Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, crystallization and preliminary X-ray studies of the recombinant PTB domain of mouse dok1 protein

Ning Shi,[a,b] Yiwei Liu,[b] Minghao Ni,[b] MaoJun Yang,[a] Jing Wu,[a] Ying Peng,[b] Feng Gao,[b] Fei Sun,[b] Xiaozhong Peng,[a] Boqin Qiang,[a] Zihe Rao[a,b]* and Jiangang Yuan[a]

[a]National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, National Human Genome Center, Beijing 100005, People's Republic of China, and [b]Laboratory of Structural Biology, Tsinghua University and National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing, People's Republic of China

Correspondence e-mail: raozh@xtal.tsinghua.edu.cn

Received 11 August 2003
Accepted 20 November 2003

The PTB domain of mouse dok1 fusion protein has been overexpressed in *Escherichia coli* and crystallized in a form suitable for X-ray crystallographic study. Crystals have been obtained using the vapour-diffusion method and belong to space group $P2_12_12_1$. X-ray diffraction data were collected in-house to 2.5 Å resolution. A selenomethionine (SeMet) dok1 PTB fusion-protein derivative was expressed using the same expression system, purified in a reductive environment and crystals were obtained under similar conditions. Subsequently, three different wavelength data sets from the derivative crystal were collected to 2.5 Å resolution at SPring-8.

1. Introduction

Dok1 is a common substrate of activated protein tyrosine kinases such as v-Abl (Yamanashi & Baltimore, 1997), v-Src (Shah & Shokat, 2002), BCR (Kato *et al.*, 2002), EphRs (Becker *et al.*, 2000), Ret (Murakami *et al.*, 2002) and integrin β (Calderwood *et al.*, 2003) *etc*. It is rapidly tyrosine-phosphorylated in response to receptor tyrosine activation in various cell systems, suggesting that the protein plays an important role in cellular signalling.

Five members of the the dok family are known to date (Grimm *et al.*, 2001). They all contain an amino-terminal PH domain followed by a central PTB domain and a proline- and tyrosine-rich carboxy-terminal tail. The PH domain is known to bind to acidic phospholipids and localizes protein to the plasma membrane, while the PTB domain mediates protein–protein interactions by binding to phosphotyrosine-containing motifs (Jacobs *et al.*, 2001). The carboxy-terminal part of dok1 contains multiple tyrosine-phosphorylation sites. When phosphorylated, they become potential docking sites for Src homology 2 containing proteins such as rasGAP and Nck. This leads to the inhibition of ras signalling pathway activation and also to c-jun amino-terminal kinase (JNK) and c-jun activation (Murakami *et al.*, 2002).

Proteins containing PTB domains have been classified into two major groups. The first group contains PTB domains that have primary sequences similar to that of the Shc PTB domain. The second group contains insulin receptor substrate (IRS) like proteins such as IRS, dok and SNT/FRS2; their PTB domains have limited sequence similarity to the Shc PTB domain but exhibit similar binding characteristics (Wolf *et al.*, 1995).

The PTB domain of dok1 belongs to the second group, with 17% sequence identity to the PTB domain of IRS. It is supposed to recognize the sequence containing the NKLpY motif (Songyang *et al.*, 2001). Here, we describe the crystallization and preliminary X-ray crystallographic analysis of native and selenomethionyl recombinant dok1 PTB-domain protein.

2. Materials and methods

2.1. Protein expression and purification

The coding sequence of the PTB domain of mouse dok1 (mdok1; residues 154–266) was amplified by PCR. The PCR product was restricted with *Bam*HI and *Xho*I and cloned into a modified pET28a(+) expression vector (Novagen) with an N-terminal hexahistidine sequence followed by a thrombin cleavage site. A further transformation into *Escherichia coli* BL21 (DE3) competent cells was performed and the sequence of the insert was verified by sequencing. Protein was expressed in *E. coli* BL21 (DE3) cells by induction (at $A_{600} \simeq 0.7$) with 1 m*M* isopropyl-β-D-thiogalactoside (IPTG) for 6 h at 303 K. Cells were harvested and resuspended in lysis buffer (20 m*M* Tris pH 8.0, 500 m*M* NaCl, 10 m*M* imidazole, 10% glycerol and 1 m*M* PMSF) and homogenized by sonication. The lysate was centrifuged at 20 000*g* for 20 min at 277 K to remove the cell debris. The supernatant was applied to an Ni–NTA resin (Qiagen) affinity column. The target protein was then eluted with lysis buffer containing 300 m*M* imidazole and incubated overnight at 289 K in the presence of one unit of thrombin per 2 mg of protein in order to cleave the N-terminal hexahistidine tag. The protein solution was exchanged into buffer *A*

(20 m*M* MES pH 6.0). This sample was subsequently applied to a 1 ml Resource S column (Pharmacia) and bound proteins were eluted with a 0–400 m*M* NaCl gradient in buffer *A*. The peak fractions were collected and concentrated. Further purification was accomplished by gel-filtration chromatography on Superdex G75 (Pharmacia) in buffer *A* containing 50 m*M* NaCl. The purified protein homogeneity was confirmed by 15% SDS–PAGE electrophoresis.

Expression of SeMet-substituted protein (mdok1 PTB contains two Met residues) was achieved by transformation of the mdok1 PTB construct into the methionine-auxotrophic *E. coli* strain B834 (DE3) (Novagen). The SeMet protein derivative was expressed in M9 minimal media containing 30 mg l^{-1} L-SeMet and purified as described for the native protein, except that 10 m*M* β-mercaptoethanol (or 5 m*M* DTT) and 0.5 m*M* EDTA were added to the buffer after the eluting step from the Ni–NTA affinity column to prevent SeMet oxidation.

2.2. Crystallization

The purified protein was concentrated to 5–10 mg ml^{-1} using a 5K Filtron ultrafiltration membrane in 10 m*M* MES buffer pH 6.5. Initial crystallization trials were carried out by the hanging-drop vapour-diffusion method using Hampton Research Crystal Screens I and II (Hampton Research). Each drop was formed by mixing equal volumes (1–3 µl) of protein solution and reservoir solution and was allowed to equilibrate *via* vapour diffusion over 200 µl of reservoir solution at 291 K. The protein concentration was 10 mg ml^{-1} prior to mixing with the reservoir solution. Conditions yielding small crystals were further optimized by variation of precipitants, buffer pH, protein concentration, additives and temperature.

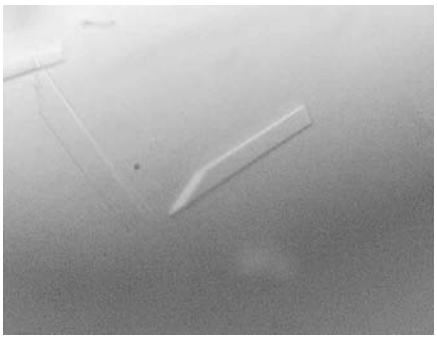

Figure 1
Native crystals of mdok1 PTB domain (the size of a single crystal is approximately 0.7 × 0.4 × 0.2 mm).

Table 1
Data-collection and processing statistics.

All data were collected on beamline 41XU at SPring-8, Japan. Values in parentheses correspond to the highest resolution shell (2.59–2.50 Å).

	Peak	Edge	Remote
Space group	$P2_12_12_1$		
Unit-cell parameters (Å, °)	a = 41.1, b = 56.2, c = 99.6, β = 90		
Wavelength (Å)	0.9798	0.9800	0.9000
Resolution (Å)	50–2.5	50–2.5	50–2.5
Mean redundancy	7.2 (4.8)	7.1 (3.9)	7.3 (2.5)
Completeness(%)	99.0 (94.5)	97.6 (89.0)	96.6 (80.1)
R_{merge}† (%)	0.106 (0.348)	0.102 (0.361)	0.108 (0.460)
Mean $I/\sigma(I)$	6.5 (1.7)	5.6 (1.6)	5.4 (1.3)
Total observations	52658	49486	55735
Unique reflections	8330 (780)	8423 (740)	8238 (670)

† $R_{merge} = \sum_h \sum_I |I_{ih} - \langle I_h \rangle| / \sum_h \sum_I \langle I_h \rangle$, where $\langle I_h \rangle$ is the mean of the observations I_{ih} of reflection h.

The purified selenomethionine derivative was concentrated to 7 mg ml^{-1}. Crystallization trials were set up based on the conditions used for the native protein.

2.3. Data collection and processing

Preliminary X-ray diffraction data were collected in-house at 100 K using a MAR Research image plate and a 4.8 kW Rigaku rotating-anode generator producing Cu $K\alpha$ radiation of wavelength 1.5418 Å, with a crystal-to-film distance of 110 mm and an oscillation range of 1°. Multiwavelength anomalous dispersion (MAD) data were collected on beamline 41XU under cryo-conditions at SPring-8. Crystals were frozen in the crystallization buffer with 20%(*v*/*v*) glycerol as cryoprotectant before data were collected.

Data processing was performed using the program *DENZO* and data sets were scaled and merged using *SCALEPACK* (Otwinowski & Minor, 1997).

The self-rotation functions (Rossmann & Blow, 1962) were calculated using the program *POLARRFN* from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

3.1. Expression and purification

A 16 kDa band corresponding to the molecular weight of mdok1 PTB-fusion protein was visible on SDS–PAGE gels after elution from Ni–NTA resin. Two peaks were observed during Resource S cation-exchange chromatography, which indicated the existence of charged isomers. Samples from the two peaks were further purified and used in crystallization trials, but yielded no crystals. After the removal of the amino-terminal His tag, only one peak was observed during Resource S cation-exchange chromatography. Further purification resulted in pure protein which was used in crystallization trials. The purification of the SeMet mdok1 PTB derivative followed the same protocol as for the native protein.

3.2. Crystallization

Initial crystals of mdok1 PTB were obtained after one week in conditions consisting of 30%(*w*/*v*) PEG MME 5000, 0.1 *M* MES pH 6.5, 0.2 *M* ammonium sulfate. After optimization of the crystallization conditions, crystals useful for crystallographic structure determination were obtained with a reservoir solution containing 28%(*w*/*v*) PEG 6000, 0.1 *M* MES pH 6.5. The best diffracting crystals grew to dimensions of 2 × 0.2 × 0.05 mm at 277 K in one week (Fig. 1). They diffracted to 2.3 Å on an in-house X-ray source. The selenomethionine-derivative crystals were obtained using similar conditions to those for native protein, except for a lower pH (0.1 *M* MES pH 6.0) and protein concentration (7 mg ml^{-1}).

3.3. Data collection and preliminary X-ray crystallographic analysis

Diffraction data to 2.5 Å resolution were collected from a native crystal on in-house X-ray equipment. The crystal belongs to space group $P2_12_12_1$, with unit-cell parameters a = 33.6, b = 51.2, c = 114.2 Å. Assuming two monomers (14.3 kDa) per asymmetric unit, the Matthews coefficient (V_M) is 2.2 $Å^3\,Da^{-1}$, corresponding to a solvent content of 32%. No significant peak was observed in the self-rotation function section corresponding to $\kappa = 180°$, indicating the absence of non-crystallographic twofold rotational symmetry or the coincidence of the crystallographic and non-crystallographic rotation axes.

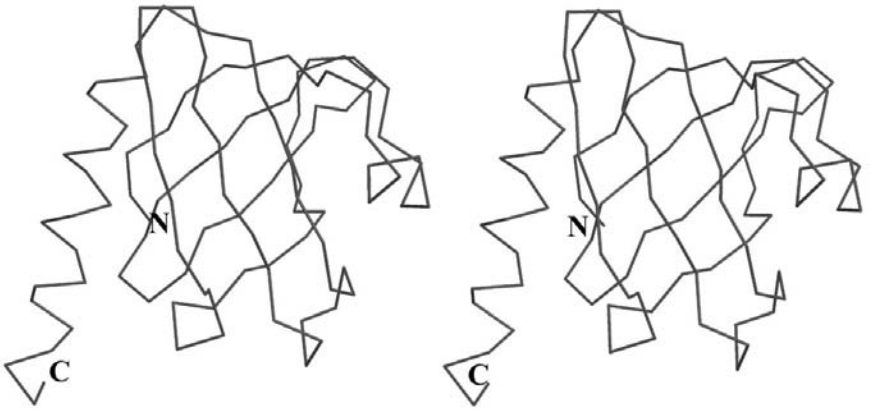

Figure 2
A stereo figure showing the C^{α} trace of the mdok1 PTB domain (generated using *MOLSCRIPT*; Kraulis, 1991). The N- and C-termini are indicated.

Attempts to solve the crystal structure of mdok1 PTB by molecular replacement using human dok5 PTB as a model (Shi *et al.*, 2002) were unsuccessful, despite the 28% sequence identity between the mdok1 and human dok5 PTB domains. MAD data were collected from a single SeMet-derivative crystal at peak (λ_1 = 0.9798 Å), inflection (λ_2 = 0.9800 Å) and remote (λ_3 = 0.9000 Å) wavelengths to 2.5 Å resolution. Statistics for MAD data collection are shown in Table 1. The SeMet-derivative crystal also belongs to space group $P2_12_12_1$, but with different unit-cell parameters (a = 41.1, b = 56.2, c = 99.6 Å) to the native crystal. Three selenium sites were located in the asymmetric unit of the protein and were sufficient for the determination of the mdok1 PTB domain crystal structure. Its fold consists of a β-sandwich composed of two nearly orthogonal seven-stranded antiparallel β-sheets and is capped at one side by a C-terminal α-helix (Fig. 2). The detailed three-dimensional structure will be described elsewhere.

We thank Dr Min Yao for assistance during data collection at beamline 41XU at SPring-8, Japan. This research was supported by the following grants: NSFC Nos. 39870174 and 39970155; Project '863' No. 2001AA233011; Project '973' Nos. G1999075602, G1999011902 and 1998051105; National Program for Key Basic Research Project ('973'; Nos. G19980510022 and 001cb510206); and the National High Technology Research and Development Program ('863'; Nos. 2001AA221041, 2002BA711A01-03); National Sciences Foundation of China (Nos. 39830070 and 3001161945).

References

Becker, E., Huynh-Do, U., Holland, S., Pawson, T., Daniel, T. O. & Skolnik, E. Y. (2000). *Mol. Cell. Biol.* **20**, 1537–1345.
Calderwood, D. A., Fujioka, Y., de Pereda, J. M., Garcia-Alvarez, B., Nakamoto, T., Margolis, B., McGlade, C. J., Liddington, R. C. & Ginsberg, M. H. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 2272–2277.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Grimm, J., Sachs, M., Britsch, S., Di Cesare, S., Schwarz-Romond, T., Alitalo, K. & Birchmeier, W. (2001). *J. Cell. Biol.* **154**, 345–354.
Jacobs, A. R., LeRoith, D. & Taylor, S. L. (2001). *J. Biol. Chem.* **276**, 40795–40802.
Kato, I., Takai, T. & Kudo, A. (2002). *J. Immunol.* **168**, 629–634.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Murakami, H., Yamamura, Y., Shimono, Y., Kawai, K., Kurokawa, K. & Takahashi, M. (2002). *J. Biol. Chem.* **277**, 32781–32790.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Shah, K. & Shokat, K. M. (2002). *Chem. Biol.* **9**, 35–47.
Shi, N., Zhou, W., Tang, K., Gao, Y., Jin, J., Gao, F., Peng, X., Bartlam, M., Qiang, B., Yuan, J. & Rao, Z. (2002). *Acta Cryst.* D**58**, 2170–2172.
Songyang, Z., Yamanashi, Y., Liu, D. & Baltimore, D. (2001). *J. Biol. Chem.* **276**, 2459–2465.
Wolf, G., Trub, T., Ottinger, E., Groninga, L., Lynch, A., White, M. F., Miyazaki, M., Lee, J. & Shoelson, S. E. (1995). *J. Biol. Chem.* **270**, 27407–27410.
Yamanashi, Y. & Baltimore, D. (1997). *Cell*, **88**, 205–211.