may not exhibit the full range of phenotypic anomalies expected in a DNA-PK_{cs} null mutation.

The ability of DNA-PK to phosphorylate substrates implicated in the regulation of the cell cycle[5], and the sequence homology of the catalytic subunit of DNA-PK to the phosphatidylinositol-3-OH kinase family and the ataxia telangiectasia gene product[28], support a role for DNA-PK in signal transduction and cell-cycle control. The fact that DNA damage in *Ku80*$^{-/-}$ MEFs induces delays in cell-cycle progression suggest that Ku80 is not required for recognizing DNA lesions, nor for transmitting signals that couple DNA damage to cell-cycle arrest. Consistent with the preservation of cell-cycle checkpoints, Ku80-deficient mice are not predisposed to early tumorigenesis (our unpublished observations). On the other hand, the presence of unrepaired DNA breaks that arise during normal cellular metabolism or during VDJ recombination might prevent *Ku80*$^{-/-}$ cells from resuming the cell cycle. Thus, arrested lymphocyte development and proportional dwarfism in *Ku80*$^{-/-}$ mice may be the result of abnormal DNA metabolism and a subsequent loss of cells with damaged or incomplete genomes. In conclusion, our findings provide new insights into the physiological function of Ku80, and a model to study DNA double-strand-break repair *in vivo*. □

Received 8 May; accepted 18 July 1996.

1. Jeggo, P. A., Taccioli, G. A. & Jackson, S. P. *BioEssays* **17,** 949–956 (1995).
2. Weaver, D. T. *Trends Genet.* **11,** 388–392 (1995).
3. Getts, R. C. & Stamato, T. D. *J. Biol. Chem.* **269,** 15981–15984 (1994).
4. Lees-Miller, S. P. *et al. Science* **267,** 1183–1185 (1995).
5. Anderson, C. W. *Trends Biochem. Sci.* **18,** 433–437 (1993).
6. Giffin, W. *et al. Nature* **380,** 265–268 (1996).
7. Kuhn, A., Gottlieb, T. M., Jackson, S. P. & Grummt, I. *Genes Dev.* **9,** 193–203 (1995).
8. Li, G. C. *et al. Proc. Natl Acad. Sci. USA* **92,** 4512–4516 (1995).
9. Smider, V., Rathmell, W. K., Lieber, M. R. & Chu, G. *Science* **266,** 288–291 (1994).
10. Taccioli, G. E. *et al. Science* **265,** 1442–1445 (1994).
11. Kirchgessner, C. U. *et al. Science* **267,** 1178–1183 (1995).
12. Blunt, T. *et al. Cell* **80,** 813–823 (1995).
13. Errami, A. *et al. Mol. Cell. Biol.* **16,** 1519–1526 (1996).
14. Chen, F., Peterson, S. R., Story, M. D. & Chen, D. J. *Mut. Res.* **362,** 9–19 (1996).
15. Jeggo, P. A. & Kemp, L. M. *Mut. Res.* **112,** 313–327 (1983).
16. Boubnov, N. V. *et al. Proc. Natl Acad. Sci. USA* **92,** 890–894 (1995).
17. Weirich-Schwaiger, H., Weirich, H. G., Gruber, B., Schweiger, M. & Hirsch-Kauffman, M. *Mut. Res.* **316,** 37–48 (1994).
18. Mombaerts, P. *et al. Cell* **68,** 869–877 (1992).
19. Shinkai, Y. *et al. Cell* **68,** 855–867 (1992).
20. Peterson, S. R. *et al. Proc. Natl Acad. Sci. USA* **92,** 3171–3174 (1995).
21. Costa, T. E. F., Suh, H. & Nussenzweig, M. C. *Proc. Natl Acad. Sci. USA* **89,** 2205–2208 (1992).
22. Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M. & Peng, A. *Genes Dev.* **7,** 2520–2532 (1993).
23. Carroll, A. M. & Bosma, M. J. *Genes Dev.* **5,** 1357–1366 (1991).
24. Carroll, A. M., Slack, J. K. & Chang, W.-T. *Mol. Cell. Biol.* **13,** 3623–3640 (1993).
25. Porter, S. E., Greenwell, P. W., Ritchie, K. B. & Petcs, T. D. *Nucleic Acids Res.* **24,** 582–585 (1996).
26. Finnie, N. J., Gottlieb, T. M., Blunt, T., Jeggo, P. A. & Jackson, S. P. *Proc. Natl Acad. Sci. USA* **92,** 320–324 (1995).
27. Rathmell, W. K. & Chu, G. *Mol. Cell. Biol.* **14,** 4741–4748 (1994).
28. Hartley, K. O. *et al. Cell* **82,** 849–856 (1995).
29. Zhu, C. & Roth, D. B. *Immunity* **2,** 101–112 (1995).

ACKNOWLEDGEMENTS. We thank E. L. Kuff for mouse Ku80 cDNA; A. Nagy for R1 ES cells; D. Kim and L. Wu for Ku antiserum; H. Ouyang for technical assistance; P. Burgman and L. Li for radiation experiments; T. Deloherey for FACS analysis; and S. Gong, H. Petrie D. Roth, A. Koff, K. Manova, T. de Lange and E. Besmer for suggestions. C.C., A.N. and G.C.L. are especially grateful to C. C. Ling and Z. Fuks for advice and support. M.C.N. is an associate investigator in the Howard Hughes Medical Institute. This work was supported in part by grants from the NIH, and by the HHMI.

CORRESPONDENCE and requests for materials should be addressed to G.C.L. (e-mail: g-Li@ski.mskcc.org).

Structural basis for inhibition of receptor protein-tyrosine phosphatase-α by dimerization

Alexandrine M. Bilwes*, Jeroen den Hertog†, Tony Hunter‡ & Joseph P. Noel*

* Structural Biology Laboratory and ‡ Molecular Biology and Virology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA
† The Netherlands Institute for Developmental Biology, Uppsalaan 8, 3584 CT Utrecht, The Netherlands

RECEPTOR-LIKE protein-tyrosine phosphatases (RPTPs), like their non-receptor counterparts, regulate the level of phosphotyrosine-containing proteins derived from the action of protein-tyrosine kinases[1]. RPTPs are type-I integral membrane proteins which contain one or two catalytic domains in their cytoplasmic region[2]. It is not known whether extracellular ligands regulate the activity of RPTPs. Here we describe the crystal structure of the membrane-proximal catalytic domain (D1) of a typical RPTP, murine RPTPα. Significant structural deviations from the PTP1B fold reside within the amino-terminal helix–turn–helix segment of RPTPαD1 (residues 214 to 242) and a distinctive two-stranded β-sheet formed between residues 211–213 and 458–461. The turn of the N-terminal segment inserts into the active site of a dyad-related D1 monomer. On the basis of two independent crystal structures, sequence alignments, and the reported biological activity of EGF receptor/CD45 chimaeras[3], we propose that dimerization and active-site blockage is a physiologically important mechanism for downregulating the catalytic activity of RPTPα and other RPTPs.

All RPTPs share a modular primary structure that includes widely varying extracellular domains likely to bind ligands[4], a single membrane-spanning segment, and one or two intracellular catalytic domains termed D1 and D2. The first catalytic domain of RPTPα, D1, was expressed as a glutathione-*S*-transferase (GST) fusion protein in *Escherichia coli*[5], purified and cleaved with trypsin to generate a fragment containing residues 202–503. Crystals of this fragment were obtained in three different space groups ($C222_1$, $P2_1$ and $P2_12_12$). The RPTPαD1 structure was solved by molecular replacement in space group $C222_1$ and refined to 2.3 Å resolution in the related space group $P2_1$. Refinement to 1.7 Å resolution in space group $P2_12_12$ is in progress (Table 1). The identical RPTPαD1 dimer (Fig. 1) is observed in the three space groups despite considerable differences in the macromolecular packing between space groups $P2_1$ (44% solvent) and $P2_12_12$ (58% solvent), indicating that this oligomeric state is structurally relevant. The root-mean-square deviation between the Cα positions of the two dimers found in the $P2_1$ and $P2_12_12$ space groups is 0.6 Å. At the dyad-related interface of the RPTPαD1 dimer, the buried surface area measures 1,500 Å^2, a value comparable to that seen with other higher affinity protein–protein interfaces[6]. Oligomerization occurs around a two-fold rotational axis that fixes both N-terminal extensions on the same surface of each monomer, presumably facing the plasma membrane (see Fig. 1*a*).

Using dynamic light scattering[7] (data not shown) and gel filtration chromatography at RPTPαD1 concentrations of 0.1 to 5.0 mg ml^{-1}, we observe monomers, dimers and higher oligomers of RPTPαD1. In similar experiments with RPTPαD2 (residues 504–793), we only observe monomeric species. We speculate that the dimer and higher oligomers observed in solution and in the two independent crystal forms partly replicate the oligomerization of intact RPTPα in the membrane. The increase in local concentration of RPTPα in the membrane, caused by the binding of multivalent ligands to extracellular domains, would significantly increase oligomerization. The additional spatial constraint that the N terminus of each RPTPαD1 monomer is anchored in the membrane (Fig. 1*a*) would enhance the dimerization potential of RPTPαD1 *in vivo* by fixing the orientation of the D1 monomer and restricting the diffusion of RPTPα to two dimensions.

Whereas the overall fold of RPTPαD1 closely resembles the fold previously observed for PTP1B (ref. 8) and one *Yersinia* protein-tyrosine phosphatase (PTP), Yop51 (ref. 9), the quaternary organization of the RPTPαD1 monomers in two independent

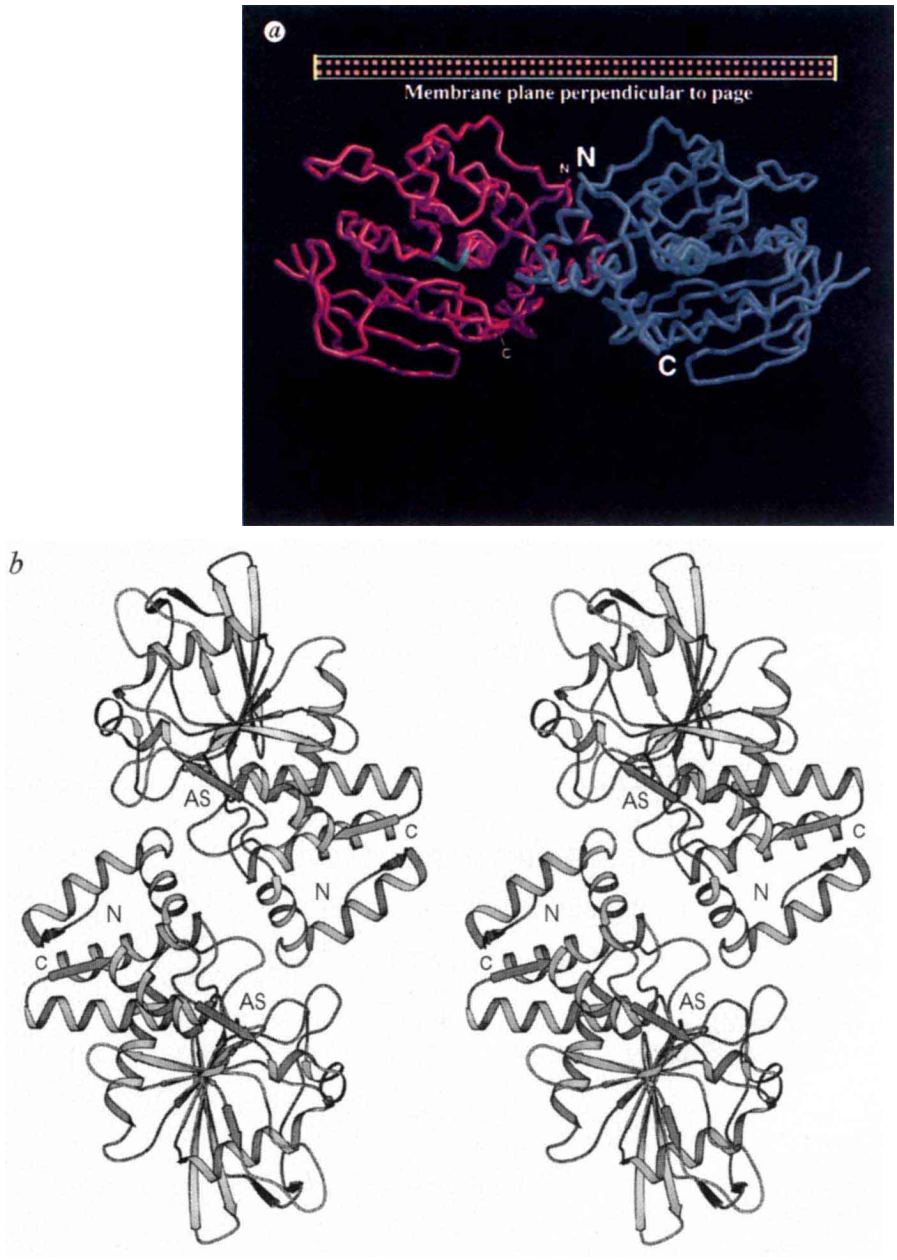
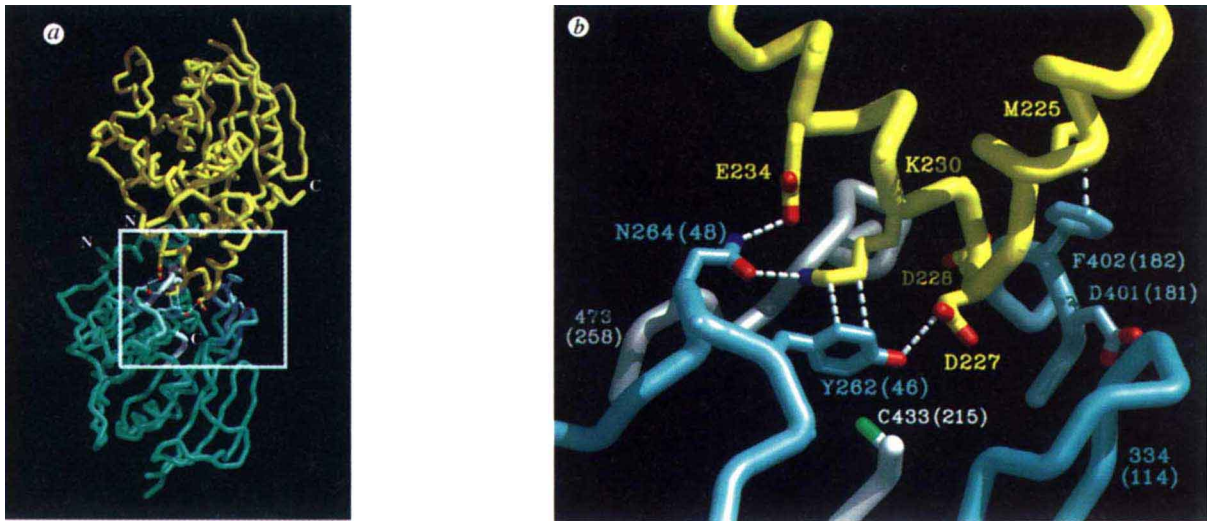


FIG. 1 Global view of the RPTPαD1 dimer. *a*, Cα trace of the RPTPαD1 dimer. The dyad axis is oriented vertically in the plane of the figure. The N and C termini are labelled for clarity. The active-site cysteine of the pink monomer is highlighted in green. A possible orientation of the catalytic domains with respect to the plasma membrane is emphasized schematically with the orange squares boxed in yellow. This orientation with an additional 41 N-terminal residues would link D1 to the transmembrane helix and thus the extracellular domain. *b*, Stereo ribbon diagram of the RPTPαD1 dimer. The dyad axis has been rotated 90° towards the reader from the orientation shown in *a*. The C termini now extend away from the reader. The label AS placed near the catalytically essential Cys 433 emphasizes the active site of each monomer. Both of these views accurately represent the dimeric form of RPTPαD1 observed in two independent crystal forms. *a* was produced with the conic option[24] of MIDAS[25], and *b* with SETOR[26].

crystal forms is unique. In the RPTPαD1 dimer, the amino-terminal segment of each monomer forms a helix–turn–helix structural wedge (interhelical angle ~80°) tucked into the active site of the opposing monomer (Fig. 2*a*, *b*). Four loops, 258–264 (L1), 330–340 (L6), 398–404 (L13 or WpD) and 472–482 (L17) delineate the entrance to the active site defined by catalytically essential Cys 433 (refs 8, 10). There are hydrogen bonds and van der Waals interactions between the residues from the N-terminal wedge of one monomer and residues from the L1 and L13 (WpD) loops of the dyad-related monomer (Fig. 2*b*). These interactions restrain the mobile L13 or WpD loop of each monomer in the 'open' position[11]. In its present conformation, the structural wedge precludes the large movement of L13 (WpD) towards the active-site cysteine as observed in both a PTP1B-peptide complex[11] and the Yop51–sulphate complex[10]. In a least-squares structural superimposition of the RPTPαD1 monomer with the sulphate-bound form of Yop51 (Brookhaven Protein Data Bank code 1YTS), the backbone atoms of the closed WpD (L13) loop of Yop51 clash with the Asp–Asp motif (residues 227–228) of the N-terminal wedge of the dyad-related RPTPαD1 monomer (Fig. 2*c*). In the RPTPαD1 dimer, both the putative general acid, Asp 401 (ref. 12), and the phosphotyrosyl-binding group, Phe 402 (ref. 11), are firmly fixed on L13 and directed away from the active site (Fig. 2*b*, *c*). In addition, several side chains from the second helix of this N-terminal wedge interact with Tyr 262 and Asn 264 on L1 in the dyad-related monomer. In the PTP1B–peptide complex[11], the two equivalent residues, Tyr 46 and Asp 48, form critical interactions with the phosphotyrosine-containing peptide. This stereochemical arrangement indicates that the dimeric form of RPTPαD1 described here is unable to bind phosphotyrosine-containing substrates and is therefore catalytically inactive. Within one monomer, the highly conserved proline-rich sequence Pro-Pro-Leu-Pro (210–213) preceding the N-terminal helix participates with residues 458–461 in three inter-strand hydrogen bonds (Fig. 3*a*). We believe that this latter structural link, together with the rigidity conferred by three proline residues, constrain the conformation of the N-terminal wedge.

The unique structural arrangement of the amino-terminal wedge and residues 458–461 of dimeric RPTPαD1 contrasts sharply with the corresponding segments in the non-receptor PTP1B, which are unrelated in sequence (Fig. 3). Compared with the human PTP1B primary sequence, RPTPαD1 contains a



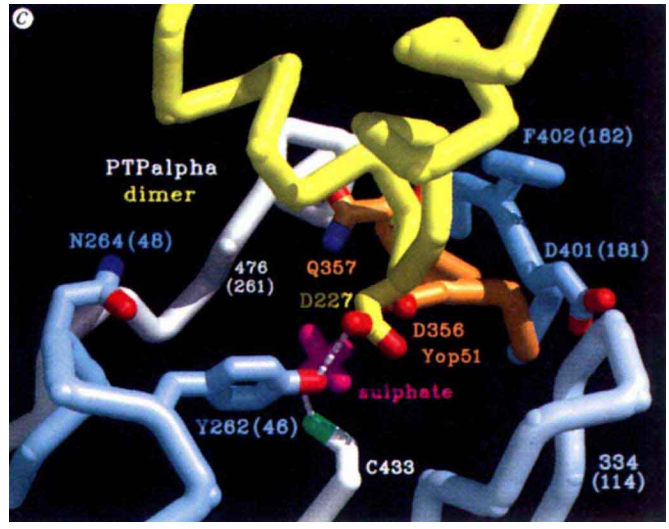


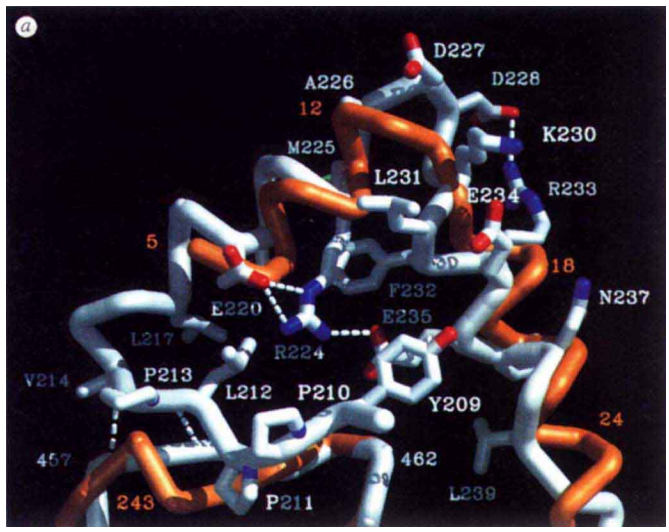
FIG. 2 Dimer specifying interactions. *a*, Global representation of the RPTPαD1 dimer. The dyad axis is nearly horizontal, rotated approximately 20° towards the reader. For clarity the main chain is a Cα trace. Side chains involved in dimeric interactions with the active site of one monomer are shown. In the current view, the monomer occluding the active site of a dyad-related monomer is coloured yellow. The opposite monomer is coloured aqua. Active-site loops L1 and L6 are light blue, L13 is grey-blue, and L17 is grey. The atoms of side-chain residues are colour coded for clarity (oxygen is red, nitrogen is blue, and sulphur is green). *b*, Close-up view of the boxed region in panel *a*. Numbers in parentheses refer to the homologous residues in PTP1B[8]. Asn 264 on L1 (Asp 48 in PTP1B) specifically interacts with Lys 230 and Glu 234 on the paired monomer through specific hydrogen bonds. In turn, Lys 230 through the side-chain methylene groups packs against the aromatic ring of Tyr 262 on L1(Tyr 46 in PTP1B). The carboxyl group of Asp 228 contacts the main-chain amide of Gly 403 (not visible in the figure) whereas the side chain of Met 225 packs against the side chain of Phe 402. Finally, the carboxyl group of Asp 227 at the tip of the turn on the paired monomer completes the stereochemical entrapment of Tyr 262 and the L1 loop of the dyad-related monomer. Tyr 262 (Tyr 46 in PTP1B) on L1 and Phe 402 (Phe 182 in PTP1B) on L13 clamp the phosphotyrosine side chain of bound substrates in PTP1B[11]. The hydrophobic nature of residue 230 is highly conserved amongst D1 primary sequences. Asp 227 also is highly conserved. *c*, View as in *b* with the closed WpD (L13) loop and sulphate ion of the Yop51–sulphate complex aligned by superimposing their phosphatase signature loops ((I/V)HCXAGXGR(S/T)G; single-letter amino-acid code). Interestingly, Asp 227 of the N-terminal wedge and Asp 356 of Yop51 occupy nearly the same position in the active site. *a*, *b* and *c* were produced with the conic option[24] of MIDAS[25].

two-residue insertion in the turn linking the first and second N-terminal helices, α1′ and α2′ (Fig. 3*b*). Both of the inserted residues, Asp 227 and Asp 228, participate in interactions across the dimer interface of RPTPαD1 (Fig. 2*b*). The two-stranded β-sheet between the proline-rich sequence, βx, and residues 458–461 in the carboxy-terminal half of RPTPαD1, βy, has no structural counterpart in PTP1B. PTP1B lacks the N-terminal extension encompassing the prolines and contains a helical three-residue insertion in the second segment, located towards the C terminus.

Using the consensus sequence for the N-terminal structural wedge (residues 208–242) generated from a small subset of D1 sequences illustrated in Fig. 3*b*, we retrieved all available D1 primary sequences from a database search. In contrast, no D2 catalytic domains of RPTPs or catalytic domains of non-receptor PTPs were retrieved. Consequently, we propose that this structurally constrained N-terminal wedge is a functionally conserved motif present in the large family of RPTPs. The sites for protein kinase C (PKC)-mediated phosphorylation of RPTPα, Ser 180 and Ser 204, reside near the dimer interface, suggesting a possible role for PKC-mediated phosphorylation in the regulation of RPTPα oligomerization[13].

A question that remains is whether other RPTPs dimerize like D1. The existence of the conserved N-terminal extension is consistent with the possibility that all RPTPs can dimerize and the variable residues in this region could provide the necessary dimerization specificity. Experimentally, there is evidence that CD45, a PTP required for antigen-dependent activation of T cells through the T-cell receptor, is negatively regulated by dimerization. The existence of CD45 homodimers has been detected using a dithio-bis-succinimidyl propionate crosslinking agent in YAC-1 cell lysates[14]. Moreover, epidermal growth factor (EGF)-induced dimerization of an artificial EGF receptor (extracellular and transmembrane portion)/CD45 (intracellular portion) chimaera expressed in a CD45-deficient T-cell line causes loss of antigen-dependent activation, whereas co-expression of the EGF receptor/CD45 chimaera gene with the gene encoding only the extracellular and transmembrane portions of the EGF receptor restored T-cell signalling even in the presence of an EGF receptor ligand[3]. Given that phosphatase activity of the D1 domain of CD45 is required for antigen-dependent activation, EGF-induced dimerization of the intracellular portion of CD45 is likely to occur in a manner analogous to that of the RPTPαD1 dimer described herein, thus explaining the loss of catalytic activity.

FIG. 3 Comparisons of RPTPαD1 with other tyrosine phosphatases. *a*, Main-chain representation of the N-terminal wedge and residues 456 to 463 (238 to 248 in PTP1B) of RPTPαD1 and PTP1B. D1 is grey. PTP1B (Brookhaven Protein Data Bank code 2HNP) is orange. Side chains of D1 residues comprising the consensus RPTP N-terminal wedge are included. Many of the conserved residues in D1 define the conformation of the wedge. In particular, a cluster of hydrophobic contacts provided by the side chains of Leu 217, Ile 221, and Phe 232, and the methylenes of Arg 224 and Glu 235 structurally define the interhelical angle. A salt bridge connecting Arg 224 to Glu 235 acts as a tether across the interhelical wedge. The Pro-Pro-Leu-Pro sequence (residues 210–213) forms a two-stranded β-sheet with residues 457 to 462 of RPTPαD1. This rigid structural element closes off the amino-terminal wedge and anchors it to the rest of the molecule. Produced with the conic option[24] of MIDAS[25]. *b*, Sequence alignment of the N-terminal wedge of RPTPs' membrane-proximal catalytic domains[27]. Line one schematically illustrates the secondary structural elements observed in RPTPαD1. Line two gives the consensus sequence conserved across the large family of D1 domains. The numbering scheme refers to the short variant of murine RPTPα[28] with the first residue of the signal peptide numbered as 1. The stars indicate residues involved in active-site-directed dimeric interactions. Residues boxed in black are common to at least five aligned D1 sequences. (*m* is mouse, *h* is human, *r* is rat, and *d* is *Drosophila*). The first 14 homologous sequences contrast sharply with the lack of sequence conservation in the final segments. The most distinguishing feature of the RPTP N-terminal wedge involves a two-amino-acid insertion (Asp 227 and Asp 228) into the tip of the non-receptor-like



PTP N-terminal segment. The second to the last line corresponds to the numbering scheme of human PTP1B. The final line emphasizes the structural elements observed in the PTP1B crystal structure[8].

TABLE 1 Refinement statistics for crystal form $P2_1$

Resolution shell (Å)	4.29	3.50	3.08	2.81	2.62	2.47	2.35	2.25	Total
Number of reflections	3,331	3,312	3,268	3,256	3,233	2,931	1,799	564	21,694
Per cent complete	99.9	99.6	99.0	99.1	97.5	89.8	56.6	17.4	82.4
$\langle I/\sigma(I)\rangle$	34	28	23	17	13	10	8	7	22
R_{sym}*	0.042	0.051	0.038	0.040	0.056	0.067	0.091	0.104	0.046
R-factor ±	0.244	0.159	0.180	0.191	0.213	0.230	0.240	0.266	0.202
Free *R*-factor ±	0.277	0.206	0.264	0.269	0.292	0.296	0.301	0.300	0.262
R.m.s. deviations		Bond lengths 0.009 Å				Bond angles 1.4°			

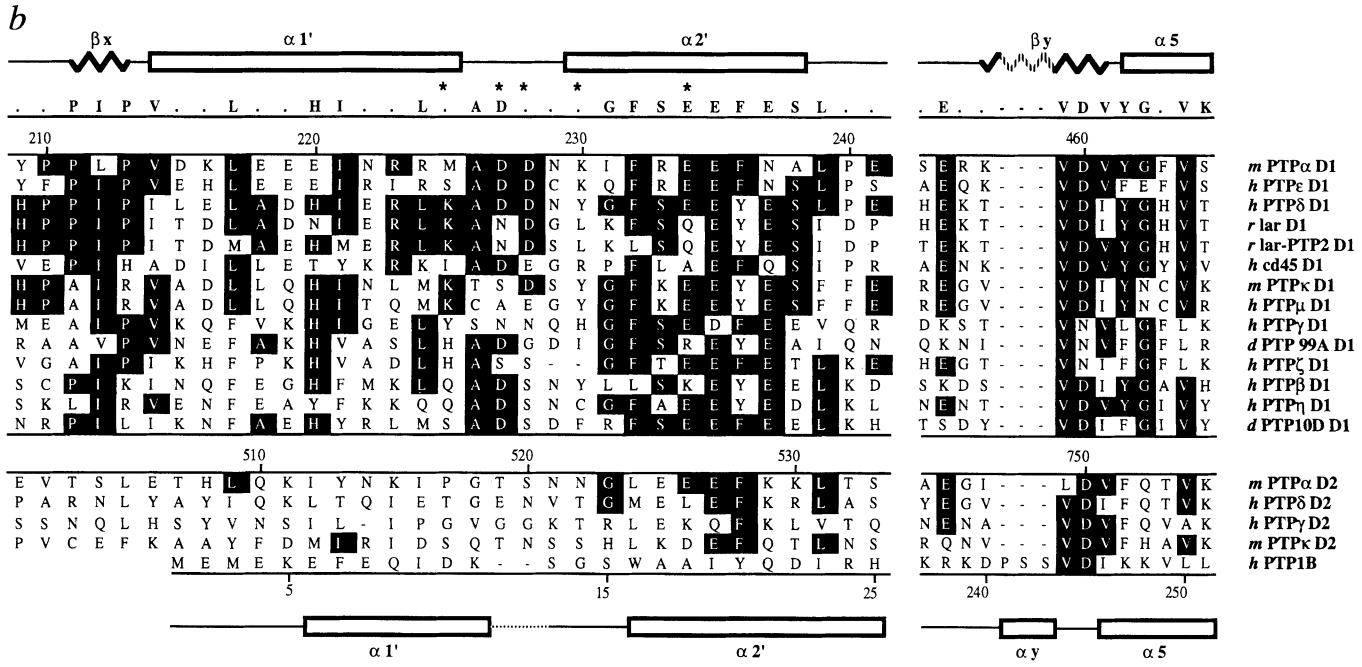
Truncated D1 (residues 202–503) was obtained by digestion of immobilized GST–RPTPαD1, residues 167 to 503 (ref. 5) with 1 mg of trypsin per 50 mg of GST–RPTPαD1 in 50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM dithiothreitol, for 1 h at 4 °C. D1 was further purified with a Pharmacia MonoQ column (pH 8.0, NaCl gradient). D1 crystals were grown by vapour diffusion in hanging drops at a protein concentration of 15 to 20 mg ml^{-1} in 50 mM sodium acetate, pH 5.0, 0.3 M ammonium acetate, 12% (w/v) PEG 8000. Crystals were flash-frozen with 20% (v/v) PEG 400. Diffraction data were collected on a DIP2000 imaging plate system equipped with double focusing Pt/Ni coated mirrors (Mac-Science Corporation, Japan). Data were processed with DENZO[15], scaled with SCALEPACK[15] and further processed with CCP4 programs[16]. Molecular replacement was performed in the space group $C222_1$ ($a = 42.3$ Å, $b = 115.4$ Å, $c = 119.9$ Å, one molecule per asymmetrical unit) using a trimmed model of PTP1B (PDB entry code 2HNP; 34% sequence identity) with AMoRe (ref. 17) integrated in an AVS environment[18,19]. After rigid-body refinement (one body) the *R*-factor dropped to 0.46 (data 20.0 to 4.5 Å). The model was adjusted using O (ref. 20) and refined with X-PLOR (ref. 21). Higher quality crystals were then obtained (same conditions) in the related space group $P2_1$ ($a = 42.2$ Å, $b = 119.8$ Å, $c = 61.2$ Å, $\beta = 110.1°$, two molecules per asymmetrical unit). In this crystal form, the second molecule of the dimer is generated by a non-crystallographic two-fold which corresponds to the crystallographic two-fold parallel to *a* of $C222_1$. X-PLOR refinement of the dimer and subsequent introduction of water molecules using the ARP software[22] allowed the unambiguous modelling of residues 209 to 217. The current model comprises 2 × 280 residues (Lys 208–Tyr 495) and 395 ordered water molecules. Residues 379 to 386 are disordered in the model; secondary tryptic cleavage sites exist at Arg 385 and Lys 386. The Ramachandran plot[23] of the final model shows that 89% of the residues are in most favoured regions. The final root-mean-square-deviation between the search model and the refined structure is 1.0 Å for 246 Cα atoms deviating by less than 3.5 Å. A second crystal form of D1 ($P2_12_12$, $a = 80.0$ Å, $b = 116.5$ Å, $c = 42.2$ Å, one molecule per asymmetrical unit) was grown using the same crystallization conditions as described above in the presence of an excess of the non-hydrolysable phosphonate ester analogue of the peptide used in the PTP1B–peptide complex[11]. Diffraction data to 1.7 Å were collected at the Stanford Synchrotron Radiation Laboratory, beamline 7–1 ($\lambda = 1.08$ Å) on an MAR imaging plate system. Data processing was performed as described previously. The second crystal form was solved using the D1 model with AMoRe. After rigid-body refinement, the *R*-factor was 36% (data 20.0 to 3.2 Å). Further refinement is in progress. So far, no electron density for the peptide has been observed.

* $R_{sym} = \Sigma_h|I_h - \langle I_h\rangle|/\Sigma_h I_h$ where $\langle I_h\rangle$ is the average intensity of reflection h for its symmetry and Friedel equivalents. ±*R*-factor $= \Sigma_h(|F_o(h) - F_c(h)|)/\Sigma_h F_o(h)$. Free *R*-factor is calculated using 10% of the data that has never been used for refinement. Unless specified, the *R*-factor is calculated using the remaining 90% of the data with no σ cutoff.

The two independent RPTPαD1 crystal structures, the primary sequence conservation of the dimer interface, and the CD45/EGF chimaera study[3] link the oligomeric structure of RPTPαD1 with a simple regulatory scheme in which dimerization inhibits RPTPα activity. As such, it should serve as a model for the functional and structural features governing the negative regulation of RPTP activity. □

Received 1 May; accepted 17 June 1996.

1. Hunter, T. *Cell* **80,** 225–236 (1995).
2. Mourey, R. J. & Dixon, J. E. *Curr. Opin. Genet. Dev.* **4,** 31–39 (1994).
3. Desai, D. M., Sap, J., Schlessinger, J. & Weiss, A. *Cell* **73,** 541–554 (1993).
4. Peles, E. *et al. Cell* **82,** 251–260 (1995).
5. den Hertog, J., Tracy, S. & Hunter, T. *EMBO J.* **13,** 3020–3032 (1994).
6. Janin, J. & Chothia, C. *J. biol. Chem.* **265,** 16027–16030 (1990).
7. Schmitz, K. S. *An Introduction to Dynamic Light Scattering by Macromolecules* (Academic, Boston, MA, 1990).
8. Barford, D., Flint, A. J. & Tonks, N. K. *Science* **263,** 1397–1404 (1994).
9. Stuckey, J. A. *et al. Nature* **370,** 571–575 (1994).
10. Schubert, H. L., Fauman, E. B., Stuckey, J. A., Dixon, J. E. & Saper, M. *Protein Sci.* **4,** 1904–1913 (1995).
11. Jia, Z., Barford, D., Flint, A. J. & Tonks, N. K. *Science* **268,** 1754–1758 (1995).
12. Zhang, Z. Y., Wang, Y. & Dixon, J. E. *Proc. natn. Acad. Sci. U.S.A.* **91,** 1624–1627 (1994).
13. Tracy, S., van der Geer, P. & Hunter, T. *J. biol. Chem.* **270,** 10587–10594 (1995).
14. Takeda, A., Wu, J. J. & Maizel, A. L. *J. biol. Chem.* **267,** 16651–16659 (1992).
15. Otwinowski, Z. *Data Collection and Processing* (SERC Daresbury Laboratory, Warrington, 1993).
16. Collaborative Computing Project No. 4 *Acta crystallogr.* **D50,** 760–763 (1994).
17. Navaza, J. *Acta crystallogr.* **A50,** 157–163 (1994).
18. Upson, C. *et al. IEEE Comput. Graphics Appli.* **9**(4), 30–42 (1989).
19. Wild, D. L., Tucker, P. A. & Choe, S. *J. molec. Graphics* **13,** 291–298 (1995).



20. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* A**47,** 110–119 (1991).
21. Brunger, A. T. X-PLOR, Version 3.1. (Yale Univ. Press, New Haven, CT 1992).
22. Lamzin, V. S. & Wilson, K. S. *Acta crystallogr.* D**49,** 127–149 (1993).
23. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. appl. Crystallogr.* **26,** 283–291 (1993).
24. Huang, C. C., Pettersen, E. F., Klein, T. E. & Langridge, R. *J. molec. Graphics* **9,** 230–236 (1991).
25. Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. *J. molec. Graphics* **6,** 13–27 (1988).
26. Evans, S. V. *J. molec. Graphics* **11,** 134–138 (1993).
27. Goldstein, B. J. *Phosphoprotein Phosphatases 1: Tyrosine Phosphatases* (Academic, London, 1995).
28. Sap, J., D'Eustachio, P., Givol, D. & Schlessinger, J. *Proc. natn. Acad. Sci. U.S.A.* **87,** 6112–6116 (1990).

ACKNOWLEDGEMENTS. We thank S. Choe, R. Ranganathan and G. V. Louie for discussion; S. M. Redford and C. M. Starks for assistance during data collection at beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL); M. Park for N-terminal sequencing; D. Wild, M. E. Bowman, W. Yang, F. D. Bushman, M. B. Harrington, and the staff at SSRL for technical help; and N. P. H. Møller and Novo Nordisk for peptides and financial support (A.M.B.). The work performed at SSRL was supported by the NIH and DOE. Work in the structural biology laboratory at the Salk Institute was supported in part from institutional funds and funds from the Lucille P. Markey Charitable Trust. J.d.H. is supported by a grant from the Dutch Cancer Society. T.H. is an American Cancer Society Professor.

CORRESPONDENCE and requests for materials should be addressed to J.P.N. (e-mail: noel@sbl.salk.edu). Coordinates for the RTPTαD1 dimer will be submitted to the Brookhaven Protein Data Bank.

Recognition of DNA by designed ligands at subnanomolar concentrations

John W. Trauger, Eldon E. Baird & Peter B. Dervan

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

SMALL molecules that specifically bind with high affinity to any predetermined DNA sequence in the human genome would be useful tools in molecular biology and potentially in human medicine. Simple rules have been developed to control rationally the sequence specificity of minor-groove-binding polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids. Two eight-ring pyrrole–imidazole polyamides differing in sequence by a single amino acid bind specifically to respective six-base-pair target sites which differ in sequence by a single base pair. Binding is observed at subnanomolar concentrations of ligand. The replacement of a single nitrogen atom with a C-H regulates affinity and specificity by two orders of magnitude. The broad range of sequences that can be specifically targeted with pyrrole–imidazole polyamides, coupled with an efficient solid-phase synthesis methodology, identify a powerful class of small molecules for sequence-specific recognition of double-helical DNA.

For side-by-side complexes of pyrrole–imidazole polyamides in the minor groove of DNA, the DNA-binding sequence specificity depends on the sequence of side-by-side amino-acid pairings[1–3]. A pairing of imidazole (Im) opposite pyrrole (Py) targets a G·C base pair, whereas pyrrole opposite imidazole targets a C·G base-pair[1–3]. A pyrrole/pyrrole combination is degenerate and targets both T·A and A·T base pairs[1–5]. Specificity for G,C base pairs results from the formation of a hydrogen bond between the imidazole N3 and the exocyclic amino group of guanine[1–3]. The generality of these pairing rules has been demonstrated by targeting a wide variety of sequences and is supported directly by several NMR structure studies[1–10].

In parallel with the elucidation of the scope and limitations of the pairing rules already described, efforts have been made to increase the DNA-binding affinity and specificity of pyrrole–imidazole polyamides by covalently linking polyamide subunits[11–16]. The polyamide ImPyPy-γ-PyPyPy-dimethylaminopropylamide (Dp) , which contains a 'turn' amino acid γ-aminobutyric acid (γ) was shown to bind the five-base-pair target site 5′-TGTTA-3′ in a 'hairpin' conformation with an equilibrium association constant, $K_a = 8 \times 10^7\ M^{-1}$, an increase of 300-fold relative to unlinked three-ring polyamide dimers. A key issue was to determine whether low-molecular-weight (M_r 1,200) pyrrole–imidazole polyamides could be constructed that would bind DNA at subnanomolar concentrations without compromising sequence selectivity.

We report the DNA-binding affinities of two eight-ring hairpin polyamides, ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**) and ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**), which differ by a single amino acid, for two 6-base-pair (bp) target sites, 5′-AGTACT-3′ and 5′-AGTATT-3′, which differ by a single base pair. Based on the pairing rules for polyamide–DNA complexes, the sites 5′-AGTACA-3′ and 5′-AGTATT-3′ are for polyamide **1** 'match' and 'single-base-pair mismatch' sites, respectively, and for polyamide **2** 'single-base-pair mismatch' and 'match' sites, respectively (Figs 1 and 2).