The test for the significance of the regression was based on a comparison of the regression and deviations from regression mean squares. Confidence limits were based on a residual mean square that pools deviations from regression and within-groups components.
19. Only three snakes performed satisfactory concertina locomotion during the oxygen consumption experiments [mass = 107.1 ± 6.2 (SE) g]. Two of these were used during lateral undulation experiments as well.
20. A. F. Bennett, in *Functional Vertebrate Morphology*, M. Hildebrand, D. M. Bramble, K. F. Liem, D. B. Wake, Eds. (Belknap, Cambridge, MA, 1985), pp. 173–184.
21. J. A. Ruben, *J. Comp. Physiol.* **109**, 147 (1976).
22. Burst speed at 30°C was measured as snakes crawled over a 4-m track, 21 cm wide and covered with artificial grass and surface projections (as above). We vigorously tapped the tail with a soft brush to elicit rapid movement and used the timer feature of the video system to determine velocity. Three trials were conducted for each snake, and the fastest 50-cm interval of all three trials was selected as the snake's maximum burst speed.
23. Using the four individuals with $\dot{V}O_2$ data at each of the four sustainable speeds of lateral undulation (0.2, 0.3, 0.4, and 0.5 km hour^{-1}), we calculated the cost per cycle for each 1-min interval within the 2- to 5-min records of $\dot{V}O_2$ used to calculate the average values shown in Fig. 1A. These values were compared by means of a two-way analysis of variance (ANOVA) with speed as a fixed effect and individual as a random effect [M. J. Norusis, *Advanced Statistics, SPSS/PC+* (Statistical Package for the Social Sciences (SPSS), Inc., Chicago, 1986)]. Following the guidelines of J. H. Zar [*Biostatistical Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1984], we calculated F statistics using comparisons among mean squares (MS) as follows: $F_{speed} = MS_{speed}/MS_{speed \times individual}$ and $F_{individual} = MS_{individual}/MS_{error}$.
24. A. F. Bennett, in *Biology of the Reptilia*, C. Gans and F. H. Pough, Eds. (Academic Press, New York, 1982), vol. 13, pp. 155–199; C. F. Herreid, in *Locomotion and Energetics of Arthropods*, C. F. Herreid and C. R. Fourtner, Eds. (Plenum, New York, 1981), pp. 491–526; R. J. Full, B. D. Anderson, C. M. Finnerty, M. E. Feder, *J. Exp. Biol.* **138**, 471 (1988); M. Walton and B. D. Anderson, *ibid.* **136**, 273 (1988).
25. C. R. Taylor *et al.*, *J. Exp. Biol.* **97**, 1 (1982); R. J. Full, in *Energy Transformations in Cells and Organisms*, W. Weiser and E. Gnaiger, Eds. (Thieme Verlag, New York, 1989), pp. 175–182.
26. H. B. John-Alder, T. Garland, A. F. Bennett, *Physiol. Zool.* **59**, 523 (1986).
27. R. Lande, *Evolution* **32**, 73 (1978).
28. Financial support was provided by NSF grants BSR 8600066 and DCB 8812028 to A.F.B. and NSF grant BNS 8919497 to B.C.J. We are grateful to J. Davis for technical assistance and to B. Full for thoughtful comments on an early draft of this report.

16 January 1990; accepted 27 April 1990

Design, Activity, and 2.8 Å Crystal Structure of a C_2 Symmetric Inhibitor Complexed to HIV-1 Protease

JOHN ERICKSON,* DAVID J. NEIDHART, JOHN VANDRIE, DALE J. KEMPF, XIU CHUN WANG, DANIEL W. NORBECK, JACOB J. PLATTNER, JUDITH W. RITTENHOUSE, MARY TURON, NORMAN WIDEBURG, WILLIAM E. KOHLBRENNER, ROBERT SIMMER, ROSALIND HELFRICH, DEBORAH A. PAUL, MARK KNIGGE

A two-fold (C_2) symmetric inhibitor of the protease of human immunodeficiency virus type–1 (HIV-1) has been designed on the basis of the three-dimensional symmetry of the enzyme active site. The symmetric molecule inhibited both protease activity and acute HIV-1 infection in vitro, was at least 10,000-fold more potent against HIV-1 protease than against related enzymes, and appeared to be stable to degradative enzymes. The 2.8 angstrom crystal structure of the inhibitor-enzyme complex demonstrated that the inhibitor binds to the enzyme in a highly symmetric fashion.

HUMAN IMMUNODEFICIENCY VIRUS type–1 (HIV-1) the causative agent of acquired immunodeficiency syndrome (AIDS) (*1*), is a member of the retrovirus family (*2*). The *gag* and *pol* genes of HIV-1 encode the viral structural and replicative enzymes that are translated as polyprotein precursors: Pr55*gag* and the ribosomal frameshift product Pr160*gag-pol* (*3*). The polyproteins are proteolytically processed by the action of a virus-encoded protease (*4*). The activity of the protease is essential for the proper assembly and maturation of fully infectious virions for HIV-1 (*5*) as well as for other retroviruses (*6*). Thus, the HIV-1 protease has become an important target for the design of antiviral agents for AIDS.

Retroviral proteases were tentatively assigned to the aspartic proteinase family on the basis of putative active site sequence homology (*7*), but are only about one-third the size of the two-domain, cellular enzymes (*8*). For this reason, the retroviral proteases were hypothesized to function as dimers in which each monomer contributes one of the two conserved aspartates to the active site (*9*). This hypothesis was verified by the crystal structure determinations of Rous sarcoma virus (RSV) protease (*10*) and recombinant (*11*) and chemically synthesized (*12*) HIV-1 protease. Furthermore, these results firmly established the structural relatedness of the retroviral and cellular enzymes. Both viral enzyme structures are highly twofold symmetric; in the case of HIV-1 protease, the dimer exhibits exact crystallographic, twofold rotational (C_2) symmetry. As predicted, the structural similarity between these enzymes is strongest in the active site region. The cellular proteases contain an extended β-hairpin structure, or so-called flap (because of its flexibility), that tightly embraces the substrate in the active site (*13*). The retroviral proteases contain an analogous region that is disordered in the crystals of RSV protease (*10*). The flap is well ordered in the native HIV-1 protease structure, but crystal packing forces maintain it in a conformation that makes it unavailable for substrate binding (*11, 12*). The crystal structure of HIV-1 protease complexed with a reduced peptide inhibitor has been determined (*14*). The flap has undergone a major structural rearrangement in the complex to make favorable van der Waals and hydrogen-bonding interactions with the inhibitor.

Strategies that have been developed for the design of inhibitors for renin (*15*), an aspartic proteinase that is an important target for the design of antihypertensive agents, are now being applied to the design of inhibitors for HIV-1 protease. Current drug discovery approaches are based on the screening of renin inhibitors against HIV-1 protease (*16*) and on the synthesis of peptide substrate analogs in which the scissile P1-P1′ amide bond has been replaced by a nonhydrolyzable isostere with tetrahedral geometry (*17*). Two hydroxyethylene-containing substrate analogs have been reported to inhibit processing of HIV-1 polyproteins and to inhibit virus infection in tissue culture (*18*). However, the development of peptide-based inhibitors into effective drugs has been hampered by the inherently poor pharmacologic properties of peptides and peptide-like pharmacophores: for example, poor oral absorption, poor stability, and rapid metabolism (*19*). For this reason, we

J. Erickson and D. J. Neidhart, Department of Computer-Assisted Molecular Design, Abbott Laboratories, Abbott Park, IL 60064.
D. J. Kempf, X. C. Wang, D. W. Norbeck, J. J. Plattner, J. W. Rittenhouse, M. Turon, N. Wideburg, W. E. Kohlbrenner, Department of Anti-Infectives Research, Abbott Laboratories, Abbott Park, IL 60064.
R. Simmer and R. Helfrich, Department of Molecular and Cellular Biology, Abbott Laboratories, Abbott Park, IL 60064.
D. A. Paul and M. Knigge, Department of Hepatitis/AIDS Research, Abbott Laboratories, North Chicago, IL 60064.
J. VanDrie, Meridien Research, Box 3429, Oak Park, IL 60303.

*To whom correspondence should be addressed.

set out to design a novel, structurally symmetric class of HIV-1 protease inhibitors that would exhibit minimal peptidic character.

The concept that led to our initial design of symmetric inhibitors for HIV-1 protease was the hypothesis, based on sequence homology and structure prediction studies (*7, 9*), that retroviral proteases are symmetric homodimers. This hypothesis had two major consequences for inhibitor design. First, the structure of the enzyme active site should exhibit C_2 symmetry, as is only approximately the case for the cellular aspartic proteinases (*8*). Second, the structures of the S1 and S1′ subsites should be indistinguishable in the unliganded retroviral enzyme in contrast to the case for the cellular enzymes, which exhibit little sequence homology in these regions. Therefore, we directed our design efforts toward inhibitors that would embody the predicted symmetry characteristics of the enzyme active site.

The design of a symmetric HIV-1 protease inhibitor must satisfy two major constraints. First, for a productive and symmetric interaction between the inhibitor and enzyme to occur, their C_2 axes should coincide. Second, the inhibitor should be able to fill the enzyme subsites that normally interact with the side chains of an asymmetric peptide substrate. When we initiated this study the HIV-1 protease structure was unavailable. However, the crystal structure determination of the RSV protease (*10*) allowed us to perform a modeling experiment to verify that a C_2 symmetric molecule could indeed conform to the active site of this enzyme (which we predicted would be similar to that of the HIV-1 enzyme) (Fig. 1). In the modeling experiment, the structure of the reduced peptide inhibitor D-His-Pro-Phe-His-Phe-ψ[CH_2-NH]-Phe-Val-Tyr, derived from the x-ray crystal structure of a rhizopuspepsin-inhibitor crystal complex (*20*), was "docked" into the active site of RSV protease by superposing the active sites of the two enzymes. Next, the P′ portion of the reduced peptide was truncated after the reduced methylene carbon of the P1 Phe. Lastly, we operated on the remaining structure of the inhibitor by the RSV protease dyad axis. This generated an artificial, twofold symmetric molecule that had nearly equivalent interactions with respect to the enzyme subsites. The symmetry operation nearly superimposed the terminal, tetrahedral carbons of the two halves of the inhibitor that were located close to the symmetry axis. On the basis of the central "core" structure of the modeled inhibitor, and on the fact that aromatic amino acid side chains are prevalent in the P1 position of naturally occurring substrates for HIV-1 protease (*4*), the prototype compound, A-74702, was synthesized (Fig. 2). This mole-

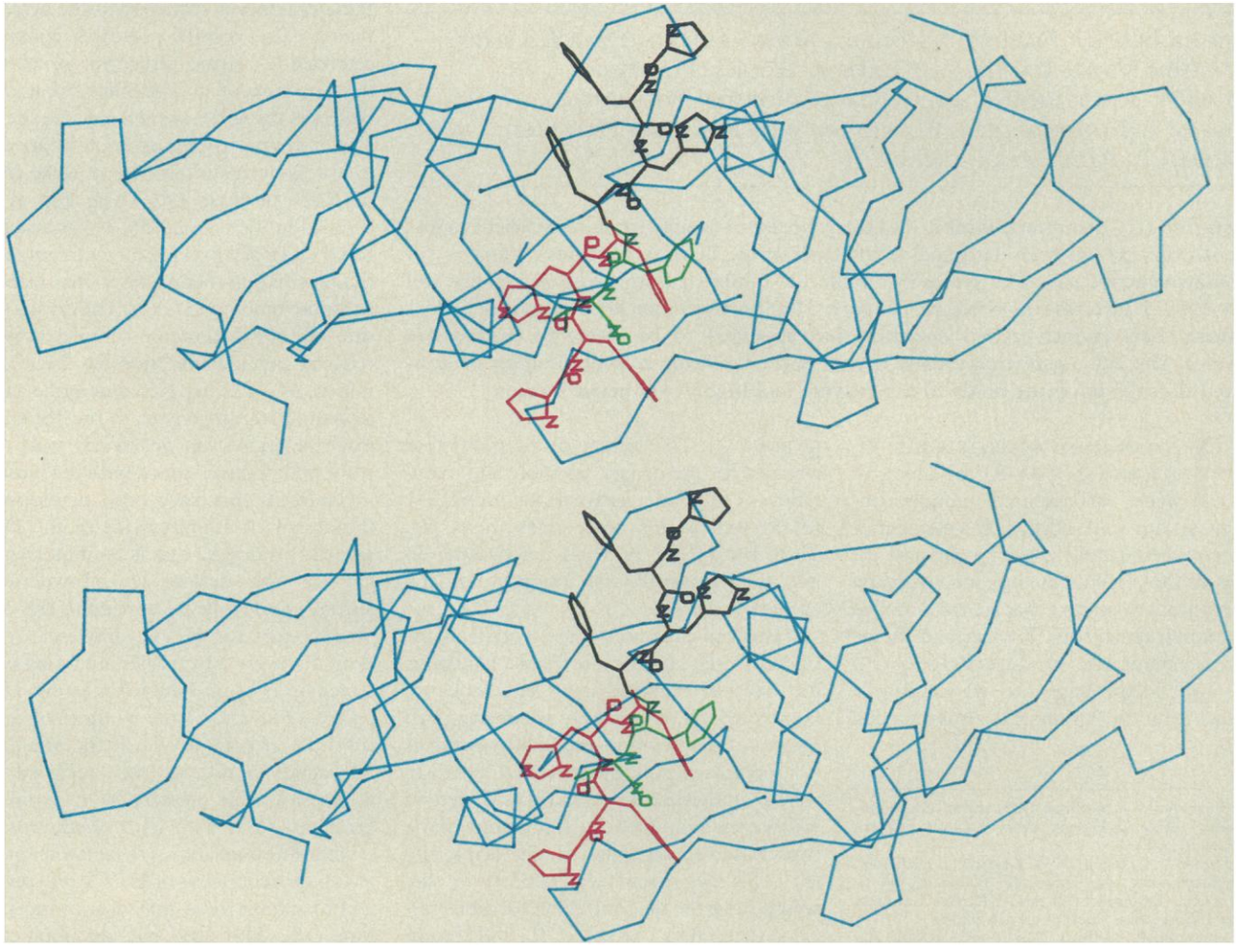

Fig. 1. Modeled structures of an asymmetric and an artificial symmetric inhibitor docked into the active site region of RSV protease. The RSV protease Cα backbone tracing (blue) is based on coordinates from Brookhaven file 2RSP (*10*). Coordinates of the reduced peptide inhibitor (green) were taken from Brookhaven file 3APR (*20*). An artificial, C_2 symmetric inhibitor (red) was produced by operating on the NH_2-terminal portion of the reduced peptide (which was truncated after the reduced methylene carbon atom) by the enzyme dyad. The approximate twofold symmetry axis (κ = 178.0°) for RSV protease was determined on the basis of least-squares superposition analysis of the two subunits (0.41 Å rms for 113 Cα pairs). The left halves of both molecules appear black because of the exact superposition of these regions; the peptide orientations of the right halves of the molecules are opposite to each other and are easily distinguished.

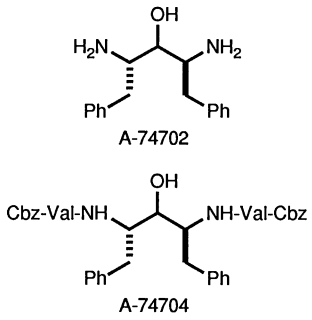

Fig. 2. Structures of C_2 symmetric inhibitors for HIV-1 protease. The abbreviations used were as follows: Ph, phenyl; Val, valine; and Cbz, carbobenzyloxy. Details of the chemical synthesis of these compounds will be presented elsewhere (*28*).

cule is essentially C_2 symmetric except for the secondary OH group on the central carbon atom. A-74702 was a weak inhibitor of HIV-1 protease [median inhibitory concentration (IC_{50}) > 200 μM] and did not exhibit significant anti-HIV activity in vitro as measured by HIV antigen production in H9 cells.

The next stage in our symmetric inhibitor design entailed increasing the binding potency to HIV protease. The usefulness of our modeled RSV protease-inhibitor complex was limited by the absence of an ordered flap. However, careful examination of the substrate binding site region of this structure indicated that only the P1 and P2 substituents would likely be buried in subsites in the enzyme. The P3 residues were likely to be exposed, because the retroviral protease is narrower across its substrate binding site than the corresponding region for the larger cellular enzymes. These considerations led us to extend our inhibitor structure by the symmetric addition of NH_2-blocked amino acids. These efforts culminated in A-74704, which had a carbobenzyloxy-valine (Cbz-Val) attached to both ends of the core (Fig. 2).

The inhibitory activity of A-74704 against HIV-1 protease was measured by the fluorogenic assay reported previously (*21*). A Dixon plot analysis indicated that A-74704 is a competitive inhibitor with an inhibition constant (K_i) of 4.5 nM (Fig. 3A). In contrast, A-74704 had no significant inhibitory activity against either avian myeloblastosis virus protease, human renin, or porcine pepsin (0% inhibition at 10 μM). Thus, the selectivity of A-74704 for HIV-1 protease over cellular aspartic proteinases can be estimated to be at least 10^4. These results show that symmetric compounds can be potent and selective inhibitors of HIV-1 protease.

The in vitro anti-HIV activity of A-74704 was assessed in H9 cells by monitoring the inhibition of viral core antigen in the supernatant (Fig. 3B). The IC_{50} value for this compound was 0.4 μM when a virus inoculum of 200 median tissue culture infectious dose ($TCID_{50}$) units was used. A-74704 inhibited the processing of Pr55*gag* in HIV-1–infected cells as measured by radioimmunoprecipitation of *gag* products (*22*). The in vitro cytotoxic effect (median lethal dose, LD_{50}) of A-74704 was assessed by trypan blue uptake in uninfected H9 cells and found to be ~200 μM. A-74704 was completely resistant to proteolytic degradation as measured by analytical high-performance liquid chromatography (HPLC) after a 3-hour incubation at 37°C with a rabbit renal cortex homogenate. (Many peptide analogs are typically 100% digested within a 10-min incubation period under similar conditions). The favorable therapeutic index of this compound (ratio of LD_{50}:IC_{50} = 500) combined with the marked stability of this compound toward proteolysis argues favorably for the use of symmetry in the targeted design of nontoxic HIV-1 protease inhibitors.

Recombinant HIV-1 protease and A-74704 were cocrystallized under conditions similar to those described for the crystallization of synthetic HIV-1 protease with MVT-101, an asymmetric, reduced peptide inhibitor (*14*). That we obtained crystals under these conditions was fortuitous, because the amino acid sequence of our enzyme differs in six positions from that of the synthetic enzyme (*23*) and our inhibitor is C_2 symmetric. However, our crystals grew in the hexagonal space group $P6_1$ (pseudo $P6_122$) instead of in the $P2_12_12_1$ orthorhombic crystal form described for the MVT-101 complex (*14*). The A-74704 complex crystals exhibited strong twofold crystallographic symmetry, and we could not unambiguously distinguish between the space groups $P6_122$ and $P6_1$ from the intensity data alone. This implied that the complex must exhibit nearly exact crystallographic twofold symmetry and that we need only solve the structure of a monomer and half of the inhibitor. However, we chose to solve and refine the structure of the complete dimer in the lower symmetry space group, $P6_1$, for several reasons: (i) a simulated *hk0* precession photograph, which was constructed with area detector intensity data, hinted at the breakdown of crystallographic twofold symmetry beyond about 3.5 Å resolution; (ii) solving the structure of the complete complex provided an independent assessment of the fidelity of the intramolecular symmetry, which is noncrystallographic and therefore not a constraint in space group $P6_1$; and (iii) refinement of the inhibitor structure was facilitated by the

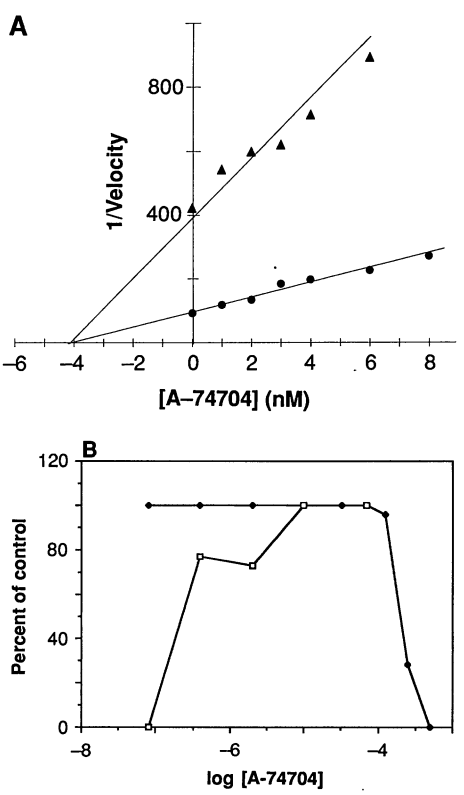

Fig. 3. Biological activity of A-74704. (**A**) Dixon plot of A-74704 inhibition of HIV-1 protease. Enzyme was assayed at the indicated inhibitor concentrations with the fluorogenic substrate Dabcyl-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln-EDANS at 0.6 μM (▲) and 2.6 μM (●). Assay conditions were as described previously (*21*). Substrate levels were well below the Michaelis constant (K_m = 100 μM); thus, the point of intersection of the lines is expected to be close to the *x*-axis for a competitive inhibitor. (**B**) Inhibition of HIV-1 infection as monitored by p24 core antigen production. A mixture of 0.1 ml of H9 cells [4×10^6 cells per milliliter in RPMI-1640 growth medium containing 10 mM Hepes buffer, 1 mM minimum essential medium (MEM) sodium pyruvate, 2 mM L-glutamine, penicillin (50 U/ml), streptomycin (50 μg/ml), and 10% fetal calf serum] and 0.1 ml of HIV-1, strain III_B (100 $TCID_{50}$ units) was incubated on a shaker for 2 hr at 37°C. The resulting culture was washed three times to remove residual virus and then resuspended in fresh medium containing A-74704 [diluted from a dimethyl sulfoxide (DMSO) stock solution to give 0.5% final concentration] at the drug concentrations indicated. Cells were then transferred to 24-well tissue culture plates and grown at 37°C under 5% CO_2. A control culture was treated in an identical manner except that DMSO without inhibitor was added to the resuspending medium. Cells were fed at day 4 with 1.0 ml of fresh media with or without inhibitor. Portions of culture supernatants were removed after 7 days and assayed for p24 antigen levels with the Abbott HIV-1 antigen enzyme immunoassay (*29*). Cell viability was determined by trypan blue dye exclusion. Percent inhibition of HIV by A-74704 was determined as the ratio of antigen in supernatants of drug-treated versus control infected cells. Anti-HIV activity is expressed as the concentration (IC_{50}) of compound that gives 50% inhibition of antigen production.

explicit treatment of the entire inhibitor in the lower symmetry space group.

The crystal structure of the protease–A-74704 complex was solved by molecular replacement with the coordinates for the HIV-1 protease derived from the MVT-101 complex (*14*) as a starting model (*24, 25*). The structure was refined at 2.8 Å to an *R* factor of 17.7%. Overall, the structure of the HIV-1 protease dimer in the A-74704 complex was very similar to that of the MVT-101 complex: the root-mean-square (rms) deviation of 198 Cα pairs for the two structures was 0.65 Å. The refined structure of the A-74704–protease complex, in which all atoms of the dimer were refined independently, exhibited a nearly exact molecular twofold axis ($\kappa = 179.9°$), which structurally superposed the 99 Cα atom pairs for both subunits to within 0.42 Å rms. The position and orientation of this twofold axis is nearly exactly coincident with the [010] crystallographic twofold axis ($\psi = 90.0°$, $\phi = 30.0°$) of the related $P6_122$ unit cell.

The fit of the inhibitor to the electron density map was excellent for the final refined structure (Fig. 4A). The overall symmetry of bound A-74704 was evident from the initial $2F_o$-F_c electron density map calculated after the first cycle of energy-refinement ($R = 22.8\%$). This map, which had no phasing contribution from the inhibitor, clearly revealed the structure of bound inhibitor. A buried water molecule in the active site could easily be seen (>3.0 σ peak) in the F_o-F_c difference electron density map calculated after the second round of energy refinement of the complex ($R = 18.2\%$) (Fig. 4B). In addition, difference electron density was evident for the side chain of Leu^{63}; this residue is a Pro in the synthetic enzyme (*23*).

The overall structure of the complex, viewed down the molecular dyad, illustrates the stairstep-like conformation of the inhibitor (Fig. 5A). This conformation is a consequence of the inhibitor's attempt to conform to the two rigid ψ loop structures, which form the base of the active site, by making hydrogen bonds to the polypeptide backbone of the protein (Fig. 5B). The ψ loops, which contain the active site aspartates, are highly conserved among both the retroviral and cellular aspartic proteinases (*26*). The opposite side of the binding site is formed by the flaps that make hydrogen-bonding interactions to the inhibitor both directly and indirectly through the buried water molecule. This water lies within about 0.2 Å of the molecular twofold axis and is tetrahedrally coordinated to the amide NH atoms of Ile^{50} and $Ile^{50'}$ of the flaps and to the carbonyl oxygens of the P2 valyl groups of A-74704. The other atoms of the protease that are hydrogen-bonded to A-74704 include Asp^{29} NH and $Asp^{29'}$ NH, Gly^{27} CO and $Gly^{27'}$ CO, Gly^{48} CO and $Gly^{48'}$ CO, and the catalytic carboxylates from Asp^{25} and $Asp^{25'}$. Except for the latter, these same groups are hydrogen-bonded to MVT-101 (*14*) despite the dissimilarity of the two inhibitor structures (Fig. 5A). In A-74704, the amide bonds have opposite orientations in the two halves of the molecule. This results in there being the same number of bonds (five), in the two inhibitors, separating the carbonyl groups that coordinate with the buried water molecule; however, there are only three bonds (versus five in MVT-101) between the P1 and P1′ NH groups. Nonetheless, the shorter spacing still allows A-74704 to make weak hydrogen bonds with the carbonyl groups of Gly^{27} and $Gly^{27'}$. These results demonstrate the importance, for inhibitor design, of satisfying the hydrogen-bonding potential of the peptide backbone in the active site region. In contrast to the conserved backbone interactions, the carboxylates of both Asp^{29} and $Asp^{29'}$ hydrogen bond to the NH_2- and COOH-terminal groups of MVT-101 (*14*), but make no direct hydrogen bonds with A-74704. These residues are involved in salt bridges in both the native enzyme and

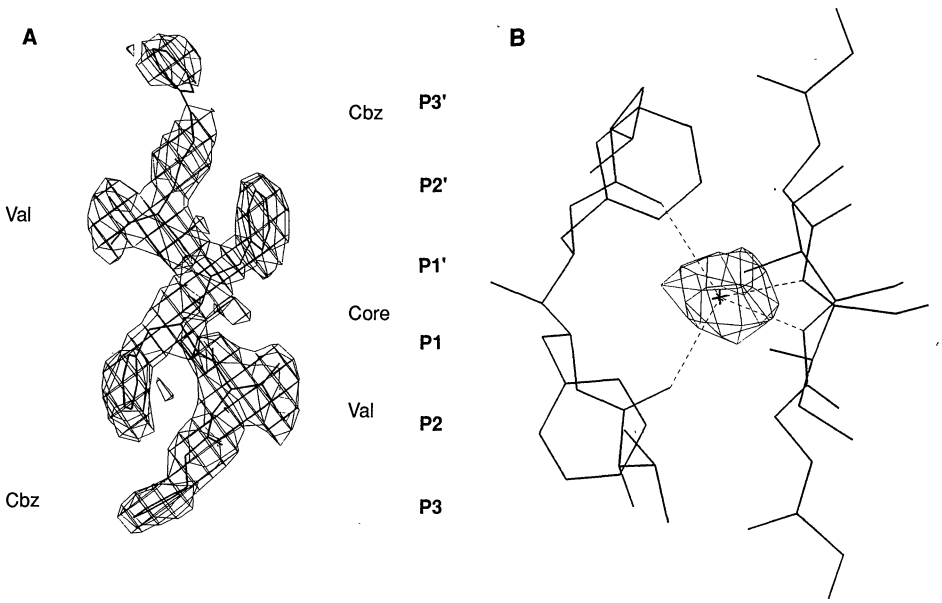

Fig. 4. Views of the electron density and refined structures of the inhibitor and buried water molecule in the active site of HIV-1 protease. (**A**) A $2F_o$-F_c, α_c electron density map is shown contoured at 1.0 σ. Phases were calculated from the refined structure of the complex. The inhibitor is drawn in thick lines and residue labels and position numbers (*30*) are in bold lettering. The P and P′ designations in this case are arbitrary since A-74704 is symmetric about the central OH group. (**B**) A F_o-F_c difference electron density map of the active site region is shown contoured at 3.0 σ. This map shows the presence of a water molecule tetrahedrally coordinated through hydrogen bonds (dashed lines) to carbonyl groups of the inhibitor (left) and amide groups of the flaps (right).

Table 1. Interactions between HIV-1 protease and A-74704. Shown are residues that form the binding subsites for A-74704 in the HIV-1 protease–A-74704 complex. Subsite residues are defined as those having atoms that lie within a 4.2 Å radius of any atom on the designated group for the inhibitor (*27*). The structure of the inhibitor core is illustrated in Fig. 2.

Enzyme subsite	Inhibitor/ group	Enzyme residue
S3	Cbz	Gly^{27}, Ala^{28}, Asp^{29}, Asp^{30}, Gly^{48}, Met^{46}, Ile^{47}
S3′	Cbz′	$Gly^{27'}$, $Ala^{28'}$, $Asp^{29'}$, $Asp^{30'}$, $Gly^{48'}$, Arg^{8}
S2	Val	Ala^{28}, Val^{32}, Ile^{47}, Gly^{48}, Gly^{49}, Ile^{50}, Ile^{84}
S2′	Val′	$Ala^{28'}$, $Val^{32'}$, $Ile^{47'}$, $Gly^{48'}$, $Gly^{49'}$, $Ile^{50'}$, Ile^{84}, $Asp^{29'}$, Ile^{50}
S1	Core	Leu^{23}, Asp^{25}, Gly^{27}, Ala^{28}, Gly^{49}, Ile^{50}, Val^{82}, Ile^{84}
S1′	Core	$Leu^{23'}$, $Asp^{25'}$, $Gly^{27'}$, $Ala^{28'}$, $Gly^{49'}$, $Ile^{50'}$, $Val^{82'}$, $Ile^{84'}$, $Pro^{81'}$, Arg^{8}
	Central OH	Asp^{25}, Gly^{27}, Ala^{28} $Asp^{25'}$, $Gly^{27'}$
	Buried H_2O	Gly^{49}, Ile^{50}, $Gly^{49'}$, $Ile^{50'}$ Val, Val′, Core (inhibitor groups)

the A-74704 complex, and evidently little interaction energy is gained by forming competing hydrogen bonds.

Examination of the solvent-accessible surface of the enzyme revealed that the binding site forms a pocketed tunnel where the pockets serve as subsites. The S1 and S2 pockets, which bind the benzyl and isopropyl side chains, respectively, are mainly hydrophobic and do not provide obvious opportunities for enhancing inhibitor solubility. In contrast, the S3 binding pockets are less defined, and the benzyl moieties of the Cbz groups protrude through the tunnel openings to the exterior surface of the enzyme. The enzyme subsites are each composed of six to nine residues (Table 1), but a total of only 29 residues (14 from one subunit, 15 from the other) interact with the inhibitor. Thus, many residues participate in more than one subsite. This situation differs from that with the MVT-101 complex, in which only 21 residues interact with the inhibitor (which is approximately equal in size to A-74704) in subsites that are composed of between three and seven residues (*14*). The increased number of contacts between A-74704 and the protease, including the added hydrogen-bonding capacity of the alcohol group on the inhibitor, may help to explain its lower K_i (4.5 nM) compared with that for MVT-101 (700 nM), although different conditions were used to perform the enzyme inhibition assays.

The global deviations from symmetry produced by the inhibitor in the MVT-101 complex were not as evident in the A-74704 structure. The twofold related, intrasubunit ionic interactions between Arg^{8}, Arg^{87}, and Asp^{29} observed in the native enzyme are preserved in the A-74704–protease structure. One of these interactions was disrupted in the MVT-101 complex by the presence of an Arg in the P3′ position, which forms two hydrogen bonds to $Asp^{29'}$. Additional regions of asymmetry in the MVT-101 complex were centered around residues 16, 39, 50, 54, 79, and 94 (rms deviation of 1.6 Å for these six Cα atom pairs within the superposed subunits). Residues 50 and 79 are adjacent to the S2 subsite. These regions of the structure all became more symmetric during the refinement of the A-74704 complex (rms deviation of 0.6 Å for the same residues). These results reflect a more symmetric binding mode for the pseudo-C_2 symmetric inhibitor. Analysis of the symmetry of the inhibitor structure revealed that all but the terminal Cbz benzyl groups, for a total of 20 nonhydrogen atom pairs, could be superposed to within 0.36 Å rms by an approximate dyad (κ = 177.9°) that lies within a few tenths of an angstrom from the position of the enzyme twofold axis. However, the orientation of the inhibitor twofold axis was inclined by about 6° relative to the enzyme axis. This is probably a consequence of the asymmetry of the inhibitor engendered by the central OH group.

Local deviations from symmetry were observed at the interfaces between dimers in the crystal and involved different rotamers for Phe^{53} and $Phe^{53'}$ in the flap and for the two Cbz benzyl moieties in the inhibitor. These distortions may be a result of an intrinsic asymmetry of the dimer in the bound state or of crystal packing forces, or they may be a subtle consequence of the

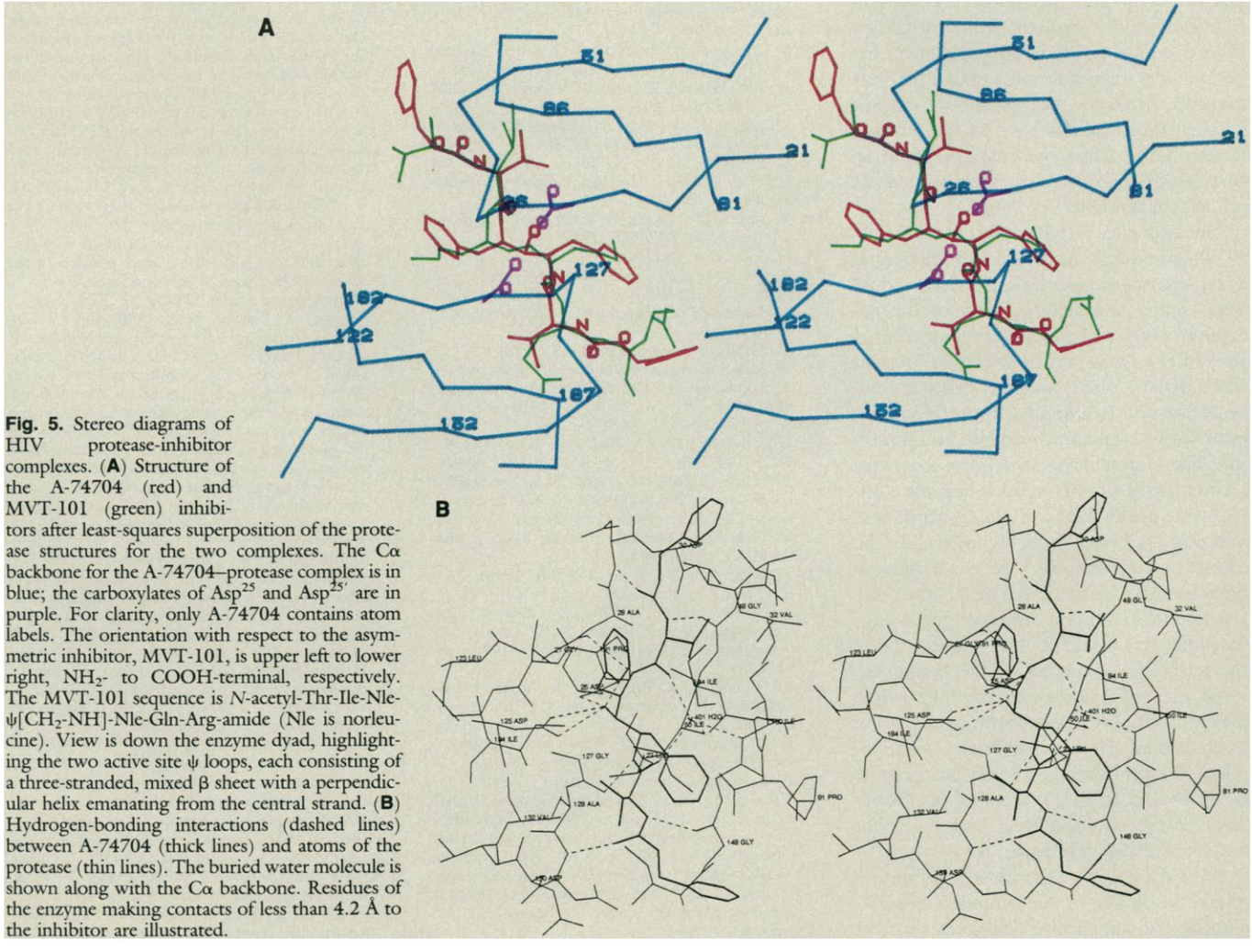

Fig. 5. Stereo diagrams of HIV protease-inhibitor complexes. (**A**) Structure of the A-74704 (red) and MVT-101 (green) inhibitors after least-squares superposition of the protease structures for the two complexes. The Cα backbone for the A-74704–protease complex is in blue; the carboxylates of Asp^{25} and $Asp^{25'}$ are in purple. For clarity, only A-74704 contains atom labels. The orientation with respect to the asymmetric inhibitor, MVT-101, is upper left to lower right, NH_2- to COOH-terminal, respectively. The MVT-101 sequence is *N*-acetyl-Thr-Ile-Nle-ψ[CH_2-NH]-Nle-Gln-Arg-amide (Nle is norleucine). View is down the enzyme dyad, highlighting the two active site ψ loops, each consisting of a three-stranded, mixed β sheet with a perpendicular helix emanating from the central strand. (**B**) Hydrogen-bonding interactions (dashed lines) between A-74704 (thick lines) and atoms of the protease (thin lines). The buried water molecule is shown along with the Cα backbone. Residues of the enzyme making contacts of less than 4.2 Å to the inhibitor are illustrated.

slightly asymmetric position of the central OH group of A-74704 with respect to the two active-site carboxylates. This group lies at an average distance of 2.8 and 2.5 Å from the carboxylate oxygens of Asp^{25} and $Asp^{25'}$, respectively. The slight asymmetry in inhibitor binding might be expected to give rise to two binding orientations that could be averaged out over the crystal structure. However, we did not detect any evidence in the 2.8 Å electron density map for such statistical disorder. The relatively minor asymmetries in the structure of the A-74704 complex accounts for the strong pseudo-$P6_122$ symmetry of these crystals.

A major element of asymmetry in the MVT-101 complex was the hydrogen bond between Gly^{51} NH and $Ile^{50'}$ CO in the flap. The peptide conformations for the two monomers differed for these residues and thus did not allow for the formation of the symmetric hydrogen bond. In the structure of the A-74704 complex, the peptide conformations of residues 50 and 51 were equivalent. This resulted in the absence of any hydrogen bonds between the two flaps, which appear to be stabilized by intramolecular van der Waals interactions and by multiple interactions with the inhibitor. However, in model building and refinement experiments we could not unambiguously distinguish between the alternative peptide conformations on the basis of the electron density alone. Thus, the final interpretation of this region of the structure must await refinement at higher resolution.

The biological activity and structure of A-74704 confirmed our initial hypothesis that a symmetric, nonpeptidic inhibitor could bind tightly and symmetrically to the enzyme and display antiviral activity. Comparison of the structures of the A-74704 and MVT-101 complexes suggests that the enzyme attempts to enforce symmetric binding even with a structurally asymmetric inhibitor. The largest deviations from symmetry seen in the MVT-101 peptide-enzyme complex were not evident with the C_2 symmetric inhibitor, and were probably induced in the former by unfavorable side chain interactions and, to a lesser extent, by an inherently asymmetric backbone structure. This suggests that symmetric inhibitors that embody the active site symmetry of the bound enzyme may offer enhanced binding, while they also simplify the problem of structural optimization. There have been several reports of tight-binding, asymmetric peptide inhibitors of HIV-1 protease (*17*). Structures of the corresponding complexes are not yet available. However, we may speculate that the ability of asymmetric inhibitors to take advantage of the symmetric binding tendency of the enzyme will likely depend on two factors: the flexibility of the inhibitor, which will determine the ease with which the inhibitor may conform to the active site symmetry, and the degree to which the inhibitor is complementary to the enzyme subsite pockets. Of course, the hydrogen-bonding potential of the active site tunnel should be satisfied with any good ligand. The existence of a buried, fully hydrogen-bonded water in the two different inhibitor complexes suggests that this solvent molecule has a role in the energetics of inhibitor binding. Consequently, the functional replacement of this water affords a logical and challenging target for future inhibitor design.

REFERENCES AND NOTES

1. R. C. Gallo and L. Montagnier, *Sci. Am.* **259**, 40 (October 1988).
2. H. Varmus and R. Swanstrom, in *RNA Tumor Viruses: Molecular Biology of Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1985), vol. 1, pp. 369–512.
3. T. Jacks *et al.*, *Nature* **331**, 280 (1988).
4. L. E. Henderson *et al.*, *J. Virol.* **62**, 2587 (1988).
5. N. E. Kohl *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4686 (1988); C. Peng, B. Ho, T. Chang, N. Chang, *J. Virol.* **63**, 2550 (1989); H. Gottlinger, J. Sodroski, W. Haseltine, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5781 (1989).
6. S. Crawford and S. P. Goff, *J. Virol.* **53**, 899 (1985); I. Katoh *et al.*, *Virology* **145**, 280 (1985).
7. H. Toh, M. Ono, K. Saigo, T. Miyata, *Nature* **315**, 691 (1985).
8. J. Tang, M. N. G. James, I. N. Hsu, J. A. Jenkins, T. L. Blundell, *ibid.* **271**, 618 (1978).
9. L. N. Pearl and W. R. Taylor, *ibid.* **329**, 351 (1987); I. Katoh, T. Yasunaga, Y. Ikawa, Y. Yoshinaka, *ibid.*, p. 654.
10. M. Miller, M. Jaskolski, J. K. M. Rao, J. Leis, A. Wlodawer, *ibid.* **337**, 576 (1989).
11. M. Navia *et al.*, *ibid.*, p. 615.
12. A. Wlodawer *et al.*, *Science* **245**, 616 (1989).
13. M. N. G. James, A. Sielecki, F. Salituro, D. H. Rich, T. Hofmann, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6137 (1982).
14. M. Miller *et al.*, *Science* **246**, 1149 (1989).
15. M. Szelke, in *Aspartic Proteinases and Their Inhibitors*, V. Kostka, Ed. (de Gruyter, New York, 1985), p. 421.
16. A. D. Richards *et al.*, *FEBS Lett.* **247**, 113 (1989).
17. G. B. Dreyer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9752 (1989); M. L. Moore *et al.*, *Biophys. Biochem. Res. Commun.* **159**, 420 (1989); A. G. Tomasselli *et al.*, *Biochemistry* **29**, 264 (1990); N. A. Roberts *et al.*, *Science* **248**, 358 (1990); D. H. Rich, J. Green, M. V. Toth, G. R. Marshall, S. B. H. Kent, *J. Med. Chem.* **33**, 1285 (1990).
18. T. D. Meek *et al.*, *Nature* **343**, 90 (1990); T. J. McQuade *et al.*, *Science* **247**, 454 (1990).
19. J. J. Plattner and D. W. Norbeck, in *Drug Discovery Technologies*, C. R. Clark and W. H. Moos, Eds. (Ellis Horwood, Chichester, England, 1990), pp. 92–126.
20. K. Suguna, E. A. Padlan, C. W. Smith, W. D. Carlson, D. R. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7009 (1987).
21. E. D. Matayoshi, G. T. Wang, G. A. Krafft, J. Erickson, *Science* **247**, 954 (1990).
22. D. A. Paul *et al.*, unpublished data.
23. The sequence of our enzyme (based on the BH102 strain of HIV-1) differs in four positions [at residues 14 (Lys versus Arg), 40 (Arg versus Lys), 62 (Leu versus Pro), and 63 (Ile versus Val)] from the natural sequence of the protease for the SF2 strain. In addition, Cys^{67} and Cys^{95} were replaced by α-amino-*n*-butyric acid residues in the synthetic SF2 protease.
24. A synthetic gene for HIV-1 protease (HIV-1 BH102 sequence) was cloned and expressed in *Escherichia coli* and the mature enzyme was purified from the soluble fraction of cell extracts (J. W. Rittenhouse *et al.*, in preparation). The enzyme was concentrated to 2.0 mg/ml in sodium acetate buffer, pH 5.4. For crystallization, a stock solution of 10 mg/ml A-74704 in DMSO was added to the enzyme to give a final ratio of inhibitor:enzyme of 10:1 in 10% DMSO. Crystals were grown at room temperature in hanging drops by vapor diffusion against an unbuffered, 30% saturated ammonium sulfate solution. Droplets were prepared by diluting a small volume of protein solution with an equal volume of 30% saturated ammonium sulfate solution. Crystals appeared in 7 to 14 days as thin rods with maximal dimensions 0.1 mm by 0.1 mm by 1.0 mm. Diffraction was observed to at least 2.5 Å, and the crystals belonged to the hexagonal space group $P6_1$, with unit cell dimensions $a = b = 63.3$ Å, $c = 83.6$ Å, and $\gamma = 120°$. Diffraction data were measured from a single crystal with a Siemens X-100A multiwire area detector with a three-axis camera. X-rays were generated by a Rigaku RU200H 12-kW rotating anode x-ray generator equipped with a 0.3 by 3.0 mm filament and operating at 50 kV and 100 mA. X-rays were monochromatized with Frank's focusing mirrors. A total scan of 90° in ω, at a fixed χ setting of 45°, was performed in a single sweep as a series of 0.25° oscillation frames. Crystal-to-detector distance was 14.0 cm at a swing angle (2θ) of 15°; exposure per frame was 300 s. A total of 12,175 raw intensities measured between 10.0 and 2.8 Å resolution were processed and scaled with the software package XENGEN [A. J. Howard, *J. Appl. Crystallogr.* **20**, 383 (1987)]; these data reduced to 3951 unique observations [amplitude $F > 2.0\ \sigma(F)$], which represent 85% of the total theoretically possible number of reflections in this resolution range. The merging *R* factor for symmetry-related reflections was 0.10 on intensities. The structure of the complex was solved by molecular replacement with the coordinates of the HIV protease dimer derived from the crystal structure of the HIV-1 protease–MVT-101 complex [(*14*); entry 4HVP in the Brookhaven Protein Data Bank)]. A Lattman cross-rotation function analysis implemented in MERLOT [P. M. D. Fitzgerald, *J. Appl. Crystallogr.* **21**, 53 (1988)] yielded the correct orientation of the dimer from the highest peak in the search (4.5 σ versus 1.0 σ for the next highest peak) with data from 10.0 to 4.0 Å. The correct position of the dimer was determined with a correlation coefficient search implemented in XPLOR [A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987)], which yielded a peak of 9.8 σ (the next highest peak was only 5.6 σ) with 6.0 to 3.0 Å data. The correct choice of space group. enantiomorph $P6_1$ (versus $P6_5$) was determined by graphical and analytical examination of the crystal packing of the dimer with the interactive docking program PROPACK, which also gave an independent solution to the translation problem (*25*). As a further check, the translation search was also performed in the $P6_5$ cell and gave a peak maximum of less than half of that obtained from the $P6_1$ search. The *R* factor for the molecular replacement solution was 0.379 for all data from 10.0 to 2.8 Å. Rigid body refinement of the two monomers in the dimer with XPLOR dropped the *R* factor to 0.323. This was followed by energy refinement of all atomic positions against 6.0 to 2.8 Å data, which further reduced the *R* factor to 0.228. Electron density maps ($2F_o$-F_c and F_o-F_c) were calculated and the inhibitor structure was readily discerned and fitted to the map with FRODO [T. A. Jones, *Methods Enzymol.* **115**, 157 (1985)]. The complex was energy-refined further to an *R* value of 0.182. An additional cycle of map fitting, with the addition of a single water molecule in the active site, followed by energy-refinement and a cycle of simulated annealing refinement at 4000° with a slow cooling protocol (A. Brunger, A. Krukowski, J. Erickson, *Acta Crystallogr.*, in press) resulted in a structure that gave a final *R* factor of 0.177 with an overall *B* factor of 12.0 $Å^2$, and overall deviations from ideal geometry of 0.020 Å for bond distances and 4.0° for bond angles. Coordinates will be deposited with the Brookhaven Protein Data Bank.

25. J. VanDrie and J. Erickson, in preparation. This program evaluates the overlapping volumes between docked surfaces. It allows the user to view various crystal packings on an interactive graphics display. The packing is continuously varied by translating the asymmetric unit of the crystal with simultaneous adjustments of the symmetry-related molecules. Overlapping regions of the structures are highlighted, and a quantitative measure of the total overlap volume is continuously updated and displayed.
26. J. Erickson, M. J. K. Rao, C. Abad-Zapatero, A. Wlodawer, in *Viral Proteinases as Targets for Chemotherapy*, H. G. Krausslich, S. Oroszlan, E. Wimmer, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989), pp. 191–201.
27. Sali *et al.*, *EMBO J.* **8**, 2179 (1989).
28. D. J. Kempf, in preparation.
29. D. A. Paul *et al.*, *J. Med. Virol.* **22**, 357 (1987).
30. Nomenclature according to I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.* **27**, 157 (1967); P and P′ refer to residues on the NH_2- and COOH-terminal sides, respectively, of the scissile amide bond of a substrate that interacts with the corresponding S and S′ subsites in the enzyme.
31. We thank A. Wlodawer and his colleagues at the National Cancer Institute for making available to us their crystallographic coordinates of the RSV protease, synthetic HIV-1 protease, and synthetic HIV-1 protease–MVT-101 complex before publication; without their generosity this study could not have been completed. We also thank H. Stein for performing the renin assay, E. Devine for performing the stability studies, and A. Craig-Kennard for technical assistance. We are grateful to J. Greer and C. Abad-Zapatero for critically reading the manuscript. Supported by NIH grant AI 27220 to J.E.

23 April 1990; accepted 14 June 1990

Identification of a Sequence in the PEPCK Gene That Mediates a Negative Effect of Insulin on Transcription

RICHARD M. O'BRIEN, PETER C. LUCAS, CLAUDE D. FOREST, MARK A. MAGNUSON, DARYL K. GRANNER*

Phosphoenolpyruvate carboxykinase (PEPCK) governs the rate-limiting step in gluconeogenesis. Glucocorticoids and adenosine 3′,5′-monophosphate (cAMP) increase PEPCK gene transcription and gluconeogenesis, whereas insulin has the opposite effect. Insulin is dominant, since it prevents cAMP and glucocorticoid-stimulated transcription. Glucocorticoid and cAMP response elements have been located in the PEPCK gene and now a 15–base pair insulin-responsive sequence (IRS) is described. Evidence for a binding activity that recognizes this sequence is presented.

INSULIN STIMULATES THE EXPRESSION of several genes including those encoding glyceraldehyde-3-phosphate dehydrogenase, c-Fos, glucokinase, gene 33 product, and α-amylase (*1, 2*). In contrast, insulin inhibits the expression of the PEPCK, adipsin, and growth hormone genes (*3*). Whereas cAMP and glucocorticoids increase PEPCK gene transcription (*4–6*), insulin inhibits transcription of this gene in rat liver and the H4IIE rat hepatoma cell line. Inhibition by insulin predominates in the presence of cAMP and glucocorticoids. The basal promoter elements, the cAMP-responsive element (CRE), and the glucocorticoid-responsive elements (GREs) in the PEPCK gene have been mapped in an effort to understand this complex interplay (*7*). The DNA sequences responsible for the dominant, negative effect of insulin on PEPCK have remained elusive, as have the sequences responsible for the regulation of transcription of other genes by insulin. Our goal was to identify such an element.

A fusion gene (PEPCK-CAT), containing the PEPCK promoter sequence between −600 and +69, relative to the transcription start site, ligated to the chloramphenicol acetyltransferase (CAT) reporter gene, is responsive to insulin when transiently expressed in H4IIE cells. Insulin inhibited basal PEPCK-CAT expression and suppressed the stimulatory effects of cAMP and dexamethasone (dex) (*8*). Attempts to define an IRS by the use of PEPCK-CAT fusion genes containing progressively shorter PEPCK promoter segments yielded inconsistent results in the transient expression system (*9*). Because the presence of the normal chromatin environment might favorably influence the function of one or more IRSs in the intact promoter, we isolated a series of H4IIE cell lines stably transfected with various PEPCK-CAT fusion genes. Results obtained from stable cell lines made using four such constructs are shown in Fig. 1A. Although insulin generally inhibited basal expression of the chimeric gene in each case, CAT expression was stimulated with the use of dex and cAMP in order to emphasize the magnitude of the insulin effect. Insulin almost completely suppressed cAMP- plus dex-stimulated CAT expression in the HL9, HL32, and FBG32 clones, just as it inhibits transcription of the endogenous gene (*6*). However, in the HL45 clones, which contain the region between −271 and +69 of the PEPCK promoter, the effect of insulin, though still present, was clearly reduced. Two other HL45 clones showed the same partial effect of insulin (*9*). The CAT assay results were confirmed by primer extension analysis in which insulin caused 86 and 39% reductions of CAT mRNA in HL9 and HL45 cells, respectively (Fig. 1B). These results suggest that the PEPCK promoter contains at least two IRSs, one located between −468 and −271, and another or others between −271 and +69.

To further delineate the distal IRS, we made use of a vector containing the herpes simplex virus thymidine kinase (TK) promoter. A series of overlapping, double-stranded oligomers spanning the region between −453 and −271 of the PEPCK promoter were synthesized and each was ligated into the Bam HI site of TKC-VI (Fig. 2A), a vector similar to that previously used to identify the inhibitory sterol-responsive element in the low-density lipoprotein receptor gene promoter (*10*). Each oligomer contained 38 bp of PEPCK sequence with Bam HI–compatible sticky ends (GATC) at each end. Previous studies show that the insertion of a 42-bp random DNA sequence into the Bam HI linker in this TK promoter has little effect on transcription (*11*). The effect of insulin on CAT expression by these constructs was initially analyzed by the use of transient expression of the chimeric DNA in H4IIE cells (Fig. 2B). The vector lacking an insert, and most of the constructs spanning the region from position −403 to −270, showed a slight stimulation of CAT expression in response to insulin. Constructs containing either the PEPCK sequence between −453 and −415 or that between −433 and −396 showed an insulin-responsive inhibition of CAT expression (Fig. 2B). Since the −433 to −396 construct gave the insulin effect of greatest magnitude, and worked well in both orientations, the search for an IRS was directed to this region of the PEPCK gene.

H4IIE cells stably transfected with either the insulin-responsive −433 to −396 construct or, as a negative control, the −403 to −366 construct were used to confirm that this effect was not simply a function of the transient transfection system. Two cell lines containing each construct were obtained. Both −433 to −396–containing clones (O and P) showed a concentration-dependent inhibition of CAT expression in response to insulin; the median effective concentration (EC_{50}) was about 0.1 nM and maximal inhibition was obtained with 10 nM insulin (Fig. 3), concentrations that provide similar

Department of Molecular Physiology and Biophysics, Vanderbilt University Medical School, Nashville, TN 37232–0615.

*To whom correspondence should be addressed.