a) to prevent overfitting (Table I).

The protein kinase catalytic cores in both structures (residues 5932–6199 in pJK4; 43–310 in TWK-43) have a typical bi-lobal protein kinase topology (Knighton *et al.*, 1991; Goldsmith and Cobb, 1994). The smaller N-terminal lobe is dominated by an antiparallel β-sheet, and the larger C-terminal lobe is built mainly of α-helices (Figure 2). The two lobes are oriented differently in twitchin and cAMP-dependent protein kinase (cAPK) (Knighton *et al.*, 1991) with the small domain of twitchin rotated ~30° relative to the corresponding lobe in cAPK, similar to the insulin receptor (Hubbard *et al.*, 1994). The C-terminal autoregulatory sequence (residues 6202–6260 in pJK4; 312–370 in TWK-43) is positioned in the cleft between the two lobes and traverses the entire substrate binding groove.

The structure of the Ig-26 domain in pJK4 (Figure 2B) is very similar to that of telokin (Holden *et al.*, 1992), the analogous domain in MLCK (Figure 1); the r.m.s. deviation of 94 pairs of C^{α} atoms upon superposition with telokin is 1.46 Å (the sequences share 36% identity) (Figure 2C). In addition to the N- and C-terminal regions, the largest deviations between the structures of Ig-26 and telokin occur in the loops around residues 6273, 6292, 6308, 6316, 6327 and 6331 (Figure 2C). There is a deletion in the Ig-26 sequence when compared with telokin just before residue 6327, resulting in a shift of the preceding β-strand.

Table I. Refinement statistics

	TWK-43	pJK4
Resolution (Å)	40–2.3	40–3.3
No. of reflections ($F>0$)	33 532	17 405
Completeness (%)	88.1	83.4
R values: resolution (Å)	6–2.3	6–3.3
R_{cryst}[a] (%) ($F>0$)	19.99	27.19
$[F>3\sigma(F)]$	17.93	21.42
R_{free}[b] (%) ($F>0$)	28.71[c]	34.25
$[F>3\sigma(F)]$	25.95[c]	29.02
No. of non-hydrogen protein atoms	5746	3617
No. of water molecules	326	0
Average B-factor (Å^2)	31.4	88.5
R.m.s. deviation of bond lengths[d] (Å)	0.013	0.015
R.m.s. deviation of bond angles[d] (°)	1.68	2.18
R.m.s. deviation of improper torsion angles[d] (°)	1.49	1.90
R.m.s. deviation of pooled χ[e](°)	14.8	19.9
Residues in most favourable regions of Ramachandran plot[e]	87.5	80.5
Residues in disallowed regions of Ramachandran plot[e]	0	0
Mean coordinate error (Å)		
Luzzati plot (Luzzati, 1952)	0.28	0.55
R_{free} Luzzati plot (Kleywegt *et al.*, 1994)	–	0.70
SIGMAA method (Read, 1986)	0.39	0.89
Average RS-fit[f]	0.84	0.86
Residues with pep-flip r.m.s. >2.5[f]	8	10
Residues with rotamer side chain values >2[f]	25	8

[a]$R_{cryst} = \Sigma hkl(||F_{obs}hkl|-|F_{calc}hkl||)/|F_{obs}hkl|$, where $|F_{obs}hkl|$ and $|F_{calc}hkl|$ are the observed and calculated structure factor amplitudes.
[b]R_{free} is equivalent to R_{cryst} but calculated with 10% of reflections that are omitted from the refinement process.
[c]The last model before all reflections were included in the refinement (see Materials and methods).
[d]Calculated with the program X-PLOR (Brünger, 1992b).
[e]Calculated with the program PROCHECK (Laskowski *et al.*, 1993).
[f]Calculated with the program O (Jones *et al.*, 1991).

Insights into structural and functional modularity of giant protein kinases and myosin light chain kinases

Twitchin kinase and related molecules are modular multidomain proteins. The structures of some individual domains are known (Holden *et al.*, 1992; Hu *et al.*, 1994; Pfuhl and Pastore, 1995; Improta *et al.*, 1996), but little is understood about their interdomain relationships. While there is evidence that the Ig-like domains of the elastic I-band region of titin are interacting only weakly (Politou *et al.*, 1996), the protein kinase domain is a part of the globular head of the molecule (Labeit *et al.*, 1992) consistent with stronger interdomain interactions in this region.

In the structure of pJK4, Ig-26 contacts the protein kinase domain opposite from the active site (Figure 2B). Ig-26 interacts with the autoinhibited Ser/Thr protein kinase domain through loops spatially clustered near the N-terminus of the Ig domain (residues 6264 at the N-terminus of the domain; 6267 and 6269 at the beginning of β-strand A; 6290–6292 in the B–C loop; and 6344 in the F–G loop). The complementary surface of the protein kinase domain is formed by the sequences 6055–6056 and 6058 at the end of helix H5, 6089–6090 and 6092–6093 at the end of strand S11 and 6188–6189 at the beginning of helix H10. There are four putative hydrogen bonds in the interface. The 720 Å^2 of surface area buried between the two domains (segments 5915–6260 and 6264–6361 were considered separate domains in the calculation; the number is 1006 Å^2 if the linker 6261–6263 is considered a part of the Ig domain) is 54% hydrophobic and 15% charged. The linker sequence 372–374 is in a 3_{10}-helical conformation. The buried surface area and the number of contacts are larger than in most module–module interfaces in modular proteins (Leahy *et al.*, 1996, and references therein), suggesting a relatively rigid association of the two domains.

The regulatory subunit of cAPK is thought to be juxtaposed to the catalytic subunit similarly to Ig-26 in twitchin, with its inhibitory sequence protruding into the substrate binding groove (Gibbs *et al.*, 1992). Furthermore, cyclin is positioned in a similar manner close to the Thr-rich activation loop in the cyclin A–cyclin-dependent kinase 2 complex (Jeffrey *et al.*, 1995); the Ig domain, however, is not required for activation of twitchin kinase by Ca^{2+}/S100A1$_2$ (Heierhorst *et al.*, 1996).

The arrangement of the protein kinase and adjacent Ig domain is similar in both titin and smMLCK (Figure 1). Because of the high structural similarity of MLCK and the corresponding region of twitchin, the architectures and certain functional aspects of these molecules are most likely conserved. Based on the structure of the twitchin fragment pJK4, we have modelled the corresponding domains in smMLCK (Figure 3). The model provides important clues as to how MLCK may anchor to myosin. Of the 12 putative myosin binding residues in telokin inferred from a sequence comparison of several homologous Ig domains of the myosin binding protein-C (Okagaki *et al.*, 1993), the six exposed residues cluster on the same face of the Ig domain (Figure 3), appropriate for possible interactions with myosin. The titin binding site identified in the tail of the skeletal muscle myosin heavy chain corresponds to a 17 residue sequence LEARV-RELEAEVESEQK (Houmeida *et al.*, 1995). In chicken gizzard myosin, the corresponding sequence motif

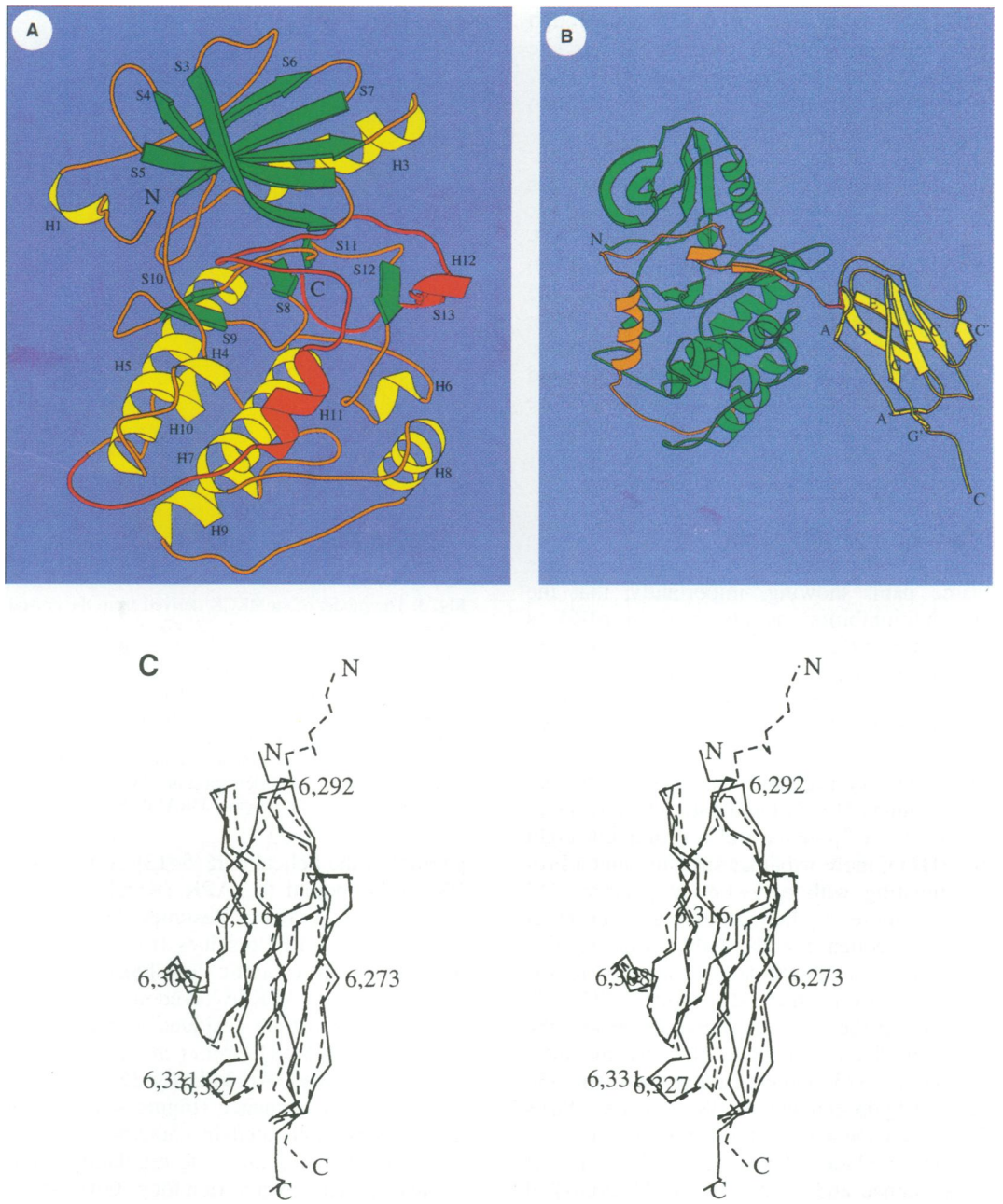

Fig. 2. The structures of TWK-43 and pJK4. (**A**) Ribbon diagram (Kraulis, 1991) of the three-dimensional structure of TWK-43. α-Helices are coloured yellow, β-strands green and the autoregulatory sequence red. Elements of secondary structure are labelled to resemble the designation in nematode twitchin (Hu *et al.*, 1994): α-helices: H1, 33–36; H3, 89–102; H4, 135–138; H5, 148–167; H6, 219–222; H7, 229–245; H8, 255–264; H9, 279–286; H10, 299–303; H11, 325–332; H12, 359–362; β-strands: S3, 53–62; S4, 65–72; S5, 78–85; S6, 113–118; S7, 122–128; S8, 170–171; S9, 180–182; S10, 190–192; S11, 199–200; S12, 207–210; S13, 364–367. The hinge region between the two lobes consists of two adjacent Gly residues (132–133 in TWK-43, 6021–6022 in pJK4). After separate superposition of individual lobes, the r.m.s. deviations of C^{α} atoms of the catalytic core of TWK-43 (residues 43–310) and pJK4 (residues 5932–6199) with the corresponding residues of cAPK (residues 34–299) are 2.4 Å for 258 residue pairs (1.1 Å for 205 residue pairs of TWK-43 and 196 residue pairs of pJK4 that are <2 Å apart). N-terminal to the catalytic core, the structures diverge from cAPK at residue 49 of TWK-43 and 5938 of pJK4 (40 of cAPK); the sequences N-terminal to these residues contain a short 3_{10}-helix and a short α-helix (H1) in both TWK-43 and pJK4. (**B**) Ribbon diagram of the three-dimensional structure of pJK4. The protein kinase domain is coloured green, the autoregulatory sequence brown, Ig-26 yellow and the linker between the protein kinase and Ig-26 domains (residues 6261–6263) red. Elements of secondary structure in Ig-26 are labelled to resemble the designation in similar Ig domains (Harpaz and Chothia, 1994): β-strands: A, 6267–6270; A′, 6276–6278; B, 6283–6290; C, 6296–6301; C′, 6304–6305; E, 6320–6325; F, 6336–6342; G, 6345–6351; G′, 6354–6356. Ig-26 belongs to the I set of the Ig superfamily (Harpaz and Chothia, 1994) and contains 18 of the 20 key residues that characterize the I set fold (the two discrepancies are Tyr at position D2 and Val at position F5, both parts of the hydrophobic core; Harpaz and Chothia, 1994). These structures consist of two major antiparallel β-sheets packed against each other. The corresponding β-sheets in Ig-26 comprise strands A–B–E and G–F–C–C′, respectively. Usually, the first sheet contains an additional strand D; although Ig-26 contains an elongated structure in the sequence 6313–6316 that packs against strand E, we have not assigned it as a part of the sheet (Kabsch and Sander, 1983). Furthermore, the strand A′ is usually a part of the second sheet. In Ig-26, however, the strands A′ and G′ form a third, two-stranded β-sheet. Residues 6352–6353 separate β-strand G from G′, thus isolating the third sheet from the second. (**C**) Stereo diagram of the superimposed C^{α} traces of Ig-26 (residues 6261–6371; full lines) and telokin (Holden *et al.*, 1992) (dashed lines). Indicated are the N- and C-termini of both fragments and the Ig-26 residues in the regions deviating most from the structure of telokin.

LEAKIASLEEQLEQEAR may anchor the MLCK through its C-terminal Ig domain. In such a case, the binding site would be distal to the myosin head, causing the MLCK catalytic core to be juxtaposed to a different myosin molecule further along the thick filament.

In smooth muscle, MLCK is also attached to the actin-containing thin filament. The actin binding domain resides in the N-terminal region of smMLCK (Kanoh *et al.*, 1993). We note that this sequence 1MDFRANLQRQVKP-KTLSEERKVHAPQQVDFRSV34 (Olson *et al.*, 1990) has features of the repetitive actin binding sites of the giant actin binding protein nebulin (Pfuhl *et al.*, 1996); the underlined residues match the consensus sequence of the nebulin repeats, and additional hydrophobic and positively charged residues conform to the signature features of these actin binding sequences. This suggests that smMLCK may exploit the same mechanism of interaction with actin as nebulin does in striated muscle.

The mechanism of autoinhibition

The autoregulatory sequences in both twitchin fragments follow the same path, showing, importantly, that the mechanism of autoinhibition as observed in pJK4 is conserved in the shorter fragments, and also conserved in twitchin molecules from evolutionally distant organisms sharing only 60% identity. The autoinhibitory mechanism will be discussed in the higher resolution structure of TWK-43.

The structure of the autoregulatory sequence starts with a turn centred at residues 313–314 and a stretch of irregular structure followed by a 3_{10}-helix and a prominent eight residue α-helix (H11); more irregular structure and a Pro-rich region terminating with a *cis*-Pro at position 343 continue into two more 3_{10}-helices and another short α-helix (H12); the sequence ends with a four residue β-strand (S13) bound in an antiparallel manner to the strand S12 and a turn centred at residues 372–373. The interface between the autoregulatory region and the catalytic core is remarkably extensive. Between the autoregulatory region (312–375) and the catalytic domain (43–310), there are 40 hydrogen bonds, 388 van der Waals contacts and 18 water-mediated interactions (Figure 4A). The direct contacts include 47 residues (179 atoms) of the regulatory sequence and 77 residues (193 atoms) of the catalytic domain. The accessible surface area buried in the interface between the autoregulatory sequence and the catalytic core amounts to 6647 Å^2 (21% of the surface of the catalytic core and 47% of the surface of the regulatory sequence are buried). Of this buried surface, 64% is non-polar and 15% is charged. Of the charged buried surface area of the regulatory sequence, 74% is positively charged. The buried accessible surface area would correspond to a decrease in the calculated free energy of folding (Eisenberg and McLachlan, 1986) of 20 and 22 kcal/mol for the catalytic and the regulatory region, respectively. The degree of shape complementarity (Lawrence and Colman, 1993) between the interacting surfaces of the catalytic and regulatory regions is comparable with protein–protein interfaces. Many hydrophobic and basic residues of the regulatory region form highly complementary surfaces interacting with the catalytic domain (Figure 4B and C).

Helix H11 (325–332) in the autoregulatory sequence is

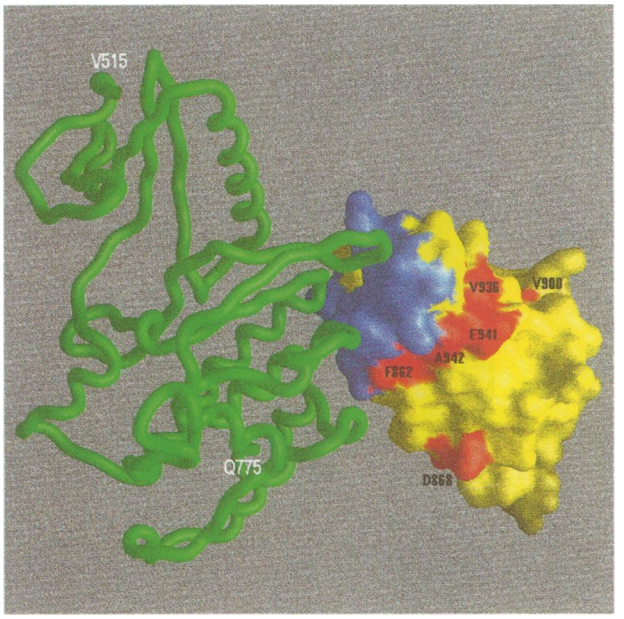

Fig. 3. The model of smMLCK derived from the crystal structure of pJK4. The catalytic core (515–775) is shown in green in a worm representation and the Ig (telokin) domain is shown in a surface representation. The surface formed by the residues of the Ig domain putatively interacting with the catalytic core is coloured blue; the surface formed by the residues putatively involved in myosin binding site is coloured red. The autoregulatory sequence connecting the catalytic core to the Ig domain is omitted. The orientation of the model is similar to the orientation of pJK4 in Figure 2B. The figure was prepared with the program GRASP (Nicholls *et al.*, 1991).

parallel to the helical part (5–13) of the inhibitory peptide PKI(5–24) bound to cAPK (Knighton *et al.*, 1991) but located ~10 Å away. Residues 330–331 are positioned similarly to the PKI residues 16–17. Glutamates 134, 178 and 215 in the catalytic core [corresponding to Glu127, 170 and 203 of cAPK involved in electrostatic interactions with basic substrate residues at the P-3, P-2 and P-6 sites, respectively (Kemp *et al.*, 1994)] form close polar interactions with Arg347, Lys334 and Lys330 of the autoregulatory sequence (Figure 4A and D). Essentially all residues implicated in catalysis (Taylor and Radzio-Andzelm, 1994) contact the regulatory sequence (Figure 4A and E). The glycine-rich loop (Gly60–Gly65) involved in anchoring the β-phosphate of ATP contacts many residues between 338 and 362 of the regulatory region. The side chain nitrogen of Lys82, equivalent to Lys72 of cAPK that neutralizes the charges of α- and β-phosphates of ATP, contacts the backbone carbonyls of the inhibitory residues 348 and 351. Glu98 (Glu91 of cAPK involved in correctly positioning Lys72) hydrogen-bonds Ser353 of the regulatory region. The catalytic base Asp174 (Asp166 of cAPK) forms a salt link with Arg355 in the autoregulatory sequence. Asp194 (Asp184 in cAPK involved in Mg^{2+} binding) binds Phe351. The TWK-43 regulatory region is buried deep in the ATP binding site and forms close contacts with practically all residues that interact with ATP in cAPK (Bossemeyer *et al.*, 1993) (Figure 4A and E). The regulatory sequence also makes extensive contacts with the Thr-rich activation loop (residues 194–220) that is the target of regulation by phosphorylation in many protein kinases, but not twitchin, MLCK and CaMK-II (Figure 4A and F); consequently, the activation loop

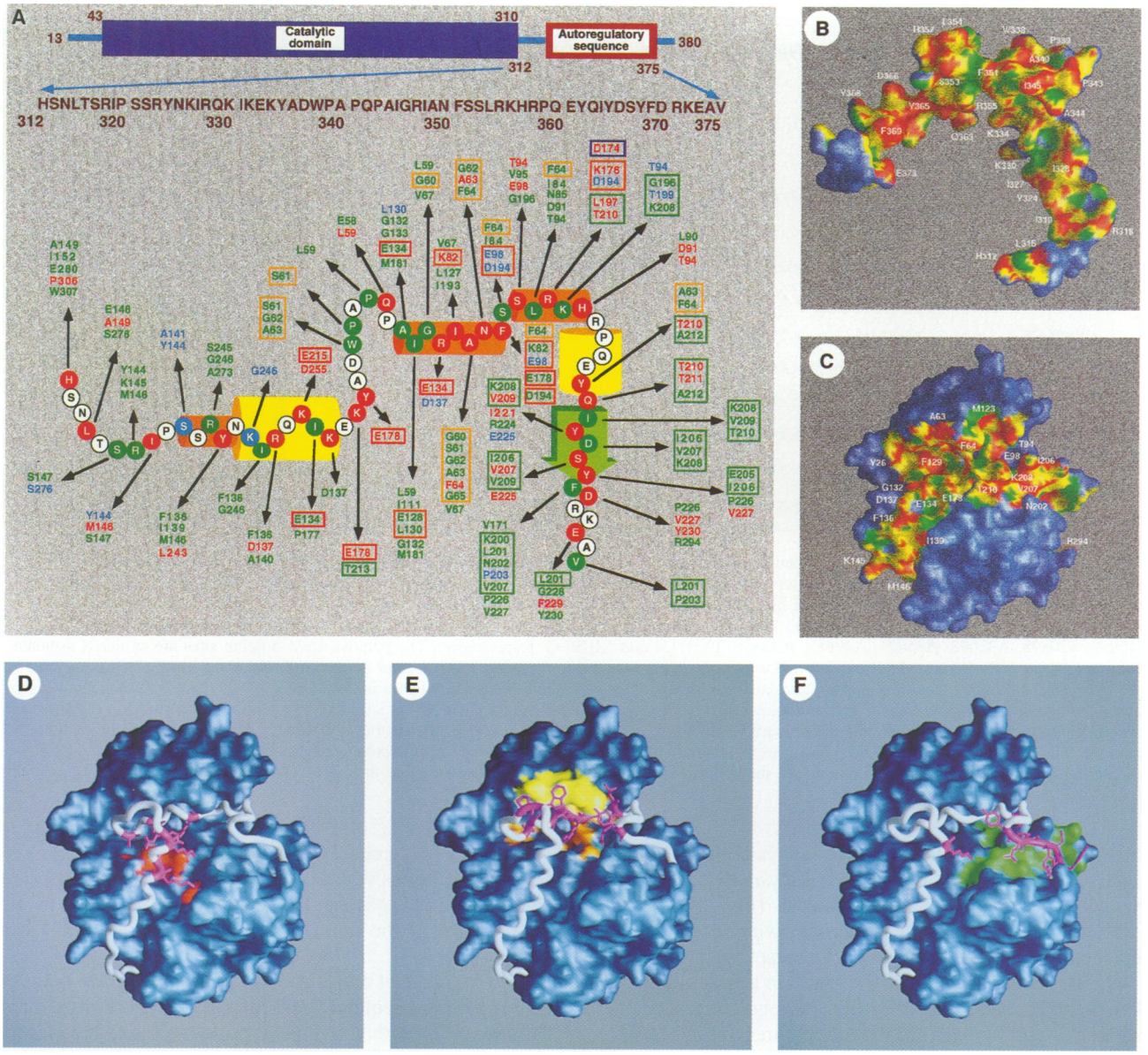

Fig. 4. The mechanism of autoinhibition. (**A**) Contacts between the autoregulatory sequence (312–375) and catalytic core (43–310) of TWK-43. The single-letter code is used for amino acids. Elements of secondary structure are depicted as wide yellow cylinders (α-helices), narrow orange cylinders (3_{10}-helices) and green arrows (β-strands). The residues coloured red form H bonds, the residues coloured green van der Walls contacts but no H bonds, and the blue residues form only water-mediated contacts. Residues of the catalytic core are shown boxed in red if involved in substrate binding (Kemp *et al.*, 1994), boxed in orange if involved in ATP binding (Bossemeyer *et al.*, 1993), boxed in yellow if a part of the phosphate anchoring loop, boxed in green if a part of the substrate anchoring (activation) loop and boxed in blue if involved in catalysis (Taylor and Radzio-Andzelm, 1994). (**B**) Molecular surface (probe radius 1.4 Å) of the regulatory region of TWK-43 colour-coded according to surface complementarity (Lawrence and Colman, 1993). Red, S_c >0.76; yellow, 0.76 >S_c >0.3; green, 0.3 >S_c >–0.3; light blue, S_c <–0.3. Selected residues are labelled. The shape correlation statistic for the regulatory sequence–catalytic domain interface, S_c equals 0.65; S_c is 1.0 for two perfectly complementary surfaces, around zero for surfaces with no complementarity and between 0.64 and 0.68 for antibody–antigen interfaces (Lawrence and Colman, 1993). (B)–(F) were prepared with GRASP (Nicholls *et al.*, 1991). (**C**) Molecular surface of the catalytic domain of TWK-43 colour-coded according to surface complementarity as in (B). The orientation is related to (B) by a 180° rotation around the *y* axis. (**D**) Interactions of the autoregulatory sequence of TWK-43 (white worm; interacting residues in magenta) with substrate binding residues (red) of the catalytic core (shown in blue in a surface representation). (**E**) Interactions of the autoregulatory sequence of TWK-43 (white worm; interacting residues in magenta) with ATP binding residues (Gly-rich loop in yellow, others in orange) and the catalytic Asp174 (cyan) of the catalytic core (shown in blue in a surface representation). (**F**) Interactions of the autoregulatory sequence of TWK-43 (white worm; interacting residues in magenta) with the Thr-rich loop (green) of the catalytic core (shown in blue in a surface representation).

adopts a different conformation from that in active protein kinases (Goldsmith and Cobb, 1994).

While the major autoinhibitory contacts are conserved in the *Aplysia* and *C.elegans* twitchin kinase structures, the following differences illustrate well the adaptability of the autoinhibitory mechanism. In *Aplysia* twitchin, Ile331 forms an unusual contact with Glu134 (P-3 site), but in nematode twitchin the corresponding Ile interacts with the equivalent of Glu178, an alternative substrate binding residue (P-2 site). Instead, the nematode equivalent of Glu134 (Glu6023) interacts with the hydroxyl of Tyr6241 in the autoregulatory sequence. In *Aplysia*, this

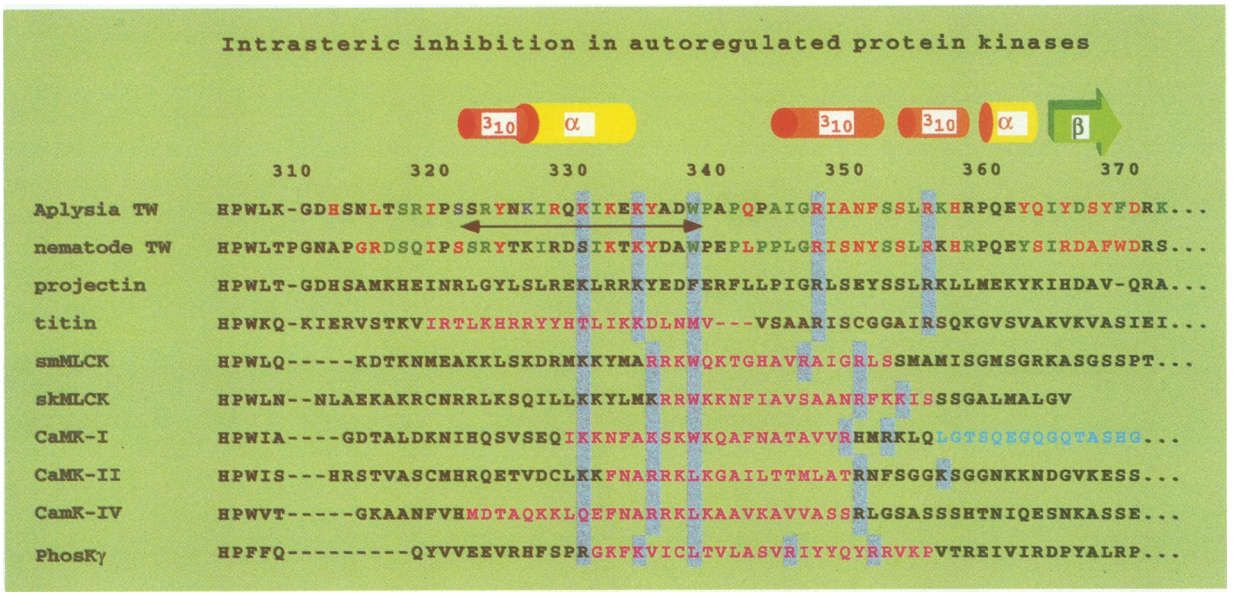

Fig. 5. Comparison of the autoregulatory sequences of Ca^{2+}/S100- and Ca^{2+}/CaM-regulated protein kinases. The following sequences are compared: *Aplysia* twitchin (Aplysia TW), *C.elegans* twitchin (nematode TW), *Drosophila* projectin, human titin, myosin light chain kinases from chicken smooth muscle (smMLCK) and rabbit skeletal muscle (skMLCK), rat calmodulin-dependent kinases-I (CaMK-I), -II (CaMK-II) and -IV (CaMK-IV), and phosphorylase kinase γ-subunit from rabbit skeletal muscle (PhosKγ) (Protein Identification Resource accession numbers S49128, S07571, A40985, S20898, A35093, A35021, A49682, A30355, TVRTC4 and KIRBFG, respectively). The putative CaM binding sites are coloured magenta (Dasgupta *et al.*, 1989; Ohmstede *et al.*, 1991; Ikura *et al.*, 1992; Meador *et al.*, 1992, 1993; Gautel *et al.*, 1995; Yokokura *et al.*, 1995); only the first of two CaM binding sites in phosphorylase kinase is shown. The S100 binding site in twitchin is underlined. All the CaBP binding sites have been identified using peptide ligands. It is not known how precisely they reflect the actual binding sites in intact proteins. The twitchin residues forming contacts with the catalytic core are coloured red (H bonds), green (van der Waals contacts but no H bonds) or blue (only water-mediated contacts). The sequence of CaMK-I not present in the crystal structure of the fragment (1–320) is coloured cyan. Some key autoinhibitory residues in twitchin with putatively conserved functions are shaded (see text). Elements of secondary structure in *Aplysia* twitchin are shown above its sequence as described in Figure 4A.

Tyr is replaced by Phe351 and is unavailable to contact Glu134. Furthermore, Lys330 of *Aplysia* twitchin binds the substrate binding residue Glu215 (P-6 site) of the catalytic core, whereas in nematode twitchin this Lys is replaced by a serine and does not contact the core at all.

The structural analysis of the twitchin kinase fragments demonstrates that contacts by the autoregulatory sequence exploit essentially all prominent features of the protein kinase active site, including the binding sites for protein substrate, ATP and Mg^{2+}, as well as catalytic residues and, particularly, the Thr-rich activation loop. The cooperative nature of these autoinhibitory contacts is in marked contrast to the structure of cAPK complexed with the inhibitory peptide, PKI_{5-24}, which is simply competitive with protein substrates and does not make the extensive contacts with residues involved in catalysis and ATP binding (Knighton *et al.*, 1991). Mutagenesis studies with smMLCK show that substrate and autoinhibitor binding are affected differentially by substitution of acidic residues on the protein kinase surface (Krueger *et al.*, 1995).

Structural implications for S100 regulation of giant protein kinases

Twitchin kinase represents the first enzyme whose activity is enhanced dramatically by a member of the S100 family (Heierhorst *et al.*, 1996), and our crystal structures provide a structural basis for this activation. The kinetic analyses of the activation process are consistent with the structural features of the autoinhibition. The predominant effect of the S100-dependent twitchin kinase activation on V_{max} is obvious with the extensive contacts between the autoregulatory sequence and the active site; the diversity of contacts explains why excess substrates (ATP and peptide) alone cannot activate the enzyme and points to the importance of the inhibitory contacts with the activation loop.

The number of contacts and the buried surface between the regulatory sequence and the catalytic core of twitchin are much larger than in typical protein–protein recognition sites (Janin, 1995), suggesting a very strong interaction. In the structures of CaM bound to peptides corresponding to the CaM binding regions of CaM-regulated protein kinases (Ikura *et al.*, 1992; Meador *et al.*, 1992, 1993), the number of intermolecular contacts is only one half of those observed between the regulatory and catalytic regions of TWK-43 [185 in chicken smMLCK (796–815)–CaM complex (Meador *et al.*, 1992), 135 in CaMK-II (290–314)–CaM complex (Meador *et al.*, 1993)]; only one half as much surface area is buried (Ikura *et al.*, 1992). It is not known how many contacts are broken when the inhibitory sequence is removed from the active site upon activator binding, and how this interaction affects other autoinhibitory contacts. Likewise, little is known about the molecular mechanisms of S100–target protein interactions (Schäfer and Heizmann, 1996). Structurally, $S100A1_2$ is a homodimer with each subunit containing two EF-hand Ca^{2+} binding loops (Potts *et al.*, 1995). Therefore, it is distinct from CaM where the two Ca^{2+} binding lobes, each containing two Ca^{2+} binding loops and thus analogous to one S100 subunit, are linked by a central helix (Babu *et al.*, 1985). The S100 binding site in twitchin (Heierhorst *et al.*, 1996) was shown previously

to bind CaM (Heierhorst *et al.*, 1994), although CaM binding did not result in activation. Given the structural similarity of CaM and S100, the role of hydrophobic interactions between the CaM Ca^{2+} binding lobes and residues in the helical target peptides (Ikura *et al.*, 1992; Meador *et al.*, 1992, 1993) may also be important for S100 binding. However, the different architectures of CaM and S100 proteins may result in substantial differences in the way the hydrophobic patches are exploited and explain why S100 activates twitchin and CaM does not.

Structural implications for other autoregulated protein kinases

The conservation of the structures of the autoregulatory regions in pJK4 and the shorter fragments of twitchin kinases indicates that the observed structures are physiologically relevant and not merely an outcome of crystallizing small fragments of giant protein kinases. The structure of the autoregulatory sequence in CaMK-I fragment (1–320) (Goldberg *et al.*, 1996) is strikingly similar to twitchin kinase in having a prominent helix in the substrate binding groove and close contacts with the Gly-rich phosphate anchoring loop. However, the structures differ dramatically in that the CaMK-I autoregulatory sequence exits the active site in a different manner (to the left of the small lobe), avoiding the catalytic residues. The activation loop in the CaMK-I structure is disordered. CaMK-I is activated by phosphorylation of Thr177 in the activation loop by a CaMK-I kinase upstream in the phosphorylation cascade (Haribabu *et al.*, 1995, and references therein). Phosphorylation by the kinase kinase depends on the removal of the autoinhibitory sequence by CaM binding or truncation mutagenesis (Haribabu *et al.*, 1995). Why CaM binding would be required for phosphorylation of Thr177 is not clear from the structure of CaMK-I; in that structure, the activation loop is available for interaction with the upstream kinase kinase. The biochemical properties are entirely consistent with the autoinhibitory mechanism observed for twitchin kinase, where the autoregulatory sequence covers the activation loop. The segment of CaMK-I autoregulatory sequence corresponding to the sequence in twitchin that makes multiple contacts with the activation loop has been truncated in the crystallized CaMK-I fragment (Goldberg *et al.*, 1996) (Figure 5). One interpretation of the discrepancy is, therefore, that the truncation itself has disturbed the structure of CaMK-I due to the loss of contacts anchoring the activation loop. A fragment of nematode twitchin of a length similar to the crystallized CaMK-I is not fully autoinhibited (Lei *et al.*, 1994), indicating that the C-terminal, Thr-rich loop binding segment of the sequence is crucial for autoinhibition. The crystallized, dephosphorylated CaMK-I fragment may also show substantial activity were Thr177 phosphorylated. It is also puzzling as to why ATP would interfere with crystallization of the described CaMK-I fragment (Goldberg *et al.*, 1996).

Inspection of the autoregulatory sequences of the CaM- and S100-regulated protein kinase subfamily reveals both similarities and differences (Figure 5). The prominent helix and substrate-like contacts observed in both twitchin and CaMK-I seem likely to be a feature of all members of Ca^{2+} binding protein-regulated kinase family. Some differences may reflect the diversity of the activation mechanisms (activation by CaM, or CaM plus phosphorylation or $S100A1_2$) and other properties; for example, while the regulatory sequence occupies the ATP binding site in twitchin, labelling studies suggest that this site is accessible in the homologous skeletal muscle MLCK (Kennely *et al.*, 1991) as well as in the more distant CaMK-II (Katoh and Fujisawa, 1991). Even the titin autoregulatory sequence differs from twitchin in the Pro-rich sequence (PxPx-PxxxR, where x is any amino acid) that directs the autoregulatory sequence deep into the ATP binding site (Figure 5). Still, the autoregulatory sequences share important common features (Figure 5). The positively charged residues in the vicinity of twitchin residues 330, 334 and 347 may interact with conserved glutamates implicated in substrate binding in most autoregulated protein kinases. Arginines corresponding to Arg355 of twitchin that interacts with the conserved catalytic Asp174 are also well conserved in giant protein kinases, and can be found in the vicinity in the autoregulatory sequences of other autoregulated protein kinases. In addition, the bulky hydrophobic group corresponding to the solvent-exposed side chain of Trp338 in twitchin [this residue is at the C-terminus of the $CaM/S100A1_2$ binding region (Heierhorst *et al.*, 1996)] may be an important conserved recognition site for the activator proteins, as has also been suggested for the analogous Trp303 in CaMK-I (Goldberg *et al.*, 1996).

Materials and methods

Structure determination of TWK-43

TWK-43 was expressed in *Escherichia coli*, purified as described (Heierhorst *et al.*, 1995) and concentrated to 2.4 mg/ml. Dynamic light scattering (DynaPro-801, Protein Solutions) indicated a hydrodynamic radius of 3.2 nm, consistent with a monomeric species. Crystallization conditions were screened by the hanging drop vapour diffusion technique (McPherson, 1982; Jancarik and Kim, 1991); crystals were obtained with polyethylene glycol (PEG) 4000 as the precipitant at pH values between 5 and 8.5. The crystal (0.5×0.15×0.15 mm) used for data collection was grown by mixing 4 µl of protein solution with 4 µl of reservoir solution containing 15% PEG 4000, 100 mM sodium citrate (pH 5) and 1 mM dithiothreitol (DTT) on a siliconized glass coverslip inverted over 0.95 ml of reservoir solution. The diffraction data were collected from a single crystal with a Weisenberg camera (Sakabe, 1991) at the Photon Factory synchrotron, Tsukuba, Japan using radiation with a wavelength of 1.0 Å, and auto-indexed and processed with the programs DENZO and SCALEPACK (Otwinowski, 1993); they have the symmetry of the space group $P2_12_12_1$ with the unit cell dimensions a = 62.3 Å, b = 87.8 Å and c = 152.9 Å. A total of 153 335 observations were measured and reduced to 52 664 unique reflections with R_{merge} = 0.118 for $I > -3\sigma(I)$ between infinity and 1.65 Å resolution [$R_{merge} = \Sigma_{hkl}(\Sigma_i(|I_{hkl,i} - \langle I_{hkl}\rangle|))/\Sigma_{hkl,i}\langle I_{hkl}\rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices h, k and l, and $\langle I_{hkl}\rangle$ is the mean intensity of that reflection]. The structure was determined by molecular replacement [program package AMoRe (Navaza, 1994)] using the structure of nematode twitchin kinase (Hu *et al.*, 1994) as the starting model. Data between 15 and 4 Å resolution and a radius of 30 Å were used for rotation searches. The translation function of the top rotation function solutions resulted in two outstanding peaks with correlation coefficients of 28.1 and 27.6% (the highest false peak 13.1%); both had an R_{cryst} of 50.9% (see Table I for a description of R factors). The two translation function solutions corresponded to the two highest maxima of the rotation function. For refinement [program package X-PLOR (Brünger, 1992b); initially, 10% of reflections were omitted for the calculation of R_{free} (Brünger, 1992a)], the two properly positioned molecules were reduced to polyalanine and subjected to rigid-body and positional refinement. Electron density maps clearly showed density corresponding to side chains, therefore the side chains were modelled (Jones *et al.*, 1990) according to the *Aplysia* sequence. The two molecules per unit cell were modelled as identical and refined as

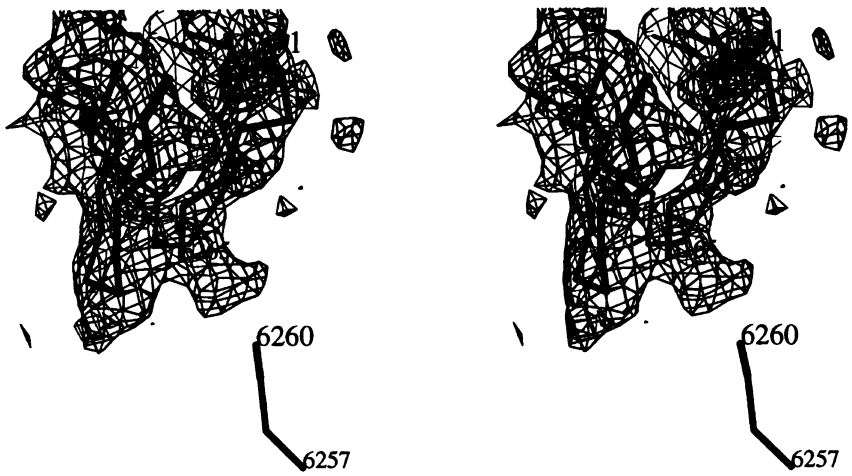

Fig. 6. Structure determination of pJK4. Stereo diagram of the electron density map of pJK4 (Jones *et al.*, 1990) calculated with coefficients $|F_{obs}|-|F_{calc}|$ (20–4 Å resolution) and phases derived from the protein kinase portion of the molecule (5890–6262) after its positional refinement, contoured at one standard deviation, in the vicinity of the N-terminus of the Ig-26 domain. Superimposed is the model of pJK4 (residues 6257–6260 of the autoregulatory region and residues 6265–6271, 6287–6298, 6316–6322, 6338–6348 of Ig-26) after refinement as a rigid body.

rigid bodies. Subsequently, strong non-crystallographic symmetry (n.c.s.) restraints were imposed; after several cycles of positional refinement, refinement of individual isotropic temperature factors, simulated annealing and manual rebuilding, the n.c.s. restraints were excluded. Further positional refinement, refinement of individual isotropic temperature factors, addition of solvent molecules and manual rebuilding resulted in an R_{cryst} of 23.5% and an R_{free} of 30.3% using data between 40 and 2.3 Å resolution with a bulk solvent correction and $F > 2\sigma(F)$. In the absence of n.c.s. restraints, refinement by simulated annealing caused an increase in R_{free} and was therefore not used. Candidate water molecules (program WATER by S.R.Sprang) were kept in the model if they were within 3.5 Å of a polar atom and had *B*-factors <70 Å^2 and the correlation coefficients of the real-space fit parameters (Jones *et al.*, 1990) >0.5. Finally, all data were included in the refinement, resulting in an improvement of the electron density maps. Positional refinement, refinement of individual isotropic temperature factors, addition of more solvent molecules and manual rebuilding yielded the final model of TWK-43 with an R_{cryst} of 21.45% using all data between 40 and 2.3 Å resolution with a bulk solvent correction (R_{cryst} = 19.99% using all data between 6 and 2.3 Å resolution). To evaluate R_{free} after the completion of refinement, 10% of data were again randomly omitted, and the model was subjected to a round of simulated annealing with an initial temperature of 3000 K, followed by 100 cycles of positional refinement. This resulted in an R_{cryst} of 21.1% and an R_{free} of 28.8% for all data between 40 and 2.3 Å resolution and a bulk solvent correction. The final electron density was of high quality; contoured at 1σ, the electron density contained no regions of discontinuous density in the backbone of the protein, except for 22 N-terminal and 11 C-terminal residues (12 for molecule B) that are presumably disordered. There are two molecules of TWK-43 in the asymmetric unit, related by a non-crystallographic 2-fold rotation axis; the final model contains 354 protein residues of molecule A, 353 protein residues of molecule B and 326 water molecules. The r.m.s. deviation of 353 C^{α} atoms of the two molecules is 0.4 Å; because there are no significant differences between molecules A and B all the discussion in the text applies to molecule A. Using 3D-1D analysis (Lüthy *et al.*, 1992) in a 21 residue window, no regions with values <0.1 were indicated; the overall score for 353 residues is 158.6. The r.m.s. deviation of the 348 C^{α} atoms after superposition of TWK-43 and nematode twitchin (Hu *et al.*, 1994) is 0.88 Å. Crystallization of TWK-43 in the presence of 2.5 mM $CaCl_2$ or soaking the crystals of TWK-43 in the presence of 20 mM $ZnCl_2$ did not produce any significant conformational changes, providing additional evidence that, during $S100A1_2$ activation, the Ca^{2+} and Zn^{2+} effects on the protein kinase activity are mediated entirely through $S100A1_2$.

Structure determination of pJK4

pJK4 (residues 5890–6380 of *C.elegans* twitchin) was expressed in *E.coli* and purified as described for other fragments (Hu *et al.*, 1994). The protein was concentrated to 5 mg/ml (Bio-Rad protein assay) in 20 mM Tris (pH 7.4), 200 mM NaCl and 2 mM DTT. Crystallization conditions were screened by the hanging drop vapour diffusion technique (McPherson, 1982; Jancarik and Kim, 1991) using Crystal Screen (Hampton Research); crystals grew after 1 month to the size 0.2×0.2×0.1 mm by mixing 2 µl of protein solution with 2 µl of reservoir solution containing 4 M sodium formate on a glass coverslip and inverting it over 0.7 ml of reservoir solution. The initial X-ray diffraction data were collected using a MARRESEARCH image plate detector with CuK_{α} radiation from a Rigaku RU-200 rotating anode generator and auto-indexed and processed with the programs DENZO and SCALEPACK (Otwinowski, 1993); they have the symmetry consistent with the space groups I23 and $I2_13$; the unit cell axis equals 197.7 Å. A total of 39 131 observations were measured and reduced to 10 528 unique reflections with R_{merge} = 0.283 for I >−3σ(I) between 40 and 4 Å resolution [R_{merge} = 0.119 for I >3σ(I)]. In cross-rotation and translation searches using twitchin kinase (Hu *et al.*, 1994) as the starting model, program packages AMoRe (Navaza, 1994) and X-PLOR (Brünger, 1992b) yielded significant and consistent solutions. Data between 15 and 5 Å resolution and vectors within a sphere with a 30 Å radius were used. With AMoRe, a rotation search yielded a correlation coefficient of 16% for the highest peak (referred to hereafter as the correct orientation) and 15.4% for the next peak (highest false orientation). Translation searches were performed for both space groups I23 and $I2_13$. Assuming I23, a translation search of the molecule in the correct orientation resulted in a correlation coefficient of 37.9% and an R_{cryst} of 48%; the numbers were 34.6 and 48.6% for the highest false orientation. Assuming $I2_13$, a translation search of the molecule in the correct orientation resulted in a correlation coefficient of 39.8% and an R_{cryst} of 44.8%; the numbers were 28.9 and 47.9% for the highest false orientation. Only the top solution of the translation search in the space group $I2_13$ resulted in a reasonable packing in the unit cell. With X-PLOR (Brünger, 1992b), the top 50 maxima of the rotation function were filtered by Patterson correlation (PC) refinement (Brünger, 1990) consisting of 10 steps of conjugate gradient minimization of the orientation of the molecule. Two outstanding peaks were found with correlation coefficients >0.05, both corresponding to the same orientation of the molecule and to the correct orientation as found in AMoRe; the correlation coefficient of the highest false peak was <0.04. A translation function search of the model in the corresponding orientation assuming the space group $I2_13$ yielded the top solution of 11σ that was consistent with the calculations in AMoRe. Searches starting with false orientations and/or assuming space group I23 resulted in an unreasonable packing in the unit cell.

Cross-rotation and translation searches using the structure of telokin (Holden *et al.*, 1992) as the starting model did not yield a solution, most likely because the Ig-26 domain represents <25% scattering matter in the crystal. Therefore, the partial structure corresponding to the twitchin kinase molecule was subjected to rigid-body and positional refinement (program X-PLOR, Brünger, 1992b) [R_{cryst} = 32.5% and R_{free} = 40.2% between 10 and 4 Å resolution and $F > 2\sigma(F)$] and the resulting $2|F_{obs}|-|F_{calc}|$ and $|F_{obs}|-|F_{calc}|$ electron density maps were inspected (Jones *et al.*, 1990). Electron density resembling the telokin molecule was evident (Figure 6); there was no evidence for additional electron density

corresponding to another molecule of pJK4 in the asymmetric unit. The model of telokin was positioned into the electron density corresponding to the Ig-26 domain on the graphics display; in spite of the low resolution of the maps, there was only one unique way to position the telokin molecule. In retrospect, the top solution of the cross-rotation search in AMoRe with telokin as the starting model (correlation coefficient 15.9%; the next highest peak had a correlation coefficient of 13.8%) was the correct orientation of the telokin molecule; the translation searches, however, did not position the molecule correctly. Refinement of twitchin kinase and telokin molecules as rigid bodies resulted in an R_{cryst} of 40.9% and an R_{free} of 38.3% between 10 and 4 Å resolution and $F > 2\sigma(F)$. We were convinced that the positions of the domains in the model were correct based on the following: (i) two different methods yielded the same solution for the position of the protein kinase domain; (ii) electron density maps phased by the properly positioned protein kinase domain revealed the electron density of the Ig-26 domain; this confirmed the correct position of the protein kinase domain; and (iii) there was only one unique way to position the telokin model into the electron density of the Ig-26 region; other orientations would either not fit the electron density to any appreciable extent or would result in impossible connectivity between the protein kinase and the Ig-26 domain.

Subsequently, we collected a dataset from a crystal flash-frozen at 100 K (Oxford Cryosystems) with a Weisenberg camera (Sakabe, 1991) at the Photon Factory synchrotron, Tsukuba, Japan using radiation with a wavelength of 1.0 Å. Prior to data collection, the crystal was transferred into 4 M sodium formate and 10% glycerol. The data were processed with the programs DENZO and SCALEPACK (Otwinowski, 1993). A total of 110 980 observations were measured and reduced to 23 129 unique reflections with R_{merge} = 0.157 for $I > -3\sigma(I)$ between infinity and 3.0 Å resolution. The data extend to 3.7 Å resolution with some reflections present at 3.3 Å. Because of the low resolution limit and no n.c.s. and because torsion angle dynamics refinement (Rice and Brünger, 1994) was not yet made available to us, we used a conservative refinement (program package X-PLOR, Brünger, 1992b; 10% of reflections were omitted for the calculation of R_{free}) protocol (Kleywegt and Jones, 1995) and used R_{free} (Brünger, 1992a) to assess possible overfitting. The protein kinase and Ig domains from the molecular replacement model were first refined as separate rigid bodies [R_{cryst} = 45.4%, R_{free} = 45.3% for data between 10 and 4 Å resolution and $F > 2\sigma(F)$]. We then introduced the correct sequence of the Ig domain and proceeded with positional refinement, refinement of an overall *B*-factor and refinement of grouped *B*-factors (each residue is described by two groups corresponding to main chain and side chain atoms) that all resulted in substantial drops of R_{free}; simulated annealing resulted in an increase in R_{free} and was therefore not used; restraining the atomic positions to their respective positions in the 2.8 Å resolution structure of the shorter fragment (Hu *et al.*, 1994) resulted in no improvement of *R* values and was also abandoned. At this stage, we could also build the linker peptide connecting the two domains. Several rounds of refinement and model building (Jones *et al.*, 1990) and the gradual introduction of all data between 40 and 3.3 Å resolution with a bulk solvent correction resulted in an R_{cryst} of 25.1% and an R_{free} of 35.1%. Although refinement of atomic *B*-factors further decreased the *R*-factors (R_{cryst} = 23.8%, R_{free} = 34.8%), we chose the model refined with grouped *B*-factors as our final model. The model is of very high quality accounting for the low resolution limit (Table I). The main reason for the poor diffraction quality of the crystals appears to be high mobility of the protein in the crystals, as reflected in a high average *B*-factor of 88.5 Å^2 [the value estimated from the Wilson plot (Wilson, 1949) is ~100 Å^2]. Despite that, the electron density is continuous at 1σ for the backbone of the entire fragment between residues 5915 and 6361, and the fit of the model to the density is high. Twenty five N-terminal and 19 C-terminal residues do not have any interpretable electron density. A 3D-1D analysis (Lüthy *et al.*, 1992) in a 21 residue window indicated no regions with values <0.05; the lowest values occur in the autoregulatory region, suggesting a restrained structure that may be altered more easily upon binding of the activator protein; the overall score for 447 residues is 202.7. The r.m.s. deviation of C^α atoms of residues 5915–6260 between the fragments pJK4 and the shorter nematode fragment (Hu *et al.*, 1994) is 0.7 Å; the biggest deviations are due mostly to involvement in crystal contacts and occur in loops.

Modelling of MLCK

The model of the chicken gizzard smMLCK core (residues 515–775) was generated from the crystal structure of nematode twitchin kinase (residues 5891–6196; there is 52% sequence identity between the corresponding regions of these two proteins with no insertions or deletions) (Hu *et al.*, 1994). The side chain atoms of the twitchin kinase structure were replaced with smMLCK side chain atoms using the program HOMOLOGY (Biosym Technologies). The model was subjected to energy minimization and simulated annealing (1000–300 K) with the program X-PLOR (Brünger, 1992b), gradually loosening the harmonic restraints. The r.m.s. deviation of the C^α atoms of the final MLCK model and twitchin crystal structure was 0.5 Å. The correctness of the model was evaluated using the 3-D profile approach (Lüthy *et al.*, 1992); the 3D-1D score averaged over 21 consecutive amino acids did not drop below 0.1. A stereochemical analysis of the model with the program PROCHECK (Laskowski *et al.*, 1993) gave values either similar to or better than expected for crystal structures determined at 3.0 Å resolution.

The C-terminal Ig domain of smMLCK was modelled (Jones *et al.*, 1990) using the crystal structure of telokin (Holden *et al.*, 1992). Telokin was positioned with respect to the smMLCK core by superposition with the pJK4 crystal structure.

The autoregulatory and linker regions (776–850) connecting the protein kinase and immunoglobulin domains have not been modelled. Although the linker region of smMLCK contains an insertion of 23 residues when compared with twitchin, the relative juxtapositions of the two domains in MLCK and twitchin are most likely conserved. Our model shows a highly complementary interface between the two domains with plausible salt bridges and other polar interactions and no steric clashes.

Structure analyses

Structure analyses and model building were performed on a graphics workstation using the program package O (Jones *et al.*, 1990). The electron density maps were weighted with the program SIGMAA (Read, 1986). Interactions were analysed with the programs X-PLOR (Brünger, 1992b) and CONTACT (CCP4, 1994). Two atoms were considered to be in contact if <4 Å apart; putative hydrogen bonds and salt links were considered if acceptor–donor distances were <3.5 Å. Accessible surface areas were calculated with the program SURFACE (CCP4, 1994); group radii were used, hydrogen atoms were excluded from the calculations and a probe radius was 1.4 Å. Surfaces were analysed graphically with the program GRASP (Nicholls *et al.*, 1991). Surface complementarity was analysed by the method of Lawrence and Colman (1993). Secondary structure was assigned with the program DSSP (Kabsch and Sander, 1983). The coordinates and structure factors have be deposited in the Brookhaven Protein Data Bank (ID codes 1KOA, 1KOB, R1KOASF and R1KOBSF).

Acknowledgements

We thank Jamie Rossjohn for his expert help with cryocrystallography and synchrotron data collection, and Matthew Wilce, Colin House, Aaron Oakley, Bill Probst and Xuexin Tang for help. This work was supported by grants from the NH and MRC and National Heart Foundation to B.E.K., the NIH to K.R.W., and fellowships from the Human Frontiers Science Program Organization and Deutsche Forschungsgemeinschaft to J.H. and the Australian Research Council to B.K. G.M.B. is an Established Investigator of the American Heart Association.

References

Ayme-Southgate,A., Southgate,R., Saide,J., Benian,G.M. and Pardue, M.L. (1995) Both synchronous and asynchronous muscle isoforms of projectin (the *Drosophila bent* locus product) contain functional kinase domains. *J. Cell Biol.*, **128**, 393–403.

Babu,Y.S., Sack,J.S., Greenhough,T.J., Bugg,C.E., Means,A.R. and Cook,W.J. (1985) Three-dimensional structure of calmodulin. *Nature*, **315**, 37–40.

Benian,G.M., Kiff,J.E., Neckleman,N., Moerman,D.G. and Waterston, R.H. (1989) Sequence of an unusually large protein implicated in regulation of myosin activity in *C.elegans*. *Nature*, **342**, 45–50.

Benian,G.M., L'Hernault,S.W. and Morris,M.E. (1993) Additional sequence complexity in the muscle gene, *unc-22*, and its encoded protein, twitchin, of *Caenorhabditis elegans*. *Genetics*, **134**, 1097–1104.

Bossemeyer,D., Engh,R.A., Kinzel,V., Postingl,H. and Huber,R. (1993) Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn^{2+} adenylyl imidodiphosphate and inhibitor peptide PKI(5–24). *EMBO J.*, **12**, 849–859.

Brünger,A.T. (1990) Extension of molecular replacement: a new search strategy base on Patterson correlation coefficient. *Acta Crystallogr.*, **A46**, 46–57.

Brünger,A.T. (1992a) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Brünger,A.T. (1992b) *X-PLOR Version 3.1*. Yale University, New Haven, CT.

CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Dasgupta,M., Honeycutt,T. and Blumenthal,D.K. (1989) The γ-subunit of skeletal muscle phosphorylase kinase contains two noncontiguous domains that act in concert to bind calmodulin. *J. Biol. Chem.*, **264**, 17156–17163.

Dibb,N.J., Brown,D.M., Karn,J., Moerman,D.G., Bolten,S.L. and Waterston,R.H. (1985) Sequence analysis of mutations that affect the synthesis, assembly and enzymatic activity of the *unc-54* myosin heavy chain of *Caenorhabditis elegans*. *J. Mol. Biol.*, **183**, 543–551.

Eisenberg,D. and McLachlan,A.D. (1986) Solvation energy in protein folding and binding. *Nature*, **319**, 199–203.

Gautel,M., Castiglione Morelli,M.A., Pfuhl,M., Motta,A. and Pastore,A. (1995) A calmodulin-binding sequence in the C-terminus of human cardiac titin kinase. *Eur. J. Biochem.*, **230**, 752–759.

Gibbs,C.S., Knighton,D.R., Sowadski,J.M., Taylor,S.S. and Zoller,M.J. (1992) Systematic mutational analysis of cAMP-dependent protein kinase identifies unregulated catalytic subunits and defines regions important for the recognition of the regulatory subunit. *J. Biol. Chem.*, **267**, 4806–4814.

Goldberg,J., Nairn,A.C. and Kuriyan,J. (1996) Structural basis for the auto-inhibition of calcium/calmodulin-dependent protein kinase I. *Cell*, **84**, 875–887.

Goldsmith,E.J. and Cobb,M.H. (1994) Protein kinases. *Curr. Opin. Struct. Biol.*, **4**, 833–840.

Haribabu,B., Hook,S.S., Selbert,M.A., Goldstein,E.G., Tomhave,E.D., Edelman,A.M., Snyderman,R. and Means,A.R. (1995) Human calcium–calmodulin dependent protein kinase I: cDNA cloning, domain structure and activation by phosphorylation at threonine-177 by calcium–calmodulin dependent protein kinase I kinase. *EMBO J.*, **14**, 3679–3686.

Harpaz,Y. and Chothia,C. (1994) Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.*, **238**, 528–539.

Heierhorst,J., Probst,W.C., Vilim,F.S., Buku,A. and Weiss,K.R. (1994) Autophosphorylation of molluscan twitchin and interaction of its kinase domain with calcium/calmodulin. *J. Biol. Chem.*, **269**, 21086–21093.

Heierhorst,J., Probst,W.C., Kohanski,R.A., Buku,A. and Weiss,K.R. (1995) Phosphorylation of myosin regulatory light chains by the molluscan twitchin kinase. *Eur. J. Biochem.*, **233**, 426–431.

Heierhorst,J., Kobe,B., Feil,S.C., Parker,M.W., Benian,G.M., Weiss,K.R. and Kemp,B.E. (1996) Ca^{2+}/S100 regulation of giant protein kinases. *Nature*, **380**, 636–639.

Holden,H.M., Ito,M., Hartshorne,D.J. and Rayment,I. (1992) X-ray structure determination of telokin, the C-terminal domain of myosin light chain kinase at 2.8 Å resolution. *J. Mol. Biol.*, **227**, 840–851.

Houmeida,A., Holt,J., Tskhovrebova,L. and Trinick,J. (1995) Studies of the interaction between titin and myosin. *J. Cell Biol.*, **131**, 1471–1481.

Hu,S.-H., Parker,M.W., Lei,J.Y., Wilce,M.C.J., Benian,G.M. and Kemp,B.E. (1994) Insights into autoregulation from the crystal structure of twitchin kinase. *Nature*, **369**, 581–584.

Hubbard,S.R., Wei,L., Ellis,L. and Hendrickson,W.A. (1994) Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature*, **372**, 746–754.

Ikura,M., Clore,G.M., Gronenborn,A.M., Zhu,G., Klee,C.B. and Bax,A. (1992) Solution structure of a calmodulin–target peptide complex by multidimensional NMR. *Science*, **256**, 632–638.

Improta,S., Politou,A.S. and Pastore,A. (1996) Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure*, **4**, 323–337.

Jancarik,J. and Kim,S.-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.*, **24**, 409–411.

Janin,J. (1995) Elusive affinities. *Proteins*, **21**, 30–39.

Jeffrey,P.D., Russo,A.A., Polyak,K., Gibbs,E., Hurwitz,G., Massague,J. and Pavletich,N.P. (1995) Mechanism of CDK activation revealed by the structure of a cyclinA–CDK2 complex. *Nature*, **376**, 313–320.

Jones,T.A., Bergdoll,M. and Kjeldgaard,M. (1990) O: a macromolecule modeling environment. In Bugg,C.E. and Ealick,S.E. (eds), *Crystallographic and Modeling Methods in Molecular Design*. Springer-Verlag, New York, pp. 189–195.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kamm,K.E. and Stull,J.T. (1985) The function of myosin and myosin light chain kinase phosphorylation in smooth muscle. *Annu. Rev. Pharmacol. Toxicol.*, **25**, 593–620.

Kanoh,S., Ito,M., Niwa,E., Kawano,Y. and Hartshorne,D.J. (1993) Actin-binding peptide from smooth muscle myosin light chain kinase. *Biochemistry*, **32**, 8902–8907.

Katoh,T. and Fujisawa,H. (1991) Calmodulin-dependent protein kinase II. Kinetic studies on the interaction with substrates and calmodulin. *Biochim. Biophys. Acta*, **1091**, 205–212.

Kemp,B.E., Faux,M.C., Means,A.R., House,C., Tiganis,T., Hu,S.-H. and Mitchelhill,K.I. (1994) Structural aspects: pseudosubstrate and substrate interactions. In Woodgett,J.R. (ed.), *Protein Kinases*. IRL Press at Oxford University Press, Oxford, pp. 30–67.

Kennely,P.J., Colburn,J.C., Lorenzen,J., Edelman,A.M., Stull,J.T. and Krebs,E.G. (1991) Activation mechanism of rabbit skeletal muscle myosin light chain kinase. *FEBS Lett.*, **286**, 217–220.

Kleywegt,G.J. and Jones,T.A. (1995) Where freedom is given, liberties are taken. *Structure*, **3**, 535–540.

Kleywegt,G.J., Bergfors,T., Senn,H., Le Motte,P., Gsell,B., Shudo,K. and Jones,T.A. (1994) Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-*trans*-retinoic acid and a synthetic retinoid. *Structure*, **2**, 1241–1258.

Knighton,D.R., Zheng,J., Eyck,L.F.T., Ashford,V.A., Xuong,N.-H., Taylor,S.S. and Sowadski,J.M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, **253**, 407–414.

Kraulis,P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Krueger,J.K., Padre,R.C. and Stull,J.T. (1995) Intrasteric regulation of myosin light chain kinase. *J. Biol. Chem.*, **270**, 16848–16853.

Labeit,S. and Kolmerer,B. (1995) Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science*, **270**, 293–296.

Labeit,S., Gautel,M., Lakey,A. and Trinick,J. (1992) Towards a molecular understanding of titin. *EMBO J.*, **11**, 1711–1716.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lawrence,M.C. and Colman,P.M. (1993) Shape complementarity at protein–protein interfaces. *J. Mol. Biol.*, **234**, 946–950.

Leahy,D.J., Aukhil,I. and Erickson,H.P. (1996) 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell*, **84**, 155–164.

Lei,J., Tang,X., Chamber,T.C., Pohl,J. and Benian,G.M. (1994) The protein kinase domain of twitchin has protein kinase activity and an autoinhibitory region. *J. Biol. Chem.*, **269**, 21078–21085.

Luzzati,P.V. (1952) Traitements statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr.*, **5**, 802–810.

Lüthy,R., Bowie,J.U. and Eisenberg,D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83–85.

McPherson,A. (1982) *Preparation and Analysis of Protein Crystals*. John Wiley and Sons, New York.

Meador,W.E., Means,A.R. and Quiocho,F.A. (1992) Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin–peptide complex. *Science*, **257**, 1251–1255.

Meador,W.E., Means,A.R. and Quiocho,F.A. (1993) Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. *Science*, **262**, 1718–1721.

Navaza,J. (1994) *AMoRe*: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

Ohmstede,C.-A., Bland,M.M., Merrill,B.M. and Sahyoun,N. (1991) Relationship of genes encoding Ca^{2+}/calmodulin-dependent protein kinase Gr and calspermin: a gene within a gene. *Proc. Natl Acad. Sci. USA*, **88**, 5784–5788.

Okagaki,T., Weber,F.E., Fischman,D.A., Vaughan,K.T., Mikawa,T. and Reinach,F.C. (1993) The major myosin-binding domain of skeletal muscle MyBP-C (C protein) resides in the COOH-terminal, immunoglobulin C2 motif. *J. Cell Biol.*, **123**, 619–626.

Olson,N.J., Pearson,R.B., Needleman,D.S., Hurwitz,M.Y., Kemp,B.E. and Means,A.R. (1990) Regulatory and structural motifs of chicken gizzard myosin light chain kinase. *Proc. Natl Acad. Sci. USA*, **87**, 2284–2288.

Otwinowski,Z. (1993) Data collection and processing. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Proceedings of the CCP4 Study Weekend*. SERC, Daresbury Laboratory, Daresbury, UK, pp. 56–62.

Pfuhl,M. and Pastore,A. (1995) Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the I set. *Structure*, **3**, 391–401.

Pfuhl,M., Winder,S.J., Castiglione Morelli,M.A., Labeit,S. and Pastore,A. (1996) Correlation between conformational and binding properties of nebulin repeats. *J. Mol. Biol.*, **257**, 367–384.

Politou,A.S., Gautel,M., Improta,S., Vangelista,L. and Pastore,A. (1996) The elastic I-band region of titin is assembled in a 'modular' fashion by weakly interacting Ig-like domains. *J. Mol. Biol.*, **255**, 604–616.

Potts,B.C.M., Smith,J., Akke,M., Macke,T.J., Okazaki,K., Hidaka,H., Case,D.A. and Chazin,W.J. (1995) The structure of calcyclin reveals a novel homodimeric fold for S100 Ca^{2+}-binding proteins. *Nature Struct. Biol.*, **2**, 790–796.

Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr.*, **A42**, 140–149.

Rice,L.M. and Brünger,A.T. (1994) Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, **19**, 277–290.

Sakabe,N. (1991) X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl. Instr. Methods Phys. Res.*, **A303**, 448–463.

Schäfer,B.W. and Heizmann,C.W. (1996) The S100 family of EF-hand calcium-binding proteins: functions and pathology. *Trends Biochem. Sci.*, **21**, 134–140.

Shirinsky,V.P., Vorotnikov,A.V., Birukov,K.G., Nanaev,A.K., Collinge,M., Lukas,T.J., Sellers,J.R. and Watterson,D.M. (1993) A kinase-related protein stabilizes unphosphorylated smooth muscle myosin minifilaments in the presence of ATP. *J. Biol. Chem.*, **268**, 16578–16583.

Taylor,S.S. and Radzio-Andzelm,E. (1994) Cyclic AMP-dependent protein kinase. In Woodgett,J.R. (ed.), *Protein Kinases*. IRL Press at Oxford University Press, Oxford, pp. 1–29.

Trinick,J. (1994) Titin and nebulin: protein rulers in muscle. *Trends Biochem. Sci.*, **19**, 405–409.

Waterston,R.H., Thomson,J.N. and Brenner,S. (1980) Mutants with altered muscle structure of *Caenorhabditis elegans*. *Dev. Biol.*, **77**, 271–302.

Wilson,A.J.C. (1949) The probability distribution of X-ray intensities. *Acta Crystallogr.*, **2**, 318–321.

Yokokura,H., Picciotto,M.R., Nairn,A.C. and Hidaka,H. (1995) The regulatory region of calcium/calmodulin-dependent protein kinase I contains closely associated autoinhibitory and calmodulin-binding domains. *J. Biol. Chem.*, **270**, 23851–23859.

Received on July 10, 1996; revised on September 3, 1996