```
          Rat 5'DI (1523) TTTATATTTG TTTATGATGG .TCACAGTGT AAAGTTCACA CAGCTGTG.A CTTGATTTTT ..AAAAATGT CGGGAAGA
        Human 5'DI (1726) TACATATTTG TTTATGATGG .CCACAGCCT AAAGTACACA CGGCTGTG.A CTTGATTCAA AAGAAAATGT TATAAGA.
    5'DI Consensus        T..ATATTTG TTTATGATGG ..CACAG..T AAAGT.CACA C.GCTGTG.A CTTGATT... ...AAAATGT ....A...

          Rat GPX (1004) GGGAGGTTTT TCCATGACGG TGTTTCCTCT AAATTTACAT GGAGAAAC.A CCTGATTTCC AGAAAAATCC CCTCAGAT
        Mouse GPX (1168) GGGCGGTTCT TCCATGATGG TGTTTCCTCT AAATTTGCAC GGAGAAAC.A CCTGATTTCC AGGAAAATCC CCTCAGAT
        Human GPX  (942) GGGGTTTTCA TCTATGAGGG TGTTTCCTCT AAACCTACGA GGGAGGAACA CCTGATCTTA CAGAAAATAC CACCTCGA
        Bovine GPX (895) AGGGATTTTG CCCATGAAGG TGTTCCCTCT AAACCTACGT GGAGGAAT.G CCTGATGTCC AGGAAAAATCC CCTGAGGT
    GPX consensus         .GG...TT.. .C.ATGA.GG TGTT.CCTCT AAA..T.C.. GG....A... CCTGAT.T.. ...AAAAT.C C.......

    5'DI consensus        T..ATATTTG TTTATGATGG ..CACAG..T AAAGT.CACA C.GCTGTG.A CTTGATT... ...AAAATGT ....A...
    GPX consensus         .GG...TT.. .C.ATGA.GG TGTT.CCTCT AAA..T.C.. GG....A... CCTGAT.T.. ...AAAAT.C C.......
SECIS consensus           ......TT.. ...ATGA.GG .........T AAA..... .......... C.TGAT.... ...AAAAT.. ........
```

FIG. 4 Sequence similarities in the stem–loop regions of the rat and human 5'DI, and mammalian GPX cDNAs. Analysis was done using the LINEUP program[14]. Nucleotide numbers are shown in parentheses.

and quantitated the 27 and 14K [$^{35}$S] methionine-labelled products after SDS-PAGE. The ratio of 27 to 14K protein was $0.17 \pm 0.03$ for the rat WT, $0.08 \pm 0.04$ for the human M1, and $0.013 \pm 0.004$ for the rat M2 construct. These results establish the UGA recognition function of the 3'ut regions.

The 3'uts of the various selenocysteine-encoding mRNAs have little primary sequence similarity. Alignment of the proposed stem–loop regions of the rat and human 5'DI mRNAs with the 3'ut regions of the mammalian GPX mRNAs identifies the conserved nucleotides shown in Fig. 4. Inversion of the sequences between 1,245 and 1,615 in the rat 5'DI cDNA resulted in loss of 5'DI activity (Fig. 1*a*). The predicted stem–loop in this construct is similar to the wild-type, but as the sequence is complementary, none of the conserved bases are present. The 3'ut of human plasma GPX[5] contains no significant similarity to the sequences in Fig. 4, however, this sequence is not full length, as it lacks a polyadenylation signal sequence and poly (A) tail. The recently reported cDNA for rat selenoprotein P (ref. 6) contains a 3'ut region of >1,600 nucleotides, and secondary structure analyses predict 15 stem–loops with free energies of −20 kcal or less. Six of these contain UAAA or AAA sequences in the loop. Because of the numerous potential stem–loops in this long 3'ut, the identification of specific regions involved in selenocysteine insertion may require deletion mapping or other functional analyses.

Previous studies of the *Escherichia coli* FDH mRNA established that only the sequences immediately adjacent to the UGA codon are required for its recognition as a selenocysteine codon[8]. We have not examined the role of such sequences in the 5'DI mRNA. Our results establish that a roughly 200 nucleotide segment located more than 1 kb downstream of the UGA in the 5'DI mRNAs is essential for insertion of selenocysteine into this protein. Although most of our results were obtained using transient expression techniques, we have demonstrated that these sequences are required for *in vivo* and *in vitro* translation of the intact, fully functional protein. Thus we may term this segment of these mRNAs a selenocysteine-insertion sequence (SECIS) motif. The requirement for a SECIS motif in the 5'DI mRNAs for successful translation of this protein, and the presence of sequences with similar function in the GPX mRNAs, identifies a previously unrecognized regulatory step in the expression of genes encoding eukaryotic selenocysteine-containing proteins. Such motifs may be required in eukaryotic expression vectors for insertion of the more reactive selenium in place of sulphur in sulphydryl active-site proteins for purposes of biochemical or structural analyses. How they function remains to be determined. □

Received 30 May; accepted 29 July 1991.


1. Zinoni, F., Birkmann, A., Stadtman, T. & Bock, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4650–4654 (1986).
2. Garcia, G. E. & Stadtman, T. C. *J. Bact.* **174**, 2093–2098 (1991).
3. Chambers, I. *et al. EMBO J.* **5**, 1221–1227 (1986).
4. Mullenbach, G. T. *et al. Protein Eng.* **2**, 239–246 (1988).
5. Takahashi, K. *et al. J. Biochem.* **108**, 145–148 (1990).
6. Hill, K. E., Lloyd, R. S., Yang, J. G., Read, R. & Burk, R. F. *J. biol. Chem.* **266**, 10050–10053 (1991).
7. Berry, M. J., Banu, L. & Larsen, P. R. *Nature* **349**, 438–440 (1991).
8. Zinoni, F., Heider, J. & Bock, A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4660–4664 (1990).
9. Bock, A. *Molec. Microbiol.* **5**, 515–520 (1991).
10. Berry, M. J., Kieffer, J. D., Harney, J. W. & Larsen, P. R. *J. biol. Chem.* **266**, 14155–14158 (1991).
11. Ho, Y.-S., Howard, A. J. & Crapo, *J. Nucleic Acids Res.* **16**, 5207 (1988).
12. Aruffo, A. & Seed, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8573–8577 (1987).
13. Brent, G. A. *et al. Molec. Endocrinol.* **3**, 1996–2004 (1989).
14. Devereux, J., Haeberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
15. Higuchi, R. in *PCR Protocols* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 177–183 (Academic, New York, 1990).


ACKNOWLEDGEMENTS. We thank B. Seed and D. D. Moore for critical reading of the manuscript. This work was supported by the NIH.

# Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crystallography and NMR spectroscopy

**Jörg Kallen, Claus Spitzfaden*, Mauro G. M. Zurini, Gerhard Wider*, Hans Widmer, Kurt Wüthrich* & Malcolm D. Walkinshaw†**

Preclinical Research, Sandoz Pharma AG, 4002 Basel, Switzerland
* Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, 8093 Zürich, Switzerland

**THE protein cyclophilin is the major intracellular receptor for the immunosuppressive drug cyclosporin A (ref. 1). Cyclosporin A acts as an inhibitor of T-cell activation and can prevent graft rejection in organ and bone marrow transplantation[2]. Cyclophilin may be responsible for mediating this immunosuppressive response. Cyclophilin also catalyses the interconversion of the *cis* and *trans* isomers of the peptidyl–prolyl amide bonds of peptide and protein substrates[3,4]. Here we report the X-ray crystal structure of human recombinant cyclophilin complexed with a tetrapeptide and the identification, by nuclear magnetic resonance spectroscopy, of the specific binding site for cyclosporin A. Cyclophilin has an eight-stranded antiparallel β-barrel structure. The prolyl isomerase substrate-binding site is coincident with the cyclosporin-binding site. These results may help to provide a structural basis for rationalizing the immunosuppressive function of the cyclosporin–cyclophilin system and will also be important in the design of improved immunosuppressant drugs.**

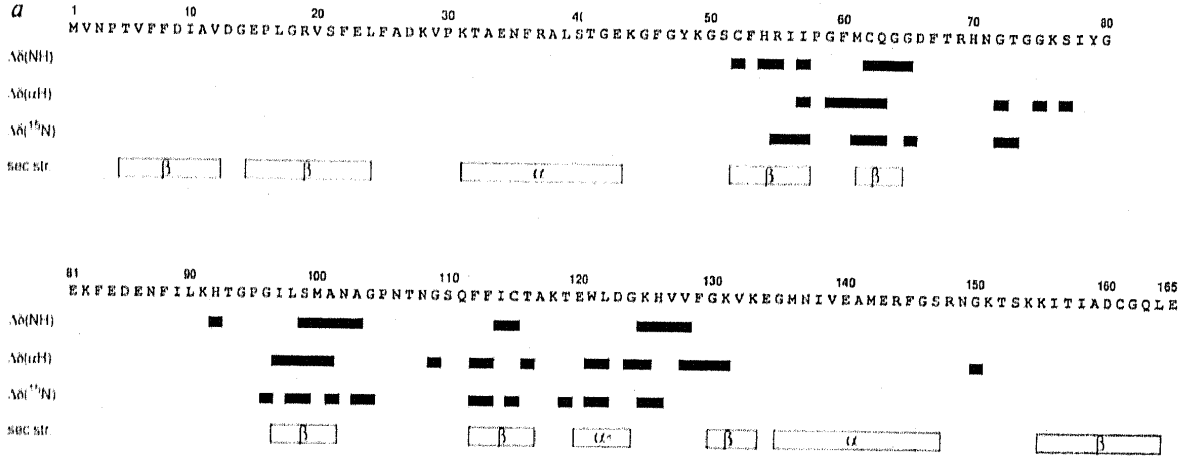
Cyclophilin (relative molecular mass 17,800 ($M_r$, 17.8K)) consists of a single polypeptide chain with 165 amino-acid residues (Fig. 1*a*). Cyclosporin A (CsA) is a cyclic undecapeptide with the sequence c-(MeBmt-Abu-Sar-MeLeu-Val-MeLeu-Ala-D-Ala-MeLeu-MeLeu-MeVal), where the prefix Me indicates *N*-methylation and the uncommon amino acids in positions 1 and 2 are (4R)-4-((E)-2-butenyl)-4,*N*-dimethyl-L-

† To whom correspondence should be addressed.

TABLE 1 Crystallographic and multiple isomorphous replacement (MIR) data

| Data set | Space group $P2_12_12_1$ ($a$=108.2 Å, $b$=123.0 Å, $c$=35.8 Å) Native (1) | Native (2) | EMTS* | PHMPS | $K_2PtCl_4$ (1) | $K_2PtCl_4$ (2) |
|---|---|---|---|---|---|---|
| Number of collected reflections | 43,082 | 42,788 | 43,308 | 26,385 | 23,713 | 23,600 |
| $R_{sym}$(I) in parentheses† | (6.9%) | (7.6%) | (7.2%) | (8.0%) | (10.2%) | (8.7%) |
| Number of unique reflections‡ | 11,254 | 12,303 | 9,602 | 8,354 | 8,156 | 7,891 |
| Maximum resolution Å | 2.6 | 2.6 | 3.0 | 3.0 | 3.2 | 3.2 |
| Mean fraction isomorphous change§ | | | 28.4% | 27.8% | 16.8% | 17.5% |
| Number of heavy atom sites | | | 2 | 4 | 6 | 6 |
| $R_c$‖ | | | 55.8% | 50.7% | 61.4% | 57.9% |
| Correlation¶ | | | 0.47 | 0.59 | 0.43 | 0.49 |
| Number of centric reflections | | | 1,611 | 1,455 | 1,415 | 1,323 |
| Overall phasing power** | | | 1.26 | 1.47 | 1.90 | 1.93 |

EMTS is ethyl-mercuri-thiosalicylate; PHMPS is *p*-hydroxymercuri-phenyl-sulphonate.
* Data set collected on image-plate area detector at the EMBL outstation, DESY, Hamburg.
† $R_{sym}=\Sigma_i|I_i-\bar{I}|/\Sigma\bar{I}$, where $I_i$ is the intensity of the reflection and $\bar{I}$ is the mean intensity of the i observations.
‡ Number of unique reflections phased: 9,898 (15 Å–2.9 Å); overall figure of merit=0.78.
§ Versus merged data set Native(1)+Native(2).
‖ $R_c$ is Cullis *R*-factor for centric reflections[19].
¶ Correlation between $F_{obs}$ and $F_{calc}$ for heavy atom structure (program REFINE in CCP4-package[18]).
** Phasing power is $F_H/E_{r.m.s.}$, where $F_H$ is heavy-atom structure factor and $E_{r.m.s.}$ is residual lack of closure[19].



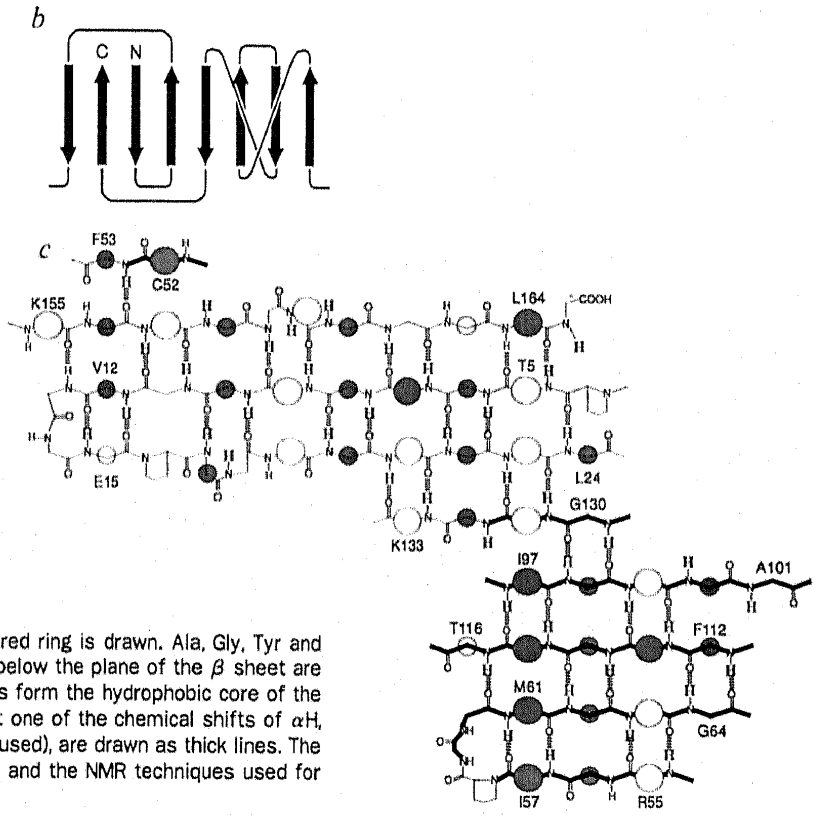




FIG. 1 Summary of the NMR results defining the CsA-binding site and the regular secondary structure and topology of cyclophilin. *a*, Amino-acid sequence of human cyclophilin, sequence locations of the regular $\alpha$-helical or $\beta$-sheet secondary structures ($\alpha$* indicates that this structure could either be a short $\alpha$ helix or a turn-like structure) and the chemical shift data used to delineate the binding site for CsA. The residues showing significant backbone chemical shift variations on binding of cyclosporin are identified with black bars. Chemical shift differences are indicated if they exceed the following limits: $\delta$(NH), 0.10 parts per million (p.p.m.); $\delta(\alpha H)$, 0.05 p.p.m.; $\delta(^{15}N)$, 0.50 p.p.m. *b*, Schematic representation of the antiparallel $\beta$ sheet in cyclophilin showing a global (+1, −3, −1, −2, +1, −2, −3) topology of the $\beta$ sheet. *c*, Drawing showing the sequence positions of the individual $\beta$ strands, the hydrogen bonds, the distribution of the hydrophobic side chains and the location of residues with large chemical shift changes on binding of cyclosporin. The two ends of each $\beta$ strand are labelled with the one-letter amino-acid code and the sequence location. Slowly exchanging amide protons are printed in bold. The presence of hydrogen bonds is identified with dashed lines. The amino-acid side chains are represented with circles at the C$\alpha$ carbon positions using the following code: big circles, side chains toward the reader; small circles, side chains pointing away from the reader; shaded circles, hydrophobic residues (C, I, V, L, F, M); empty circles, polar and charged residues (S, T, N, D, Q, E, R, K). For Pro a five-membered ring is drawn. Ala, Gly, Tyr and His have no symbol at the C$\alpha$ position. Essentially all side chains below the plane of the $\beta$ sheet are hydrophobic. In the X-ray structure (Figs 2 and 3) these side chains form the hydrophobic core of the $\beta$ barrel. Residues that experience a significant change of at least one of the chemical shifts of $\alpha$H, NH or $^{15}$N after complexation with cyclosporin (see *a* for the limits used), are drawn as thick lines. The sequence-specific NMR assignments for the polypeptide backbone and the NMR techniques used for the secondary structure determination are described in ref. 7.

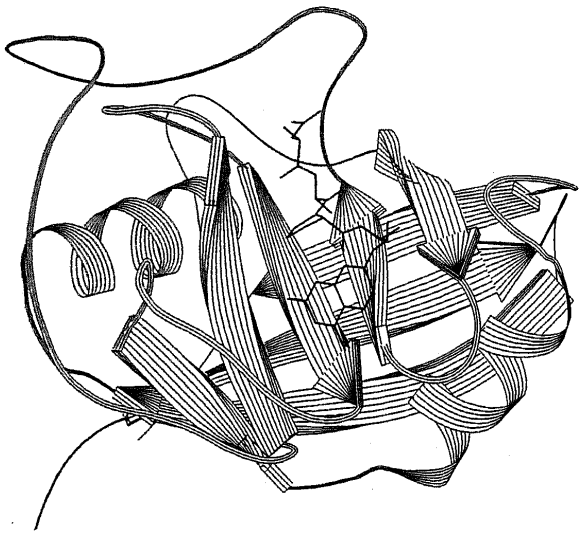


FIG. 2 A ribbon plot[18] illustrating the overall fold of human cyclophilin. The model tetrapeptide prolyl isomerase substrate (*N*-acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin) is also shown. The crystal structure was solved for a complex of human recombinant cyclophilin with this linear tetrapeptide using phases determined with the multiple isomorphous replacement (MIR) method from two heavy-atom derivatives (Table 1). There are two molecules in the asymmetric unit related to each other by a noncrystallographic twofold rotation axis. The MIR phases provided an electron density map[18,19] good enough for chain tracing[20] for all amino acids of both molecules. No solvent-flattening procedures were used. The secondary structure of cyclophilin as determined by NMR (ref. 7) was of considerable help in the chain tracing and model-building. The two independent molecules differ only slightly in loop regions involved in lattice contacts; the Cα atoms currently have an r.m.s. fit of 0.8 Å. Noncrystallographic symmetry was not imposed during refinement using X-PLOR (ref. 21). The present *R*-factor using data between 2.5 Å and 8 Å is 25.7%. Water molecules have not yet been included and further refinement is underway.

threonine and L-α-aminobutyric acid, respectively. The three-dimensional structure of cyclophilin-bound CsA has recently been determined by nuclear magnetic resonance (NMR) (refs 5, 6). Complete sequence-specific $^{1}H$ and $^{15}N$ NMR assignments for the cyclophilin backbone have also been determined and used to define the regular secondary structure[7], which consists of two well defined α helices and an eight-stranded antiparallel β sheet (Fig. 1*b*).

Various crystal forms of human cyclophilin complexed with either a tetrapeptide or CsA have been grown[8]. The crystal complex of cyclophilin with CsA has six molecules in the asymmetric unit and is being studied. The legend for Fig. 2 contains details of the X-ray crystallographic structure determination of a complex of human recombinant cyclophilin with the tetrapeptide *N*-acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin. This peptide is a model substrate for the prolyl isomerase, which catalyses the *cis–trans* conversion of the alanyl–prolyl amide bond.

Cyclophilin is a roughly spherical molecule with a radius of about 17 Å (Fig. 2). The main structural feature is the eight-stranded antiparallel β barrel that has a +1, −3, −1, −2, +1, −2, −3 topology (Fig. 1*b*). The barrel consists of two roughly perpendicular four-stranded β sheets (Fig. 1*c*) connected by short junctions at residues Leu98–Gly130 and Phe53–Ile156. Inside the barrel, a tightly packed core contains most of the hydrophobic side chains (Fig. 1*c*). Other hydrophobic residues are located in the contact region of the two amphipathic helices with the β barrel and in the cyclosporin-binding site.

There is a structural resemblance between cyclophilin and the superfamily of proteins involved in ligand transport including retinol-binding protein (RBP), bilin-binding protein and β-lactoglobulin[9]. Most of these molecules encapsulate their ligand in the β-barrel core. By contrast, the barrel core in cyclophilin is tightly packed with hydrophobic residues (Fig. 1*c*), and the putative ligand-binding site is on the outside of the barrel. The topology of cyclophilin also differs from the simple $(+1)_n$ up-and-down fold found in the RBP class of proteins, or the (−3, +1, +1) Greek key topology that is most frequently found in antiparallel β-barrel proteins. In particular the two crossover connections Gly64–Ile97 and Thr116–Gly130 represent an uncommon topological feature. As both loops lie on the outside of the barrel, the [−2, +1, −2] topology requires that the two loops cross each other. The unusual left-handed connection of the 116–130 loop may be rationalized by the length of the loop which may accommodate a variety of local conformations.

Sequence-specific NMR assignments for the $^{1}H$ and $^{15}N$ spins of the polypeptide backbone of cyclophilin were also obtained for a complex formed with a water-soluble CsA derivative. The very similar nuclear Overhauser effect (NOE) patterns observed for free and CsA-bound cyclophilin indicate that there is no change in secondary structure on CsA binding. Nonetheless, comparison with free cyclophilin showed that there are residues for which cyclosporin binding caused significant chemical shift changes (Fig. 1*a*). These chemical shift data were used to delineate the CsA-binding site in the three-dimensional cyclophilin structure (Fig. 3).

When the cyclophilin residues thus identified as being involved in cyclosporin binding (Fig. 1*a*) are mapped onto the three-dimensional structure, they are found to cluster on one side of the molecule and incorporate the tetrapeptide-binding site (Figs 1*c*, 2 and 3). In the crystal structure, the tetrapeptide ligand binds in a long deep groove located on the protein surface between one face of the β barrel and the Thr116–Gly130 loop. The best fit in the current model shows the Ala–Pro amide bond in the *trans* conformation. Site-directed mutagenesis studies[10,11] have been used to study enzymatic activity. Both Cys115 and Cys62 are near the peptide-binding groove, however replacing each Cys individually by Ala did not affect prolyl isomerase activity[10]. Site-directed mutagenesis studies have also shown that the sole tryptophan (Trp121), which sits in the middle of the binding loop, is implicated in CsA binding, but has little effect on prolyl isomerase activity[11]. In the crystal structure, this

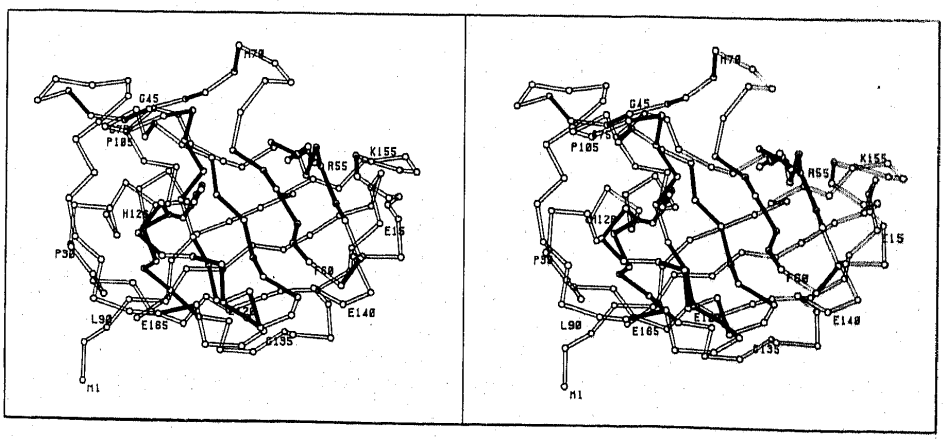



FIG. 3 Stereo PLUTO plot[18] showing the Cα skeleton of cyclophilin. Selected Cα atoms have been labelled with the sequence number and the one-letter amino-acid code. Residues which are implicated in CsA binding based on the chemical shift data in Fig. 1*a*, are shown in this diagram with filled Cα–Cα bonds. The side chains of Arg55 and His126, which may be involved in the prolyl isomerase mechanism, are also shown.

tryptophan is close to the coumarin ring of the tetrapeptide, and intermolecular NOEs observed in the cyclophilin–CsA complex in solution showed that it is in close contact with residues 9 and 11 of CsA (ref. 5). His126 sits on the same binding loop as Trp121, with $N^{\delta}$ about 6 Å from the carbonyl carbon of the prolyl amide. A possible mechanism for prolyl isomerase activity could involve His126 acting as proton acceptor for a water molecule involved in nucleophilic attack on the carbonyl carbon of the Ala–Pro amide bond. Further activation of the amide carbonyl group could be provided by the hydrogen bond between the carbonyl oxygen atom and the guanidinium group of Arg55.

The macrolide FK506 is chemically unrelated to CsA but is also a potent immunosuppressant with a very similar biological profile[12]. FK506 binds to the specific cytosolic immunophilin protein receptor FK-binding protein (FKBP) (ref. 12). Human recombinant FKBP has a chain length of 107 amino acids ($M_r$, 11.8K) and has no significant sequence homology with cyclophilin. The recently determined X-ray and NMR structures of human recombinant FKBP (refs 13–15) show that it exists in the crystal and in solution as a five-stranded antiparallel $\beta$-sheet which wraps around a short helix. There is thus no readily apparent similarity with the three-dimensional cyclophilin structure. The FK506 class of drugs do not bind to cyclophilin and CsA does not bind to FKBP. Nonetheless the mode of action of CsA and FK506 in the cell seem to be very similar[12,16]. FKBP also shows a *cis–trans* prolyl isomerase activity. But for both FK506 and CsA, suppression of prolyl isomerase activity alone seems insufficient to explain the biological activity, as drug concentrations necessary to inhibit T-cell activation would not saturate the abundant prolyl isomerases[12,17].

Our results are in line with the earlier NMR reports on CsA bound to cyclophilin[5]: intermolecular NOEs determined between cyclophilin and CsA implicate residues along one face of cyclosporin (residues 9, 10, 11, 1 and 2) as being important for binding. These complementary structural studies using both NMR spectroscopy in solution and X-ray crystallography thus provide new insights into the binding of CsA and, more generally, the molecular basis of immunosuppressive activity.

*Note added in proof*: The current *R*-factor is 21.8% for all data from 8 to 2.3 Å. An independent X-ray structure of unliganded cyclophilin has been determined[22]. Despite different unit cells, the molecular architecture seems similar. □

Received 5 July; accepted 1 August 1991.


1. Handschumacher, R. E., Harding, M. W., Rice, J., Drugge, R. J. & Speicher, D. W. *Science* **226**, 544–547 (1984).
2. Borel, J. F. *Pharmac. Rev.* **41**, 259–371 (1989).
3. Takahashi, N., Hayano, T. & Suzuki, M. *Nature* **337**, 473–475 (1989).
4. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. *Nature* **337**, 476–478 (1989).
5. Weber, C. *et al. Biochemistry* **30**, 6563–6574 (1991).
6. Fesik, S. W. *et al. Biochemistry* **30**, 6574–6583 (1991).
7. Wüthrich, K., Spitzfaden, C., Memmert, K., Widmer, H. & Wider, G. *FEBS Lett.* (in the press).
8. Zurini, M. *et al. FEBS Lett.* **276**, 63–66 (1990).
9. Cowan, S. W., Newcomer, M. E. & Jones, T. A. *Proteins* **8**, 44–61 (1990).
10. Liu, J., Albers, M. W., Chen, C., Schreiber, S. L. & Walsh, C. T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2304–2308 (1990).
11. Liu, J., Chen, C. & Walsh, C. T. *Biochemistry* **30**, 2306–2310 (1991).
12. Schreiber, S. L. *Science* **251**, 283–287 (1991).
13. Michnick, S. W., Rosen, M. K., Wandless, T. J., Karplus, M. & Schreiber, S. L. *Science* **252**, 836–839 (1991).
14. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. *Science* **252**, 839–842 (1991).
15. Moore, J. M., Peattie, D. A., Fitzgibbon, M. J. & Thomson, J. A. *Nature* **351**, 248–250 (1991).
16. Tropschug, M., Barthelmess, I. B. & Neupert, W. *Nature* **342**, 953–955 (1989).
17. Bierer, B. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9231–9235 (1990).
18. *CCP4; A Suite of Programs for Protein Crystallography* (SERC Daresbury Laboratory, Warrington, UK, 1986).
19. Blundell, T. L. & Johnson, L. N. *Protein Crystallography* 357 (Academic, Oxford, 1976).
20. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
21. Brunger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
22. Ke, H. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).


ACKNOWLEDGEMENTS. We thank R. Wenger for supplying the water-soluble cyclosporin and K. Wilson and C. Betzel of the EMBL outstation, Hamburg, for their help with data collection using synchrotron radiation. This work was supported by the Schweizerischer Nationalfonds (to K.W.).