tions at either terminus did not induce NF-κB activity (Fig. 1C), implying that two DED modules are required for the activation.

2.2 Interaction of PDCasp8/10 with IKK, RIP, TRAF2, and NIK

In the TNFR1-mediated NF-κB activation pathway, RIP and IKK are thought to be essential components, although the importance of IKKα in this pathway has been challenged recently by the study of mice lacking IKKα [22, 23]. NF-κB-inducing kinase (NIK) has been reported to interact with TRAF2 and IKKα, and to induce NF-κB activity when overexpressed [24]. The involvement of NIK in the TNFR1-mediated NF-κB activation pathway, however, is controversial [25]. RICK (also called RIP2 and CARDIAK) is a homolog of RIP with a CARD domain in its C terminus instead of a death domain, and it is responsible for the NF-κB activation mediated by Toll-like receptors [26–29]. Next, we investigated the interactions of PDCasp8/10 with these components using the GST pull-down assay. Among the proteins described above, PDCasp8 interacted with RIP, NIK, and TRAF2 but not with IKKα, IKKβ and RICK (Fig. 2A, top panel). Considering the nature of this experiment, these molecules are indicated to directly interact with PDCasp8. A similar result was obtained using GST-PDCasp10 (Fig. 2A, middle panel).

Having demonstrated that both of the DED domains are required for NF-κB activation, we next examined the requirement of these domains for the interaction with binding partners. Again, when 30 or more amino acids were deleted from either the N or C terminus, all of the interactions with RIP, TRAF2 and NIK were abolished, demonstrating that the interaction with binding partners well-correlated with NF-κB activation (Fig. 2B).

2.3 Functional analysis of RIP deletion mutants

Originally identified as a Fas-interacting protein, RIP is a serine-threonine kinase that contains an N-terminal kinase domain, an α-helical intermediate domain, and a C-terminal death domain [30]. We next constructed deletion mutants of RIP as depicted in Fig. 3A and analyzed them for their ability to interact with PDCasp8/10. As shown in Fig. 3B, PDCasp-8/10 bound to all of the deletion mutants except for RIPΔKI, indicating that PDCasp8/10 interact with both kinase and intermediate domains. Interestingly, RIPΔK bound to PDCasp8/10 with a much higher affinity than other mutants, suggesting an inhibitory effect of the kinase domain on the intermediate domain for the binding to PDCasp8/10.

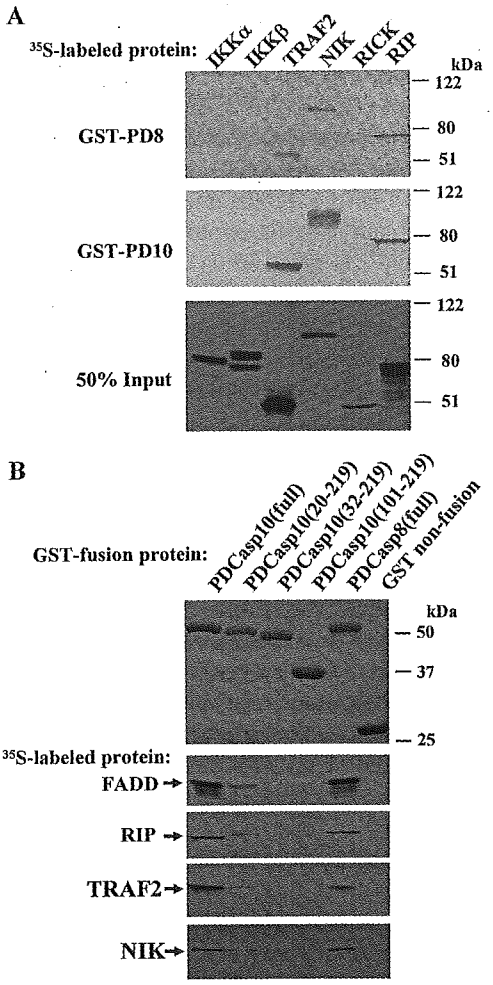

Fig. 2. Interaction of PDCasp8/10 with proteins implicated in the NF-κB pathway. (A) PDCasp8 and PDCasp10 fused with GST (GAT-PD8 and GST-PD10) were expressed in *E. coli* and purified. GST fusion proteins (10 μg) were immobilized on glutathione-Sepharose and incubated with 10 μl of the L-[^{35}S]methionine-labeled *in vitro*-translated proteins indicated at the top. After extensive washing, beads were boiled in Laemmli buffer. Eluted fractions were analyzed by SDS-PAGE and detected by autoradiography. Five microliter aliquots of *in vitro*-translated proteins were electrophoresed directly in the gel as a control and the autoradiograph is shown at the lower panel. (B) PDCasp10 and its deletion mutants were expressed in *E. coli* and purified. Two micrograms of each fusion protein stained with Coomassie brilliant blue after SDS-PAGE is shown at the upper panel. *In vitro* binding assay was performed as described in (A). FADD, a known binding partner of caspase-8 and -10, was also translated *in vitro* and used as a positive control.

We then tested these mutants for the ability to activate NF-κB by luciferase assay. As shown in Fig. 3C, the RIP wild-type and K45A (kinase-dead) mutant showed similar abilities to induce NF-κB activation indicating that the

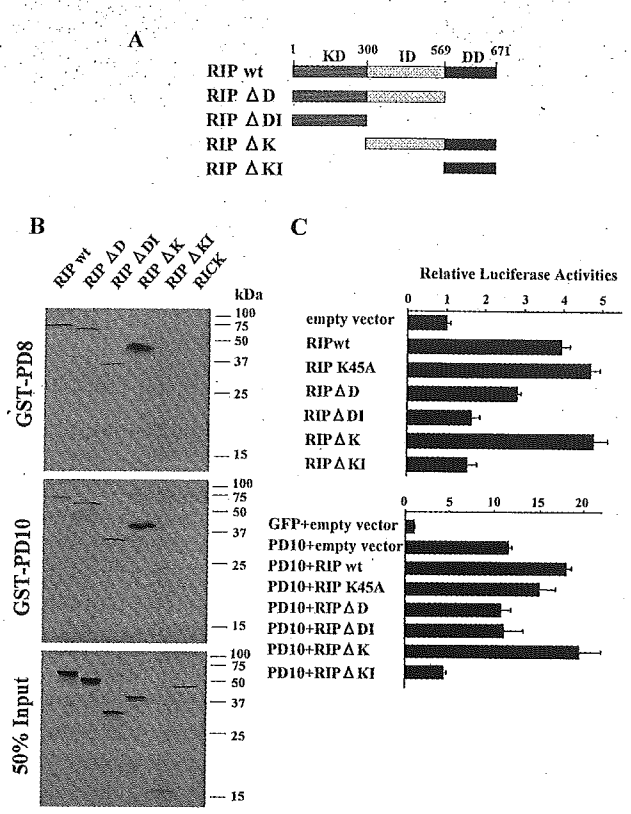

Fig. 3. Functional analysis of RIP deletion mutants. (A) Various deletion mutants used in (C) are depicted. Numbers refer to amino acid positions. KD, ID and DD denote the kinase domain, intermediate domain and death domain, respectively. All proteins were tagged with HA. (B) PDCasp8 and PDCasp10 fused with GST (GAT-PD8 and GST-PD10) were expressed in *E. coli* and purified. GST pull-down assays were performed as in Fig. 2 using the *in vitro*-translated proteins described in (A) as well as myc-tagged RICK. (C) 293 cells were transfected with the indicated expression plasmids as in Fig. 1A, and luciferase activities were measured. A plasmid for EGFP non-fusion protein (GFP) was used as a control, because PDCasp10 (PD10) was expressed as a GFP fusion.

kinase activity is dispensable. In contrast, RIPΔDI and RIPΔKI were defective in NF-κB activation suggesting an essential role of intermediate domain for this process as described previously [31]. We next investigated the effect of RIP and its deletion mutants on PDCasp10-induced NF-κB activation (Fig. 3C, lower graph). The wild type of RIP enhanced NF-κB activity induced by PDCasp10. However, RIPΔD and ΔDI did not have any effects on PDCasp10-induced NF-κB activation. In contrast, RIPΔKI, known as a dominant-negative mutant of wild-type RIP [31], remarkably inhibited NF-κB activation induced by PDCasp10. Although RIPΔK bound to PDCasp8/10 with a higher affinity, it did not have any effect on the synergy with PDCasp10 in terms of NF-κB activation, when compared with wild-type RIP. Taken together, these results suggest that the binding to PDCasp8/10 is necessary but not sufficient for RIP to enhance PDCasp8/10-induced NF-κB activation, and more importantly, RIP is required for PDCasp8/10 to activate NF-κB.

2.4 NIK and IKKα are required for the PDCasp8/10-induced NF-κB activation pathway

To determine the role of other upstream kinases of the NF-κB pathway, we used dominant-negative mutants of these kinases in NF-κB reporter assays. NIK, but not IKKα or IKKβ, synergistically activated NF-κB with PDCasp8/10 (Fig. 4A). A kinase-negative mutant of NIK completely inhibited the PDCasp8/10-induced NF-κB activation. Interestingly, IKKα mutant (mt), but not IKKβmt, both of which are kinase negative, had a similar effect to NIKmt, although IKKαmt was a weak inhibitor compared with NIKmt. These results indicated that NIK and IKKα are required for PDCasp8/10-mediated NF-κB activation. Gel shift experiments with nuclear extracts from cells transfected with various combinations of expression plasmids further confirmed the involvement of RIP, NIK and IKKα in PDCasp8-mediated NF-κB activation (Fig. 4B). However, PDCasp8 was a relatively weak stimulator of the NF-κB pathway compared with TNFR1 as previously described [15].

Since the inhibition of NF-κB activity by IKKαmt was partial, to confirm the requirement for IKKα, we employed small interfering RNA (siRNA) technology to shut down the expression of IKKα. The siRNA for IKKα and IKKβ successfully down-regulated the expressions of proteins translated from the respective genes, confirming the specificity of the experiment (Fig. 5A). The shutoff of IKKα, but not IKKβ, expression led to a partial but significant reduction of NF-κB activity induced by PDCasp8, again implying that IKKα is, at least in part, indeed required for this process (Fig. 5B). A partial inhibition may imply that low levels of IKKα expression still contribute to NF-κB activation. Alternatively, there may be a redundant pathway from PDCasp8/10 to NF-κB.

3 Discussion

Both caspase-8 and -10 have been identified as initiator caspases that are involved in a death receptor-mediated, also called an extrinsic, pathway of apoptosis. However, evidence is accumulating that these caspases do not always play roles solely associated with the induction of apoptosis. Both caspase-8 and -10 have isoforms that lack protease domains analogous to PDCasp8/10

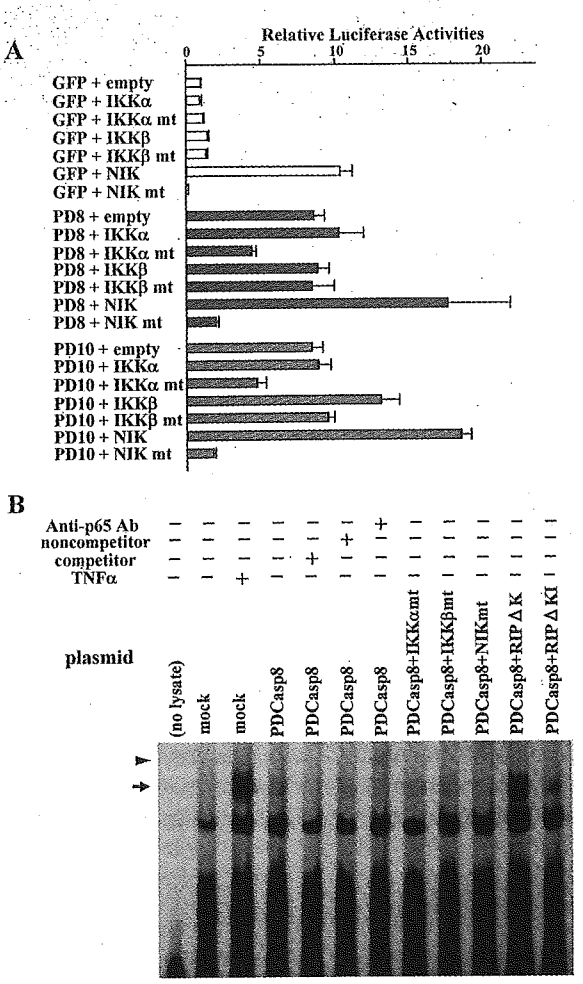

Fig. 4. NIK and IKKα are required for the PDCasp8/10-induced NF-κB activation pathway. (A) 293 cells were transfected with various combinations of the indicated expression plasmids as in Fig. 3C, and luciferase activities were measured. A plasmid for EGFP non-fusion protein (GFP) was used as a control because PDCasp8 (PD8) and PDCasp10 (PD10) were expressed as GFP fusions. (B) Radiolabeled oligonucleotide containing the consensus κB site was incubated with nuclear extracts from cells transfected with various combinations of the indicated expression plasmids in the presence or absence of a 50-fold excess of cold competitor containing the κB site or an irrelevant sequence or anti-p65 antibody. Nuclear extract from TNF-α-treated cells served as a positive control. The specific NF-κB-DNA complex and that supershifted by anti-p65 are indicated by an arrow and arrowhead, respectively.

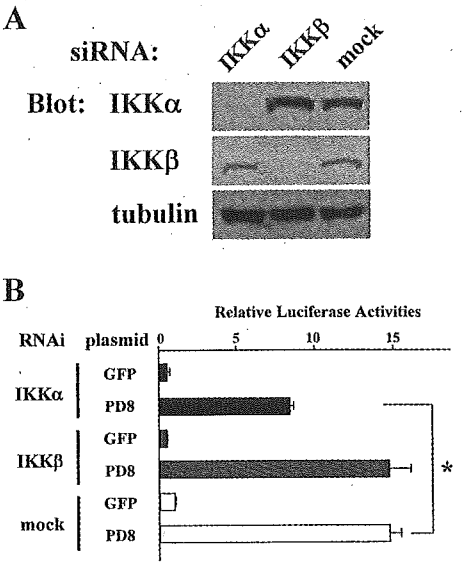

Fig. 5. Effects of reduced IKKα/β gene expression on PDCasp8-mediated NF-κB activation. (A) 293 cells were mock-transfected or transfected with siRNA for IKKα or IKKβ. At 48 h after the transfection, cells were harvested and subjected to SDS-PAGE followed by immunoblotting to verify the reduced expression by siRNA. Probing the blot with anti-tubulin antibody provided a control for protein loading. (B) Cells were transfected with siRNA as in (A). After 24 h, cells were subjected to a second transfection with plasmids for GFP or GFP-PDCasp8 (PD8) along with p3X-κB-luciferase reporter construct and pCMVβGal. At 24 h after the second transfection, cells were harvested and subjected to a luciferase assay. *$p<0.01$.

employed in our experiments, and the expression of the caspase-10 isoforms is reported to be highly variable among normal tissues, as well as among various tumor cell lines [2, 17]. Moreover, in a recent report, a novel caspase-8 isoform lacking the protease domain was identified. The expression level of this isoform was shown to be significantly decreased in the peripheral blood lymphocytes of some SLE patients compared to the healthy population [32]. In patients with SLE, apoptosis of the PBL has been reported to be increased compared with that of healthy controls, so it may be relevant to consider that this isoform has an anti-apoptotic effect. From these findings, along with our data that the expression of PDCasp8/10 affected the sensitivity of death receptor-mediated apoptotic signals [18], the mRNA expression profiles of caspase-8 and -10 isoforms may have an effect on the regulation of cell death. Their imbalance may lead to tumorigenesis or immunological disorders.

CASP8-deficient mice reportedly exhibit phenotypes, such as impaired heart muscle development, that cannot be explained solely by insufficient induction of apoptosis [33]. In addition, recently, a germ-line mutation in the *CASP8* gene has been reported to result not only in defective lymphocyte apoptosis, but also in pleiotropic defects in lymphocyte activation, again, indicating that caspase-8 has non-apoptotic signaling roles yet to be uncovered [34].

However, the molecular nature of these anti- or non-apoptotic signaling pathways mediated by caspase-8/10 has been elusive. Here, we demonstrated that PDCasp8/10 mediate NF-κB activation by binding to a selective set of proteins upstream of NF-κB. Although full-length caspase-8 and -10 do induce NF-κB activation as shown in Fig. 1, when their proapoptotic activities are blocked by a caspase inhibitor, we employed prodomain-only versions of caspase-8/10 to evaluate NF-κB activation more clearly by eliminating cell death. RIP is a serine-threonine kinase that interacts with multiple proteins, including Fas, TNFR1, TRADD, TRAF2, IKKγ and RAIDD [30, 31, 35, 36]. Recently, it was reported that caspase-8 cleaves RIP and prompts TNF-induced apoptosis [37]. Here, we provided evidence to show additional physical and functional interactions between RIP and caspase-8. PDCasp8/10 interacted with both the kinase and the intermediate domains and by using the death domain of RIP, previously reported to act as a dominant-negative inhibitor, we demonstrated the requirement of RIP for PDCasp8/10-mediated activation of NF-κB.

Previously, Chaudhary et al. [15] reported that caspase-8 and IKKα/β were coimmunoprecipitated with each other through the prodomain of caspase-8. Considering our result that the two do not physically associate using the GST pull-down assay, this interaction is suggested to be indirect. They also demonstrated that IKKβK44M but not IKKαK44M inhibited NF-κB activation induced by caspase-8/10 [15], while, according to Hu et al. [16], both of these mutants can inhibit this activation. Overexpression of dominant-negative mutants sometimes leads to contradictory results due to the various levels of expressed proteins or transfection efficiencies. To circumvent this problem, we employed siRNA technology to knock out IKKα or IKKβ to determine their roles under more physiological conditions. We, thus, conclude that IKKα but not IKKβ is required for caspase-8/10-induced NF-κB activation at least in our experimental systems.

Several lines of evidence reported recently suggest that there are two pathways leading to NF-κB activation, canonical and noncanonical pathways, the former requiring IKK complex and IκB degradation, the latter requiring NIK, IKKα and the processing of p100 [38]. PDCasp8/10 required NIK and IKKα, but not IKKβ, for the activation of NF-κB. However, the processing of p100 was not observed during this process (data not shown), making it difficult to determine which of the pathways are involved in PDCasp8/10-induced NF-κB activation. Although a more detailed analysis is obviously needed, this implies that it is not always possible to clearly separate these two pathways, or there may be cross-talk between these pathways. In any case, our study provides a potential mechanism to explain the unexpected phenotypes observed in mice or humans with a defective caspase-8 gene. Since PDcasp8/10 were identified to interact with TRAF2, the MAPK pathway would be another plausible target for future investigation.

4 Materials and methods

4.1 Cells

The human embryonic kidney cell line 293 was cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 50 U/ml penicillin and 0.1 mg/ml streptomycin at 37°C in a humidified atmosphere of 5% CO_2 and 95% air.

4.2 Plasmids

To generate plasmids for GST-PDCasp8, GST-PDCasp10 and their deletion mutants, DNA fragments encoding the respective prodomains were excised from the plasmids for EGFP-caspase fusion proteins [18] and subcloned into pGEX-5X-1 (Amersham Pharmacia Biotech). The p3X-κB-luciferase was provided by Dr. B. Sugden (University of Wisconsin). Plasmids for Flag-IKKα, -IKKαmt (K44A), -IKKβ, and -IKKβmt (K44A) were provided by Dr. Y. Nakano (Juntendo University, Tokyo, Japan). The pcDNA3-RICK-Myc was provided by Dr. N. Inohara (University of Michigan Medical School). The pcDNA3-Flag-NIK and -NIKmt (KK429–430AA) were provided by Dr. D. Wallach (The Weizmann Institute of Science). The cDNA for RIP and TRAF2 were gifts from Dr. D. V. Goeddel (Tularik Inc.) and Dr. T. Watanabe (University of Tokyo), respectively. To construct plasmids for TRAF2, RIP and RIP deletion mutants, RIPΔD (amino acid residues 1–568), RIPΔDI (1–300), RIPΔK (301–671) and RIPΔKI (569–671), PCR products for the respective fragments were subcloned into pcDNA-3XHA [18] after digestion with the appropriate enzymes. RIPK45A was constructed by PCR-mediated mutagenesis as described previously [39]. All other constructs have been described previously [18, 21].

4.3 Transfection and luciferase assay

293 cells were transfected with various combinations of expression constructs as indicated along with 3X-κB-luciferase and pCMVβ-gal plasmid using Effectene reagent (Qiagen). At the time of transfection, z-VAD-fmk (50 μM) was used for all samples to prevent cell loss. The cells were harvested at 24 h after transfection without prior washing of the cell surface, and used for a luciferase assay. Luciferase activities were measured as described previously and normalized for transfection efficiency based on β-galactosidase activities [18].

4.4 Western blotting

Western blot analysis was performed as described previously [18]. In brief, 30 μg of the cell lysate was subjected to SDS-PAGE and transferred onto a nitrocellulose membrane. The membrane was incubated with anti-IKKα polyclonal antibody (Santa Cruz), anti-IKKβ monoclonal antibody (Upstate Biotechnology), or anti-tubulin monoclonal antibody (Sigma-Aldrich) followed by horseradish peroxidase-conjugated anti-rabbit immunoglobulins (Sigma) or horseradish peroxidase-conjugated anti-mouse immunoglobulins (DAKO). The proteins were visualized by using enhanced chemiluminescence immunoblotting detection reagents (Amersham).

4.5 *In vitro* binding assay

The *in vitro* binding assay using GST-fusion proteins was described previously [40]. Briefly, ^{35}S-labeled proteins were generated using the TNT T7-coupled reticulocyte lysate system (Promega) according to the manufacturer's protocol. Glutathione-Sepharose beads were added to 10 μg of each GST fusion protein, and washed 4-times with HKM buffer (10 mM Hepes pH 7.2, 142.5 mM KCl, 5 mM $MgCl_2$, 1 mM EGTA and 0.2% NP40). Then the beads were combined with the *in vitro* translation lysate in 150 μl HKM buffer and incubated at 4°C for 2 h. The beads were again washed four times with HKM buffer and the bound proteins were resolved in SDS polyacrylamide gels and visualized by autoradiography.

4.6 Electrophoretic mobility shift assay

293 cells were cultured for 4 h in the presence or absence of TNF-α (10 ng/ml). For some experiments, 293 cells were transfected with the plasmid for PDCasp8 or PDCasp10. Nuclear extracts were prepared according to a previous report [41] and incubated with an end-labeled double-stranded oligonucleotide probe containing the sequence for a consensus κB site (5'-ATCAGGGACTTTCCGCTGGGGACTTTCCG-3'). The reaction was performed in 15 μl binding buffer [4% glycerol, 1 mM $MgCl_2$, 0.5 mM EDTA, 50 mM NaCl, 10 mM Tris-HCl pH 7.5, and 0.05 mg/ml poly(dI-dC)] for 20 min at room temperature. For competition experiments, a 50-fold excess of unlabeled, double-stranded oligonucleotide, containing either a κB site or glucocorticoid response element (5'-GATCAGAACACAGTGTTCTCTA-3') as an irrelevant sequence, was included in binding reactions. Supershift analysis was performed by including anti-p65 monoclonal antibody (Santa Cruz) in binding reactions. Samples were fractioned on a non-denaturing 6% polyacrylamide gel and visualized by autoradiography.

4.7 siRNA analysis

The 21-nucleotide siRNA were custom-synthesized (JBioS, Asaka, Japan). The siRNA sequence targeting IKKα and IKKβ corresponded to the coding regions 588–606 and 742–760 relative to the first nucleotide of the start codon, respectively. For annealing of siRNA, 50 μM single strands were incubated in annealing buffer (100 mM potassium acetate, 30 mM Hepes-KOH, pH 7.4, and 2 mM magnesium acetate) for 1 min at 90°C followed by 1 h at 37°C.

293 cells were seeded at a density of 2×10^5 cells/well using six-well plates and cultured for 24 h. Cells were then transfected with 4 μg siRNA using GeneSilencer transfection regent (Gene Therapy Systems, Inc.) according to the manufacturer's protocol. At 24 h after the introduction of siRNA, 1.5 μg of expression plasmid for GFP or GFP-PDCasp8 along with 0.9 μg of the p3X-κB-luciferase reporter construct and 1.0 μg pCMVβGal was transfected using Effectene reagent (Qiagen). After another 24 h, the cells were harvested without prior washing of the cell surface, and subjected to a luciferase assay as described before [18]. The results were normalized for transfection efficiency based on β-galactosidase activities.

Acknowledgments: We are grateful to Drs. B. Sugden, Y. Nakano, N. Inohara, D. Wallach, D. V. Goeddel and T. Watanabe for providing plasmids. We thank Yuko Ohtsuka and Mami U for excellent technical assistance. We also thank Kayoko Saito for secretarial assistance. This study was supported in part by grants for Brain Research and for Genome Research from the Ministry of Health, Labour and Welfare, a Grant-in-Aid for Scientific Research and a Grant for Organized Research Combination System from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1 **Kischkel, F. C., Hellbardt, S., Behrmann, I., Germer, M., Pawlita, M., Krammer, P. H. and Peter, M. E.,** Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *EMBO J.* 1995. **14:** 5579–5588.

2 **Boldin, M. P., Goncharov, T. M., Goltsev, Y. V. and Wallach, D.,** Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* 1996. **85:** 803–815.

3 **Fernandes-Alnemri, T., Armstrong, R. C., Krebs, J., Srinivasula, S. M., Wang, L., Bullrich, F., Fritz, L. C., Trapani, J. A., Tomaselli, K. J., Litwack, G. and Alnemri, E. S.,** *In vitro* activation of CPP32 and Mch3 by Mch4, a novel human apoptotic cysteine protease containing two FADD-like domains. *Proc. Natl. Acad. Sci. USA* 1996. **93:** 7464–7469.

4 **Muzio, M., Chinnaiyan, A. M., Kischkel, F. C., O'Rourke, K., Shevchenko, A., Ni, J., Scaffidi, C., Bretz, J. D., Zhang, M., Gentz, R., Mann, M., Krammer, P. H., Peter, M. E. and Dixit, V. M.,** FLICE, a novel FADD-homologous ICE/CED-3-like protease,

is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* 1996. **85:** 817–827.

5 **Vincenz, C. and Dixit, V. M.,** Fas-associated death domain protein interleukin-1β-converting enzyme 2 (FLICE2), an ICE/Ced-3 homologue, is proximally involved in CD95- and p55-mediated death signaling. *J. Biol. Chem.* 1997. **272:** 6578–6583.

6 **Nagata, S.,** Apoptosis by death factor. *Cell* 1997. **88:** 355–365.

7 **Alnemri, E. S., Livingston, D. J., Nicholson, D. W., Salvesen, G., Thornberry, N. A., Wong, W. W. and Yuan, J.,** Human ICE/CED-3 protease nomenclature. *Cell* 1996. **87:** 171.

8 **Hsu, H., Xiong, J. and Goeddel, D. V.,** The TNF receptor 1-associated protein TRADD signals cell death and NF-κB activation. *Cell* 1995. **81:** 495–504.

9 **Ashkenazi, A. and Dixit, V. M.,** Death receptors: signaling and modulation. *Science* 1998. **281:** 1305–1308.

10 **Kelliher, M. A., Grimm, S., Ishida, Y., Kuo, F., Stanger, B. Z. and Leder, P.,** The death domain kinase RIP mediates the TNF-induced NF-κB signal. *Immunity* 1998. **8:** 297–303.

11 **Lee, S. Y., Reichlin, A., Santana, A., Sokol, K. A., Nussenzweig, M. C. and Choi, Y.,** TRAF2 is essential for JNK but not NF-κB activation and regulates lymphocyte proliferation and survival. *Immunity* 1997. **7:** 703–713.

12 **Yeh, W. C., Shahinian, A., Speiser, D., Kraunus, J., Billia, F., Wakeham, A., de la Pompa, J. L., Ferrick, D., Hum, B., Iscove, N., Ohashi, P., Rothe, M., Goeddel, D. V. and Mak, T. W.,** Early lethality, functional NF-κB activation, and increased sensitivity to TNF-induced cell death in TRAF2-deficient mice. *Immunity* 1997. **7:** 715–725.

13 **Wang, C. Y., Mayo, M. W., Korneluk, R. G., Goeddel, D. V. and Baldwin, A. S. J.,** NF-κB anti-apoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. *Science* 1998. **281:** 1680–1683.

14 **Karin, M.,** The beginning of the end: IκB kinase (IKK) and NF-κB activation. *J. Biol. Chem.* 1999. **274:** 27339–27342.

15 **Chaudhary, P. M., Eby, M. T., Jasmin, A., Kumar, A., Liu, L. and Hood, L.,** Activation of the NF-κB pathway by caspase 8 and its homologs. *Oncogene* 2000. **19:** 4451–4460.

16 **Hu, W. H., Johnson, H. and Shu, H. B.,** Activation of NF-κB by FADD, Casper, and caspase-8. *J. Biol. Chem.* 2000. **275:** 10838–10844.

17 **Ng, P. W., Porter, A. G. and Janicke, R. U.,** Molecular cloning and characterization of two novel pro-apoptotic isoforms of caspase-10. *J. Biol. Chem.* 1999. **274:** 10301–10308.

18 **Shikama, Y., Shen, L., Yonetani, M., Miyauchi, J., Miyashita, T. and Yamada, M.,** Death effector domain-only polypeptides of caspase-8 and -10 specifically inhibit death receptor-induced cell death. *Biochem. Biophys. Res. Commun.* 2002. **291:** 484–493.

19 **Bertin, J., Nir, W. J., Fischer, C. M., Tayber, O. V., Errada, P. R., Grant, J. R., Keilty, J. J., Gosselin, M. L., Robison, K. E., Wong, G. H., Glucksmann, M. A. and DiStefano, P. S.,** Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-κB. *J. Biol. Chem.* 1999. **274:** 12955–12958.

20 **Willis, T. G., Jadayel, D. M., Du, M. Q., Peng, H., Perry, A. R., Abdul-Rauf, M., Price, H., Karran, L., Majekodunmi, O., Wlodarska, I., Pan, L., Crook, T., Hamoudi, R., Isaacson, P. G. and Dyer, M. J.,** Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. *Cell* 1999. **96:** 35–45.

21 **Shikama, Y., U, M., Miyashita, T. and Yamada, M.,** Comprehensive studies on subcellular localizations and cell death-inducing activities of eight GFP-tagged apoptosis-related caspases. *Exp. Cell. Res.* 2001. **264:** 315–325.

22 **Hu, Y., Baud, V., Delhase, M., Zhang, P., Deerinck, T., Ellisman, M., Johnson, R. and Karin, M.,** Abnormal morphogenesis but intact IKK activation in mice lacking the IKKα subunit of IκB kinase. *Science* 1999. **284:** 316–320.

23 **Takeda, K., Takeuchi, O., Tsujimura, T., Itami, S., Adachi, O., Kawai, T., Sanjo, H., Yoshikawa, K., Terada, N. and Akira, S.,** Limb and skin abnormalities in mice lacking IKKα. *Science* 1999. **284:** 313–316.

24 **Malinin, N. L., Boldin, M. P., Kovalenko, A. V. and Wallach, D.,** MAP3K-related kinase involved in NF-κB induction by TNF, CD95 and IL-1. *Nature* 1997. **385:** 540–544.

25 **Yin, L., Wu, L., Wesche, H., Arthur, C. D., White, J. M., Goeddel, D. V. and Schreiber, R. D.,** Defective lymphotoxin-β receptor-induced NF-κB transcriptional activity in NIK-deficient mice. *Science* 2001. **291:** 2162–2165.

26 **Inohara, N., del Peso, L., Koseki, T., Chen, S. and Nunez, G.,** RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. *J. Biol. Chem.* 1998. **273:** 12296–12300.

27 **McCarthy, J. V., Ni, J. and Dixit, V. M.,** RIP2 is a novel NF-κB-activating and cell death-inducing kinase. *J. Biol. Chem.* 1998. **273:** 16968–16975.

28 **Thome, M., Hofmann, K., Burns, K., Martinon, F., Bodmer, J. L., Mattmann, C. and Tschopp, J.,** Identification of CARDIAK, a RIP-like kinase that associates with caspase-1. *Curr. Biol.* 1998. **8:** 885–888.

29 **Kobayashi, K., Inohara, N., Hernandez, L. D., Galan, J. E., Nunez, G., Janeway, C. A., Medzhitov, R. and Flavell, R. A.,** RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. *Nature* 2002. **416:** 194–199.

30 **Stanger, B. Z., Leder, P., Lee, T. H., Kim, E. and Seed, B.,** RIP: a novel protein containing a death domain that interacts with Fas/APO-1 (CD95) in yeast and causes cell death. *Cell* 1995. **81:** 513–523.

31 **Hsu, H., Huang, J., Shu, H. B., Baichwal, V. and Goeddel, D. V.,** TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. *Immunity* 1996. **4:** 387–396.

32 **Horiuchi, T., Himeji, D., Tsukamoto, H., Harashima, S., Hashimura, C. and Hayashi, K.,** Dominant expression of a novel splice variant of caspase-8 in human peripheral blood lymphocytes. *Biochem. Biophys. Res. Commun.* 2000. **272:** 877–881.

33 **Varfolomeev, E. E., Schuchmann, M., Luria, V., Chiannilkulchai, N., Beckmann, J. S., Mett, I. L., Rebrikov, D., Brodianski, V. M., Kemper, O. C., Kollet, O., Lapidot, T., Soffer, D., Sobe, T., Avraham, K. B., Goncharov, T., Holtman, H., Lonai, P. and Wallach, D.,** Targeted disruption of the mouse caspase 8 gene ablates cell death induction by the TNF receptors, Fas/Apo1, and DR3 and is lethal prenatally. *Immunity* 1998. **9:** 267–276.

34 **Chun, H. J., Zheng, L., Ahmad, M., Wang, J., Speirs, C. K., Siegel, R. M., Dale, J. K., Puck, J., Davis, J., Hall, C. G., Skoda-Smith, S., Atkinson, T. P., Straus, S. E. and Lenardo, M. J.,** Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. *Nature* 2002. **419:** 395–399.

35 **Duan, H. and Dixit, V. M.,** RAIDD is a new 'death' adaptor molecule. *Nature* 1997. **385:** 86–89.

36 **Inohara, N., Koseki, T., Lin, J., del Peso, L., Lucas, P. C., Chen, F. F., Ogura, Y. and Nunez, G.,** An induced proximity model for NF-κB activation in the Nod1/RICK and RIP signaling pathways. *J. Biol. Chem.* 2000. **275:** 27823–27831.

37 **Lin, Y., Devin, A., Rodriguez, Y. and Liu, Z. G.,** Cleavage of the death domain kinase RIP by caspase-8 prompts TNF-induced apoptosis. *Genes Dev.* 1999. **13:** 2514–2526.

38 **Pomerantz, J. L. and Baltimore, D.,** Two pathways to NF-κB. *Mol. Cell* 2002. **10:** 693–695.

39 **Imai, Y., Matsushima, Y., Sugimura, T. and Terada, M.,** A simple and rapid method for generating a deletion by PCR. *Nucleic. Acids. Res.* 1991. **19:** 2785.

40 **U, M., Miyashita, T., Ohtsuka, Y., Okamura-Oho, Y., Shikama, Y. and Yamada, M.,** Extended polyglutamine selectively interacts with caspase-8 and -10 in nuclear aggregates. *Cell Death Differ.* 2001. **8:** 377–386.

41 **Pfeffer, K., Matsuyama, T., Kundig, T. M., Wakeham, A., Kishihara, K., Shahinian, A., Wiegmann, K., Ohashi, P. S., Kronke, M. and Mak, T. W.,** Mice deficient for the 55 kDa tumor necrosis factor receptor are resistant to endotoxic shock, yet succumb to *L. monocytogenes* infection. *Cell* 1993. **73:** 457–467.

Correspondence: Toshiyuki Miyashita, Department of Genetics, National Research Institute for Child Health and Development, 3-35-31 Taishido, Setagaya-ku, Tokyo 154-8567, Japan
Fax: +81-3-3414-3208
e-mail: tmiyashita@nch.go.jp

J Hum Genet (2003) 48:410–414
DOI 10.1007/s10038-003-0047-x

ORIGINAL ARTICLE

Atsushi Miyahara · Yuko Okamura-Oho
Toshiyuki Miyashita · Akinori Hoshika
Masao Yamada

Genomic structure and alternative splicing of the insulin receptor tyrosine kinase substrate of 53-kDa protein

Received: 2 April 2003 / Accepted: 26 May 2003 / Published online: 16 July 2003

Abstract Insulin receptor tyrosine kinase substrate of 53-kDa protein (IRSp53) is now known to be a key factor in cytoskeleton reorganization. The human IRSp53 was identified as a binding partner with DRPLA protein, a product of the gene responsible for a neurodegenerative disorder, dentatorubral pallidoluysian atrophy, as well as a binding partner with brain-specific angiogenesis inhibitor 1. Previous studies identified at least four isoforms (L-, M-, S- and T-forms) in human, where 511 amino acid residues from the N-terminus were identical, followed by unique sequences of 9–41 amino acid residues. As each isoform had a distinct function, the unique sequences at the C-terminus had a vital role in its function. Here we report that these isoforms were indeed generated by alternative splicing, which was established by experimental and computational studies on human and rodent genomes. Previous biochemical reports suggested that rodents may lack one of the isoforms (L-form). This study solved this issue, as a nucleotide substitution occurred at a splice donor site followed by a large deletion in the rodent genome compared with human, which made the generation of the L-form impossible. This study also revealed overlapping of the *IRSp53* and *AATK* genes coded for by complementary strands.

Keywords Alternative splicing · Genomic organization · Insulin receptor tyrosine kinase substrate · IRSp53 · AATK

Accession numbers: AB104726, AB104729, AB105194, AB105195, AB105196

A. Miyahara · Y. Okamura-Oho · T. Miyashita · M. Yamada (✉)
Department of Genetics, National Research Institute for Child Health and Development,
3-35-31 Taishido, Setagaya-ku,
Tokyo 154-8567, Japan
E-mail: myamada@nch.go.jp
Tel.: +81-3-34160181
Fax: +81-3-34143208

A. Miyahara · A. Hoshika
Department of Pediatrics, Tokyo Medical University,
6-7-1 Nishishinjuku, Shinjuku-ku, Tokyo 160-0023, Japan

Introduction

An insulin receptor tyrosine kinase substrate of 53/58-kDa protein was originally identified in hamster cells through biochemical studies after insulin and/or IGF-I treatment (Yeh et al. 1996). It is phosphorylated upon stimulation with insulin and/or IGF-I, but differs from other members of the well-known insulin receptor substrate groups (namely, human *IRS1*, *IRS2* and *IRS4*) in terms of conserved amino-acid sequence motifs and other features (Hubbard and Till 2000). The human homologue was identified as a binding partner with *DRPLA* (Okamura-Oho et al. 1999), in which CAG triplet repeat expansion in the coding region causes a neurodegenerative disorders, dentatorubral pallidoluysian atrophy (Naito and Oyanagi 1982; Nagafuchi et al. 1994a, 1994b; Koide et al. 1994). The human homologue was also identified as a binding partner with a serpentine receptor, brain-specific angiogenesis inhibitor 1 (*BAI1*), and named as BAI1-associated protein 2 (*BAIAP2*) (Oda et al. 1999). The human homologue not only has a sequence similarity with hamster IRSp53/58, but also has been demonstrated to be phosphorylated upon stimulation with insulin and/or IGF-I (Okamura-Oho et al. 1999). IRSp53 is now highlighted as a key factor in the cytoskeleton reorganization: IRSp53 functions as an adaptor that binds Rho family GTPases (Rho, Rac and cdc42) and their effectors (mDia, WAVE2 and Mena), and mediates the activation of these molecules (Miki et al. 2000; Krugmann et al. 2001; Miki and Takenawa 2002). The cdc42 protein controls the formation of actin bundles in membrane ruffling and filopodia formation at the cellular periphery. IRSp53 is also known to localize at postsynaptic density of the central nerve system, which suggests a role in neurite outgrowth (Abbott et al. 1999; Soltau et al. 2002).

To date, at least four isoforms of IRSp53 have been identified in human. We identified IRSp53-L and IRSp53-S, consisting of 552 and 521 amino acid residues, respectively, as binding partners with DRPLA protein (Okamura-Oho et al. 1999). Oda and co-workers

(1999) identified two isoforms, named as BAIAP2-α and -β, which were composed of 521 and 520 amino acids, respectively. BAIAP2-α is identical to IRSp53-S, while BAIAP2-β is unique. Accordingly, we use IRSp53-T in this report rather than BAIAP2-β. The fourth isoform (IRS-58), with 534 amino acid residues, was identified during a cloning process of binding partners with cdc42 (Govind et al. 2001). As the relationship between protein isoforms of 53 or 58 kDa and mRNA isoforms is still uncertain, IRS-M is used in this report for the isoform. The four mRNA isoforms have been repeatedly confirmed in RT-PCR by both others and ourselves, as well as in many expression sequence tags (EST). The four IRSp53 transcripts generate respective protein isoforms sharing the identical 511 amino acid residues from the N-terminus and differing only in short peptide sequences at their C-terminus. Each isoform has distinct functions; for example, IRSp53-L and -S were phosphorylated with insulin but not with IGF-I in transfected cultured cells, while IGF-I phosphorylated only the T-form (Okamura-Oho et al. 2001). Thus, the unique short peptide sequences at the C-terminus have a vital role in its function probably through regulating accessibility to functional sites by intra-molecular binding. This is quite important as there are several discordance results in functional analyses with IRSp53 expression vectors. These isoforms are supposed to be generated by alternative splicing, but it has been proved yet. Here, we report that the four isoforms are indeed generated by alternative splicing by experimental and computational studies. This study on human and rodent genomes solved the issue of whether rodents lack one of the isoforms (L-form).

Materials and methods

DNA analyses

Human genomic DNA was isolated from peripheral leukocytes with the standard phenol/chloroform extraction method. Genomic DNA of mouse and Sprague-Dawley (SD) rat was prepared from tail tissues with the Dneasy Tissue Kit (QUIAGEN, Hilden, Germany). Polymerase chain reaction (PCR) was conducted as previously described (Tadokoro et al. 1992). Briefly, the reaction mixture consisted of 10 ng genomic DNA, 0.5 μM primers, 200 μM of each dNTP and 0.5 U Taq DNA polymerase (Takara, Shiga, Japan) in standard reaction buffer. The PCR conditions consisted of one cycle at 94 °C for 4 min, 30 cycles at 94 °C for 1 min for denaturation, 56 °C for 1 min for annealing and 72 °C for 4 min for extension, and then another step at 72 °C for 6 min to ensure complete extension. The following primer sets were used to amplify the exon 16-*AATK* region: mouse 5′-TGCAGTCCTGTGCCTTGCGA-3′ (forward) and 5′-AGAGATGCCCTCTGCAGGGTAGT-3′ (reverse); rat 5′-CAGGAATCCCTTCGCCAACGTC-3′ (forward) and 5′-AGATGCCCTCTGCAGGGTAGT-3′ (reverse). PCR products were purified with QIAXII (QIAGEN, Hilden, Germany), and subjected to a direct sequence analysis with CEQ2000 Dye Terminator Cycle Sequencing with the Quick Start Kit (Beckman Coulter, Fullerton, Calif., USA) and a Beckman CEQ2000 automatic sequencer (Beckman Coulter, Fullerton, Calif., USA). Sequence analyses were done with computer software, Genetyx-Win (Genetyx, Tokyo, Japan).

Results and discussion

Genomic organization of human *IRSp53*

When we started this study, few genomic sequences for *IRSp53* were detectable in public databases, which made it impossible to determine the genomic organization of the *IRSp53* gene only with computational analyses. Thus, we attempted to isolate genomic clones especially covering the C-terminal region with primer pairs designed based on the cDNA sequences. As the order of exons and their boundaries were unknown at that time, multiple combinations of primers were used to try to clone intronic sequences. Several sets of primers successfully generated DNA fragments by PCR and the nucleotide sequences were determined, some of which were deposited in a public database (see Comments). After more genomic sequences were deposited in public databases, along with the progress of human genome project, it became easier to identify genomic sequences covering the *IRSp53* gene by BLAST searches with the cDNA sequences as well as the genomic sequences we determined. The representative clones and sequences covering the *IRSp53* gene turned out to be PR11-149I9 (AC115099) and RP13-1277B16 (AC129919). These sequences had no annotation for *IRSp53* to date.

Comparing the cDNA (accession numbers NM_017450, NM_017451 and NM_006340 for the S-, L- and T-forms) and genomic sequences, the human *IRSp53* gene spanned about 82 kb, and consisted of 17 exons (Fig. 1). Except for the transcriptional termini (exons 14, 16 and 17), all the exon-intron boundaries were accorded for the consensus GT-AG rule (Fig. 2). The common part of the four isoforms was encoded by exon 1 through 13 (Fig. 3). The S-form went through to exon 14 and ended with a polyadenylation (polyA) signal. The nucleotide sequence we previously determined (AB017120) had 2,033 bp followed by the polyA tail, while the NM_017450 sequence had additional 135 bp.

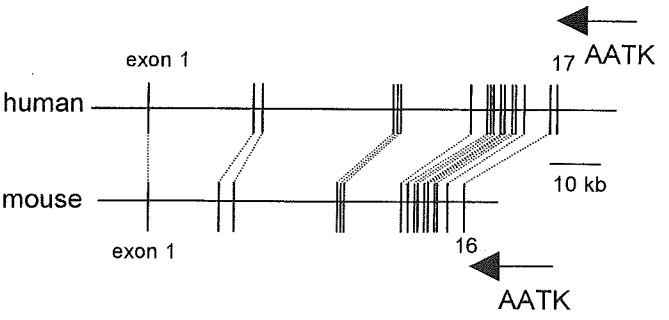

Fig. 1 Genomic organization of human and mouse *IRSp53* genes. Schematic illustrations showing that the genomic organization is generally well conserved between human and mouse, except for exon 17. Note that although we have tried to illustrate it as faithfully as possible, the exons and several introns are too small to draw in scale

human

intron		exon		intron
---tggttcgggt	CCGCT--	1	--ATAAG	gtgagcgccc---
22544	1		147	22692
---tttcttccag	ACCAT--	2	--GGCAG	gtggaactgc---
41041	148		223	41118
---tctctctcag	GTGTG--	3	--ACTCG	gtgagacccc---
45254	224		310	45342
---tgcttcccag	GAGAC--	4	--AAATG	gtgagtccac---
72205	311		372	72268
---ccatgtccag	CTGAA--	5	--TGAGT	gtaagtgcac---
73027	373		444	73100
---tgaatctcag	GCTGC--	6	--TGCAG	gtaggcccgc---
73816	445		582	73955
---ccctccacag	TACAT--	7	--CCAAG	gtgaggcggc---
87301	583		735	87455
---tccacttcag	GGCAA--	8	--TGGGG	gtgagtctgt---
90875	736		957	91098
---ttccctgcag	CGGAT--	9	--CACCA	gtaagggctc---
91280	958		1159	91483
---gtcccagcag	CCGAG--	10	--AAGAT	gtgagtgttt---
91887	1160		1361	92090
---tctcttccag	GCGGG--	11	--ATGAG	gtgagctctg---
93451	1362		1430	93521
---gtccctacag	CCTGC--	12	--CCCAG	gtcagtgggc---
94118	1431		1593	94282
---ctgcccccag	GGCCT--	13	--AGCAG	gtaaggggac---
95848	1594		1628	95884
---tgtttcacag	TGGCA--	14	--ATTGC	acgagttggg---
96341	S 1629		S 3168	97887
---ttccttgcag	CGCCG--	15	--AGTTA	gtaagttgcc---
98287	T 1629		T 1674	98334
---tctctttcag	GAATC--	16	--CCTGC	accagGTGTG---
103143	T 1675		T 2129	103599
---CCTGCaccag	GTGTG--	17	--ACAAT	aacttaaaat---
103603	L 1677		L 2877	104807

mouse

---tgcagttgtc	GCTTT--	1	--ACAAG	gtgagtttcc---
31401786	1		115	31401902
---tcttttccag	ACCAT--	2	--GGCAG	gtatagctgg---
31415754	116		191	31415831
---ctcccctcag	GTGTC--	3	--ACTTG	gtaagaccct---
31418809	192		278	31418897
---tttcctccag	GGGAC--	4	--AGACG	gtgagtttgg---
31439446	279		340	31439509
---ctgtgtccag	CTGAA---	5	--TAAGT	gtaagtacag---
31440050	341		412	31440123
---gggtctccag	GCTGC--	6	--TGCAG	gtaggtctgc---
31440646	413		550	31440785
---ctgtccccag	TACAT--	7	--CCAAG	gtgagctagg---
31452354	551		703	31452508
---tccactccag	GGCAA--	8	--TGGGG	gtgagtcctg---
31455181	704		928	31455407
---cttcctccag	CGGAT--	9	--CACCA	gtaagggcct---
31455547	929		1130	31455750
---gtctcggcag	CTGAG--	10	--AAGAT	gtgagcaccc---
31456147	1131		1332	31456350
---tttcttctag	GCGGG--	11	--ATGAG	gtaagcatac---
31457292	1333		1401	31457362
---cgccccgcag	CCTGC--	12	--CTCAG	gtgaggcctg---
31457793	1402		1564	31457957
---ttgttcccag	GGTCT--	13	--AGCAG	gtaagaggtt---
31459210	1565		1599	31459246
---tgttccacag	TGGCA--	14	--GCTCT	ctgcgcccct---
31459732	1600		1798	31459932
---ttcctcacag	CGCGG--	15	--AGTTA	gtaagttgcc---
31461696				31461745
---ctttccccag	GAATC--	16	--TCACC	atgtgtagtg---
31465018				31465438

Fig. 2 Exon-intron boundaries of the human and mouse *IRSp53* genes. The boundaries were defined by alignment of the cDNA and genomic sequences. It should be noted that the downstream boundaries for exons 14, 16 and 17 are the end of the cDNA sequences indicated, and does not necessarily mean the position of the polyA tail. *Upper and lower* cases indicate exon and intron sequences, respectively, and the position of the boundary nucleotide in the given sequences are indicated. The accession numbers of referenced sequences are AC115099 for human genome, NM_017450 for human S-form cDNA, NM_017451 for human L-form cDNA, NM_006340 for human T-form cDNA, NT_039521 for mouse genome and AF390178 for mouse cDNA

There was a typical polyA signal in the genomic sequence near the end of AB017120, but also a continuous A sequence as well in the genome (at 96747 in AC115099). Thus, the exact termini of transcripts are

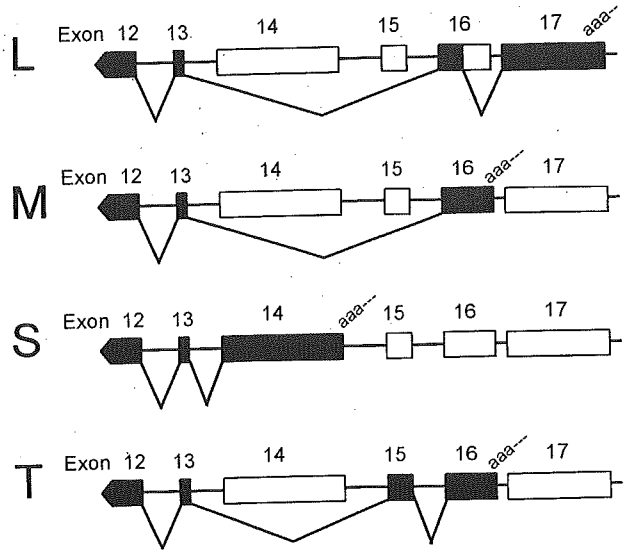

Fig. 3 Pattern of alternative splicing of the human *IRSp53* gene to generate four isoforms. The *black shaded* regions are used by respective isoforms; *aaa* polyA tails

somewhat uncertain, and we use a longer transcript in comparison in Fig. 2. The T-form skipped exon 14 and used exons15 and 16. Both the L and M forms skipped exons14 and 15 and reached to exon 16. The M-form ended with the polyA signal near the downstream boundary of exon 16. In contrast, the L form left exon 16 halfway, just ahead of the stop codon in frame, and resumed at exon 17, resulting in further extension of amino acid coding. The splice donor site in exon 16 for the L-form was also accorded for the consensus GT-AG rule, but slightly irregular as intronic +3 and 4 was CT (see below). It should be noted that there were only 5 bp between exons 16 and 17, but the downstream boundary of exon 16 did not provide a splice donor site.

Genomic organization in rodents

Mouse and human *IRSp53* were well conserved despite of long evolutional history. When compared with the S-form cDNA, they were 96% identical over the 522 amino acids and 87% identical at the nucleotide level over the entire coding region. Although three isoforms (S, T and M) were identified in rodents (including mouse, rat and hamster), it has been argued whether the L-form existed in rodents (Alvarez et al. 2002). We determined mouse and rat genomic sequences covering the region downstream of exon 16 (AB105194 and AB105195, respectively). Rat has been frequently used in studies on IRSp53 as its brain, one of the main expression sites, is larger than mouse. Recently, a mouse genomic sequence covering the entire coding region of *IRSp53* was deposited in the public database (NT_039521). Comparing the cDNA (AF390178) and genomic sequences, the mouse *IRSp53* gene spanned about 64 kb, and consisted of 16 exons (Fig. 1). The

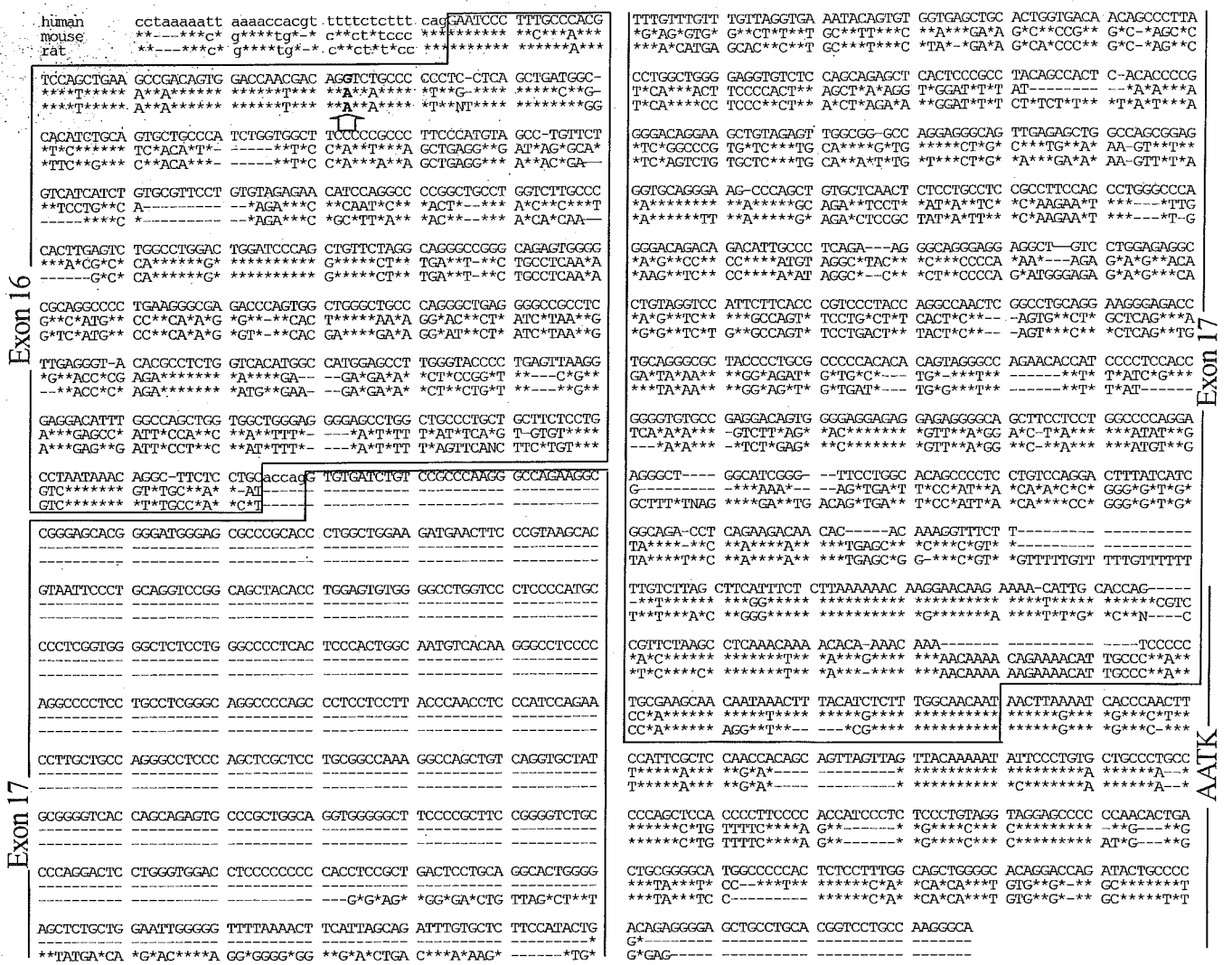

Fig. 4 Alignment of human, mouse and rat nucleotide sequences covering exons 16 and 17 of *IRSp53*. *Upper and lower cases* indicate exon and intron sequences, respectively. The identical nucleotides and gaps are indicate with an *asterisk* and a *dash*, respectively. The nucleotide change affecting the generation of the L-form is indicated by an *arrow*, and is also shown in Fig. 5. The accession numbers of the sequences are AC115099 for human, NT_039521 for mouse and AC105195 for rat. The regions for exon 16 and 17 for *IRSp53* are *boxed*. As the 3′ terminus of the *AATK* gene is unknown, it is not boxed but just indicated

genomic organization was also well conserved between human and mouse, although the size of several intron sequences varied. The exon-intron boundaries were similar (Fig. 2), and the generation mechanism of the M, S and T-forms was identical to human. However, there was a notable difference between human and rodent genomes, which affected the generation of the L-form. Mouse and rat sequences were shorter by about 400 bp in the region corresponding to human exon 17. In addition, there were many discordant nucleotides in the distal half of exon 16 and most parts of exon 17 when these were aligned (Fig. 4). The G nucleotide situated at the position corresponding to the splice donor site within exon 16 in human, was replaced with A in rodents (Fig. 4 *arrow*). Although several computer programs to predict splice sites (Splice view, http://l25.itba.mi.cnr.it/~webgene/wwwspliceview.html and Splice Site Prediction by Neural Network in Berkley *Drosophila* Genome Project http://www.fruitfly.org/seq_tools/splice.html) poorly recognized the splice donor site in human, the nucleotide substitution in rodents further decreased the possibility as the substitution abolished the GT consensus sequence at the boundary. Together with the finding of lack of the coding sequence specific to the L-form, we conclude that rodents do not generate the L-form.

We previously detected each protein isoform with specific antibodies recognizing the unique amino acid sequences at the C-terminus, where the L-form-specific antibody recognized some protein species in rat brain tissues (Okamura-Oho et al. 2001). Based on the study described here, the detected protein species were not derived from the IRSp53 L-form, although other results are still valid. The L-form-specific amino acid sequence may be coded for by another gene in rodents, as suggested (Alvarez 2002). The fifth isoform was recently

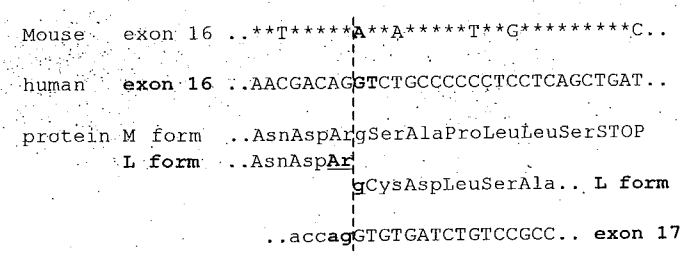

Fig. 5 Nucleotide change affecting in the generation of the L-form of *IRSp53*. *Upper and lower cases* indicate exon and intron sequences, respectively. The identical nucleotides between human and mouse are indicated with an *asterisk*. The amino acid sequences in the M-form, which reads through the indicated point, and the L-form, which is generated by splicing, are indicated *below* the nucleotide sequences

reported in rodents, which lacks 40 amino acids encoded by exon 9 (Alvarez et al. 2002). This is generated by the use of an additional splice acceptor site within exon 9 instead of the start position of exon 9 described in this report. A similar mechanism may be possible in human; however, no such transcripts detectable in human by RT-PCR were reported. The isoforms with and without the 40 amino acids may reflect the size difference in protein between 53 and 58 kDa. However, there is still confusion regarding the identity of protein species, some of which may due to different conditions for SDS-PAGE and as yet unknown posttranscriptional modifications. Therefore further studies will be required to identify protein species.

As regards evolution, it is interesting whether humans (or ancestor species) have gained the additional L-form or rodents have lost one of the isoforms. The former is plausible as each isoform of IRSp53 is involved in fine-tuning of its function, which may contribute to advancement of the central nerve system. This should be clarified by examination of other mammalian genomes, as well as by functional analyses of each isoform.

Finally, the downstream sequence (about 130 bp) of *IRSp53* was overlapped with the *AATK* gene, which encoded apoptosis-associated tyrosine kinase (Gaozza et al. 1997) (Fig. 4). The orientation of the transcripts was opposite (thus, encoded by the complementary strands) and the overlap occurred in their 3′-non-coding regions. This is highlighted by homology between human and rodent sequences in the vicinity of the end of exon 17, although the region was not used for *IRSp53* in rodents. The overlap is one of the examples for enriched gene distribution in a particular region of genome.

Acknowledgements We thank A. Asaka and Y. Ohtsuka for technical assistance, and K. Saito for preparing the manuscript. This study was supported in part by Grants for Human Genome, for Brain Science and for Pediatric Research from the Ministry of Health, Labor and Welfare, a Grant from the Organization for Pharmaceutical Safety and Research (Millennium Genome Project-5) and a Grant for Organized Research Combination System from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

Abbott MA, Wells DG, Fallon JR (1999) The insulin receptor tyrosine kinase substrate p58/53 and the insulin receptor are components of CNS synapses. J Neurosci 19:7300–7308

Alvarez CE, Sutcliffe JG, Thomas EA (2002) Novel isoform of insulin receptor substrate p53/p58 is generated by alternative splicing in the CRIB/SH3-binding region. J Biol Chem 227:24728–24734

Gaozza E, Baker SJ, Vora RK, Reddy EP (1997) AATYK: a novel tyrosine kinase induced during growth arrest and apoptosis of myeloid cells. Oncogene 15:3127–3135

Govind S, Kozma R, Monfries C, Lim L, Ahmed S (2001) Cdc42Hs facilitates cytoskeletal reorganization and neurite outgrowth by localizing the 58-kD insulin receptor substrate to filamentous actin. J Cell Biol 152:579–594

Hubbard SR, Till JH (2000) Protein tyrosine kinase structure and function. Annu Rev Biochem 69:373–398

Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T, Saito M, Tomoda A, Miike T, Naito H, Ikuta F, Tsuji S (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–13

Krugmann S, Jordens I, Gevaert K, Driessens M, Vandekerckhove J, Hall A (2001) Cdc42 induces filopodia by promoting the formation of an IRSp53:Mena complex. Curr Biol 11:1645–1655

Miki H, Yamaguchi H, Suetsugu S, Takenawa T (2000) IRSp53 is an essential intermediate between Rac and WAVE in the regulation of membrane ruffling. Nature 408:732–735

Miki H, Takenawa T (2002) WAVE2 serves a functional partner of IRSp53 by regulating its interaction with Rac. Biochem Biophys Res Commun 293:93–99

Modrek B, Lee C (2002) A genomic view of alternative splicing. Nat Genet 30:13–19

Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, Takeda T, Tadokoro K, Kondo I, Murayama N, Tanaka Y, Kikushima H, Umino K, Kurosawa H, Furukawa T, Nihei K, Inoue T, Sano A, Komure O, Takahashi M, Yoshizawa T, Kanazawa I, Yamada M (1994a) Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. Nat Genet 6:14–18

Nagafuchi S, Yanagisawa H, Ohsaki E, Shirayama T, Tadokoro K, Inoue T, Yamada M (1994b) Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). Nat Genet 8:177–182

Naito H, Oyanagi S (1982) Familial myoclonus epilepsy and choreoathetosis: Hereditary dentatorubral-pallidoluysian atrophy. Neurology 32:798–807

Oda K, Shiratsuchi T, Nishimori H, Inazawa J, Yoshikawa H, Taketani Y, Nakamura Y, Tokino T (1999) Identification of BAIAP2 (BAI-associated protein 2), a novel human homologue of hamster IRSp53, whose SH3 domain interacts with the cytoplasmic domain of BAI1. Cytogenet Cell Genet 84:75–82

Okamura-Oho Y, Miyashita T, Ohmi K, Yamada M (1999) Dentatorubral-pallidoluysian atrophy protein interacts through a proline-rich region near polyglutamine with the SH3 domain of an insulin receptor tyrosine kinase substrate. Hum Mol Genet 8:947–957

Okamura-Oho Y, Miyashita T, Yamada M (2001) Distinctive tissue distribution and phosphorylation of IRSp53 isoforms. Biochem Biophys Res Commun 289:957–960

Soltau M, Richter D, Kreienkamp HJ (2002) The insulin receptor substrate IRSp53 links postsynaptic shank1 to the small G-protein cdc42. Mol Cell Neurosci 21:575–583

Tadokoro K, Fujii H, Ohshima A, Kakizawa Y, Shimizu K, Sakai A, Sumiyoshi K, Inoue T, Hayashi Y, Yamada M (1992) Intragenic homozygous deletion of the *WT1* gene in Wilms' tumor. Oncogene 7:1215–1221

Yeh TC, Ogawa W, Danielsen AG, Roth RA (1996) Characterization and cloning of a 58/53-kDa substrate of the insulin receptor tyrosine kinase. J Biol Chem 271:2921–2928

J Hum Genet (2004) 49:97–101
DOI 10.1007/s10038-003-0117-0

ORIGINAL ARTICLE

Kazuaki Nagao · Katsunori Fujii · Masao Yamada
Toshiyuki Miyashita

Identification of a novel polymorphism involving a CGG repeat in the *PTCH* gene and a genome-wide screening of CGG-containing genes

Received: 2 September 2003 / Accepted: 20 November 2003 / Published online: 21 January 2004

Abstract Mutations in the human homologue of the *Drosophila patched* gene (*PTCH*) are responsible for the hereditary disorder called nevoid basal cell carcinoma syndrome (NBCCS). *PTCH* has a CGG triplet repeat located 4 bp upstream of the first methionine codon. Here we report a novel polymorphism involving the number of the CGG-repeat. The major allele (86.3%) contained a repeat size of seven, whereas the minor allele contained eight. No significant difference in the distributions of genotypes was observed between normal and NBCCS individuals. However, when the repeat was inserted between a heterologous promoter and the luciferase gene, the longer repeats tended to induce higher luciferase activities, suggesting that the repeat length potentially affects the levels of gene expression. A genome-wide screening revealed that 68 and 146 genes contained a CGG/CCG repeat in the coding region and in the 5'-untranslated region (5'-UTR), respectively. None of the genes had this repeat in 3'-UTR. Interestingly, the number of genes with a CGG repeat in the 5'-UTR was significantly higher than that with a CCG repeat in the 5'-UTR. The localization of a CGG/CCG repeat in *PTCH* is quite unique in that only four other genes have been found in which the repeat is localized up to 4 bp upstream of the first methionine.

Keywords PTCH · Nevoid basal cell carcinoma syndrome · Gorlin syndrome · Polymorphism · Triplet repeat

K. Nagao · M. Yamada · T. Miyashita (✉)
Department of Genetics,
National Research Institute for Child Health and Development,
3-35-31 Taishido, Setagaya-ku,
Tokyo 154-8567, Japan
E-mail: tmiyashita@nch.go.jp
Tel.: +81-3-34140181
Fax: +81-3-34143208

K. Fujii
Department of Pediatrics,
Graduate School of Medicine,
Chiba University, Chiba, Japan

Introduction

Mutations in the *PTCH* gene are responsible for the hereditary disorder called nevoid basal cell carcinoma syndrome (NBCCS; MIM# 109400) (Hahn et al., 1996; Johnson et al., 1996). NBCCS, also called Gorlin syndrome, is an autosomal dominant neurocutaneous disorder characterized by developmental abnormalities and tumorigenesis such as palmar and plantar pits, jaw cysts, calcification of the falx cerebri, skeletal anomalies, basal cell carcinoma, ovarian fibroma, and medulloblastoma (Gorlin, 1987). *PTCH* (MIM # 601309) is a human homologue of the *Drosophila* segment polarity gene *patched*. It has been mapped to 9q22.3-q31 and consists of 23 exons encoding a protein with 1,447 amino acid residues. The PTCH protein is a receptor for a secreted molecule Sonic hedgehog and has twelve transmembrane domains. At least two forms of PTCH protein are known to exist, reflecting the use of alternative exon 1a versus 1b (Hahn et al., 1996; Wicking et al. 1997a). Mutations in exon 1b have not been investigated so far due to, at least in part, the extreme GC-rich sequence (Wicking et al. 1997a; Fujii et al. 2003a). In the course of analyzing mutations in exon 1b, using a new set of primers and a PCR condition, we discovered a novel polymorphism involving a CGG trinucleotide repeat immediately upstream of the first in-frame methionine codon. We compared allele frequencies between healthy individuals and NBCCS patients. We also investigated the effect of the repeat length on the gene expression using a heterologous reporter gene. In addition, the results of a genome-wide screening of CGG/CCG-containing genes are demonstrated.

Materials and methods

DNA samples

After informed consent was obtained from 51 healthy, unrelated individuals and 14 patients with NBCCS, total genomic DNAs were isolated from peripheral leukocytes by the standard phenol/

chloroform extraction method. Patients were diagnosed as having NBCCS according to the clinical criteria (Kimonis et al. 1997). All studies were approved by the local ethnic committee. Among 14 patients with NBCCS, *PTCH* mutations were found in 11. Some of the mutations have already been reported (Fujii et al. 1999, Fujii et al. 2003a, Fujii et al. 2003b) and some will be reported elsewhere.

Polymerase chain reaction and sequencing

The genomic region of *PTCH* including the 5'-untranslated region (5'-UTR) and exon 1b was amplified by using the forward primer, 5'-CGCGCAATGTGGCAATGGAA-3', and the reverse primer, 5'-AGAGGAGGGAAGAGAAAGTG-3'. The polymerase chain reaction (PCR) was carried out in a 20 μl reaction volume by using LA Taq with GC Buffer (TaKaRa) according to the manufacturer's instruction. PCR was run for 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 3 min on a Program Temp Control System PC-800 (ASTEC, Fukuoka, Japan). Both the sense and antisense strands of the PCR products were directly sequenced by using the same primers as described above. PCR products purified by a QIAquick PCR Purification Kit (QIAGEN) were used as the template DNA for cycle sequencing with a CEQ DTCS Quick-Start Kit (Beckman Coulter). Sequencing analysis was performed on a CEQ 8000 Genetic Analysis System (Beckman Coulter) according to the manufacturer's instructions.

Plasmid construction

Luciferase constructs containing the sequence of *PTCH* 5'-UTR were generated by a PCR-mediated method described previously (Imai et al. 1991) using pGV-P2 (Wako Chemicals, Osaka, Japan) as a template. The authenticity of all constructs was confirmed by sequencing.

Luciferase assay

The human embryonic kidney cell line 293 growing on six-well culture plates were cotransfected using Effectene reagent (QIAGEN) with 0.5 μg of luciferase plasmid and 0.5 μg of pCMVβGal. The cells were harvested at 24 h after the transfection and used for a luciferase assay. Luciferase activities were measured as described previously and normalized for transfection efficiency based on β-galactosidase activities (Shikama et al. 2001).

Real-time quantitative RT-PCR

Total RNA was extracted from the transfected cells described above using TRIzol reagent (Invitrogen). One-step RT-PCR was performed with a 7700 ABI PRISM Sequence Detector System (Perkin Elmer-Applied Biosystems) using primers 5'-TCTGGATCTACTGGTCTGCCTAA-3' and 5'-GCGCACTTTGAATCTTGTAATCCTG-3'. To normalize the expression of luciferase, the *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) housekeeping gene was also amplified, using primers 5'-GAAGGTGAAGGTCGGAGT-3' and 5'-GAAGATGGTGATGGGATTTC-3'. Fluorogenic probes 5'-CAAATCATTCCGGATACTGC-3' and 5'-CAAGCTTCCCGTTCTCAGCC-3' carrying 5' 6-carboxy-fluorescein as a reporter dye and 3' 6-carboxy-tetramethyl-rhodamine as a quencher dye were used to detect the PCR product of luciferase and *GAPDH*, respectively. In every experiment, *GAPDH* was amplified using a series of dilutions of a known amount of the standard RNA supplied by Perkin Elmer to prepare a standard curve.

Computational screen

Human mRNA sequences that contain more than seven repeats of CGG were downloaded from NCBI nucleotide databases using the search program termed as "Search for short, nearly exact sequences" with $(CGG)_7$ as a query (http://www.ncbi.nlm.nih.gov/BLAST/). A full screening of the genes of interest was confirmed because genes with exact matches (bit score 42) were followed by the genes with partial matches (bit score less than 42). Genes for unidentified coding sequences were excluded from further study.

Statistical analysis

Genotype distributions and allele frequencies of CGG repeat numbers were compared between cases and controls by means of the χ^2 test. Odds ratios (OR) and 95% confidence intervals (95% CI) were calculated by Wolf's method.

Results and discussion

The *PTCH* gene has two alternative first exons—exon 1a and exon 1b (Fig. 1A). Exon 1b contains the first in-frame methionine codon, while exon 1a is a noncoding exon. We noticed a CGG trinucleotide repeat located 4 bp upstream of the first methionine codon. Although exon 1b is a coding exon, mutations in this exon have not been reported. In the course of analyzing mutations in exon 1b using samples from NBCCS individuals that do not have a mutation elsewhere in *PTCH*, we discovered a novel polymorphism involving the CGG trinucleotide repeat (Fig. 1B). The major allele contained

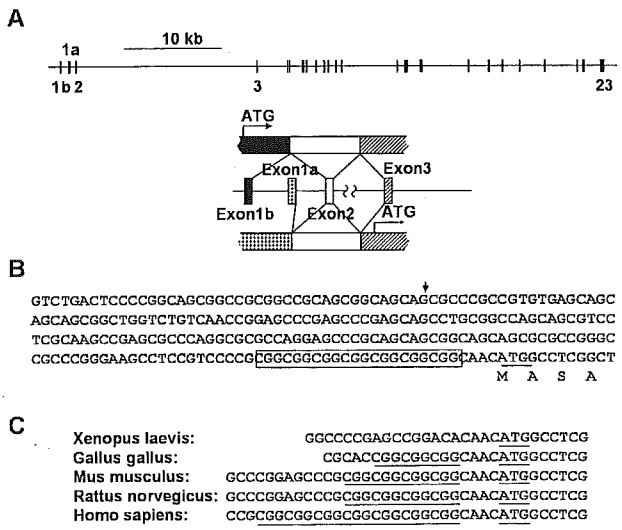

Fig. 1 **A** Genomic organization of human *PTCH*. The *PTCH* locus based on the sequence AL161729 is shown at the *top*. Two cDNA sequences, GenBank U43148 and U59464, are generated by alternative splicing using exon 1a and 1b, respectively, as schematically depicted at the *bottom*. **B** Nucleotide sequence of the human *PTCH* gene including 5'-UTR and exon 1b. The first methionine codon is *underlined*. Polymorphic CGG repeat is *boxed*. The putative transcription start site is indicated by an *arrowhead*. **C** Nucleotide sequence alignment of the *PTCH* genes. CGG repeat and the first methionine codon are underlined

seven repeats of CGG, while the minor one contained eight (Table 1). As far as we examined, we did not find repeat numbers other than seven or eight. The repeat is conserved among vertebrates, since chicken, mouse, and rat *PTCH* contain four or five repeats of CGG (Fig. 1C). However, the repeat has not been found in Xenopus *PTCH*, indicating it is not conserved in amphibians.

Abnormal expansion of the CGG triplet repeat in the 5'-UTR of the *fragile X mental retardation-1* (*FMR1*) gene is responsible for fragile X syndrome, in which the repeat is abnormally hypermethylated, resulting in the silence of the *FMR1* (reviewed by Jin et al. 2000). Since CGG repeat in *PTCH* is immediately upstream of the first in-frame methionine codon, the repeat number may influence the efficiency of translation as well as of transcription. To address this issue, various lengths of $(CGG)_nCAAC$ were subcloned into the luciferase plasmid pGV-P2 between the SV40 promoter and the coding sequence for luciferase (Fig. 2A), and luciferase assays were performed. Luciferase activities gradually increased with the number of CGG repeats, at least within the range we examined. The highest level of luciferase activity was obtained when cells were transfected with the plasmid pGV-$(CGG)_{19}CG(CGG)_6$, which was generated by chance during PCR reaction (Fig. 2B). These results suggest that individuals with $(CGG)_8/(CGG)_8$ have higher levels of PTCH protein expression than those with $(CGG)_7/(CGG)_7$. This is contradictory to the case of *FMR1*. However, it should be noted that in fragile X syndrome, the repeat is massively expanded over 230, and the repeat is located more than 50 bp upstream of the first methionine codon.

To address the question of whether the difference in luciferase activity is transcriptional or translational, the levels of luciferase RNA expression were quantified by a real-time RT-PCR. As shown in Fig. 2C, in contrast to the activities of luciferase, no significant difference in luciferase transcription was observed. Moreover, unexpectedly, the cells transfected with the plasmid pGV-$(CGG)_{19}CG(CGG)_6$ expressed significantly lower levels of luciferase RNA. Therefore, the increase in luciferase activities with the expansion of the CGG repeat is due to the increased efficiency of translation.

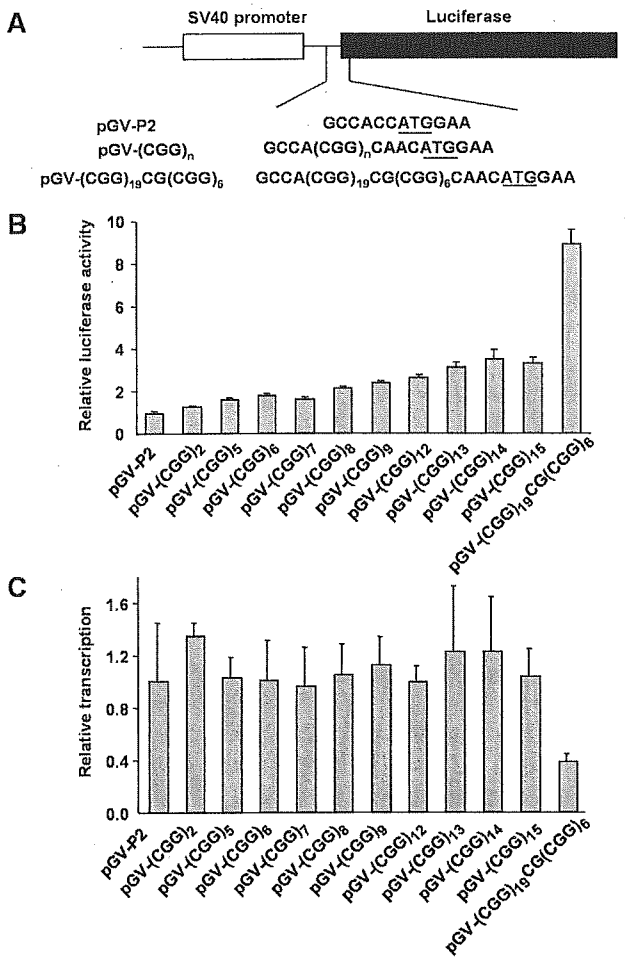

Fig. 2 **A** Schematic depiction of reporter gene constructs used for a luciferase assay. Nucleotide sequences inserted between SV40 promoter and the luciferase gene are indicated at the *bottom*. The first methionine codon of the luciferase gene is *underlined*. **B** The effect of the repeat length on luciferase activities; 293 cells transfected with plasmids indicated at the *bottom* were harvested 24 h after the transfection and subjected to a luciferase assay. **C** The effect of the repeat length on luciferase transcriptions. Total RNA was extracted from 293 cells transfected with plasmids indicated at the *bottom* and subjected to a real-time RT-PCR. Luciferase transcriptions were normalized by those of *GAPDH*

Table 1 Genotype data of a $(CGG)_n$ on the *PTCH* gene

	Controls	NBCCS
Major allele (repeat number: 7) [%]	88 [86.3]	25 [89.3]
Minor allele (repeat number: 8) [%]	14 [13.7]	3 [10.7]
Total	102 [100.00]	28 [100.00]
Major homozygous [%]	39 [76.5]	12 [85.7]
Heterozygous [%]	10 [19.6]	1 [7.1]
Minor homozygous [%]	2 [3.9]	1 [7.1]
Total	51 [100.0]	14 [100.0]
χ^2 [*P*]		
Genotype frequency (2×3 table)	1.38 [0.50]	
Allele frequency (major versus minor)	0.18 [0.68]	
Major homozygous versus others	0.56 [0.46]	
Minor homozygous versus others	0.26 [0.61]	
Odds ratio [95% CI]		
Major homozygous versus others	1.85 [0.36–9.43]	
Minor homozygous versus others	1.88 [0.16–22.44]	
Major allele versus minor allele	0.75 [0.20–2.83]	

Table 2 CGG/CCG-containing genes. *UTR* untranslated region

	$(CGG)_n$	$(CCG)_n$	Total
CDS	39	29	68
5' UTR	95[a]	51[a]	146
Total	134	80	214

Genes with a repeat of seven or more CGG/CCG were downloaded from NCBI nucleotide databases:
[a]The number of CGG-containing genes are significantly higher than CCG-containing genes (χ^2 test, $P=0.00027$)

Table 3 $(CGG)_n/(CCG)_n$-containing genes in which the triplet repeat is located immediately upstream of the first in-frame methionine codon

Accession no.[a]	Gene[b]	Nucleotides between $(CGG)_n/(CCG)_n$ and ATG	CGG or CCG	n
BC029158	Clone MGC:34313 IMAGE:5198758	CCCCCGGGGC	CCG	10
NM_025075	Hypothetical protein FLJ23445	CGCACGCC	CCG	8
BC015930	Clone MGC:8881 IMAGE:3920963	CGGGGGCC	CCG	8
NM_013396	Ubiquitin specific protease 25 (USP25)	CGGGGGCC	CCG	8
NM_013272	Solute carrier family 21 (organic anion transporter), member 11 (SLC21A11)	GGGAAGG	CGG	8
NM_005360	Transcription factor C-MAF (c-maf)	CAGGAGA	CGG	7
NM_004699	DNA segment on chromosome X (unique) 9928 expressed sequence (DXS9928E) (XAP-5)	CTGCC	CCG	9
NM_000264	PTCH	CAAC	CGG	7
NM_145296	TSLC1-like 2 (TSLL2)	CACC	CGG	7
NM_002958	RYK receptor-like tyrosine kinase (RYK)	CCC	CGG	7
NM_017811	Ubiquitin-conjugating enzyme E2R 2 (UBE2R2)	CG	CCG	7
NM_173054	Reelin (RELN)	C	CGG	8–10[c]

[a]One representative accession number for one gene
[b]Genes that have less than ten nucleotides between triplet repeat and the first methionine
[c]Repeat number varies depending on deposited sequences

The distributions of genotypes that we observed in NBCCS patients and controls did not differ from the expected frequencies under the assumption of Hardy-Weinberg equilibrium (data not shown), nor were significant associations with NBCCS observed (Table 1). Thus far, no genotype-phenotype correlation between the position of mutations and major clinical features of NBCCS is evident (Wicking et al. 1997b). Since developmental defects associated with the disorder are most likely due to haploinsufficiency, and the repeat length potentially alter the expression levels of PTCH, the repeat number may have an effect on the severity of the disease. It would also be interesting to examine the association of the repeat number with sporadic or non-inherited basal cell carcinoma or medulloblastoma, since PTCH acts as a tumor suppressor in these tumors (reviewed by Hunter 1997).

In order to find other genes with CGG repeats, we next performed a genome-wide screening of CGG/CCG-containing genes from NCBI nucleotide databases. A total of 214 genes having seven or more of the repeat number were downloaded. A complete list of the CGG/CCG-containing genes can be obtained from our Web site, http://genetics.nch.go.jp/supplements.htm. Of those 214 genes, 146 (68.2%) contained the repeat in the 5'-UTR (Table 2). Interestingly, significantly more genes have CGG repeats than CCG repeats (65.1% versus 34.9%, $P=0.00027$). More significantly, none of the downloaded genes contained repeats in the 3'-UTR. The genes containing CGG/CCG repeats in close proximity to their first methionine codons are listed in Table 3. Only five genes including *PTCH* have intervening sequences of up to 4 bp between $(CGG)_n/(CCG)_n$ and ATG. In this regard, *PTCH* is quite unique in terms of the location of the repeat. Considering our results, polymorphisms of the repeat number that might exist in these genes potentially affect their expression levels.

Acknowledgements We would like to thank the patients and family members as well as the clinicians who contributed to this study. We gratefully acknowledge assistance from Mami U, Kaori Inoue, and Kayoko Saito. This study was supported in part by grants for Brain Research, Pediatric Research, Cancer Research, and Genome Research from the Ministry of Health and Welfare, and a Grant-in-Aid for Scientific Research and the Budget for Nuclear Research from the Ministry of Education, Culture, Sports, Science, and Technology.

References

Fujii K, Miyashita T, Takanashi J, Sugita K, Kohno Y, Nishie H, Yasumoto S, Furue M, Yamada M (1999) γ-irradiation deregulates cell cycle control and apoptosis in nevoid basal cell carcinoma syndrome-derived cells. Jpn J Cancer Res 90:1351–1357

Fujii K, Kohno Y, Sugita K, Nakamura M, Moroi Y, Urabe K, Furue M, Yamada M, Miyashita T (2003a). Mutations in the human homologue of *Drosophila patched* in Japanese nevoid basal cell carcinoma syndrome patients. Hum Mutat 21:451–452

Fujii K, Miyashita T, Omata T, Kobayashi K, Takanashi J, Kouchi K, Yamada M, Kohno Y (2003b) Gorlin syndrome with ulcerative colitis in a Japanese girl. Am J Med Genet 121:65–68

Gorlin RJ (1987) Nevoid basal-cell carcinoma syndrome. Medicine 66:98–113

Hahn H, Wicking C, Zaphiropoulous PG, Gailani MR, Shanley S, Chidambaram A, Vorechovsky I, Holmberg E, Unden AB, Gillies S, Negus K, Smyth I, Pressman C, Leffell DJ, Gerrard B, Goldstein AM, Dean M, Toftgard R, Chenevix-Trench G, Wainwright B, Bale AE (1996) Mutations of the human homolog of Drosophila *patched* in the nevoid basal cell carcinoma syndrome. Cell 85:841–851

Hunter T (1997) Oncoprotein networks. Cell 88, 333–346

Imai Y, Matsushima Y, Sugimura T, Terada M (1991) A simple and rapid method for generating a deletion by PCR. Nucleic Acids Res 19:2785

Jin P, Warren ST (2000) Understanding the molecular basis of fragile X syndrome. Hum Mol Genet 9:901–908

Johnson RL, Rothman AL, Xie J, Goodrich LV, Bare JW, Bonifas JM, Quinn AG, Myers RM, Cox DR, Epstein EH, Jr, Scott MP (1996). Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. Science 272:1668–1671

Kimonis VE, Goldstein AM, Pastakia B, Yang ML, Kase R, DiGiovanna JJ, Bale AE, Bale SJ (1997) Clinical manifestations in 105 persons with nevoid basal cell carcinoma syndrome. Am J Med Genet 69:299–308

Shikama Y, U M, Miyashita T, Yamada M (2001) Comprehensive studies on subcellular localizations and cell death-Inducing activities of eight GFP-tagged apoptosis-related caspases. Exp Cell Res 264:315–325

Wicking C, Gillies S, Smyth I, Shanley S, Fowles L, Ratcliffe J, Wainwright B, Chenevix-Trench G (1997a) De novo mutations of the *patched* gene in nevoid basal cell carcinoma syndrome help to define the clinical phenotype. Am J Med Genet 73:304–307

Wicking C, Shanley S, Smyth I, Gillies S, Negus K, Graham S, Suthers G, Haites N, Edwards M, Wainwright B, Chenevix-Trench G (1997b). Most germ-line mutations in the nevoid basal cell carcinoma syndrome lead to a premature termination of the PATCHED protein, and no genotype-phenotype correlations are evident. Am J Hum Genet 60:21–26

electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification and preliminary crystallographic analysis of human sorbitol dehydrogenase

Connie Darmanin, Takeshi Iwata, Deborah A. Carper, Lindsay G. Sparrow, Roland P.-T. Chung and Ossama El-Kabbani

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification and preliminary crystallographic analysis of human sorbitol dehydrogenase

Connie Darmanin,[a] Takeshi Iwata,[b] Deborah A. Carper,[c] Lindsay G. Sparrow,[d] Roland P.-T. Chung[a] and Ossama El-Kabbani[a]*

[a]Department of Medicinal Chemistry, Victorian College of Pharmacy, Monash University (Parkville Campus), Parkville, Victoria 3052, Australia, [b]National Tokyo Medical Centre, 2-5-1 Higashigaoka, Meguro, Tokyo 152-8902, Japan, [c]National Eye Institute, NIH, Bethesda MD 20892, USA, and [d]CSIRO, Health Sciences and Nutrition, Parkville, Victoria 3052, Australia

Correspondence e-mail: ossama.el-kabbani@vcp.monash.edu.au

Received 9 October 2002
Accepted 2 January 2003

Human sorbitol dehydrogenase (SDH) was expressed in *Escherichia coli* BL21 cells and purified using ammonium sulfate precipitation and anion-exchange and dye-affinity chromatography. Purified SDH was crystallized from polyethylene glycol solutions using the hanging-drop vapour-diffusion method. X-ray data were collected to 2.75 Å resolution. The crystals belong to the monoclinic $C2$ space group, with unit-cell parameters a = 145.9, b = 52.3, c = 169.0 Å, β = 101.8°. This is the first crystallization report of human sorbitol dehydrogenase.

1. Introduction

Sorbitol dehydrogenase (SDH), a member of the medium-chain dehydrogenase/reductase protein family, is the second enzyme of the polyol pathway of glucose metabolism. The polyol pathway is comprised of aldose reductase, which converts glucose to sorbitol, and sorbitol dehydrogenase, which converts sorbitol to fructose strictly using NAD^+ as coenzyme. It is believed that the increased flux of glucose through the polyol pathway during hyperglycaemia contributes to the development of diabetic complications (Kinoshita & Nishimura, 1998; The Diabetes Control and Complication Trials Group, 1993). SDH is a tetramer with a catalytic Zn atom bound in the active site (Jeffery *et al.*, 1984). Human liver SDH has been purified previously and was shown to have a molecular weight of 155 kDa, with subunits of approximately 37 kDa (Maret & Auld, 1988).

Homology-modelling studies have suggested that SDH is structurally homologous to mammalian alcohol dehydrogenases (ADH) with respect to a conserved zinc-binding motif and a hydrophobic substrate-binding pocket (Jeffery *et al.*, 1981; Eklund *et al.*, 1985). Additionally, a model of human SDH in complex with coenzyme and inhibitor, based on the structure of human β_3 alcohol dehydrogenase (26% identity; Hurley *et al.*, 1991), has been published (Darmanin & El-Kabbani, 2001). More recently, the crystal structure of rat SDH has been solved (Johansson *et al.*, 2001) and was found to be similar to the mammalian ADH structure, containing two distinct catalytic and coenzyme-binding domains. However, the zinc coordination in the active site of rat SDH was found to be different from that of mammalian alcohol dehydrogenase. In rat SDH, the coordinating ligands are His69, Cys44, Glu155 and Glu70. On the other hand, the coordinating ligands in alcohol dehydrogenase are His69, Cys44 and Cys155 (using the residue numbering for rat SDH). Glu70 in alcohol dehydrogenase is in close proximity to the Zn atom but is not a ligand (Johansson *et al.*, 2001). Interestingly, the zinc coordination found in rat SDH is similar to that found in the bacterial tetrameric NADP(H)-dependent alcohol dehydrogenase of *Clostridium beijerinckii* (Korkhin *et al.*, 1998).

While the amino-acid sequences of rat and human SDH are similar (82% sequence identity), biochemical and modelling studies have suggested that non-conserved residues may be involved in the binding of both substrate and inhibitor (Darmanin & El-Kabbani, 2001; Höög *et al.*, 1993). These include a substitution at residue 274, which was shown by modelling to bind to both substrate and inhibitor, and substitutions on residues immediately adjacent to those involved in ligand binding (residues 203, 208, 272 and 344). These residues may be responsible for the differences in substrate and inhibitor specificities between the two enzymes. Höög *et al.* (1993) compared the kinetic constants of recombinant rat SDH and human liver SDH with sorbitol, fructose, 2,3-butanediol, ribitol, L-threitol and xylitol as substrates. They found that the enzymes have similar K_M values for most of these substrates, but the k_{cat} values were about tenfold lower for the rat enzyme.

Sorbitol dehydrogenase is expressed almost ubiquitously in all mammalian tissues, including the brain, lens, erythrocytes and liver (O'Brien *et al.*, 1983; Jedziniak *et al.*, 1981; Barretto *et al.*, 1985). The human enzyme has attracted considerable interest owing to its implication in the development of diabetic complications such as cataracts, neuropathy, retinopathy and nephropathy (Obrosova *et al.*, 1999). The determination of the crystal struc-

ture of human SDH will help to elucidate the catalytic mechanism and model the interactions of the enzyme with substrate and inhibitor. In this study, the purification of human sorbitol dehydrogenase, obtained from the expression of the SORD gene, is reported. The first report of the crystallization and preliminary crystallographic analysis of human sorbitol dehydrogenase is also presented. The structure of the human enzyme may facilitate the development of drugs for the treatment of diabetic complications.

2. Experimental

2.1. Expression and purification

The coding region of human sorbitol dehydrogenase (SORD) cDNA previously isolated from liver cDNA library (Iwata *et al.*, 1995) was inserted into prokaryotic expression vector [pET23(+) Novagen, Madison, WI, USA] and transformed in *Escherichia coli* BL21 (DE3) (Novagen). The SORD expression was triggered by an addition of 1 m*M* IPTG and incubated at 310 K for 4 h. The bacteria were collected by centrifugation and the pellet was resuspended in 10 m*M* sodium phosphate buffer pH 7.4 containing 5 m*M* β-mercaptoethanol and sonicated to release the protein into the supernatant. The supernatant containing the SDH was collected by centrifugation. Ammonium sulfate fractionation of the supernatant was carried out at 30, 50 and 70% saturation, with SDH precipitating out at both the 30 and 50% saturation. These pellets were resuspended in 10 m*M* sodium phosphate buffer pH 7.4 and dialyzed in the same buffer to remove any salt present in solution. The dialyzed samples were loaded onto a Q-Sepharose column (Amersham-Pharmacia) and eluted with a stepwise sodium chloride gradient. The fractions that eluted out at 0.1 *M* salt concentration were collected and concentrated using an Omega 10 kDa membrane (Pall) and dialyzed in 20 m*M* sodium phosphate buffer pH 6.5 containing 2 m*M* magnesium chloride. The dialyzed sample was loaded onto a Green 18 dye-affinity column and the bound SDH was again eluted with a salt gradient. Peak samples were assayed for activity and the purity of SDH was checked by 12.5% SDS–PAGE. A single band corresponding to approximately 37 kDa was obtained from the 0.3 *M* salt peak (Fig. 1). The purification protocol is shown in Table 1. The fractions that showed co-purification of the 37 kDa band and SDH activity were pooled. The combined fraction was then subjected to three concentration–dilution cycles at 3200*g* with milli-Q water, using a 10 kDa Ultrafree-4 centrifugal unit (Millipore) in a Megafuge 1.0 R centrifuge (Heraeus) to reduce the salt concentration in the sample. The enzyme was finally concentrated to 13 mg ml^{-1} for crystallization.

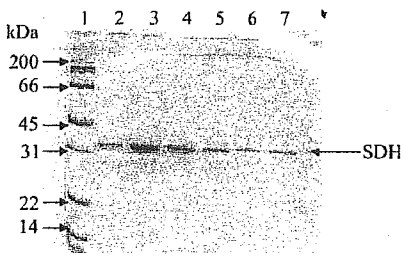

Figure 1
SDS–PAGE showing purified fractions of human SDH. Lane 1 shows standard molecular-weight markers and lanes 2–7 show purified enzyme from consecutive collection fractions.

Table 1
Purification of human sorbitol dehydrogenase.

Sample	Activity (mU ml^{-1})	Protein concentration (mg ml^{-1})	Specific activity (mU mg^{-1})	Overall purification†
Crude extract	ND‡	336.84	ND‡	–
Ammonium sulfate fraction	1.93×10^{-2}	57.19	3.34×10^{-4}	–
Q-column	3.21×10^{-3}	2.38	1.35×10^{-3}	Fourfold
Green column	1.61×10^{-3}	0.211	7.62×10^{-3}	23-fold

† Based on the specific activity of each subsequent step compared with that of the ammonium sulfate fraction (since the crude extract data was not available). ‡ ND, not determined: the initial reaction was too rapid to be followed by the conventional UV–Vis technique.

2.2. Enzyme assays and protein measurements

Sorbitol dehydrogenase activity was determined on a Shimadzu UV–Vis spectrophotometer (model UV160A) by following the increase in absorbance of NADH at 340 nm (ε = 6220 M^{-1} cm^{-1}). Each 1 ml assay sample contained 42 m*M* glycine buffer pH 9.9, 9.9 m*M* D-sorbitol and 0.5 m*M* β-NAD$^+$. The reaction commenced on the addition of the enzyme. One milliunit (mU) of activity is defined as the amount of enzyme needed to oxidize one millimole of substrate per minute under initial velocity conditions at room temperature (293 K).

Protein concentrations were routinely determined by using Coomassie Blue dye according to the method of Bradford (1976) or by measurement of the absorbance at 280 nm (ε_{280nm} at 1 mg ml^{-1} = 0.57; Maret & Auld, 1988).

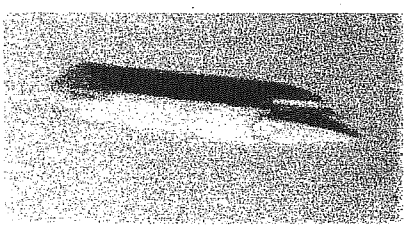

Figure 2
A crystal of human sorbitol dehydrogenase. The crystal in this photograph has dimensions of 0.3 × 0.06 × 0.06 mm.

2.3. Crystallization and X-ray data collection

Crystals of human sorbitol dehydrogenase were grown at 295 K by the vapour-diffusion method (McPherson, 1985). The enzyme and cofactor NAD$^+$, at a respective molar ratio of 1:6, were incubated at 277 K for 2 h. Each hanging drop consisted of 2.5 µl SDH holoenzyme, 2 µl well buffer (0.1 *M* Tris pH 8.6, 0.2 *M* sodium acetate and 10% PEG 3350) and 0.5 µl 30%(*v*/*v*) MPD (2-methyl-2,4-pentanediol). Crystals grew within one week to maximum dimensions of 0.3 × 0.06 × 0.06 mm (Fig. 2). The crystals were picked up with a nylon loop and flash-cooled at 100 K in a stream of gaseous nitrogen. A diffraction data set from one flash-cooled crystal was recorded at 100 K on a MAR345 image plate mounted on a Rigaku RU-300 rotating-anode X-ray generator operated at 50 kV and 90 mA. Each frame was recorded with a 1800 s exposure and 0.5° oscillation around φ. The crystal-to-detector distance was set to 250 mm so that the spots were well resolved. The data was processed and scaled using the *HKL* software package (Otwinowski & Minor, 1997).

3. Results

A near-complete set of data was collected from a single crystal to a resolution of 2.75 Å (Table 2 shows the data-collection statistics). Sorbitol dehydrogenase crystallized in the monoclinic *C*2 space group, with unit-cell parameters a = 145.9, b = 52.3, c = 169.0 Å, β = 101.8°. The Matthews coefficient was