Title
Structure of HIV-1 and SIV proteases wih implications for the catalytic mechanism and drug resistance

Author
Rose, Robert Bennett

Publication Date
1996

Peer reviewed|Thesis/dissertation

STRUCTURES OF HIV-1 AND SIV PROTEASES WITH IMPLICATIONS FOR THE CATALYTIC MECHANISM AND DRUG RESISTANCE

by

Robert Bennett Rose

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOPHYSICS

in the

GRADUATE DIVISION

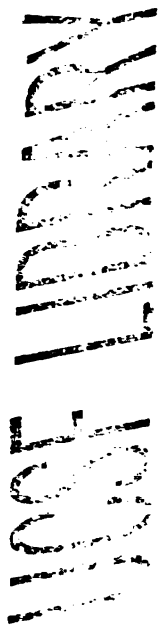
of the

UNIVERSITY OF CALIFORNIA

San Francisco

For the millions of people worldwide

living with HIV



Acknowledgements

I am periodically reminded that Science, like most human endeavors, is a communal effort. Despite the many hours spent alone focusing on a problem, some happily others in deference to the work, the contribution of the individual is only significant in the context of the whole. I have lost track of this truth many times as a graduate student, and have been reminded of it again as I've reviewed my accomplishments and failures during the writing of this thesis. This is not only true in the content of our work, but in the context. I have been privileged to work on a problem which is both important and challenging. The "sophistication" of the virus continuously astonishes and humbles me, and I hope that my work has contributed to the understanding of HIV and the eventual eradication of AIDS.

I am indebted to the many people in the Stroud laboratory who have both taught me crystallography and taught what it is to be a crystallographer, including Michael Wiener, Janet Finer-Moore, V. Ramalingan (Ramu), Sasha Kamb, Doug Freymann, Kathy Perry, Tom Stout, and Richard Morse. Sally Raguet and Larry Miercke introduced me to the art of protein purification. Larry Miercke helped get the HIV integrase project started in the Stroud lab. Earl Rutenber led me through the determination of my first structure. He has provided me with much insight, not only as a participant in the HIV project but his knowledge of crystallography, his willingness to think about problems, and his good-natured participation in all aspects of laboratory life. My fellow graduate students in the laboratory have provided both company and inspiration, particularly Bob Keenan, Paul Foster, Partho Ghosh, Stephanie Mel, Chris Schafmeister, Sherry LaPorte, Julian Chen, and Eric Fauman.

I also thank my colleagues in the Biophysics program, especially those learning crystallography with me. Julie Sohl has been a constant and generous friend. Theresa Gamble and Elizabeth Jaffe have always reminded me to listen to my own heart, and not to be afraid to grow and change.

Andy Leavett has been a wonderful collaborator on the HIV integrase project. His constant optimism has been inspiring, and his confidence in me has been much appreciated.

I am also indebted to the collaboration with Charles Craik and his laboratory. He provided guidance, advice and encouragement as well as a good environment to think about proteases. Members of his lab have contributed greatly to my and the Stroud lab's crystallographic efforts, including Fiona McPhee, Nancy Douglas, Gian-Carlo Ochoa, and Nick Endres.

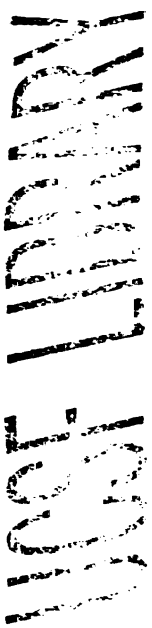
I want to thank the San Francisco AIDS Foundation for helping me to understand what it is to live with AIDS. I worked as a volunteer on the hotline for a year during graduate school.

There is a personal toll for committing to such an all-encompassing project as graduate school in ones relationship to family and friends. I am fortunate to have parents who have always supported me emotionally, and financially at times. That gift is immeasurable. To my friends, my partner Deb, and my son Asher, I hope to make up for lost time in the near future.

And of course I want to thank Bob Stroud for introducing me to the joys of crystallography and sharing some of his deep insights into the subject.

Preface

Individuals become scientists for many reasons, including ego gratification, altruism, the intellectual challenge, the social identification, and economic security. Because of economic uncertainties, the other motivations are increasingly colored by the pressure for results in research. Certainly when working in a field of research such as AIDS it is justifiable to operate with a crisis mentality, when almost 20 million people in the world are infected with the virus (ref). Public support for biomedical research is reflected in the 5% increase of the National Institute of Health budget when the mood in Congress has been to search every corner for savings to balance the budget. Justification of this funding, especially in a time of fiscal shortages, is usually in terms of short term successes. But such complicated biological problems as presented by the Human Immunodeficiency Virus (HIV) has eluded quick solutions.



A recent report by a government-appointed panel charged with reviewing AIDS research in the country criticized the organization of AIDS research as stifling of novel ideas for short term gain. As quoted by the New York Times (Thursday, March 14, 1996 p. 1): "Because of intense competition for limited money, 'many investigators are reluctant to submit novel or innovative proposals.'" The article notes that similar criticisms have been made about National Institute of Health support in other fields.

The shift in research support from government to companies has redefined what it means to be a successful research laboratory. Success is judged in terms of ties to companies as well as size of the lab. These ties come with their own demands for short-term results. The proper role of the

University, even in these difficult times, is to take a larger view of problems and, as much as possible, not be swept away by the short term pressures. For the researcher this may involve taking risks by working on projects that might not succeed, as well as working on less immediately glamorous problems demanded by rigorous and careful research. The proper balance between pragmatism and the "long term view" is difficult to achieve. I have tried in this work to keep this balance in mind.

STRUCTURES OF HIV-1 AND SIV PROTEASES WITH IMPLICATIONS FOR THE CATALYTIC MECHANISM AND DRUG RESISTANCE

by

Robert Bennett Rose

Abstract

Human Immuodeficiency Virus (HIV) protease is a major target for the development of drugs against AIDS. Although mechanism-based inhibitors have provided guides for drug design, the chemical mechanism and the determinants of substrate specificity are not yet understood. Crystal structures of product complexes were studied to elucidate the mechanism. Structures of N-terminal products complexed with HIV-1 and Simian Immunodeficiency Virus (SIV) proteases indicate that the C-terminal carboxyl group hydrogen bonds with both catalytic aspartic acids similarly. When the N- and C-terminal products were modeled into the same structure, the scissile bond could be reformed without disrupting the conformation of the products or protease significantly. Therefore substrate strain resulting from binding to the protease does not contribute to catalysis. A mechanism is proposed from modeled structures of the Michaelis complex and gemdiol intermediate based on the proper positioning of the substrate and catalytic residues. An unliganded structure of SIV protease showed rigid body rotations of domains of the protease, in agreement with an induced fit mechanism. Binding substrate stabilizes the protease in a closed conformation, positioning the catalytic aspartic acids, rotating the domains to optimize interactions with the substrate, and stabilizing the substrate in a suitable conformation for catalysis.

The development of resistance mutations by the virus is the major obstacle to designing effective anti-HIV treatments. SIV and other retroviral

protease structures indicate the resilience of the protease to mutation. Because both substrates and inhibitors bind to the protease's open conformation, the development of resistance during drug treatment of AIDS may result from mutations stabilizing the open conformation. Anti-protease drugs have targeted the closed conformation. An approach to mitigating the effects of resistance mutations is to develop covalent inhibitors of the protease that modify the catalytic aspartic acids. Crystal trials of experimental irreversible inhibitors were unsuccessful.

Finally, crystal trials of a second enzyme, HIV integrase, were unsuccessful. Binding studies indicate that integrase binds DNA non-specifically, despite its recognition of specific Long Terminal Repeat sequences of the viral chromosome.

Table of Contents

INTRODUCTION....1

References....12

CHAPTER ONE: Structure of the Protease from Simian Immunodeficiency Virus....17

Reprinted from *Biochemistry 32*, 12498-12507.

CHAPTER TWO: Irreversible inhibitors of HIV protease....29

Introduction....30

Materials and Methods....32

Results and Discussion....33

Conclusions....43

References....45

CHAPTER THREE: Three-dimensional structures of HIV-1 and SIV protease product complexes....49

Submitted for publication to *Biochemistry*

Abbreviations....51

Abstract....52

Introduction....53

Materials and Methods....55

Results....60

Discussion....69

Conclusions....76

Acknowledgements....77

References....78

Figure Legends....87

CHAPTER FOUR: Domain flexibility in retroviral proteases:
catalysis by covariant alignment....99
Submitted for publication to *Proteins*
Abbreviations....101
Abstract....102
Introduction....103
Materials and Methods....105
Results....108
Discussion....116
Acknowledgements....123
References....124
Figure Legends....133
CHAPTER FIVE: HIV integrase: solubility, crystallization trials, and
DNA binding....148
Introduction....149
Purification....149
Solubility....153
Analytical Ultracentrifugation....157
Oligonucleotides....158
Crystallization trials....160
DNA binding assays....162
Discussion....170
Acknowledgements....171
References....172

APPENDIX: Both substrate and target oligonucleotide sequences affect in vitro integration mediated by Human Immunodeficiency Virus type 1 integrase protein produced in *Saccharomyces cerevisiae*.....179
Reprinted from *J. Viol. 66*, 2359-2368.

List of Tables

exclusive of Chapter One

Chapter Two:

Table I. Statistics for the data collections of SIV protease complexed with three inhibitors: curcumin, xantho-bilirubin, and UCSF59 38

Chapter Three:

Table Ia. Statistics for HIV-1 protease product complex structures 82

Table Ib. Statistics for SIV protease product complex structures 83

Table IIa. Dimensions of HIV-1 protease product complex unit cells 84

Table IIb. Dimensions of HIV-1 protease product complex unit cells 85

Table III. Protease sidechains forming the peptide binding subsites 86

Chapter Four:

Table I. Statistics for unliganded SIV protease structure 136

Table II. HIV-1 protease structures from the Brookhaven Protein Data Bank 137

Table III. Root mean square deviations of superpositions of liganded and unliganded SIV and HIV-1 protease structures 138

Table IV. Distance between atoms of the two catalytic aspartic acids of the liganded and unliganded protease dimers 139

Table V. Movements of atoms forming the S1, S1', and S3 subsites between the open and closed SIV protease 140

List of Figures

exclusive of Chapter One

Chapter Two:

Figure 1. Derivatives of haloperidol: reversible and irreversible inhibitors of HIV protease. 47

Figure 2. Non-haloperidol inhibitors of HIV protease 48

Chapter Three:

Figure 1. 2|Fo|-|Fc| density in the binding pocket of HIV-1 protease complexed with a) Ac-S-L-N-F and b) P-I-V-NH_2 91

Figure 2. Interactions between the C-terminal carboxyl group of the N-terminal product and the catalytic aspartic acids 92

Figure 3. Tautomers of the N- and C-termini of the products interacting with the catalytic aspartic acids 93

Figure 4. Peptide products bound to the proteases compared with peptidomimetic inhibitors 94

Figure 5. 2|Fo|-|Fc| density in the binding pocket of SIV protease complexes with a) F-L-E-K and b) F(NO_2)-E-A-Nle-S 95

Figure 6. Model of the substrate from the two products 96

Figure 7. Proposal for a chemical mechanism 97

Figure 8. Model of the Michaelis complex with the catalytic water 98

Chapter Four:

Figure 1. Labeled ribbon diagram of SIV protease 141

Figure 2. Carbon alpha traces of the liganded and unliganded SIV proteases superimposed by a) dimers b) monomers c) terminal domain d) core domain 142

Figure 3. Interface between the core and terminal domains 143

Figure 4. Superposition of the carbon alpha traces of 17 HIV-1

protease structures 144

Figure 5. Axes of rotation of the core and flap domains 145

Figure 6. Superposition of the core domain of the unliganded and liganded SIV protease showing movement in the active site 146

Figure 7. Comparison of the deviation in coordinates of 18 superimposed HIV-1 structures with their average B-factors 147

INTRODUCTION

Introduction

The dream of rational drug design is to be able to devise a tight-binding inhibitor by examining the structure of a protein molecule alone. This dream is far from being realized. Our understanding of the relationship between structure and function is still in its infancy: we are no more able to divine the function of a protein from its structure than we are able to predict the fold of a protein from its amino acid sequence. The recent structure of a fragment of HIV integrase is a case in point: the structure does not identify the active site residues nor the mode of DNA binding (Dyda, Hickman et al. 1994). Even if the target site is well characterized, as with HIV protease, the *de novo* design of inhibitors has been difficult. And even when the structure of the protein-ligand complex is known, we are still unable to predict the most significant interactions. The buried surface does not contribute uniformly to the binding energy (Clackson and Wells 1995). Estimations of the electrostatic, hydrogen bonding, or van der Waals contact potentials (information most readily gleaned from crystal structures) only address the enthalpic contribution of individual residues, while the entropic contribution to the overall free energy of binding prevents any simple calculation of the binding affinity of the ligand from this information (Mark and van Gunsteren 1994). Another difficulty with this approach is that the targeted protein is often represented by a static model such as a crystal structure. This overlooks the conformational flexibility and dynamic motions of the actual protein. The inhibitor is not designed to bind to these unrepresented forms of the enzyme. And finally the inhibitor must not only bind tightly to the targeted protein, it must compete with the natural substrate *in vivo*. The efficiency with which the drug must compete depends on whether the reaction must be inhibited

completely or only partially to achieve the desired effect. Even then the issues of bioavailability and selectivity for the targeted protein over other proteins in the cell have not been addressed.

The notion of rational drug design grew from the successes of the mechanism-based approach, which applied knowledge of substrates or ligands of the targeted enzyme or receptor to design inhibitors (Weber 1989). Typically analogs of the transition state or reaction intermediates have been synthesized as initial drug leads. These lead compounds are then modified, usually by systematically screening derivatives of the initial inhibitor. Many successful drug leads are still derived from naturally occurring substrates or ligands. The best inhibitors of reverse transcriptase and HIV protease are substrate analogs. The non-nucleoside inhibitors of reverse transcriptase, found through screens of chemical libraries, are more susceptible to resistance mutations than are the nucleoside-based inhibitors (Pearl 1987). A notable tight binding non-peptide based inhibitor of the protease is a cyclic urea, which was designed from peptidomimetic-protease structures (Lam, Jadhav et al. 1994).

The three retroviral enzymes, the protease, reverse transcriptase, and integrase, are all potential targets for anti-viral drugs. The structure of reverse transcriptase was determined in 1992 (Kohlstaedt, Wang et al. 1992). The first anti-viral drugs approved by the Food and Drug Administration (FDA) were against reverse transcriptase (AZT in 1987, followed by ddI). In 1989 the structure of the protease was determined, focusing drug design efforts on anti-protease inhibitors. The first protease inhibitor, saquinavir, was approved by the FDA in 1995. The structure of the integrase is still not completed; a domain of integrase has been recently solved (Dyda, Hickman et al. 1994).

These enzymes are all essential to the viral life cycle, and disruption of any of them is fatal to the virus. Both reverse transcriptase and integrase function early in the infection of the cell. Reverse transcriptase transcribes the viral genomic RNA into double-stranded DNA. Integrase then integrates the DNA copy into the chromosome of the infected cell. The protease functions later in the life cycle, after the new virus has been assembled. The viral gag and pol proteins are first generated as polyproteins instead of the individual functional proteins. The role of the protease is to cleave apart the individual proteins. The necessity of HIV protease for the virus has been demonstrated by genetically replacing the wild type protease with an inactive mutant and by inhibiting the protease in cell culture (Kohl, Emini et al. 1988; Debouck 1992). In both cases the virus particle does not mature and is not infectious. Reduction of protease activity by between 4 and 50 fold results in uninfectious viral particles (Rosé, Babe et al. 1995).

The importance of understanding the target protein in functional as well as structural terms in order to design a drug is particularly true of HIV protein targets due to the rapid mutation rate of the virus. It has been calculated that the high rate of viral reproduction coupled with the rapid rate of mutation results in each single-point mutation of the protease occurring between 10^4 and 10^5 times a day (Roberts 1995). Resistance mutations in reverse transcriptase and protease in response to drug treatments have been identified. Some of these mutations contact the drug and substrate directly, and are easily rationalized. The functional significance of other mutations are more difficult to understand. For example, it has been proposed that mutations in reverse transcriptase distant from the nucleoside-based inhibitor's binding site might indirectly influence the geometry of this site by altering the position of the template-primer, which forms part of the binding

pocket (Tantillo, Ding et al. 1994). Of the over 20 mutations that arise in the virus in response to treatment with the protease inhibitor saquinavir, at least 10 are far from the binding pocket where both substrate and inhibitor bind (Roberts 1995) (Chong-Hwan Chan, personal communication).

The protein fold and function of the viral proteins are incredibly adaptable to sequence variation. The degree of tolerance for mutations in these proteins is indicated by the alignment of related retroviral proteins. The structures of the Rous Sarcoma Virus (RSV) protease and HIV protease (with 31 identical residues, 9 conservative amino acid substitutions) have been compared and are almost identical, despite the additional 25 amino acids of RSV protease (Weber 1989; Grinde, Cameron et al. 1992). Chapter 1 describes the structure of another relative of HIV protease: the Simian Immunodeficiency Virus (SIV) protease. The sequence of SIV protease is 50% identical to HIV-1 protease (Grant, Deckman et al. 1991) and 90% identical to HIV-2 protease (Henderson, Copeland et al. 1988). Some of the resistance mutations in HIV-1 protease that are selected for during drug treatments are modified to the SIV protease sequence (for example Met36Ile, and Met46Ile) (Roberts 1995).

One strategy to overcome the high mutation rate is to develop irreversible inhibitors that covalently modify the catalytic aspartic acids. Mutations of the aspartic acids disable the enzyme. Haloperidol had been found through a computer-based search at the University of California, San Francisco as a non-covalent inhibitor of the protease (DesJarlais, Seibel et al. 1990). Various functional groups were added to haloperidol to develop it into an irreversible inhibitor. Chapter 2 describes our efforts to crystallize these irreversible inhibitors with the protease.

Chapters 3 and 4 describe structural studies of SIV and HIV proteases aimed at understanding the functioning of these proteins. Chapter 3 investigates the mechanism of proteolysis from the vantage point of product complexes. Chapter 4 discusses the conformational flexibility of the protease.

HIV protease is an aspartyl protease, with a mechanism very similar to the monomeric aspartyl proteases. By examining the structure of monomeric aspartyl proteases, Tang, et al predicted the existence of an ancestral dimeric form which, through gene duplication and fusion, resulted in the two-domain pseudo-symmetric monomeric form (Tang, James et al. 1978). The dimeric retroviral aspartyl proteases fit this description. The domains of the monomeric aspartyl proteases are still capable of functioning independently: clones of the individual domains of pepsin are catalytically active when mixed (Lin, Lin et al. 1994).

The retroviral proteases were initially identified as aspartyl proteases through sequence alignments with the monomeric proteases, particularly the active site Asp-Thr-Gly residues (Toh, Kikuno et al. 1985; Pearl 1987). While the monomeric aspartyl proteases are composed of two domains, with one catalytic aspartic acid on each domain, the retroviral proteases are dimeric, with one aspartic acid contributed by each monomer. 56 of the Ca atoms of HIV-1 protease superimpose on one of the domains of rhizopuspepsin with an rms deviation of 1.4Å (Wlodawer, Miller et al. 1989).

Like most of the monomeric aspartyl proteases, the retroviral proteases are optimally active at acidic pH's: most eukaryotic aspartyl proteases are optimally active between pH 2 and 4 (Lin, Fusek et al. 1992) (except for human renin, which is optimally active between pH 5.5 and 7.5 (Inagami 1981), while HIV protease is optimally active between pH 4 and 6 (Meek, Rodriguez et al. 1994). Also like most aspartyl proteases, substrate specificity is determined by

up to 4 residues on either side of the scissile bond. HIV protease cleaves 8 distinct sequences while processing the viral polyprotein (Henderson, Copeland et al. 1988).

Structural and functional studies of the monomeric aspartyl proteases have led to some consensus regarding the mechanism. Because of the incorporation of ^{18}O from $H_2{}^{18}O$ water into substrates, as was demonstrated for chymotrypsin, and the high efficiency of transpeptidation, the substrate was initially proposed to form an acyl intermediate like the serine proteases (Fruton 1976). Unlike the serine proteases, a covalent intermediate was not able to be isolated. Through heavy oxygen water experiments, Antonov proved that a water molecule, and not a residue from the enzyme, served as the nucleophile to attack the carboynl of the scissile bond (Antonov, Ginodman et al. 1981). Polgár proposed that the two catalytic aspartic acids functioned in a "push-pull" mechanism, allowing one to act as a general base to pull a hydrogen from the attacking water and the other as a general acid to donate a proton to the carbonyl of the scissile bond (Polgár 1987). From liganded and unliganded structures of penicillopepsin, James and Sielecki identified an ordered water molecule ideally positioned to attack the carbonyl (James and Sielecki 1985). The unliganded HIV and monomeric protease structures show an ordered water molecule hydrogen bonded between the two catalytic aspartic acids (Suguna, Bott et al. 1987; Wlodawer, Miller et al. 1989; Sielecki, Fedorov et al. 1990). Though this water is not optimally situated for nucleophilic attack of the carbonyl carbon, it is now accepted as the catalytic water (Suguna, Padlan et al. 1987; James, Sielecki et al. 1992).

Many questions remain unanswered about the mechanism of proteolysis by this family of proteases. For example, what determines the pH optimum of the different members of the family (Ido, Han et al. 1991). How is the catalytic water positioned in order to attack the scissile carbonyl? Many authors discuss strain in the scissile bond as a result of binding to the protease (Davies 1990). How can this strain be characterized and measured? Fruton showed that Km=Kd for pepsin (Fruton 1976). Is this true for other aspartyl proteases? Measuring solvent isotope effects, Rebholz and Northrup detected a slow step of the reaction after bond cleavage (Rebholz and Northrop 1991). Because Km includes terms other than substrate binding, the simplest interpretation of the equivalence of Km and Kd, namely of a rapid equilibrium of the Michaelis complex, cannot be assumed.

Some issues are complicated by the fact that substrate specificity is contributed by 7 residues for HIV protease, four residues N-terminal of the scissile bond, and three residues C-terminal (Moore, Bryan et al. 1989). For the monomeric aspartyl proteases, six residues contribute to substrate specificity: four residues N-terminal of the scissile bond and two residues C-terminal (Davies 1990). The rate limiting step for pepsin can be product release, and not bond cleavage, resulting in high rates of transpeptidation (Blum, Cunningham et al. 1991). Meek, et al proposed that the rate limiting step for HIV-1 protease is the chemical steps of bond cleavage (Meek, Rodriguez et al. 1994). Polgár, et al showed that the rate limiting step changes with sequence (Polgár, Szeltner et al. 1994). This suggests that substrate specificity will only be understood when the sequence dependence of each potentially rate limiting step is evaluated.

The decision to crystallize product complexes of HIV and SIV proteases was inspired by the kinetic studies of Tom Meek and co-workers (Hyland,

Tomaszek et al. 1991; Hyland, Tomaszek et al. 1991; Meek, Rodriguez et al. 1994). They measured product inhibition patterns and found them to be sequence dependent, though either product could be released first. Their detection of a slow step after bond cleavage suggested that product release could be rate limiting for some substrates. We expected that structures of product complexes would offer an explanation for the order of release of products. We also thought we could understand something about the nature of flap opening by examining the complex with one product bound. A third issue was to evaluate substrate strain by examining the conformation of the products after bond cleavage occurred. With bond cleavage, the products would find the least energetic conformation attainable by that sequence. By modeling the substrate from the structures of the individual bound products, we conclude that the substrate binds to the protease without strain. These issues are discussed in Chapter 3. We did not do a systematic study of product sequences. Such a study would refine the comparison made here.

The major conformational change associated with ligand binding has been considered closing of the flap residues. Each flap consists of 10 residues from one monomer of the protease dimer that forms a beta strand. A comparison of the unliganded HIV protease structure with the protease bound to a peptidomimetic inhibitor shows that ligand binding promotes closing of the two flaps, sequestering the substrate from solvent. The flaps in the product complex structures were closed, even for the flap not associated with a product. Chapter 4 describes an unliganded SIV protease structure with the flaps open. An unliganded SIV protease structure with the flaps closed had been determined (Wilderspin and Sugrue 1994). A comparison of the unliganded SIV protease structure with the flaps-closed structure shows that ligand binding also induces rigid body rotations of a core domain of each

monomer of the protease. We propose an induced fit model for catalysis involving the rotation of the core domains to orient the catalytic aspartic acids for proteolysis. The open binding pocket promotes binding of the polyprotein substrate while the closed conformation promotes catalysis. Our conclusions from Chapters 3 and 4 favor an induced fit mechanism for proteolysis instead of a substrate strain model, in which positioning the active site and substrate properly leads to catalysis.

The domain motions of the protease may explain some of the resistance mutations observed *in vivo* in response to anti-protease therapies. Several drug resistant mutations are located at the interface of these domains. These mutations may shift selectivity towards the substrate if the drug binds the open conformation less efficiently than does the substrate.

Chapter 5 describes our attempts to crystallize HIV integrase. The lack of structural information for integrase has impeded functional studies of the protein and the application of structure-based drug design. The major impediment to crystallization is the insolubility of the protein. The domain whose structure has been determined was still only soluble enough to attain crystals after a point mutation. This chapter describes a purification protocol we developed, and our efforts to solubilize and crystallize the protein.

The DNA binding properties of integrase are not well understood. Integrase recognizes both specific and non-specific DNA sequences, but it is still not known whether a single binding site binds both of these sequences or there are two distinct binding sites. A related question is the multimeric state of the protein during integration. A DNA binding assay for integrase would address these fundamental questions. Chapter 5 also describes our development of a DNA binding assay for integrase. We planned to use this

assay to determine conditions that promoted DNA binding for crystallization, and provided a means for screening potential inhibitors. Though the results presented here are preliminary, it appears that integrase binds the specific DNA sequence with the same affinity that it binds non-specific DNA. This result is not unprecedented, but it is surprising. It raises the question of how integrase recognizes the specific DNA sequences to which it reacts.

References:

Antonov, V. K., L. M. Ginodman, et al. (1981). "Studies on the mechanisms of action of proteolytic enzymes using heavy oxygen exchange." *Eur. J. Biochem.* **117**: 195-200.

Blum, M., A. Cunningham, et al. (1991). "Mechanism and pathway of penicillopepsin-catalyzed transpeptidation and evidence for noncovalent trapping of amino acid and peptide intermediates." *J. Biol. Chem.* **266**(15): 9501-9507.

Clackson, T. and J. A. Wells (1995). "A hot spot of binding energy in a hormone-receptor interface." *Science* **267**: 383-386.

Davies, D. R. (1990). "The structure and function of the aspartic proteinases." *Annu. Rev. Biophys. Biophys. Chem.* **19**: 189-215.

Debouck, C. (1992). "The HIV-1 protease as a therapeutic target for AIDS." *AIDS Res. Hum. Retroviruses* **8**(2): 153-164.

DesJarlais, R. L., G. L. Seibel, et al. (1990). "Structure-based design of nonpeptide inhibitors specific for the Human Immunodeficiency Virus 1 protease." *Proc. Natl. Acad. Sci. U.S.A.* **87**: 6644-6648.

Dyda, F., A. B. Hickman, et al. (1994). "Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases [see comments]." *Science* **266**(5193): 1981-1986.

Fruton, J. S. (1976). "The mechanism of the catalytic action of pepsin and related acid proteinases." *Methods Enzymol.* **44:** 1-35.

Grant, S. K., I. C. Deckman, et al. (1991). "Purification and biochemical characterization of recombinant simian immunodeficiency virus protease and comparison to human immunodeficiency virus type 1 protease." *Biochemistry* **30**(34): 8424-8434.

Grinde, B., C. E. Cameron, et al. (1992). "Mutations that alter the activity of the Rous sarcoma virus protease." *J. Biol. Chem.* **267**(14): 9481-9490.

Henderson, L. E., T. D. Copeland, et al. (1988). Human Retrovirus, Cancer and AIDS: Approaches to Prevention and Therapy. New York, Liss.

Hyland, L. J., T. J. Tomaszek, et al. (1991). "Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism." *Biochemistry* **30**(34): 8454-8463.

Hyland, L. J., T. J. Tomaszek, et al. (1991). "Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by 18O isotope exchange." *Biochemistry* **30**(34): 8441-8453.

Ido, E., H. P. Han, et al. (1991). "Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S." *J Biol Chem* **266**(36): 24359-24366.

Inagami, T. (1981). . Biochemical Regulation of Blood Pressure. R. L. Soffer. New York, John Wiley & Sons, Inc.: 39-73.

James, M. N. and A. R. Sielecki (1985). "Stereochemical analysis of peptide bond hydrolysis catalyzed by the aspartic proteinase penicillopepsin." *Biochemistry* **24**(14): 3701-3713.

James, M. N., A. R. Sielecki, et al. (1992). "Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides." *Biochemistry* **31**(15): 3872-3886.

Kohl, N. E., E. A. Emini, et al. (1988). "Active human immunodeficiency virus protease is required for viral infectivity." *Proc. Natl. Acad. Sci. U.S.A.* **85**: 4686-4690.

Kohlstaedt, L. A., J. Wang, et al. (1992). "Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor." *Science* **256**(5065): 1783-1790.

Lam, P. Y., P. K. Jadhav, et al. (1994). "Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors." *Science* **263**(5145): 380-384.

Lin, X. L., Y. Z. Lin, et al. (1994). "Relationships of human immunodeficiency virus protease with eukaryotic aspartic proteases." *Methods Enzymol.* **241**: 195-224.

Lin, Y., M. Fusek, et al. (1992). "pH dependence of kinetic parameters of pepsin, rhizopuspepsin, and their active-site hydrogen bond mutants." *J. Biol. Chem.* **267**: 18413-18418.

Mark, A. E. and W. F. van Gunsteren (1994). "Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies." *J. Mol. Biol.* **240**(2): 167-176.

Meek, T. D., E. J. Rodriguez, et al. (1994). "Use of steady state kinetic methods to elucidate the kinetic and chemical mechanisms of retroviral proteases." *Methods Enzymol.* **241**(127): 127-156.

Moore, M. L., W. M. Bryan, et al. (1989). "Peptide substrates and inhibitors of the HIV-1 protease." *Biochem. Biophys. Res. Commun.* **159**(2): 420-5.

Pearl, L. H., Taylor, W.R. (1987). "A structural model for the retroviral proteases." *Nature* **329**: 351-354.

Polgár, L. (1987). "The mechanism of action of aspartic proteases involves 'push-pull' catalysis." *FEBS Lett.* **219**: 1-4.

Polgár, L., Z. Szeltner, et al. (1994). "Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease." *Biochemistry* **33**(31): 9351-9357.

Rebholz, K. L. and D. B. Northrop (1991). "Slow step after bond-breaking by porcine pepsin identified using solvent deuterium isotope effects." *Biochem. Biophys. Res. Commun.* **176**: 65-69.

Roberts, N. A. (1995). "Drug-resistance patterns of saquinavir and other HIV proteinase inhibitors." *AIDS* **9 (supplement 2)**: S27-S32.

Rosé, J. R., L. M. Babe, et al. (1995). "Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity." *J. Virol.* **69**(5): 2751-2758.

Sielecki, A. R., A. A. Fedorov, et al. (1990). "Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8Å resolution." *J. Mol. Biol.* **214**: 143-170.

Suguna, K., R. R. Bott, et al. (1987). "Structure and refinement at 1.8Å resolution of the aspartic proteinase from Rhizopus chinensis." *J. Mol. Biol.* **196**: 877-900.

Suguna, K., E. A. Padlan, et al. (1987). "Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action." *Proc. Natl. Acad. Sc.i U.S.A.* **84**(20): 7009-7013.

Tang, J., M. N. G. James, et al. (1978). "Structural evidence for gene duplication in the evolution of the acid proteases." *Nature* **271**: 618-621.

Tantillo, C., J. Ding, et al. (1994). "Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance." *J. Mol. Biol.* **243**(3): 369-87.

Toh, H., R. Kikuno, et al. (1985). "Close structural resemblance between putative polymerase of a Drosophila transposable genetic element 17.6 and pol gene product of Moloney murine leukaemia virus." *EMBO J.* **4**: 1267-1272.

Weber, I. T. (1989). "Structural alignment of retroviral protease sequences." *Gene* **85**: 565-566.

Wilderspin, A. F. and R. J. Sugrue (1994). "Alternative native flap conformation revealed by 2.3 A resolution structure of SIV proteinase." *J. Mol. Biol.* **239**(1): 97-103.

Wlodawer, A., M. Miller, et al. (1989). "Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease." *Science* **245**: 616-621.

CHAPTER ONE:

STRUCTURE OF THE PROTEASE FROM SIMIAN IMMUNODEFICIENCY VIRUS

Reprinted from (1993)*Biochemistry 32*, 12498-12507.

Preface

There were two motivations for determining the structure of SIV protease: 1) comparing the structure to HIV-1 protease, particularly considering the rapid rate of mutation of HIV protease *in vivo*. In addition, the sequence of SIV protease is 90% identical to HIV-2 protease, whose structure had not yet been determined. 2) SIV protease was easier to purify in large quantities than HIV-1 protease, and crystallized readily. It was therefore considered useful for crystallization of potential drug leads. It did not turn out to be easier to crystallize weak-binding inhibitors with SIV protease than it was with HIV-1 protease.

Jason Rosé developed the expression system for SIV Ser 4 His protease and purified the initial protein for crystallization. Rafa Salto showed the stoichiometry of binding of EPNP for SIV protease. I determined the crystallization conditions based on the HIV-1 protease conditions. I determined the structure and analyzed it. Jason helped immensely with writing the paper.

Structure of the Protease from Simian Immunodeficiency Virus: Complex with an Irreversible Nonpeptide Inhibitor[†,‡]

Robert B. Rose,[§] Jason R. Rosé,[∥] Rafael Salto,[⊥] Charles S. Craik,[§,⊥] and Robert M. Stroud*[,§,⊥]

Departments of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448

Received June 9, 1993; Revised Manuscript Received August 27, 1993[●]

ABSTRACT: A variant of the simian immunodeficiency virus protease (SIV PR), covalently bound to the inhibitor 1,2-epoxy-3-(*p*-nitrophenoxy)propane (EPNP), was crystallized. The structure of the inhibited complex was determined by X-ray crystallography to a resolution of 2.4 Å and refined to an *R* factor of 19%. The variant, SIV PR S4H, was shown to diminish the rate of autolysis by at least 4-fold without affecting enzymatic parameters. The overall root mean square (rms) deviation of the α-carbons from the structure of HIV-1 PR complexed with a peptidomimetic inhibitor (7HVP) was 1.16 Å. The major differences are concentrated in three surface loops with rms differences between 1.2 and 2.1 Å. For 60% of the molecule the rms deviation was only 0.6 Å. The structure reveals one molecule of EPNP bound per protease dimer, a stoichiometry confirmed by mass spectral analysis. The epoxide moiety forms a covalent bond with either of the active site aspartic acids of the dimer, and the phenyl moiety occupies the P1 binding site. The EPNP nitro group interacts with Arg 8. This structure suggests a starting template for the design of nonpeptide-based irreversible inhibitors of the SIV and related HIV-1 and HIV-2 PRs.

Simian immunodeficiency virus (SIV)[⊥] is a retrovirus closely related to the type 2 human immunodeficiency virus (HIV-2) and more distantly related to human immunodeficiency virus type 1 (HIV-1). Infectious clones of a strain isolated from macaque monkeys ($SIV_{mac}239$) also produce an AIDS-like disease in rhesus monkeys (Kestler et al., 1990). These infected monkeys provide an animal model for testing therapeutic agents targeting HIV-1 or HIV-2. The structure of SIV_{mac} protease (SIV PR) was determined to facilitate the incorporation of data from *in vivo* testing with efforts to improve the design of drugs targeting the HIV-2 and HIV-1 PRs.

The residues in the binding pocket of SIV PR differ from those of HIV-1 PR in 3 of 13 positions identified as major peptide binding determinants in HIV-1 PR (Miller et al., 1989). Despite these differences, SIV PR is capable of authentically processing the HIV-1 $p53^{gag}$ polyprotein *in vitro* (Grant et al., 1991). The sequence of HIV-2 PR is identical to SIV PR at these 13 positions.

It has been shown *in vivo* that the virus can develop resistance to drugs targeting reverse transcriptase (Richman, 1993). It has been proposed that the virus could develop resistance to antiprotease drugs as well (Cameron et al., 1993). A comparison of the HIV-1, HIV-2, and SIV protease structures will identify regions that are structurally conserved and may represent binding sites for inhibitors that are less susceptible to the development of drug resistance.

A variant of SIV_{mac} protease was constructed to stabilize the enzyme to autolysis, by replacing serine 4 by histidine (SIV PR S4H), at the major autolytic site (Rosé et al., 1993). This mutation was found to diminish the rate of autolysis by a factor of 4 without affecting enzymatic parameters.

1,2-Epoxy-3-(*p*-nitrophenoxy)propane (EPNP) is a covalent inhibitor of aspartyl proteases, requiring the presence of two neighboring carboxyl groups for the reaction to take place at the active site. This is the first structure of a retroviral protease bound to a covalent inhibitor. The structure reveals aspects that bear directly on the reaction mechanism of inhibition.

MATERIALS AND METHODS

Purification of Recombinant Proteins. The expression of $SIV_{mac}239$ from plasmid SOD/SIV has been described elsewhere (Rosé et al., 1993). Ser 4 of SIV PR was replaced with His to increase the stability of the protease to autoproteolysis. Soluble mature SIV PR was purified from lysates generated by sonication of induced cells harboring the plasmid. Pellets were resuspended in 14 volumes of lysis buffer [50 mM Tris, pH 7.5, 5 mM EDTA, 10% glycerol, 0.2 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (ICN)] and lysed by sonication. The lysate was clarified by centrifugation at 12 500 rpm in a SS34 rotor for 30 min and the supernatant precipitated overnight by treatment with 57.5% ammonium sulfate. The pellet was resuspended in 5 volumes of 50 mM HEPES, pH 7.5, and 1 mM EDTA and loaded on a column of pepstatin A agarose. The protease was eluted with a buffer containing 250 mM ε-aminocaproic acid, pH 10.5, 5% glycerol, 5% ethylene glycol, and 1 mM EDTA. Active fractions were pooled and loaded onto a DEAE-Sepharose column equilibrated in the pepstatin A agarose column elution buffer. The protease was eluted from the DEAE column with 50 mM HEPES, pH 7.0, and 1 mM EDTA. Fractions after the pH change were analyzed by SDS–PAGE on 15% polyacrylamide gels, and fractions of single-band purity were used for crystallography (Figure 1). Typical yields were 6–8 mg of

[†] This work was supported by NIH Grant GM 39552. R.B.R. was supported by a graduate training grant from the University of California Systemwide Biotechnology Research and Education Program, Project Title "Biotechnology Program in Program Crystallization". J.R.R. was supported by Graduate Training Grant GM 09175.

[‡] The atomic coordinates in this paper have been submitted to the Brookhaven Protein Data Bank (1Sam).

* Author to whom correspondence should be addressed.

[§] Graduate Group in Biophysics.

[∥] Department of Pharmacology.

[⊥] Department of Pharmaceutical Chemistry.

[●] Abstract published in *Advance ACS Abstracts*, October 15, 1993.

[⊥] Abbreviations: DMSO, dimethylsulfoxide; EPNP, 1,2-epoxy-3-(*p*-nitrophenoxy)propane; SIV, simian immunodeficiency virus; HIV, human immunodeficiency virus; PR, protease; DTT, dithiothreitol; HPLC, high-performance liquid chromatography; EDTA, ethylenediaminetetraacetic acid.

0006-2960/93/0432-12498$04.00/0

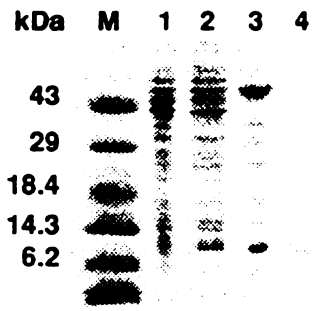


FIGURE 1: Purification of SIV PR. (Lane M) Low molecular weight markers; (lane 1) crude cell lysate; (lane 2) ammonium sulfate precipitate; (lane 3) pepstatin column elution pool; (lane 4) DEAE elution pool.

purified SIV PR from 6.5 g of *Escherichia coli*. The protease was frozen over dry ice without glycerol and stored at −70 °C.

SIV PR Inactivation with EPNP. EPNP, an active site directed irreversible inhibitor of aspartyl proteases (Tang, 1971), was purchased from Sigma and was purified by recrystallization in methanol. HIV-1 PR (15 μg/mL final concentration) and HIV-2 or SIV PR (24 μg/mL final concentration) were preincubated at 25 °C in buffer containing 250 mM sodium acetate, pH 5.4, 500 mM NaCl, 1 mM EDTA, and 10% DMSO in the presence of 1–10 mM EPNP; 500 μM DTT was added for the HIV-1 enzyme. Baseline measurements were carried out with 10% DMSO in the absence of inhibitor. At eight different times, aliquots were removed and assayed for activity. Proteases were assayed with the fluorescent substrate ABZ-Thr-Ile-Nle-Phe(p-NO_2)-Gln-Arg-NH_2 under conditions described previously (Toth & Marshall, 1990). For the calculation of the rates of inactivation, kinetic data were fit to a pseudo-first-order equation. Eight different inhibitor concentrations were used to calculate the maximum inactivation rate and $K_{inactivation}$ (K_{inact}) for each protease. K_{inact} is the concentration of inhibitor that results in half the maximal rate of inactivation. Each determination of K_{inact} was carried out in duplicate.

SIV Labeling with EPNP. SIV PR (300 μg) was treated with 10 mM EPNP at 25 °C in 1 mL of 250 mM sodium acetate, pH 5.4, containing 500 mM NaCl, 1 mM EDTA, and 10% DMSO. After 2 h, the fully inhibited enzyme was isolated by reversed-phase HPLC on a Vydac C_3 column using a 100-min linear gradient from 5 to 65% acetonitrile in 0.1% TFA. The peak of protein identified by absorbance at 220 nm was collected, concentrated under vacuum, and dissolved in a small volume of 50% acetonitrile and 1% acetic acid. Molecular weights of protease and protease–EPNP complexes formed in solution were determined by mass spectrometry, using a model Bio-Q VG electrospray mass spectrometer. Crystals of SIV PR S4H inhibited with EPNP were dissolved in pH 5.5 sodium acetate buffer (50 mM) and analyzed in an identical fashion.

Crystallization. Previously stored samples were thawed rapidly to room temperature using a water bath, and the pH was adjusted by dilution with 1 M sodium acetate, pH 5.4, to a final concentration of 300 mM sodium acetate; 10 mM EPNP was added and incubated for 3 h. Precipitated EPNP was removed by centrifugation in a microfuge. The protein was dialyzed and concentrated in an 8-mL collodion membrane (Schleicher and Schuell), using a dialysis buffer of 20 mM sodium acetate, pH 5.4, and 50 mM NaCl. Final protein concentration for crystallization was 2.5–3 mg/mL as measured by the Bio-Rad Bradford assay using IgG as a standard.

Table I: Statistics on Data Collection and Refinement

resolution (Å)	2.4
R_{sym}[a] (%)	5.4
completeness to 2.5 Å (%)	92
completeness of 2.5–2.25 Å resolution shell (%)	77
total no. of reflections $>1\sigma$	21 880
no. of independent reflections	8 824
search model based on Swain et al. (1990)	7HVP
final R factor (7–2.4 Å) (%)	19
rms deviation of bond lengths (Å)	0.015
rms deviation of bond angles (deg)	3.2
rms deviation of dihedral angles (deg)	28.3
Wilson B factor	23

[a] $R_{sym} = \Sigma_i\Sigma_j|(I_i) - I_i|/\Sigma_i\Sigma_j I_{ij}$. Data reported for $I > 1\sigma$.

Crystals were grown using the hanging drop vapor diffusion method. Four microliters of protein solution was mixed with 4 μL of well buffer: 4–6% saturated NaCl (230–330 mM) and 100 mM sodium cacodylate, pH 6.5. Large crystals (0.4 × 0.15 × 0.05 mm) were grown consistently by macroseeding pre-equilibrated drops. Crystals were plates with rhomboid faces with two opposite corners truncated by parallel sides.

Structure Determination. The space group and cell dimensions of the SIV PR S4H/EPNP crystals were determined by precession photography. I_{hkl} data were collected with an R-axis II image plate detector using X-rays generated from a Rigaku 18 KW generator. Data were reduced using software provided by Rigaku. A summary of the data collection and refinement statistics is provided in Table I. Initial phases were determined by molecular replacement using a model that was built from an HIV-1 PR structure, 7HVP (Swain et al., 1990). All nonidentical residues between SIV and HIV-1 protease were changed to alanines to generate the search model (41 of 99 residues). Positions in the search model corresponding to glycines in either the HIV-1 or SIV PR structure were defined as glycines (there are three nonconserved glycines). I62 and I66 in the HIV-1 PR structure were changed to V62 and V66 as in the SIV PR structure. The 10 flap residues, 46–55, were removed because the conformation of the flaps can be highly variable (Navia et al., 1989; Wlodawer et al., 1989; Swain et al., 1990; Rutenber et al., 1993). Because the asymmetric unit contained one protease monomer, a monomer was used as the initial search model.

A rotation search using the above model was carried out using I_{hkl} data between 15- and 3-Å resolution and a maximum vector length of 30 Å. Patterson correlation refinement distinguished the correct solution from others. In fact, the correct solution corresponded to the highest peak from the initial rotation search. The solution obtained was confirmed in part since it reconstructed the expected association of monomers in the protease dimer of HIV-1 by orientation relative to one of the crystallographic twofold axes. A translation search was carried out using F_{hkl} data between 8- and 3-Å resolution. A grid size of 1 Å was used. After rigid body minimization of the top 1000 translation search solutions, the R factor for the top sclution was 47%. The R factor for the next best search solution was 50%. Molecular replacement calculations were carried out using the program XPLOR (Brunger et al., 1987).

The search model was refined using XPLOR (Brunger et al., 1987) with manual rebuilding into $(2|F_o| - |F_c|),\alpha_c$ and $(|F_o| - |F_c|),\alpha_c$ electron density maps displayed using the molecular graphics programs FRODO (Jones, 1985) and CHAIN (Sack, 1988). After refinement to an R factor of 0.28, individual isotropic B factors were added. EPNP was

added to the model at an R factor of 0.27. Individual water molecules were added if density appeared in difference maps calculated using terms ($|F_o| - |F_c|$),α_c and if they formed reasonable hydrogen bonds to the protein (Finer-Moore et al., 1992). In the final structure there was no density for four side chains, all of which are on the protein surface and are probably disordered.

Electron density maps computed using terms ($2|F_o| - |F_c|$),α_c showed equivalent density for EPNP at each of the two active site aspartic acids, which corresponded to half-occupancy at each carboxyl. This could correspond either to only one molecule of EPNP bound per dimer—bound statistically to either site—or to statistical occupation of each site in the dimer independently. In either case, the dimeric molecule becomes *asymmetric*. However, molecules are statistically distributed, 50:50, in the crystal; thus, *statistically*, the space group reflected in the data is $C222_1$, though any particular single unit cell may not present this symmetry. To allow for the refinement of two equivalent half-occupied EPNP sites that might overlap during refinement if there were only one per dimer, interactions between monomers were not included during refinement. To assess whether there was some asymmetry in the protein due to a single bound EPNP, possibly expected in the region of the EPNP, refinement of the dimer as the asymmetric unit in a lower symmetry space group ($C2$) was carried out. At this resolution no asymmetry is yet apparent; thus, all final refinement was carried out using the monomer plus 0.5 EPNP as the asymmetric unit in space group $C222_1$.

RESULTS

Production of SIV PR. Expression and characterization of $SIV_{mac}BK28$ was as described by Grant et al. (1991). The protease clone used in this study was of $SIV_{mac}239$, which differs from $SIV_{mac}BK28$ by the substitution of Glu for Lys at position 63. This results in a significant change in the isoelectric point of the enzyme, which is predicted to be 5.1 rather than the observed value of 8.2 for $SIV_{mac}BK28$. The pI calculation used the program PREDICT (available on request). This predicted pI of 5.1 is essentially the same as that predicted for HIV-2 PR, and a similar purification protocol to that described previously for HIV-2 was effective (Rosé et al., 1993).

SIV PR S4H Mutant. Crystals were initially grown using wild-type SIV PR. However, out of concern that autoproteolysis prior to inhibition with EPNP might affect sample homogeneity, a variant of SIV PR was made to prevent cleavage at the primary autolysis site. This variant, SIV PR S4H, did not significantly affect the kinetic parameters of the enzyme but stabilized it approximately 4-fold to autolysis (Figure 2). The loss of activity seen with the variant did not correlate with the formation of autoproteolytic fragments as monitored by SDS–PAGE (data not shown). In addition, the decay curve for SIV PR S4H is fit better by a first-order than a second-order rate equation, suggesting that a process other than autoproteolysis is responsible for the observed loss of activity. It is possible that this is due to oxidation of methionine residues or thermal denaturation during the course of the assay. No difference density for the oxygens was seen for the methionines (residues 76 and 95). In crystallization trials the SIV PR S4H variant produced large crystals consistently, which were used for data collection. These crystals appeared to have the same morphology as those obtained from wild-type SIV PR.

EPNP Inactivation. The maximum rates of inactivation (V_{max}) of the SIV, HIV-1, and HIV-2 PRs by EPNP were

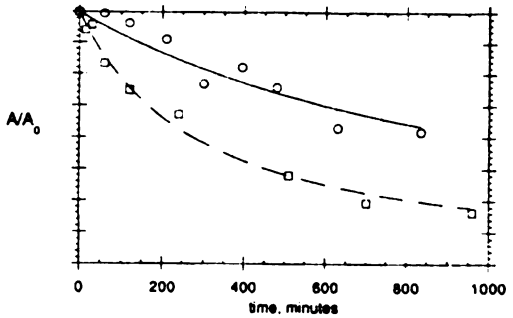


FIGURE 2: Comparison of decay curves for SIV PR (□) and SIV PR S4H (○). Starting protease concentration was 40 ng/μL. A/A_0 represents the ratio of active protein at a given time (A) to the total active protein at time = 0 (A_0). Decay curves were generated by assaying remaining protease activity at various time points and fit to either a first-order $A = A_0e^{-kt}$ or a second-order $A = A_0/(1 + A_0k_2t)$ rate equation, where t is time and k and k_2 are the rate constants for inactivation.

Table II: Inactivation of HIV and SIV PRs by EPNP

protease	V_{max}[a] (min^{-1})	K_{inact}[b] (mM)
HIV-1	0.060 ± 0.006	9.85 ± 0.98
HIV-2	0.048 ± 0.004	6.71 ± 0.58
SIV	0.060 ± 0.002	8.03 ± 0.24

[a] Maximum inactivation rate. [b] EPNP concentration at half-maximal inactivation rate, $V_{max}/2$.

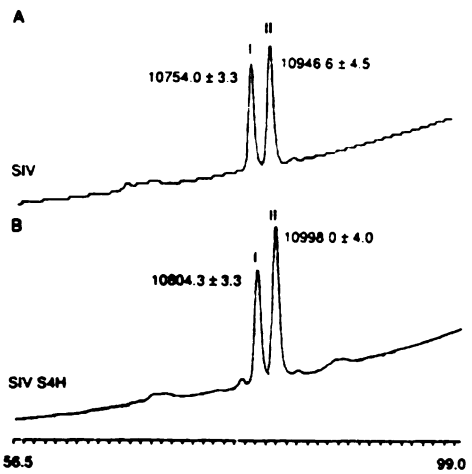


FIGURE 3: Stoichiometry of EPNP modification. SIV PR that had been inactivated in solution (A) or inactivated and crystallized SIV S4H (B) was analyzed by HPLC at 220 nm on a Vydac C_3 column using a 100-min linear gradient from 5 to 65% acetonitrile in 0.1% TFA. Molecular weights for peaks I and II were determined by electrospray mass spectrometry. The areas under the peaks were 3.92×10^5 (I) and 4.91×10^5 (II) for (A) and 4.22×10^5 (I) and 6.00×10^5 (II) for (B).

similar: approximately 0.05 min^{-1} (Table II). The similarity indicates that the mechanism of inactivation is the same for the three proteins. The stoichiometry of protease modification was determined by electrospray mass spectrometry. Samples of inactivated SIV PR in solution and of inactivated SIV PR S4H from crystals contained two major species which could be isolated by HPLC (Figure 3). The areas under peaks I and II were roughly equal for both SIV PR and SIV PR S4H, indicating that half of the protease monomers had been modified in each case. Material from peaks I and II was

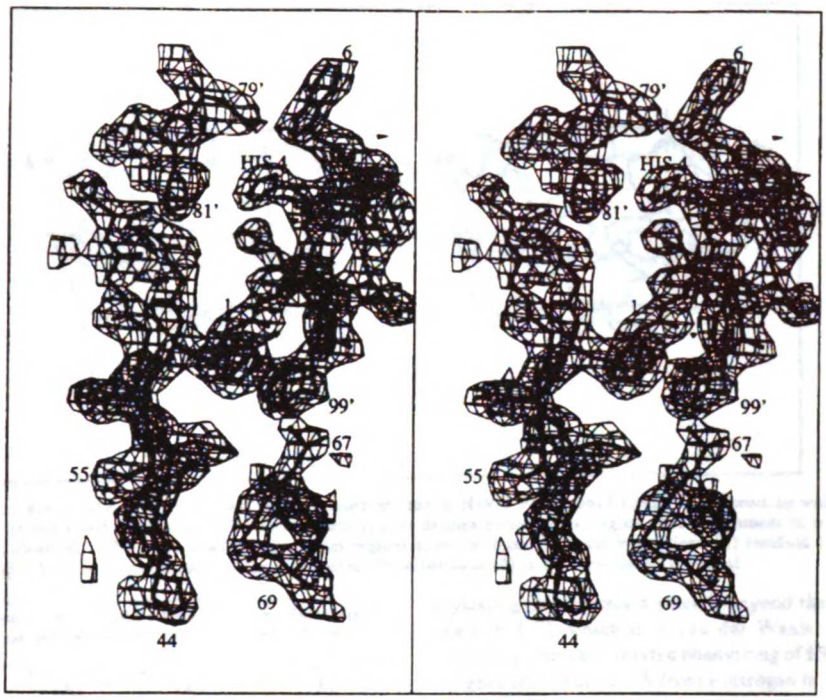


FIGURE 4: Electron density in a $(2|F_o| - |F_d|),\alpha_c$ map showing the crystal contacts formed near the flap residues, 44–55. Included in the figure are the dimer-related residues, 79′–81′, and residues from the C-centered related dimer: 1–6, 67–69, and 96′–99′. Density for the substitution S4H can be seen. Density for the side chains is well-defined, and main-chain carbonyl oxygens are visible. The density is contoured at 0.9 σ from the XPLOR output.

analyzed by electrospray mass spectrometry to determine the molecular masses of the two species. The masses of the two species differed by 194 daltons, the predicted molecular mass of EPNP. Thus, one molecule of EPNP was covalently added to each functional dimer.

Overall Structure. The crystal contained one monomer of the protease per asymmetric unit, with the twofold axis of the PR dimer coincident with the crystallographic *a*-axis. The space group was determined to be $C222_1$ with unit cell dimensions of $a = 62.7$ Å, $b = 32.2$ Å, and $c = 96.1$ Å, with an error of ±0.1 Å. In addition to the protease, the final structure contained one molecule of EPNP per protease dimer and 25 water molecules. The atomic coordinates and temperature factors have been submitted to the Brookhaven Protein Data Bank (1Sam). Electron density computed using terms $(2|F_o| - |F_d|),\alpha_c$ for the "flaps" (residues 46–55) and the C- and N-terminal residues is shown in Figure 4. Density for the His 4 side chain, the point mutation, is well-defined in the figure.

The overall fold of SIV PR follows that of HIV-1 PR, as can be seen in the superposition of SIV PR and 7HVP in Figure 5. A core of 53 residues was identified in the SIV PR dimer using difference distance matrices (Perry et al., 1990) to identify regions of the structure whose α-carbon positions overlap most closely with the HIV-1 PR structure, 7HVP. These 53 residues were used to superimpose the two structures using the program Gem [written by E. B. Fauman (University of California at San Francisco, 1993) available on request], resulting in an rms deviation of 0.24 Å for these 53 residues. With this alignment, the overall rms deviation for the α-carbon backbones of the two protease dimers is 1.16 Å. Figure 5 shows the dimers divided into three regions. The backbones of the central region are the most conserved. Excluding the three amino-terminal residues from each monomer, which have different conformations, the rms deviation for this central region is 0.64 Å. This core of the protein contains the active site aspartic acids, D25 and D25′ (primes indicate the symmetry-related monomer of the dimer), and the flaps comprising residues 46–55 and 46′–55′. The outer regions show the most deviation: β-strand residues 11–22, rms deviation of 1.2 Å; random coil residues 33–45, rms deviation of 2.1 Å; β-strand residues 63–75, rms deviation of 1.7 Å.

The positional difference of the 13 residues in contact with the peptidomimetic inhibitor in the HIV-1 structure, 7HVP (R8, L23, D25, G27, A28, D29, V32, I47, G48, I50, P81, V82, I84) in the HIV-1 structure (Miller et al., 1989), and the corresponding residues in SIV PR S4H is recorded in Table III. Three of these residues differ between these two sequences: at residues 32, 47, and 82. To assess the significance of these positional differences, Table III also shows the standard deviation of the position expected when atomic coordinates of two independently refined but otherwise identical structures are compared (Fauman, 1993), as an extension of the methods developed in Chambers and Stroud (1979) and Perry et al. (1990). This error is an empirically derived function of the *B* factor (*B*): $p_1 + p_2 \exp(B/p_3)$, with p_1, p_2, and p_3 functions of the number of atoms refined in the structure divided by the number of reflections measured in the data. In this structure determination the standard deviation of atomic position, SD, is calculated

$$SD = 0.15 + 0.012 \exp(B/10.99)$$

Density for the flaps was clearly seen in the original maps after being omitted in the molecular replacement search model. The flaps were found to be in a "closed" conformation, similar to the conformation in the 7HVP complex which contained

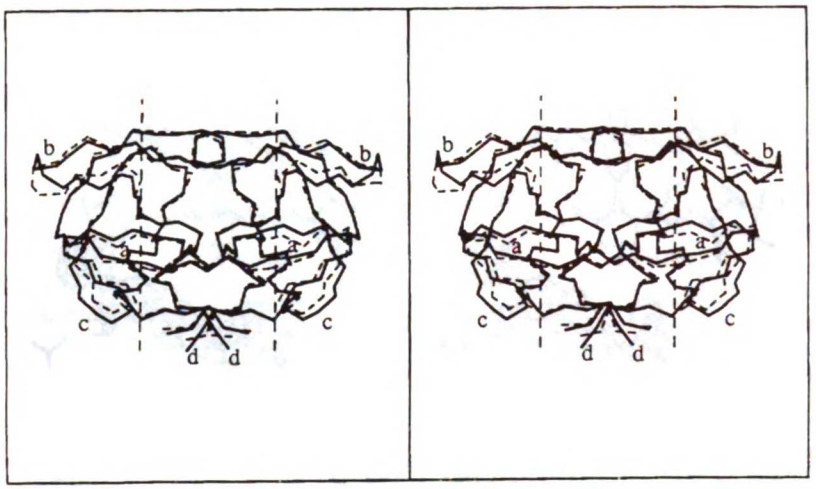


FIGURE 5: Backbone trace of SIV PR (solid lines) superimposed on that of HIV-1 PR (dashed lines). Superposition was carried out using a core of 53 residues as described in the text. The dotted vertical lines delineate the central region of the alignment in which the backbone structures are almost identical. The loops labeled in the outer regions show the most variation in position: (a) residues 11–22; (b) residues 33–45; (c) residues 63–75. The conformation of the N-terminal three residues (d) is also noticeably different.

Table III: Comparison of the α-Carbon Positions of the Substrate Binding Pocket Residues in SIV S4H PR and HIV-1 PR, 7HVP

residues[a]	Δr (1–99)[b] (Å)	Δr_r[c] (1–99)[b] (Å)	Δr (101–199)[d] (Å)	Δr_r (101–199)[c] (Å)
R8	0.2	0.4	0.4	0.3
L23	0.5	0.3	0.3	0.3
D25	0.2	0.3	0.6	0.3
G27	0.3	0.3	0.4	0.3
A28	0.2	0.3	0.5	0.3
D29	0.2	0.4	0.6	0.4
I32(V)	0.5	0.3	0.5	0.3
V47(I)	0.6	0.5	0.5	0.5
G48	0.2	0.5	0.7	0.5
I50	0.9	0.7	0.4	0.7
P81	**1.2**	**0.5**	**1.2**	**0.5**
I82(V)	**1.0**	**0.4**	**1.3**	**0.4**
I84	0.1	0.4	0.1	0.4

[a] Residues are from the SIV PR sequence, with the HIV-1 PR residue in parentheses when different from the SIV PR sequence. [b] Deviations between SIV PR and the monomer of 7HVP with residues numbered between 1 and 99. [c] Δr_r (in columns 3 and 5) is an estimate of the standard deviation in atomic position if the two structures were identical. [d] Deviations between the symmetry-related monomer of the dimer, 1′–99′ in SIV PR, and the monomer of 7HVP with residues numbered between 101 and 199. In bold are the residues whose positions vary significantly relative to the estimated error.

a peptidomimetic inhibitor. The flaps form a β-strand which is initially parallel to the protein surface and which turns 90° with a right-handed twist. The tips of the dimer-related flaps are parallel to each other and are within 3.2 Å. The rms deivation of the flaps between the SIV PR and 7HVP is 0.78 Å. The largest positional difference, 1 Å, is between the α-carbon atoms of G49. The phenyl ring of the F53 side chain in the flaps of SIV PR has a significantly different conformation from that found in the 7HVP structure, facing into solvent instead of parallel to the protein surface.

The flaps participate in a crystallographic contact with other protease dimers related by C-centering in the unit cell: one dimer contacting each of the flaps. Three crystallographically related N- and C-terminal side chains, F99, F3, and P1, are within van der Waals contact of F53 and I46 side chains in the flaps (Figure 4). The four rings of F99, F3, P1, and F53 pack together, within 3.4–4.1 Å. The I46 side chain in the flaps is 2.9 Å from the carbonyl oxygen of L67. This crystallographic contact extends beyond the flaps to the side chain of L67, which is in van der Waals contact with the crystallographically related phenyl ring of F99. The carboxyl oxygens of D79 are 3.7 Å from a nitrogen in the histidine ring of H4.

The side chains of the active site aspartic acids, D25 and D25′, are more skewed in the SIV PR S4H structure than in the 7HVP structure. If a plane is defined by the Cα and Cγ atoms of each of the aspartic acid side chains, then the angle made by the C_γ–Oδ bond with the plane describes their deviation from planarity (Figure 6c). In the 7HVP structure, this angle is +13° for Oδ1 and −4° for Oδ2. In the SIV structure these angle are +36° for Oδ1 and −16° for Oδ2. In the unliganded form, a hydrogen bond exists between D25-Oδ1 and D25′-Oδ1. The unfavorable geometry in the SIV PR S4H structure suggests that the hydrogen bond between the active site aspartic acids has been disrupted as a result of covalent modification with EPNP. A hydrogen bound to Oδ1 of one of the aspartic acids is 3.1 Å from the other aspartic acid's Oδ1, making an angle of 89° for Oδ1–H–Oδ1.

Water Molecules. A total of 25 ordered water molecules have been resolved in the structure. Water 301 in the 7HVP structure is conserved in the SIV PR structure, hydrogen-bonded to the amide nitrogens of I50 and I50′ of both flaps of the dimer (Figure 6a). This water is 0.5 Å from the site occupied by water 301 in 7HVP and is 3.05 Å from the amide nitrogens to which it is bound. The *B* factor for this water is low, 9.25, while the *B* factor for the backbone nitrogens is 40. Water 305, which appears in the unliganded structure between the two catalytic aspartic acids (Navia et al., 1989; Wlodawer et al., 1989), has been displaced by EPNP. Two other well-ordered water molecules appear in the binding cavity of the SIV structure: water 307 and water 313. Water 307 is 3.0 Å from the backbone amide of D29. Water 313 is 3.2 Å from the backbone amide of G48. Both waters occupy positions near where the peptidomimetic inhibitor in 7HVP forms hydrogen bonds (the backbone nitrogen of G48 is hydrogen-bonded to the carbonyl oxygen of S_4 or S_3' in 7HVP, and the backbone nitrogen of D29 is hydrogen-bonded to the carbonyl oxygen of S_3 or S_2' in 7HVP). The remaining ordered

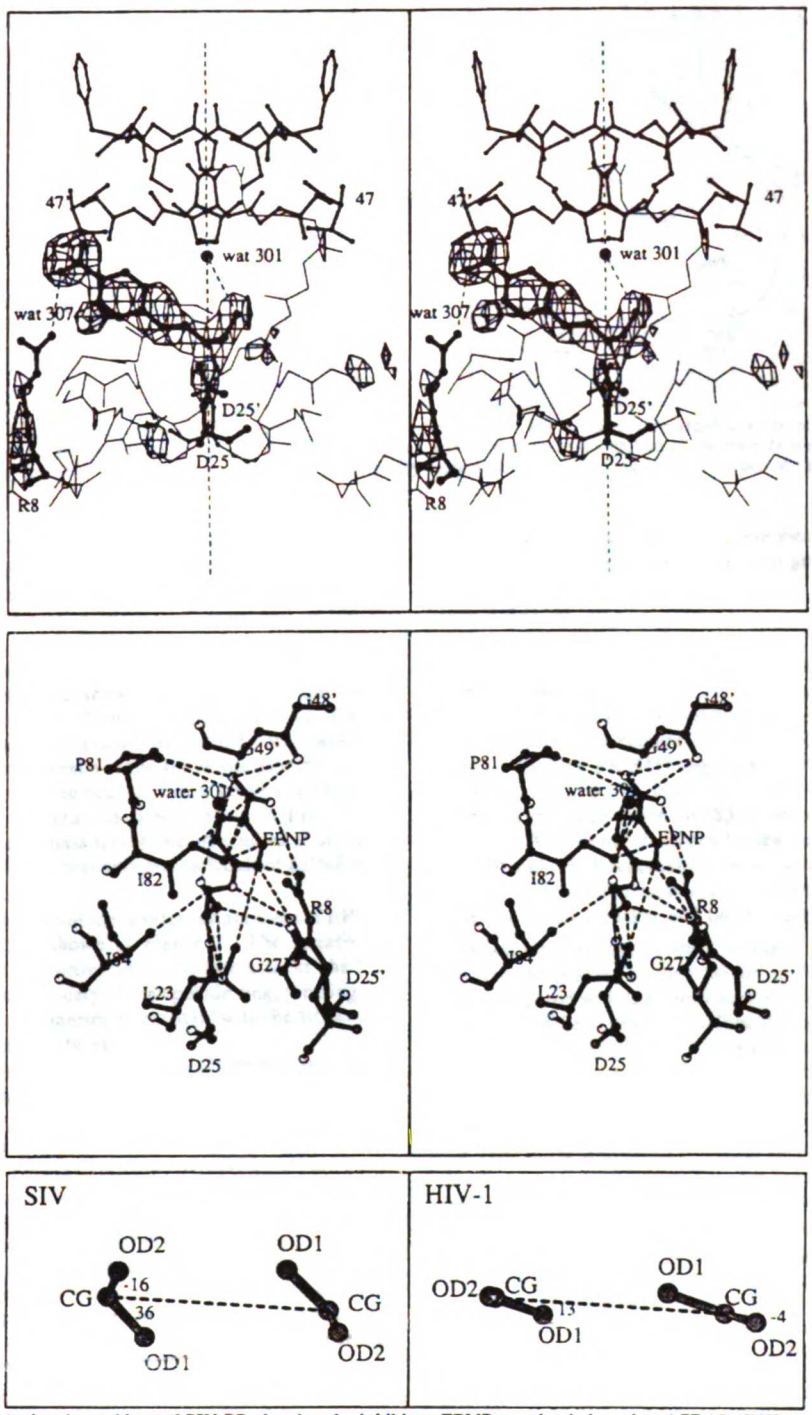


FIGURE 6: (a, top) Active site residues of SIV PR showing the inhibitor, EPNP, covalently bound to ASP 25. Difference density contoured at 1.5σ is shown after the EPNP and water 307 were omitted from the model and a "slow cool" was performed from 500 to 300 K. Drawn in thick lines are EPNP, the active site aspartic acids, and R8, which contacts the nitro group of EPNP. In thinner dark lines are the two flaps of the dimer. Also displayed is water 301, which is hydrogen-bonded to the amide backbone of I50 and I50′ of the flaps. Dotted lines to EPNP show interactions: water 301 is 3.7 Å from the EPNP hydroxyl and the guanidinium group of R8 is 3.6 Å from the nitro group of EPNP. The thin lines show the backbone trace of nearby residues. Because of the low contour level, some noise also appears in the difference map. The vertical dotted line is along the dimer axis. (b, middle) Residues within van der Waals contact (dotted lines) of EPNP in the SIV PR active site. White atoms are oxygens, gray atoms are nitrogens, and black atoms are carbons. (c, bottom) Angles formed by the carboxyl oxygens of the active site aspartic acids with a plane joining the other active site aspartic acid. These angles are formed by the Cγ–Oδ bonds and a plane through the four atoms D25 Cα, D25 Cγ, D25′ Cα, and D25′ Cγ. The angles in the SIV PR structure with EPNP bound are much greater than in the HIV-1 peptidomimetic inhibitor bound structure (7HVP).

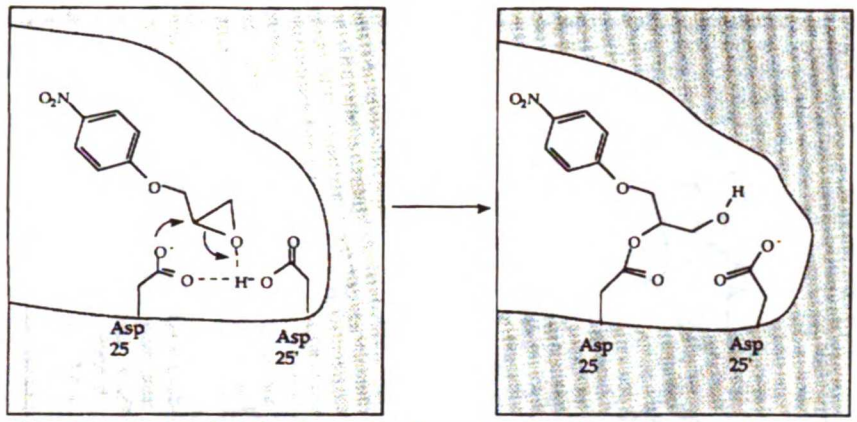


FIGURE 7: Proposed reaction mechanism of EPNP with one active site aspartic acid. The two active site aspartic acids initially share a proton. This proton is donated to the epoxide oxygen while the aspartyl oxygen attacks the secondary carbon of the epoxide ring (the C2 carbon of the propane moiety of EPNP). No reaction is observed at the primary carbon of the epoxide ring (the C1 carbon of the propane moiety of EPNP).

waters in the SIV structure are not in the binding cavity.

Inhibitor Binding Mode. Figure 6a shows difference density in the active site of SIV PR with EPNP and water 307 omitted from the model and after a "slow cool" from 500 to 300 K (Kirkpatrick et al., 1983). EPNP is shown built back into the model. With the occupancy of each orientation of EPNP set to 50%, a difference map showed no density. In addition, when the occupancy was refined, it refined to a value of 40% for each EPNP site. This is consistent with an occupancy of 50% for EPNP bound to each of the active site aspartic acids. Similar density is observed near the other active site aspartic acid of the dimer, though this is not shown in Figure 6a. Chromatographic and mass spectrometric studies of crystallized material confirmed that only one molecule of EPNP was present per dimer.

A proposed mechanism of the covalent attachment of EPNP to aspartic acid 25 is shown in Figure 7. The negatively charged oxygen of the active site aspartate attacks the C2 carbon of the propane moiety of the epoxide ring, forming an ester linkage. This mechanism is consistent with the difference density observed in Figure 6a.

Figure 8 shows both possible positions of EPNP in the active site of SIV protease, viewed down the twofold axis. The position where the peptidomimetic inhibitor from the 7HVP structure would be is also shown for comparison. From this figure, the EPNP is seen to be oriented parallel to the backbone of a bound peptide, with the phenyl group near the P1 (or P1′) sites. This figure also shows that the protease binding cavity would also accommodate EPNP if it were rotated 180 ° so that the phenyl ring occupied the position of the S2 or S2′ subsites. When two EPNP molecules were built into the model in these additional conformations and the occupancy was allowed to refine, the occupancy of the alternate binding mode refined to zero.

Figure 6b is a more detailed view of the contacts between EPNP and the protein. Nonspecific van der Waals interactions are made with residues 23, 27, 48, 49, 81, 82, and 84. The carbonyl oxygen of G27 forms a weak hydrogen bond with the oxygen of the phenoxy moiety of EPNP, 3.1 Å away. The terminal hydroxyl of EPNP, generated by opening of the epoxide ring, is oriented toward water 301, 3.7 Å away. In addition, the guanidinium group of R8 also contacts the nitro group of EPNP, 3.6 Å away. This interaction is too far for a hydrogen bond, but at the pH used for crystallization, pH 6.5, the guanidinium group is protonated and forms an electrostatic interaction with the nitro group oxygens.

DISCUSSION

Characteristics of the SIV PR S4H Structure. As anticipated, the overall fold of SIV PR is identical to that of HIV-1 PR. Gustchina et al. modeled the structure of HIV-2 PR by building from the structure of HIV-1 PR (Gustchina et al., 1991). The SIV PR S4H sequence is 87% identical to that of HIV-2 PR and 51% identical to HIV-1 PR. The model predicted an rms deviation for 185 of 198 residues in the dimer to be 0.28 Å. The rms deviation for the Cα positions between SIV PR and 7HVP is actually 1.16 Å. Most of this difference derives from two β-strands and a random coil loop farthest from the binding pocket and on the surface of the protein.

When the positions of the Cα atoms of the 13 residues forming the peptide binding pocket in HIV-1 PR (7HVP) were compared with their positions in SIV-1 PR, there were significant differences for residues P81 and I82 (see Table III). I82 is a Val in HIV-1 PR, which may contribute to its positional difference. Binding of the different inhibitors in the SIV structure and in the HIV-1 structure may account for some differences.

Unlike the peptidomimetic-bound HIV-1 structure 7HVP, the asymmetric unit of the SIV PR crystals contained one monomer of the functional dimer. The symmetry of the protease dimer allows the dimer twofold axis to be a crystallographic twofold axis. Although EPNP binding to one of the active site aspartic acids introduces local asymmetry in the dimer, the symmetry of the crystal is not affected because each aspartic acid of the dimer is modified with equal probability. A larger asymmetric inhibitor could presumably influence the crystal packing and reduce the symmetry of the crystal.

The monomer subunits of the 7HVP structure are not identical. When the α-carbon backbones are optimally superimposed, the rms deviation of the Cα atoms is 0.6 Å with a rotation of 178.6° instead of the 180° required by a crystallographic twofold axis. (If rotated by 180°, the rms deviation between the Cα atoms of the two monomers is 3.9 Å). Much of this asymmetry can be attributed to the peptide-based inhibitor, which is asymmetrical. In addition, an

FIGURE 8: α-Carbon trace of the SIV PR dimer looking down the twofold axis. Two EPNP molecules are shown as they are oriented in the active site when bound to each of the two aspartic acids (white in black outline). Superimposed is the peptidomimetic inhibitor that was crystallized with HIV-1 in the 7HVP structure (in bold black). (The superposition uses the Cα postions of the 53 core residues, as described in the text). The P4 to P3′ peptide binding sites are labeled.

asymmetrical hydrogen bond is formed between the flaps of the dimer: the amide nitrogen of G51 donating the hydrogen to the carbonyl oxygen of I50′ (I150 according to the 7HVP numbering). This is only possible because the carbonyl of I50′ points toward its dimer-related flap, while the carbonyl of I50 points away from its dimer-related flap. In the SIV structure both of these carbonyl oxygens point away from their dimer-related flaps. No density can be seen for the other conformation of the carbonyl, even if it is built into the model.

The histidine side chain present in SIV PR S4H may have had two effects on the crystallization of the protease. The mutation clearly stabilizes the protein to autolysis, thereby increasing sample homogeneity which may facilitate better crystal formation. H4 also makes up part of the surface involved in a crystal contact. The His side chain points into solvent and participates in the crystallographic contact surface formed by the N- and C-terminal β-sheet. It is therefore possible that the S4H mutation influenced the ease of crystallization of SIV PR.

The flaps of the EPNP-bound SIV structure were in their fully closed position, as in the 7HVP structure which has a peptidomimetic inhibitor bound (Swain et al., 1990). Interactions between EPNP and the flaps are weak, consisting of van der Waals interactions, and yet the conformation of the flaps is closed: the Cα of I50 shifts by 0.4–0.9 Å from the 7HVP position. It is possible that the conformation of the flaps is influenced by crystallographic contacts. Though the present crystal packing arrangement would accommodate the flaps in the open conformation, the close van der Waals interactions in the present structure may form more favorably in the flaps-closed conformation, thus favoring this conformation during crystal growth. The conformation of the flaps in the HIV-2 PR structures with peptide-based inhibitors appears also to be closed, very similar to the SIV PR structure (Mulichak et al., 1993).

Another difference in structure is the conformation of the loop preceding the flaps, residues 33–45, which may influence the conformation of the flaps. The position of the prolines in SIV PR differs in this loop from HIV-1 and HIV-2 PRs: residue 40 is a Pro in SIV PR, residue 39 is a Pro in HIV-1 PR, and there is no Pro at either of these positions in HIV-2 PR. All three sequences have a conserved Pro at residue 44. Pro 40 is responsible for the large difference in position of this loop in the SIV PR structure relative to the HIV-1 structure.

The position of the tightly bound water 301 (B factor of 9), hydrogen-bonded between the flaps, is the same as that seen in the HIV structures. The other waters in the SIV PR S4H binding pocket, 307 and 313, occupy positions near where a bound peptide would form hydrogen bonds. We would expect the carbonyl oxygens and amide nitrogens of the protein to be hydrogen-bonded to waters if possible when they are not interacting with a peptide, though no ordered water is observed near most of these in the EPNP-bound structure.

The conformation of the N-terminal three residues of SIV PR is shifted away from the body of the protein relative to the conformation in HIV-1 PR (7HVP): the Cα of the N-terminal proline is 1.9 Å from its position in 7HVP, and the Cγ position is 3.5 Å from its position in 7HVP. These

three residues participate in a four-stranded β-sheet with the other N-terminal and C-terminal residues of the dimer. Protein packing may influence the conformation of these residues. Hydrogen bonding is not disrupted in the β-sheet configuration in the SIV structure. In contrast, two hydrogen bonds are disrupted with the backbone amide and carbonyl of F3 in the HIV-2 PR structure (Mulichak et al., 1993). The N-terminal residues also participate in crystal contacts in this structure.

Binding Mode of EPNP. EPNP is a specific inhibitor of aspartyl proteases which labels the active site aspartic acids (Tang, 1971). Epoxides undergo ring opening by nucleophilic attack at either the primary or the secondary carbon. At neutral or basic pH, the less substituted carbon is attacked more readily. At acidic pH, the epoxide is protonated and the secondary carbon forms the more stable carbocation and is preferentially attacked (March, 1985). Nucleophilic attack on 1,2-epoxides in acidic solutions is known to occur preferentially at the 2-position of the propane moiety (Parker & Isaacs, 1959).

The only other crystal structure of EPNP complexed with an aspartyl protease is of penicillopepsin (James et al., 1977). This structure revealed three sites where EPNP bound to the protein: one noncovalently associated with the protein outside the active site and one bound to each of the active site aspartic acid residues. From the electron density map, James et al. concluded that one aspartic acid reacted at the C2 position of the propane moiety, while the other aspartic acid reacted at the C1 position. The difference density in the SIV PR S4H structure shows EPNP bound at the C2 position of the propane moiety. The density around the EPNP is consistent with there being one EPNP molecule bound per protease dimer.

To catalyze the ring opening of the epoxide at the C2 position, the protease must satisfy three conditions: A strong nucleophile is required; this is provided by the active site aspartic acid. The EPNP must be positioned favorably for a reaction to occur (Tang, 1971), and a proton donor must be available.

At pH 5.4 one of the SIV PR active site aspartic acids is protonated and the other is unprotonated (Grant et al., 1991). A hydrogen shared between the two aspartic acids raises the pK_a of the second proton from this pair of aspartic acids. Therefore, initially only one aspartic acid is available for a reaction with EPNP. Because the secondary carbon of the epoxide ring becomes covalently linked to the aspartic acid, the epoxide oxygen must be protonated. An appealing source for this proton is the hydrogen that was shared between the active site aspartic acids (see Figure 7). This hydrogen bond is clearly disrupted after the EPNP reacts, as demonstrated in Figure 6c. After the initial EPNP reacts, the second aspartic acid is accessible to solvent, and its pK_a probably reduces to its solvent-exposed value of 4.0. At pH 5.4 it is therefore predominantly unprotonated and could therefore act as a nucleophile. A second EPNP does not react at this site, perhaps because no proton is available to donate to the epoxide oxygen. Also, the positioning of the C2 carbon of a second EPNP at this aspartic acid is sterically unfavorable: the C1 carbon of the first EPNP would then be only 2.5 Å from the C1 carbon of the second EPNP. Reaction of the second EPNP at the C1 position of the propane moiety is not sterically prohibited but is not observed. An aspartic acid is not generally a strong enough nucleophile to perform this reaction, and it is unclear why the reaction occurs in the case of penicillopepsin.

EPNP orients itself in the binding pocket of the protease parallel to where a peptide backbone would lie, with the phenyl ring reaching into the S1 (or S1') site. Tang proposed that EPNP reacts with pepsin by occupying the hydrophobic binding site which would be occupied by a substrate (Tang, 1971). HIV and SIV proteases prefer hydrophobic residues in the S1 and S1' subsites (Poorman et al., 1991). The S1 and S1' binding pockets are composed of residues L23, D25, G27, I50, P81, V82, and I84 in HIV-1 (Miller et al., 1989). All of these except I50 form van der Waals contacts with EPNP, with distances very similar to those found for the P1 and P1' sites in the 7HVP peptidomimetic-bound structure. The side chain of I50 is out of reach of the EPNP. Residue 82 is an Ile in the SIV sequence and makes a closer contact with EPNP than there would be if it were a Val, as in HIV-1.

Other than the reactive aspartic acid, the only specific interactions with EPNP are weak: the phenoxy oxygen of EPNP and the carbonyl oxygen of G27, and electrostatic attraction between the nitro group of EPNP and the guanidinium group of R8. R8 is associated with the S3 binding pocket (Miller et al., 1989). Water 301 is too far from the EPNP hydroxyl group to form a hydrogen bond. Its hydrogens are probably oriented toward the EPNP hydroxyl since its lone-pair electrons are hydrogen-bonded to the amide nitrogens of I50 and I50' of the flaps. This would allow van der Waals interactions between the hydrogens of water 301 and the hydrogen of the hydroxyl of EPNP.

For a covalent inhibitor like EPNP, the K_{inact} reflects the noncovalent interactions with the protein. K_{inact} is the concentration of inhibitor that results in half the maximal rate of inactivation. The weakness of these interactions explains why K_{inact} is 8 mM.

This structure represents the first structure of a covalently modified protease closely related to the HIV-1 PR and can be used as a guide for drug design efforts. Starting from the known position of EPNP, derivatives may be designed in which further substituents are added onto the EPNP scaffold to take advantage of binding pockets in the enzyme. Enhancing the hydrophobic and van der Waals interactions which already exist, or building specific hydrogen-bonding interactions, should decrease the K_{inact}. Such substituents may increase both the affinity and specificity of these compounds of retroviral enzymes relative to cellular aspartyl proteases. In addition, a comparison of the conserved surfaces between binding pockets of the SIV and HIV-1 PRs can provide a template for the design of irreversible inhibitors with specificity for the HIV-1, HIV-2, and SIV proteases. Such inhibitors may be less susceptible to the development of resistance by the virus.

ACKNOWLEDGMENT

We acknowledge Earl Rutenber and Janet Finer-Moore for helpful discussions on crystallography and Julie Newdoll for assistance with graphics programs. Mass spectrometry was performed by Zhoughua Xu in the University of California San Francisco Mass Spectrometry Facility (A. L. Burlingame, Director), which is supported by National Institute of Environmental Health Sciences Grant ES 04705.

REFERENCES

Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Cameron, C. E., Grindes, B., Jacques, P., Jentoft, J., Leis, J., Wlodawer, A., & Weber, I. T. (1993) *J. Biol. Chem. 268*, 11711–11720.

Chambers, J., & Stroud, R. M. (1979) *Acta Crystallogr. B35*, 1861–1874.
Fauman, E. B. (1993) Ph.D. Thesis, University of California, San Francisco.
Finer-Moore, J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T., & Stroud, R. M. (1992) *Proteins 12*, 203–222.
Grant, S. K., Deckman, I. C., Minnich, M. D., Culp, J., Franklin, S., Dreyer, G. B., Tomaszek, T. A., Jr., Debouck, C., & Meek, T. D. (1991) *Biochemistry 30*, 8424–8434.
Gustchina, A., Weber, I. T., & Wlodawer, A. (1991) in *Structure and Function of the Aspartic Proteinases* (Dunn, B. M., Ed.) pp 549–553, Plenum Press, New York.
James, M. N. G., Hsu, I., & Delbaere, L. T. J. (1977) *Nature 267*, 808–813.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157–171.
Kestler, H., Kodama, T., Ringler, D., Marthas, M., Pedersen, N., Lackner, A., Regier, D., Sehgal, P., Daniel, M., King, N., & Desrosiers, R. (1990) *Science 248*, 1109–1112.
Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983) *Science 220*, 671–680.
March, J. (1985) in *Advanced Organic Chemistry*, Wiley, New York.
Miller, M., Jaskolski, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1989) *Nature 337*, 576–579.
Mulichak, A. M., Hui, J. O., Tomasselli, A. G., Heinrikson, R. L., Curry, K. A., Tomich, C., Thaisrivong, S., Sawyer, T. K., & Watenpaugh, K. D. (1993) *J. Biol. Chem. 268*, 13103–13109.
Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Lea, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P. (1989) *Nature 337*, 615–620.
Parker, R. E., & Isaacs, N. S. (1959) *Chem. Rev. 59*, 7373–7399.
Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F., & Stroud, R. M. (1990) *Proteins 8*, 315–333.
Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L., & Kézdy, F. J. (1991) *J. Biol. Chem. 266*, 14554–14561.
Richman, D. D. (1993) *Antimicrob. Agents Chemother. 37*, 1207–1213.
Rosé, J. R., Salto, R., & Craik, C. S. (1993) *J. Biol. Chem. 268*, 11939–11945.
Rutenber, E., Fauman, E. B., Keenan, R. J., Fong, S., Furth, P. S., Ortiz de Montellano, P. R., Meng, E., Kuntz, I. D., DeCamp, D. L., Salto, R., Rosé, J. R., Craik, C. S., & Stroud, R. M. (1993) *J. Biol. Chem. 268*, 15343–15346.
Sack, J. (1988) *J. Mol. Graphics 6*, 225.
Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B. H., & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805–8809.
Tang, J. (1971) *J. Biol. Chem. 246*, 4510–4517.
Toth, M. V., & Marshall, G. R. (1990) *Int. J. Pept. Protein Res. 36*, 544–550.
Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616–621.

CHAPTER 2:

IRREVERSIBLE INHIBITORS OF HIV PROTEASE

Introduction:

As part of the rational drug design effort at the University of California, San Francisco (UCSF) involving HIV protease, I assayed and crystallized a number of potential inhibitors of HIV protease. The major effort was to develop an irreversible inhibitor that modified the catalytic aspartic acids. The most promising inhibitors being developed at the time were reversible inhibitors, and the concern was that resistance would develop to these inhibitors through mutations in the protease as had occurred with reverse transcriptase. Resistance mutations did appear in tissue culture (Roberts 1995). An irreversible inhibitor might be less sensitive to mutations since it would only have to bind well enough to react with the catalytic aspartic acids. Because modification of the aspartic acids inactivates the protease, resistance mutations cannot develop involving these residues (Kohl, Emini et al. 1988). The structure of SIV protease with 1,2-epoxy-(*p*-nitrophenoxy)propane (EPNP) was an example of an irreversible inhibitor (Rose, Rosé et al. 1993), but EPNP inactivates all aspartyl proteases and is therefore not a suitable drug. Haloperidol derivatives had been studied at UCSF as specific inhibitors of the protease (DesJarlais, Seibel et al. 1990). They are selective for HIV protease: the thioketal derivative of haloperidol, UCSF8 (derivatives of haloperidol synthesized at UCSF have been assigned a UCSF number), is selective for HIV-1 protease by 40-100 fold over pepsin, renin, and cathepsin D, as measured by IC_{50}'s (Rutenber, Fauman et al. 1993). The IC_{50}'s of haloperidol for HIV-1 and HIV-2 proteases are 125mM and 140mM, respectively (DesJarlais, Seibel et al. 1990). Reactive groups were added to haloperidol derivatives to find an irreversible inhibitor that was also selective for HIV protease (De Voss, Sui et al. 1994). Zhonghua Yu (a graduate student in the

Alma Burlingame lab) examined the protease after reaction with the compounds by reverse phase HPLC and mass spectrometry. This technique had successfully corroborated that EPNP covalently modified the catalytic aspartic acid of HIV-1 protease (Rose, Rosé et al. 1993). Five halperidol derivatives were tested for crystallization, including two epoxides (UCSF 70 and UCSF 84), two ynones (UCSF 86 and UCSF 142), and an α,β-unsaturated ketone (UCSF 115) (Figure 1). A boronated curcumin derivative was also tested (Sui, Salto et al. 1993).

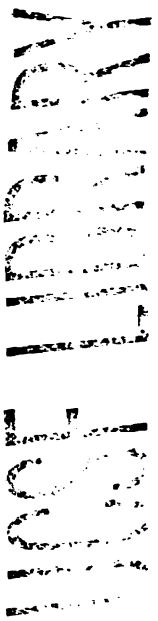
Another effort was to find new scaffolds to use for inhibitor design, since modifications of the haloperidol lead had proven difficult to improve to better than a micromolar inhibitor. A thioketal derivative of haloperidol, UCSF 59, one of the tightest binding modifications of haloperidol, was also crystallized with SIV protease. Several promising inhibitors were crystallized with the protease, including curcumin, rose bengal, and erythrocin isothiocyanate. (Figure 2). SIV protease was used in these crystal trials because it is more easily and reliably purified than HIV-1 protease. We also had some early indications that compounds that did not crystallize with HIV-1 protease did form crystals with SIV protease.

Materials and Methods

Synthesis of the haloperidol-based derivatives has been described elsewhere (De Voss, Sui et al. 1994). Rose bengal was purchased from Aldrich (91% pure). Erythrosin isothiocyanate was purchased from Sigma. Curcumin was purchased from Janssen Chimica. Boronated curcumin was synthesized as described previously (Sui, Salto et al. 1993). Biliverdin dihydrochloride was purchased (from Sigma or Aldrich) and purified by Patricia Caldera. All inhibitors were initially dissolved in DMSO.

An IC_{50} for the irreversible inhibitors was measured as the concentration of inhibitor that resulted in half the maximal rate of inactivation. First the protein was preincubated with the inhibitors in a total volume of 100ml. Unless otherwise specified, the final composition of the inhibition buffer was 1M sodium chloride, 50mM Hepes buffer, pH 8.0, 1mM EDTA. The inhibitors were dissolved in DMSO and added to the sample at less than the solubility limit so that the final DMSO concentration was 5-10%. Finally 50ml of the protein (about 100-200mg/ml as measured by the BioRad Bradford assay with BSA as a standard) was added to the sample and preincubated for various times. The activity was measured by removing 5ml of the sample and adding it to 5ml of 10mM substrate, and 1ml of reaction buffer: 100mM Hepes, pH 8.0, 1mM EDTA, 1M NaCl. The substrate, R-V-Nle-F(NO_2)-E-A-Nle-S-NH_2, is monitored over time at 300nm for changes in absorbance as a result of bond cleavage between the norleucine and para-nitro-phenyl residues (Tomaszek, Magaard et al. 1990). Absorbance was monitored with a Kontron UVIKON 860. Cleavage rates from inhibited protease were compared with the uninhibited protease.

Protein was purified by Nancy Douglas, Nick Endre, and Gian-Carlo Ochoa according to the usual protocol (Rose, Rosé et al. 1993). The crystallization procedure was as described for the unliganded SIV protease (see chapter 4). Unless otherwise specified, sodium hydroxide was added to the protein after thawing to enhance its solubility. Data was collected on the in-house R-axis II image plate detector and reduced using the R-axis software or DENZO and SCALEPACK (Otwinowski 1993). XPLOR was used for refinement (Brünger, Kuriyan et al. 1987), and CHAIN was used for visualization and manual rebuilding of structures (Sack 1988).



Results and Discussion:

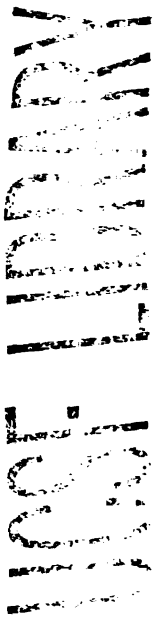
Epoxide derivatives of haloperidol:

UCSF 70 and UCSF 84 differ in the orientation of the piperdine ring: in UCSF 84, like haloperidol, the piperdine nitrogen is bound to the acyclic chain. Also like haloperidol, the cyclic substituent of the piperdine ring of UCSF 84 is chlorinated. The cyclic substituent of UCSF 70 contains an extra bond in the linker to the piperdine nitrogen and is unchlorinated. This carbon was removed from the 4-carbon alkyl chain of haloperidol. Because of the triple bond in the acyclic chain, these compounds cannot adopt the same orientation as haloperidol does when bound to HIV-1 protease (Rutenber, Fauman et al. 1993).

UCSF 70 negligibly inhibited SIV protease at 1mM concentration after 3 hours, in 10% DMSO. The IC_{50} for this inhibitor had been measured for HIV-1 and HIV-2 protease as 725mM and 350mM, respectively (De Voss, Sui et al. 1994). Our assay method would not measure reversible inhibition because the inhibited protein was diluted 200-fold prior to the determination of residual

activity. Any reversible inhibitor would re-equilibrate after dilution. Though UCSF 70 inhibits both proteases, it only irreversibly inhibits HIV-1 protease (De Voss, Sui et al. 1994; Salto, Babe et al. 1994). The implication is that UCSF 70 does not modify the catalytic aspartic acids and, like haloperidol, is reversibly inhibiting the proteases. The irreversible inhibition of HIV-1 protease may be at the two cysteine residues, which are not present in HIV-2 and SIV proteases. It is also possible that this inhibitor was unstable over time. The integrity of the inhibitor was not tested at the time of these assays.

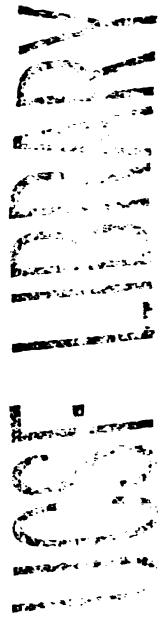
UCSF 84 inhibited SIV protease with a half time of 2 hours, at 500mM, pH 8 and 5% DMSO. We measured a half time of about 1 hour for HIV-2 protease. The rate was similar at pH 5.4 and pH 6.5. This was tested to assure that the adduct would be stable at the crystallization conditions. However, unlike the EPNP result (Rose, Rosé et al. 1993), no protein adduct could be identified by reverse phase chromatography; only a peak for the unmodified protease monomer resulted (Zhonghua Yu, personal communication). Crystals were set up after incubating the protein with the inhibitor for 12 hours at 500mM concentration. The final protein concentration for the crystal trial was 1mg/ml, which is lower than the usual concentration for crystallization. No crystals grew.



Ynone derivatives of haloperidol:

UCSF 86 is identical to haloperidol with the modification of the 4-carbon alkyl chain to an ynone and the removal of the fluorine from the fluorophenyl ring. In UCSF 142, the fluorophenyl ring has been replaced by an *m*-biphenyl group. The biphenyl at this position had been shown to improve the affinity of haloperidol for HIV-1 protease (De Voss, Sui et al. 1994).

UCSF 86 displayed no irreversible inhibition of SIV protease after 80 minutes. Though the inhibitor was not tested at the time of these assays to assure its integrity, the compound had been stored in the freezer as a powder prior to use. This inhibitor was not soluble at 1mM concentration in aqueous solution. Its activity was tested at 2 concentrations: 1mM and 400mM. We were unable to replicate the reversible inhibition of HIV-2 protease at 400mM inhibitor concentration. (The IC_{50} of UCSF 86 for HIV-1 and HIV-2 have been reported as 20mM and 33mM, respectively (De Voss, Sui et al. 1994)). UCSF 86 was reported to have irreversibly inhibited HIV-1 and HIV-2 proteases, though its rate of inactivation of HIV-1 protease was 50 times that of HIV-2 protease (De Voss, Sui et al. 1994; Salto, Babe et al. 1994). Zhonghua Yu, et al have subsequently shown that this inhibitor (referred to as compound 8 in their paper) preferentially modifies Cys 95, then Cys 67, and finally the N-terminus in about 10% of the protease monomers (Yu, Caldera et al. 1996). It is not known whether the inhibition of HIV-2 protease results from modification of the aspartic acids, though modification of the N-terminus may also explain the observed irreversible inhibition.

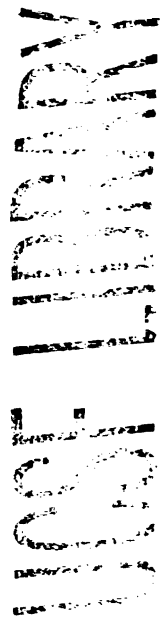


UCSF 142 did not irreversibly inhibit SIV protease at 250mM in 5% DMSO. Again the compound had been stored in the freezer as powder prior to use. Because the inhibitor was not soluble at 250mM, it was also tested at 100mM, at which it was soluble in 10% DMSO. UCSF 142 was reported to have irreversibly inhibited HIV-1 and HIV-2 proteases (De Voss, Sui et al. 1994; Salto, Babe et al. 1994), though the rate of inactivation of HIV-1 protease was reported 6-fold higher by De Voss, et al than by Salto, et al. Despite the inability to inhibit SIV protease, a minor population of the protease was modified by UCSF 142, as shown by reverse phase chromatography followed by mass spectrometry of the collected peaks (Zhonghua Yu, personal

communication). Zhonghua Yu, et al showed that UCSF 142 (referred to as compound 9 in their paper) modified the two cysteines of HIV-1 protease with no indication of other modifications (Yu, Caldera et al. 1996). The reported inhibition of HIV-2 and the modification of SIV protease presumably occurs at a site other than the catalytic aspartic acids, possibly the N-terminus.

α,β-unsaturated ketone derivatives of haloperidol:

UCSF 115 contains a double bond adjacent to a carbonyl group along the 4-carbon alkyl chain of haloperidol. The piperdine nitrogen is also modified to the N-oxide. At 1mM concentrations of UCSF 115, no inhibition of SIV protease could be observed after 3 hours, in 10% DMSO. SIV and HIV-2 proteases were tested with a separate aliquot of inhibitor at a final concentration of 10mM, in 10% DMSO. After 2 hours there was no inhibition of either protease. Irreversible inhibition of HIV-1 and HIV-2 proteases have been reported, with the rate of inactivation of HIV-1 protease 10 to 20 times faster than HIV-1 protease (De Voss, Sui et al. 1994; Salto, Babe et al. 1994). Zhonghua Yu, et al showed that the similar compound, without the oxygen of the N-oxide (compound 10 in their paper), modified the cysteines of HIV-1 protease and not the catalytic aspartic acids (Yu, Caldera et al. 1996). This compound had been stored in DMSO, and no powder form remained at the time of these assays. It is not known whether the compound had decayed during storage.



Curcumin:

Curcumin, an ingredient of the spice turmeric, has been shown to have anti-viral effects (Sui, Salto et al. 1993). Curcumin was reported to be a

reversible inhibitor of HIV-1 and HIV-2 proteases with IC_{50}'s of 100mM and 250mM, respectively. In the irreversible inhibitor assay, SIV protease was 20% inhibited by 300mM of curcumin in 150 minutes, despite reports that it reversibly inhibits HIV proteases. The same study reported that boronated curcumin is an irreversible inhibitor of HIV-1 and HIV-2 proteases, with IC_{50}'s of 6mM (Sui, Salto et al. 1993). The chloride salt of boronated curcumin (reported to be more soluble than the perchlorate salt - personal communication Zhihua Sui) was soluble in aqueous solution to greater than 20mM and less than 50mM in 5% DMSO. Boronated curcumin did not inhibit SIV protease after 90 minutes at 20mM concentration.

Gian-Carlo Ochoa (a Masters student in the Craik lab) set up crystals of SIV protease with curcumin (Ochoa 1994). Crystals grew to a size of 750 microns in length and looked pyramidal. The crystals appeared to be yellow to the eye, suggesting that curcumin was bound. Table I shows the statistics for data collection and refinement (see below for the structure determination).

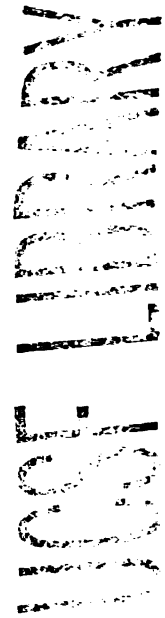
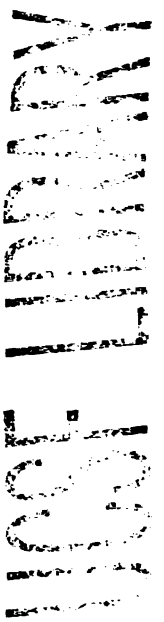


Table I: Statistics for data collection and refinement of SIV protease crystallized with curcumin, xantho-bilirubin, UCSF 59, and unliganded.

peptide	curcumin	xantho-bilirubin	UCSF 59	unliganded
space group reduced in	$P3_221$	$P3_221$	$P3_221$	$P3_221$
resolution (Å)	2.2	2.5	2.2	2.0
no. of reflections ($I \geq 1s(I)$)	31,700	7,720	29,900	35,300
no. unique reflections	10,170	3,160	6,700	7,300
completeness (%)	98	87	93	97
completeness in highest res. bin	96	88	89	94
[a] R_{symm} (%)	6.9	6.4	4.6	8.6
R-factor (from 7Å)	21	[b] 26	[b] 23	20.4
R-free (from 7Å)	31			26.6
unit cell				
a	45.1	46.3	45.3	45.4
b	45.1	46.3	45.3	45.4
c	89.1	89.5	87.9	87.9

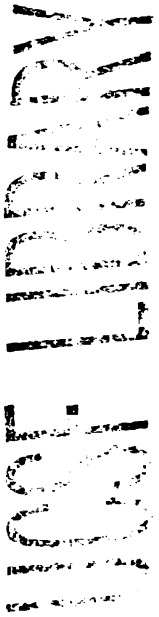
[a] $R_{symm} = \Sigma_i \Sigma_j | \langle I_i \rangle - I_{ij} | / \Sigma_i \Sigma_j I_{ij}$

[b] after rigid body minimization, without positional refinement



Biliverdin:

Biliverdin and bilirubin were identified in a computer search as potential inhibitors of HIV protease by Guy Bemis. Bemis developed a computer algorithm to search a chemical data base for compounds that matched a known ligand of a protein. The search with the Merck inhibitor L700,417 (a low nanomolar inhibitor of HIV-1 protease) identified biliverdin and bilirubin as having similar connectivity and distance constraints. Others have shown that biliverdin has anti-viral effects against HIV-1 in cell culture (Mori, Otake et al. 1991). Though biliverdin is too large to fit in the binding pocket of the protease, at least with the flaps closed, it inhibited SIV and HIV-2 proteases with an IC_{50} of 8μM, and HIV-1 protease with an IC_{50} of 2μM (Ochoa 1994).



Gian-Carlo Ochoa grew crystals of SIV protease with biliverdin (Ochoa 1994). Crystals grew to a size of 600 microns in length and looked pyramidal. The crystals appeared to be green to the eye, suggesting that biliverdin was bound. Table I shows the statistics for data collection and refinement (see below for the structure determination).

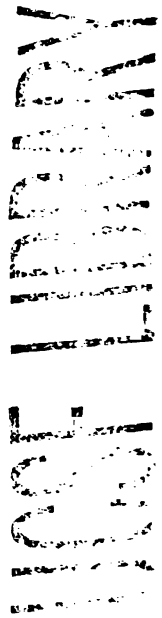
Xantho-bilirubin:

Fiona McPhee (a post-doc in the Craik lab) found that the IC_{50} of xantho-bilirubin was the same as biliverdin (Fiona McPhee, personal communication). Xantho-bilirubin is half the biliverdin molecule. Co-crystals were grown of SIV protease with xantho-bilirubin. Table I shows the statistics for data collection and refinement (see below for the structure determination). Co-crystals were also grown with HIV-1 protease. These crystals were bipyramidal, in space group $P4_12_12$, similar to the unliganded "flap open" structure of HIV-1 protease crystals (Wlodawer, Miller et al. 1989)

and with UCSF 8 bound to HIV-1 protease, in which the flaps were partially opened (Rutenber, Fauman et al. 1993). The flaps were not well defined in this structure, and there was no density in the active site for a ligand.

Rose Bengal:

Rose Bengal inactivates HIV in the presence of visible light by inhibiting viral fusion (Lenard, Rabson et al. 1993). It also irreversibly inhibits SIV protease. McPhee showed that the inhibition by rose bengal was light sensitive (personal communication). We first determined that rose bengal was soluble at 100mM in 5% DMSO. At this concentration, the compound inhibited the protease with a half time of 45 to 60 minutes. SIV protease (at about 150mg/ml) was incubated overnight in the presence of 100mM rose bengal in preparation for crystal trials. The next day, the majority of the protein had precipitated with the inhibitor. Hanging drops were set up despite there being no protein (as measured by Bradford), but no crystals grew. The rate of inactivation was approximately the same rate in colored tubes as transparent tubes, suggesting that inhibition of protease is independent of the formation of singlet oxygen (Lenard, Rabson et al. 1993). Rose bengal is very hydrophobic and may bind to the protease non-specifically in several places, causing aggregation. This behavior could appear as irreversible inhibition. To confirm this hypothesis, analytical ultracentrifugation could be used to test the size of the protease adducts.

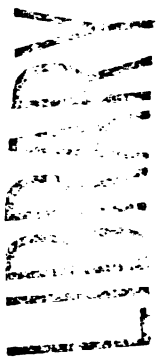


Erythrocin Isothiocyanate:

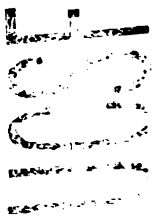
Erythrocin isothiocyanate also irreversibly inhibited SIV protease, though only in transparent tubes and not in colored tubes. McPhee also realized that inhibition by this compound was light sensitive (personal

communication). By protection experiments with a tight binding peptidomimetic inhibitor from Upjohn, U-85548, known to bind at the active site of the protease, McPhee determined that erythrocin isothiocyanate did not bind at the active site of HIV-1 protease. Similar experiments with SIV protease showed that erythrocin isothiocyanate did compete with the UpJohn inhibitor.

The compound was soluble at 100mM in 5% DMSO. At this concentration it inhibited SIV protease with a half time of less than 10 minutes. A similar incubation was set up overnight as with rose bengal in preparation for crystal trials. Like rose bengal, there was a lot of precipitate in the sample by morning. The concentration of the protein could not be determined by Bradford because the inhibitor reacted with the Bradford reagent. The sample was concentrated and hanging drops were set up, but no crystals grew.

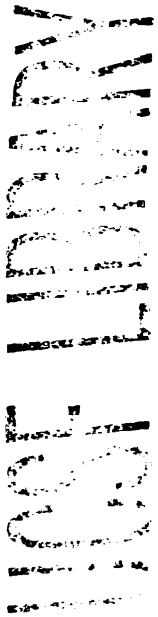


When the inhibition took place at pH 7.5, no protein eluted from the reverse phase column (Zhonghua Yu, personal communication). At pH 5.5, two peaks eluted from the column, as would be expected if only one protease monomer were labeled. When the peaks were monitored at 525nm to detect the erythrocin isothiocyanate, both peaks were labeled with the inhibitor. ESIMS mass spectrometry of the two peaks detected only protein in each of the two peaks. The implication is that erythrocin isothiocyanate irreversibly inhibits the protease by non-specifically binding it. The inhibitor must bind tightly enough to associate with the protease during reverse phase, but dissociate during the ESIMS procedure. This explanation is consistent with the precipitation of protein over time.



UCSF 59:

UCSF 59 is a thioketal-*m*-biphenyl derivative of haloperidol. It is one of the modifications of haloperidol with the tightest affinity for HIV-1 protease: IC_{50} of 7mM for HIV-1 protease and 45mM for HIV-2 protease. The IC_{50} for SIV protease was measured as 32mM (Ochoa 1994). Pyramid shaped crystals were also grown of SIV protease with UCSF 59 at a protein concentration of 10-15mg/ml, as measured by Bradford with IgG as a standard. Crystals grew to a length of 700 microns with a pyramidal morphology. Table I shows the statistics for data collection and refinement (see below for the structure determination).



Crystal structure determination with curcumin, xantho-bilirubin, and UCSF 59:

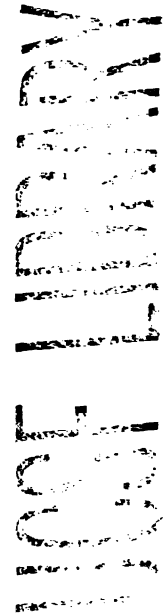
Initially we were very optimistic about the SIV protease crystallizes with biliverdin, xantho-bilirubin, curcumin, and UCSF 59. The biliverdin and curcumin crystals were colored, suggesting that inhibitor had bound. The crystals were large and diffracted well. In addition, crystal trials with HIV-1 protease had not yielded crystals. One disturbing fact was that crystals only grew at relatively low inhibitor to protein ratios: one-to-one or two-to-one at best. At higher inhibitor concentrations, crystals did not form.

The morphologies of all of these crystals were similar: pyramidal or bipyramidal instead of plates as was found for SIV protease crystallized with EPNP. Table I compares the unit cell dimensions of these crystals with those of the unliganded SIV protease structure crystallized similarly. Chapter 4 describes the space group determination and molecular replacement solution for the unliganded structure. After rigid body refinement, none of these structures showed difference density greater than 2.5s in the active site, except

for a water molecule shared between the catalytic aspartic acids. An |Fo|-|Fo| difference density map between the unliganded SIV protease and SIV protease bound to curcumin showed no difference density for the ligand.

Conclusions:

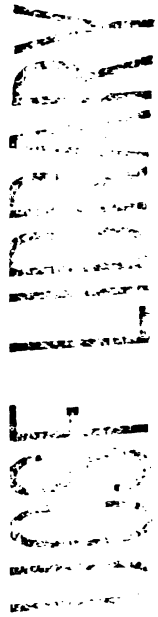
None of the irreversible haloperidol inhibitors reacted with the catalytic aspartic acids. Since the time of these experiments, Zhonghua Yu, et al have succeeded in synthesizing an epoxide that does react with the aspartic acids preferentially (Yu, Caldera et al. 1996). They conclude that α,β-unsaturated ketones are too reactive to be good inhibitors, modifying the cysteine and N-terminus of the protease instead of the aspartic acids. *Trans*-1,2-disubstituted epoxides are too unreactive, while *cis*-1,2-disubstituted epoxides are appropriately reactive. The spiro epoxides (UCSF 70 and UCSF 84) did not appear to interact with the aspartic acids. The adduct with UCSF 84 was unstable to the HPLC and mass spectrometry treatment. The ynones also reacted preferentially with the cysteines and N-terminus of the protease. One reason that the aspartic acids are not modified preferentially may result from the reversible binding of haloperidol. Because haloperidol binds the protease weakly, and in several conformations (Rutenber, Fauman et al. 1993), it may not be a good scaffold for targeting the aspartic acids.



Both rose bengal and erythrosin isothiocyanate precipitate the protease over time. They may inhibit the protease by forming nonspecific aggregates. If so, these would not make good starting points for drug design.

The crystals with curcumin, xantho-bilirubin, and UCSF 59 that did form in space group $P3_221$ were all unliganded. The biliverdin structure, refined by Earl Rutenber, had some density in the active site which was not defined enough to build the inhibitor into. Several explanations for this are

possible. The crystals do not form at high inhibitor to protease stoichiometries. They may therefore only form when the inhibitors are not bound. The protease in this crystal form is in the open conformation, which the inhibitors, such as UCSF 59, may have less affinity for than the closed conformation. And the high pH treatment of the protein may disturb the structure in a way that prevents binding of the inhibitors. If this last explanation were true, it was not apparent in the structure of the unliganded protease. Our initial optimism regarding the ability to grow crystals with SIV protease with inhibitors that would not crystallize with HIV-1 protease has not been realized. New conditions could be screened that might allow these inhibitors to crystallize with the protease. A larger problem may be that inhibitors based on haloperidol are not specific enough to be good drug leads.



References:

Brünger, A. T., J. Kuriyan, et al. (1987). *Science* **235**: 458-460.

De Voss, J. J., Z. Sui, et al. (1994). "Haloperidol-based irreversible inhibitors of the HIV-1 and HIV-2 proteases." *J. Med. Chem.* **37**: 665-673.

DesJarlais, R. L., G. L. Seibel, et al. (1990). "Structure-based design of nonpeptide inhibitors specific for the Human Immunodeficiency Virus 1 protease." *Proc. Natl. Acad. Sci. U.S.A.* **87**: 6644-6648.

Kohl, N. E., E. A. Emini, et al. (1988). "Active human immunodeficiency virus protease is required for viral infectivity." *Proc. Natl. Acad. Sci. U.S.A.* **85**: 4686-4690.

Lenard, J., A. Rabson, et al. (1993). "Photodynamic inactivation of infectivity of human immunodeficiency virus and other enveloped viruses using hypericin and rose bengal: inhibition of fusion and syncytia formation." *Proc. Natl. Acad. Sci. U.S.A.* **90**: 158-162.

Mori, H., T. Otake, et al. (1991). "In vitro anti-human immunodeficiency virus type 1 activity of biliverdin, a bile pigment." *Jpn. J. Cancer Res.* **82**(7): 755-757.

Ochoa, G.-C. (1994). Structure-based inhibition of the Simian Immunodeficiency Virus protease. Biochemistry. San Francisco, University of California, San Francisco.

Otwinowski, Z. (1993). Oscillation data reduction program. CCP4 Study weekend: Data collection and processing, 29-30 Jan 1993, SERC Daresbury Laboratory, England.

Roberts, N. A. (1995). "Drug-resistance patterns of saquinavir and other HIV proteinase inhibitors." *AIDS* **9 (supplement 2)**: S27-S32.

Rose, R. B., J. R. Rosé, et al. (1993). "Structure of the Protease from Simian Immunodeficiency Virus: Complex with an Irreversible Nonpeptide Inhibitor." *Biochemistry* **32**: 12498-12507.

Rutenber, E., E. B. Fauman, et al. (1993). "Structure of a non-peptide inhibitor complexed with HIV-1 protease." *J. Biol. Chem.* **268**: 15343-15346.

Sack, J. (1988). *J. Mol. Graphics* **6**: 225.

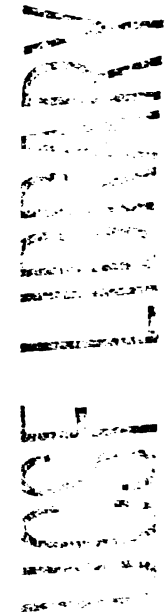
Salto, R., L. M. Babe, et al. (1994). "In vitro characterization of nonpeptide irreversible inhibitors of HIV proteases." *J. Biol. Chem.* **269**(14): 10691-8.

Sui, Z., R. Salto, et al. (1993). "Inhibition of the HIV-1 and HIV-2 proteases by curcumin and curcumin boron complexes." *Bioorg. Med. Chem.* **1**(6): 415-22.

Tomaszek, T. A., Jr., V. W. Magaard, et al. (1990). "Chromophoric peptide substrates for the spectrophotometric assay of HIV-1 protease." *Biochem. Biophys. Res. Commun.* **168**: 274-280.

Wlodawer, A., M. Miller, et al. (1989). "Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease." *Science* **245**: 616-621.

Yu, Z., P. Caldera, et al. (1996). "Irreversible inhibition of the HIV-1 protease: targeting alyklating agents to the catalytic aspartic groups." *accepted for publication.*



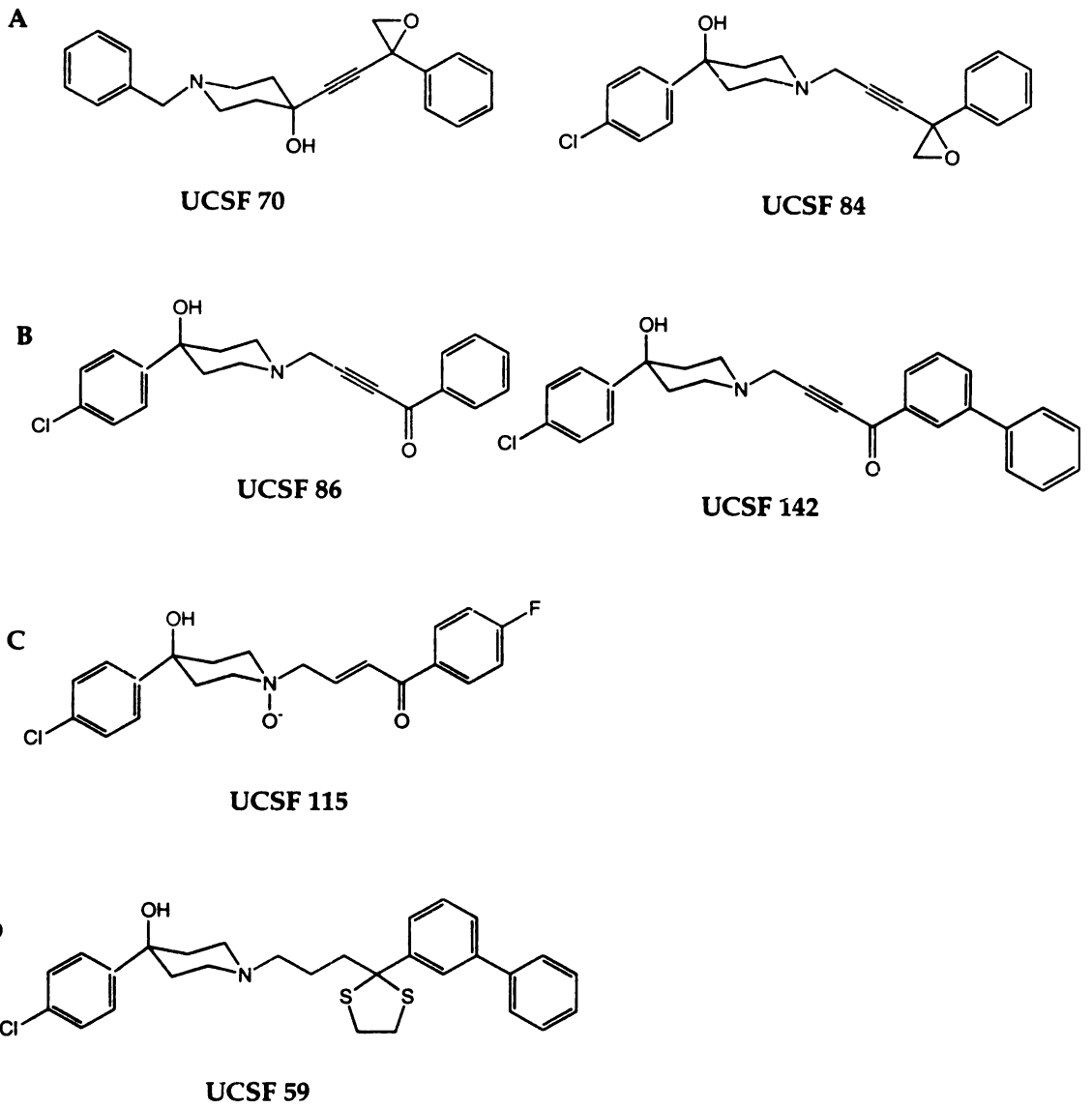
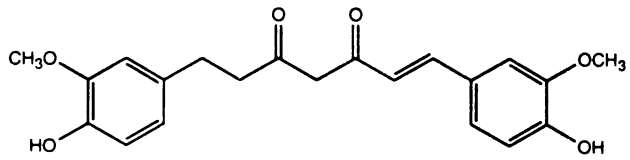
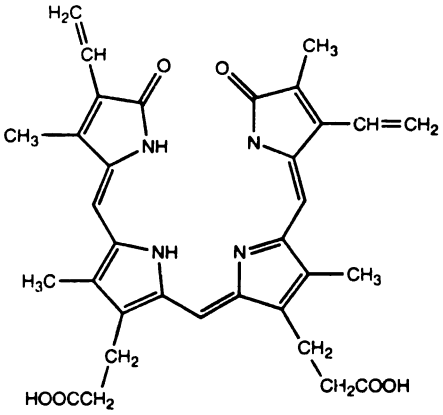


Figure 1: Derivatives of haloperidol. A. spiro epoxides, irrevesible inhibitors. B. ynones, irrevesible inhibitors. C. alpha-beta-unsaturated ketone, irreversible inhibitor. D. thioketal-m-biphenyl, reversible inhibitor

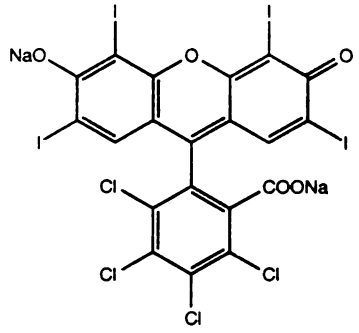
Curcumin



Biliverdin



Rose Bengal



Erythrosin B isothiocyanate

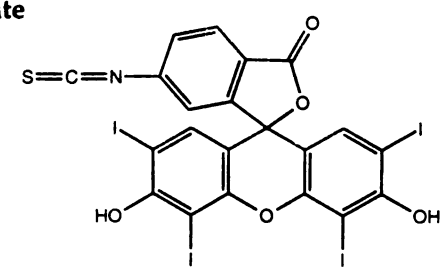
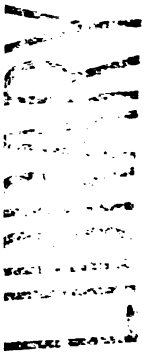


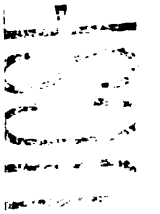
Figure 2: Non-haloperidol inhibitors

CHAPTER THREE:

THREE-DIMENSIONAL STRUCTURES OF HIV-1 AND SIV PROTEASE PRODUCT COMPLEXES



Submitted for publication to *Biochemistry*

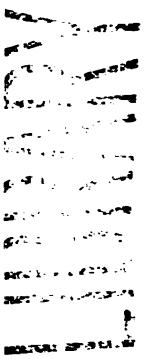


Three-Dimensional Structures of HIV-1 and SIV Protease Product Complexes†

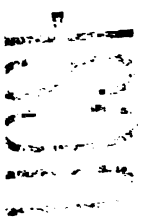
Robert B. Rose‡ ,Charles S. Craik‡§, Nancy L. Douglas§, Robert M. Stroud*‡§

‡Graduate Group in Biophysics, and §Departments of Pharmaceutical Chemistry and Biochemistry and Biophysics, The University of California, San Francisco, California 94143

* Author to whom correspondence should be addressed



†This work supported by NIH Grant GM 39552, NIH Training Grant CA-09215, and the Biotechnology Resources and Education Program Training Grant.

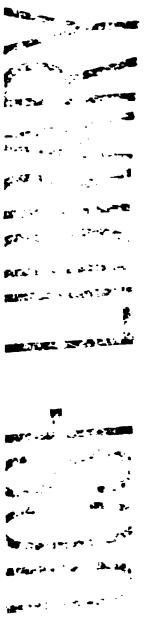


running title: HIV-1 and SIV protease product complex structures

key words: enzyme mechanism, proteinase, AIDS, strain

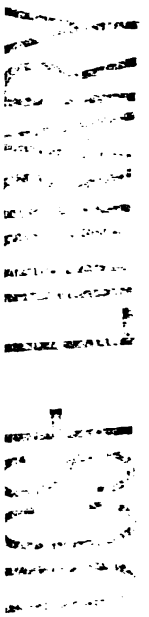
[1]Abbreviations: Ac, acetyl group; AIDS, Acquired Immune Deficiency Syndrome; DMSO, dimethyl sulfoxide; DTT, dithiothreitol; F_c, calculated structure factor; F_o, observed structure factor; HIV, Human Immunodeficiency Virus; Nle, norleucine; rms, root-mean-square; SIV, Simian Immunodeficiency Virus; TFA, trifluoroacetic acid; amino acids of proteins designated by the three-letter cod; amino acids of peptide ligands designated by the one-letter code; ↓ as part of a peptide name designates the residue interacting with the catalytic aspartic acids; / as part of a peptide name designates the scissile bond of the substrate or resulting products.

[2]The coordinates have been deposited with the Brookhaven National Laboratories Protein Data Bank with the assigned identification codes: HIV-1 protease complexed with Ac-S-L-N-F is designated ____, HIV-1 protease complexed with P-I-V-NH_2 is designated ____, SIV protease complexed with F-L-E-K is designated ____, SIV protease complexed with F(NO_2)-E-A-Nle-S is designated ____.



ABSTRACT

Eight Human Immunodeficiency Virus (HIV)-1 and Simian Immunodeficiency Virus (SIV) protease crystal structures complexed with products of the enzymatic reaction have been determined. Of these structures, two N-terminal products and two C-terminal products provide unambiguous density for the ligands. The structures range from 2.2Å to 2.6Å resolution, and 17% to 21% final R-factors. Comparison of the two C-terminal products, F-L-E-K and F(NO_2)-E-A-Nle-S, indicate that the position of the P2' residue is constrained by the protease, while the positions of the P1' and P3' residues are sequence dependent. The N-terminal product, Ac-S-L-N-F, overlaps closely with the N-terminal sequences of peptidomimetic inhibitors bound to the protease. The carboxyl group generated at the scissile bond interacts with both catalytic aspartic acids. This structure directly suggests the interactions of the gemdiol intermediate at the active site. A model of the substrate is generated by joining the N- and C-termini of the products requiring minimal conformational changes in the enzyme or peptides. Thus binding of substrate does not strain the scissile peptide bond. From these structures a catalytic mechanism is proposed by which the nucleophilic water is displaced stereospecifically by substrate binding towards one catalytic aspartic acid, while the scissile carbonyl hydrogen bonds to the other catalytic aspartic acid in position for hydrolysis.

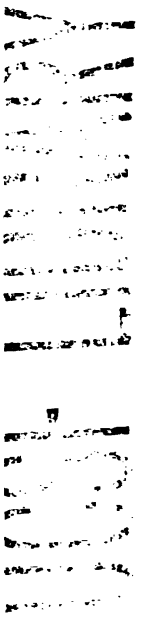


The retroviral proteases are homo-dimeric aspartyl proteases. The protease is an essential enzyme of the Human Immunodeficiency Virus (HIV)[1], processing eight distinct cleavage sites of the viral gag and gag-pol polyprotein (Henderson et al., 1988, Kohl et al., 1988, McQuade et al., 1990, Seelmeier et al., 1988). Each monomer of the protease contributes one of the catalytic aspartic acids, located inside a tunnel-shaped binding pocket. Substrate binding is associated with a large conformational change of the protease "flaps", a flexible beta-strand structure forming the "ceiling" of the binding pocket (James et al., 1982, Miller et al., 1989). Structural (James et al., 1977) and kinetic (Antonov et al., 1978, Antonov et al., 1981, Polgár, 1987) data have led to general agreement that the first step in the catalytic mechanism after formation of the Michaelis complex involves acid-base catalysis by the two aspartic acids. The unliganded structures of monomeric and retroviral aspartyl proteases show the putative catalytic water hydrogen bonded between the two aspartic acids (Sielecki et al., 1990, Suguna et al., 1987a, Wlodawer et al., 1989). The nucleophilic water attacks the substrate's carbonyl carbon, forming a gemdiol intermediate. The configuration of the catalytic water in the Michaelis complex, and the configurations of the transition state and the gemdiol intermediate are still controversial.

Peptidomimetic inhibitors with tetrahedral geometry in place of the scissile carbonyl have been proposed to mimic the transition state (Wlodawer & Erickson, 1993). Many HIV protease structures complexed with inhibitors have been determined as part of the drug design effort against AIDS. Most recently difluoroketone inhibitors, which bind to the protease as gemdiols, have been co-crystallized with different aspartyl proteases to interpret the mechanism (James et al., 1992, Parris et al., 1992, Silva et al., 1996,

Veerapandian et al., 1992). The gemdiol in these structures interact with one catalytic aspartic acid preferentially.

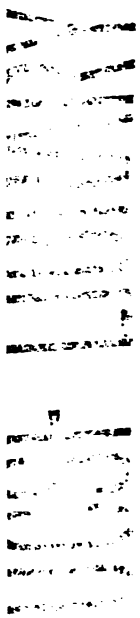
We report the first structures of product complexes of aspartyl proteases with HIV-1 and Simian Immunodeficiency Virus (SIV) proteases: two N-terminal and two C-terminal product complexes[2]. The model for the gemdiol intermediate based on the conformation of the carboxyl group of the N-terminal product interacts with both aspartic acids. These structures have the advantage, when interpreting the catalytic mechanism, of representing actual steps along the reaction pathway. From these structures we model the substrate and show that it binds in the active site without strain. A chemical mechanism is proposed based on models for the Michaelis complex and gemdiol intermediate.



MATERIALS AND METHODS

Crystallization:

HIV-1 protease (Rosé et al., 1993) and SIV protease (Rose et al., 1993) were expressed and purified as described previously. Each of these proteins has been engineered with a single point mutation to reduce autoproteolysis: Gln 7 to Lys in HIV-1 protease, Ser 4 to His in SIV protease. Protein was dialyzed and concentrated for crystallization in an 8-mL collodion membrane (Schleicher and Schuell) (Rose et al., 1993). The dialysis buffer for both HIV-1 and SIV protease crystallizations contained 20mM sodium acetate, pH 5.4, and 50mM sodium chloride, with 1mM DTT added for HIV-1 protease crystallizations. Protein was concentrated to about 1mg/ml in the collodion membrane, as measured by the Bio-Rad Bradford assay. For SIV protease, immunoglobulin G was used as a standard; for HIV-1 protease, bovine serum albumin was used as a standard. Peptides were added and the protein was further concentrated by evaporation with nitrogen until the volume was halved and quartered.

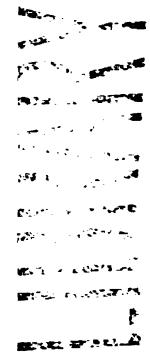


Crystals were grown using the hanging drop vapor diffusion method. Crystallization conditions for SIV protease were 4-6% saturated sodium chloride and 100mM sodium cacodylate, pH 6.5 (Rose et al., 1993) . Crystallization conditions for HIV-1 protease were 42-50% saturated ammonium sulfate, 200mM sodium acetate, pH 5.4, and 1mM DTT (McKeever et al., 1989). A high salt condition was chosen for HIV-1 protease crystallization to increase peptide binding (Hyland et al., 1991b, Richards et al., 1990)

Peptides for crystallization were synthesized by AnaSpec, Incorporated (San Jose, CA) and purified by reverse phase chromatography on a Vydac C-18

column with a water, 0.1% TFA-acetonitrile-0.1% TFA gradient. The peptides were then lyophilized. For crystallization, peptides were dissolved in DMSO and added to protein before evaporation to a final concentration of 30-50mM and a final DMSO concentration of 5% (v/v). When uncleaved substrate was co-crystallized with protease, the protein was allowed to incubate in the presence of peptide for 2 hours prior to evaporation.

SIV protease was co-crystallized with the substrate R-V-Nle/F(NO_2)-E-A-Nle-S (the slash denotes the scissile bond) and its products, as well as with the product /F-L-E-K. HIV-1 protease was co-crystallized with the substrate Ac-S-L-N-F/P-I-V-NH_2 and its products, with the N-terminus acetylated and the C-terminus amidated to neutralize the charge. HIV-1 protease was also co-crystallized with the product P-I-V and with both peptides Ac-S-L-N-F/ and /P-I-V.



Structure determination:

Data for the HIV-1 protease co-crystals with the peptide /P-I-V, the substrate Ac-S-L-N-F/P-I-V-NH_2, and with both peptides Ac-S-L-N-F/ and /P-I-V, and for the SIV protease co-crystals with the peptide /F(NO_2)-E-A-Nle-S, were collected using a MAR Research Image Plate detector with X-rays generated from a Rigaku generator operated at 3 kW. Data for the HIV-1 protease co-crystals with the peptide Ac-S-L-N-F/ was collected at the Stanford Synchrotron Radiation Laboratory. Data for the SIV protease co-crystals with the peptides ↓F-L-E-K, ↓R-V-Nle, and R-V-Nle↓F(NO_2)-E-A-Nle-S was collected with an R-axis II image plate detector using X-rays generated from a Rigaku generator operated at 9kW. Statistics for the data collections are summarized in Table I. All data was reduced using the program DENZO (Otwinowski, 1993). Refinement was carried out using the program XPLOR

(Brünger et al., 1987). The molecular graphics program CHAIN was used for displaying structures and building into electron density maps (Sack, 1988).

The HIV-1 protease crystals were isomorphous with the previously determined structures, in space group $P2_12_12_1$, with a dimer in the asymmetric unit (Table IIa) (Rutenber et al., 1993). The initial maps were generated after rigid body minimization, omitting the flaps, waters, and ligand from the HIV-1 protease structure, allowing the two monomers to refine independently.

The SIV protease crystals were isomorphous with the previously determined structure (Table IIb) (Rose et al., 1993). The data was reduced in space group C2, beta=90°, with a dimer in the asymmetric unit. The initial maps were generated after rigid body minimization, omitting the flaps, waters, and ligand from the SIV protease structure, allowing the two monomers to refine independently. If the density in the active site around the dimer two-fold axis was symmetrical, the structure was further refined in space group $C222_1$ with a monomer in the asymmetric unit.

After positional and individual B-factor refinements of the protease molecules alone and addition of water molecules, the ligands were built into density. For all structures, careful attempts were made to test alternate interpretations for the density in the binding pocket, including building peptides in backwards and filling density with water molecules. The R-free (Brünger, 1992) was used to determine when the addition of peptide atoms to the model was justified: any atoms which increased the R-free were removed. Table I lists the final R-factors, R-frees, and deviations from ideal geometry for each of the structures.

Comparison of structures:

Structures were compared by superimposing the carbon alpha atoms of the protease monomers or dimers using the program GEM (Eric Fauman, personal communication). To compare the configurations of the bound products with those of peptidomimetic inhibitors, three protease-inhibitor structures from the literature were superimposed on the product structures: 8hvp (Jaskólski et al., 1991), 7hvp (Swain et al., 1990), and 1siv (Zhao et al., 1993) (referred to by their Brookhaven Protein Databank code). 8hvp and 7hvp are HIV-1 protease structures and 1siv is an SIV protease structure. The peptide-based inhibitor sequences and inhibitor names as referred to by the authors are:

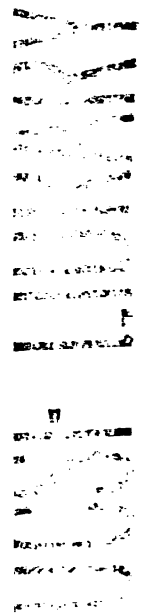
8hvp	VSQNL-Ψ[CH(OH)CH_2]-VIV	U-85548E
7hvp	Ac-SLNF-Ψ[CH(OH)CH_2N]-PIV-OMe	JG365
1siv	AAF-Ψ[(S)-CH(OH)CH_2]-GVV-OMe	SKF107457

These hydroxyethylene and hydroxyethylamine isosteres are all tetrahedral at the P1 carbon in place of the peptide carbonyl. JG365 has the same sequence as one substrate and products used in crystallization: Ac-S-L-N-F/P-I-V.

Overlapping of the bound peptides was accomplished by superimposing the carbon alpha atoms of the protease monomer most closely associated with the product. This avoids any differences in the structures resulting from dimer formation, since the SIV protease dimer is two-fold symmetrical and the HIV-1 protease dimer is not. The HIV-1 and SIV protease structures superimposed with an rms deviation of 0.5 Å or less. The HIV-1 and SIV protease structures superimposed with an rms deviation of about 1.2 Å.

Generating the substrate model:

The substrate was modeled by superimposing two protease structures with N- and C-terminal products and reconnecting the scissile bond. The HIV-1 protease dimer bound to the N-terminal product Ac-S-L-N-F/ was superimposed on that of SIV protease bound to the C-terminal product F-L-E-K/. One of the C-terminal carboxyl oxygens of Ac-S-L-N-F/ was removed and the terminal carbon was reconnected to the terminal nitrogen. The geometry of the peptide was regularized using the routine in CHAIN. The starting position of the carbonyl oxygen, as well as the atoms fixed, were varied to determine their effects on regularization.



RESULTS

Structure of the N-terminal product Ac-S-L-N-F↓:

The N-terminal product Ac-S-L-N-F↓ (the arrow denotes the amino acid oriented towards the catalytic aspartic acids in the structure) was co-crystallized with HIV-1 protease. Throughout the paper, peptides oriented with their C-termini towards the aspartic acids will be referred to as "P"-products, and peptides oriented with their N-termini towards the aspartic acids will be referred to as "Q"-products. The P-product amino acids will be identified as P1 to P4, starting from the C-terminus (Schecter & Berger, 1967). The Q-product amino acids will be identified as P1' to P4', starting from the N-terminus.

Figure 1a shows the final $2|F_o|-|F_c|$ density in the binding pocket for HIV-1 protease bound to the product Ac-S-L-N-F↓. Density for the catalytic aspartic acids, the peptide, and the conserved water 401 is displayed. Only the N-terminal acetyl group of the peptide is without density. The density in the unliganded half of the binding pocket was filled by 3 water molecules with high B-factors (between 43Å^2 and 53Å^2, compared with the average B-factor for the peptide carbon alphas of 20Å^2). These waters were not hydrogen bonded to the protein or other ordered water molecules, and did not explain all of the difference density.

The C-terminal carboxyl group oxygens are each within hydrogen bonding distance of one of the catalytic aspartic acids, forming an unusual three-carboxyl group interaction (Figure 2). Two tautomeric forms are represented in Figure 3a, though the geometry favors the first configuration. The other aspartic acid oxygens share a hydrogen bond, as well as being hydrogen bonded to the amide nitrogen of Gly 27 and Gly 27'.

All hydrogen bonds to the product are made with one monomer of the protease dimer (identified as the unprimed monomer for P-products) except for the hydrogen bonds to the catalytic aspartic acids. The backbone hydrogen bonds are the same as those formed by peptidomimetic inhibitors (Swain et al., 1990): 2 hydrogen bonds to the nitrogen and oxygen of Gly 48 of the flap, and 2 hydrogen bonds to the oxygen of Gly 27 and the nitrogen of Asp 29 at the base of the binding pocket. A fifth hydrogen bond is formed between the sidechain oxygen of Asp 29 and the backbone oxygen of the P3 Leu of the peptide. Water 401 is hydrogen bonded to the P2 Asn carbonyl of the peptide. This conserved water is hydrogen bonded to the amide nitrogens of Ile 50 and Ile 50', located on the flaps of both protease monomers, as found in all peptidomimetic inhibitor structures (Wlodawer & Erickson, 1993). Two hydrogen bonds are also formed between sidechains of Ac-S-L-N-F↓ and the protease. The P2 Asn is hydrogen bonded with Asp 29 and 30 of the protease, and the P4 Ser is hydrogen bonded with Asp 30 of the protease.

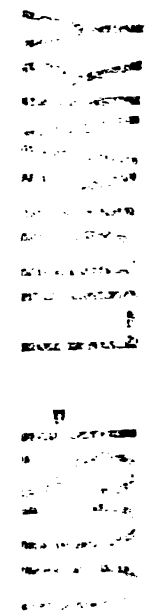


Table III lists the protease residues forming the binding pockets of the product complex structures (defined as being within 4.2Å of the residue). The sidechains of the product peptides protrude on alternate sides of the peptide backbone in a beta sheet conformation. Most of the interactions are with residues of the unprimed monomer, the same monomer forming the hydrogen bonds to the product, except for the P1 subsite.

Figure 4a shows the superposition of the Ac-S-L-N-F↓ product peptide with JG365 and U-85548E. The backbone atoms of Ac-S-L-N-F↓ and JG365 for residues P1 to P3 overlap with an rms deviation of 0.4 Å, and diverge at the P1 carboxyl carbon of Ac-S-L-N-F↓, which is 0.8 Å from the corresponding carbon of JG365. The conformation of the N-terminal acetyl group and the P4 Ser are less well conserved. The hydroxyl of the hydroxyethylene isostere

points between the carboxyl oxygens of Ac-S-L-N-F↓. The backbone atoms of Ac-S-L-N-F↓ and U-85548E for residues P1 to P3 overlap with an rms deviation of 0.4 Å. The C-terminal carbon of Ac-S-L-N-F↓ is 0.3Å from the corresponding carbon of U-85548E. According to an estimation of the errors in the atomic coordinates as a function of B-factor (Stroud & Fauman, 1995), the atomic coordinates for the backbone atoms of Ac-S-L-N-F↓ at residues P1-P3 are accurate to within 0.2Å. The atomic coordinates for the backbone atoms of JG365 at residues P1-P3 are accurate to within 0.3Å. Therefore changes of the order of 0.4Å between Ac-S-L-N-F↓ and JG365 are significant.

The individual monomers of the protease dimer are not identical once the peptide product binds. The carbon alphas from one monomer optimally superimpose on those of the other monomer with a rotation of 178° and an rms deviation of 0.53Å. Differences between corresponding atoms are as great as 1.8 Å, at residue 49 of the flaps. The unprimed monomer, which forms backbone hydrogen bonds to the peptide, is slightly more compact than the primed monomer, with the mainchain atoms shifted inward towards the bound peptide. This is particularly true of residues 27 to 30 (the beta strand forming the base of the binding pocket), residues 33 to 56 (the flap residues), residues 78 to 83 (the loop which partially forms the P1 binding pocket), and residues 86-95 (the helix). The N-terminal beta-strand of the primed monomer, residues 1'-8', are shifted in towards the peptide; Arg 8' forms part of the P3 binding pocket. A few sidechain conformations differ between monomers, mostly on the surface of the protein. Within the binding pocket, the sidechain of Ile 84' rotates to interact with the peptide P1 Phe. Also the sidechain of Asp 30 is hydrogen bonded with the peptide P4 Ser in the liganded monomer and is turned towards solvent in the unliganded monomer.

From a comparison of the temperature factors between the carbon alpha atoms of the monomers, the liganded monomer (residues 1-99) is less mobile than the unliganded monomer (residues 1'-99'). The average B-factors for the carbon alpha atoms of the liganded and unliganded monomers are 18.5 $Å^2$ and 22.8 $Å^2$, respectively. The average B-factors of both flaps are 17 $Å^2$, indicating that both flaps are stabilized in this structure.

Other N-terminal product structures:

Co-crystallization of HIV-1 protease with both products Ac-S-L-N-F/ and /P-I-V resulted in a similar structure to that of HIV-1 protease co-crystallized with Ac-S-L-N-F/ alone. The rms deviation between the peptides from these two structures was 0.23 Å.

Co-crystallization of HIV-1 protease with the whole substrate, Ac-S-L-N-F/P-I-V-NH_2 resulted in a structure of the product P-I-V-NH_2↓ bound in the catalytic site as a P-product. The orientation of the peptide was reversed with respect to the substrate binding mode. Figure 1b shows the final $2|F_o|-|F_c|$ density in the binding pocket. The C-terminal amidated carboxyl group was pointed perpendicular to the orientation of the C-terminal carboxyl group of Ac-S-L-N-F↓, similar to the orientation of the hydroxyl group of the hydroxyethylene isosteres of U-85548E and SKF107457. It was not possible to distinguish the amino group from the oxygen from the density. The amino group was built towards the active site aspartic acids, 3.1Å and 3.5Å away. The oxygen was within hydrogen bonding distance of waters 401 and 521. The amino group was considered responsible for reorienting the carboxy terminus, and the hydrogen bonds with the aspartic acids were the stronger interaction.

With only 3 residues, the peptide P-I-V-NH_2↓ makes 3 of the 4 backbone interactions observed in the Ac-S-L-N-F↓ peptide, missing the hydrogen bond to the Gly 48 oxygen of the flap. This hydrogen bond is formed by the nitrogen of the P3 proline.

The peptide /P-I-V was co-crystallized with HIV-1 protease to determine if the unamidated peptide would bind as a Q-product. There was good difference density for the isoleucine in one half of the binding pocket, but inclusion of the proline or valine on either side of the isoleucine increased the R-free. We conclude that the peptide bound in multiple conformations.

The structure of SIV protease with the product R-V-Nle↓ showed asymmetrical omit map difference density in the active site. The structure was therefore refined in space group C2 with a dimer in the asymmetric unit, allowing for different contents for the two halves of the active site. The P-product A-A-A↓ fit the density better than the Q-product ↓A-A-A because of the density for the carboxyl group near the catalytic aspartic acids. Unlike in the Ac-S-L-N-F↓ structure, the C-terminal carboxyl group in this structure was oriented towards one of the aspartic acids.

The structure of SIV protease crystallized with the complete substrate R-V-Nle/F(NO_2)-E-A-Nle-S resembled that of SIV protease crystallized with R-V-Nle/ alone. The density was best described by the peptide sequence A-A-A↓.

Structure of the C-terminal products ↓F-L-E-K and ↓F(NO_2)-E-A-Nle-S:

The structure of SIV protease crystallized with the peptide ↓F-L-E-K or ↓F(NO_2)-E-A-Nle-S was refined in space group $C222_1$ with a monomer in the asymmetric unit. Despite the appearance of peptide density in both halves of the binding pocket, the stoichiometry of binding was one peptide per protease dimer. The N-termini of the symmetry-related ↓F(NO_2)-E-A-Nle-S peptides were only 1.5Å apart, excluding the possibility that two peptides were bound to the same dimer. The N-termini of the symmetry-related ↓F-L-E-K peptides were 2.9Å apart, but occupancy refinement of the peptide resulted in 60% occupancy, near to 50% occupancy explained by the binding of one peptide per dimer. The crystals were therefore composed of dimers bound to one peptide product but statistically arranged in two orientations. Any differences between monomers that resulted from ligand binding was therefore averaged in the electron density.

Figure 5a shows the final $2|F_o|-|F_c|$ density in the active site of the ↓F-L-E-K structure. Density for the catalytic aspartic acids, the peptide, and the conserved water 401 is displayed. The peptide is in good density except for the terminal three atoms of the P3' Leu sidechain.

Figure 5b shows the final $2|F_o|-|F_c|$ density in the active site of the ↓F(NO_2)-E-A-Nle-S structure. The sequence ↓F(NO_2)-E-A-A was built into the active site; there was no density for the P4' norleucine (Nle) sidechain or for the P5' Ser. The phenyl ring of the para-nitro-phenylalanine residue was not completely in density suggesting disorder; the B-factors for the ring were low, from 3 to 6 $Å^2$. Unlike the structure with ↓F-L-E-K, there was continuous density between the aspartic acids. A water molecule built

between the aspartic acids clashes sterically with the N-terminus of the peptide.

All hydrogen bonds are made with one monomer of the protease dimer (identified as the primed monomer for the Q-products), except for the hydrogen bonds to the catalytic aspartic acids. The same four backbone hydrogen bonds formed by the P-product are formed to Gly 48' of the flaps, and Gly 27' and Asp 29' at the base of the binding pocket. The P1' carbonyl forms a hydrogen bond to the conserved water 401. The N-terminus of ↓F-L-E-K is hydrogen bonded with the catalytic aspartic acid of the unprimed monomer, 3.3Å away. The N-terminus of ↓F(NO_2)-E-A-Nle-S is hydrogen bonded with both active site aspartic acids, both 3.1Å away. Two tautomeric forms for each configuration are represented in Figures 3b and 3c.

↓F-L-E-K and ↓F(NO_2)-E-A-Nle-S each make one sidechain hydrogen bond with the protease: the P4' Lys of ↓F-L-E-K is hydrogen-bonded with Asp 58, and the P2' Glu of ↓F(NO_2)-E-A-Nle-S is hydrogen bonded with Asp 30. In addition, Arg 8 forms a salt-bridge with the P3' Glu of ↓F-L-E-K, and the P1' nitro group of ↓F(NO_2)-E-A-Nle-S.

Table III lists the protease residues forming the binding pockets for ↓F-L-E-K. Most of the interactions are with residues of the primed monomer, the same monomer forming the hydrogen bonds to the product, except for the S1' subsite. The S3′ subsite, on the same side of the peptide as the S1′ subsite, is formed partly by Arg 8 of the unprimed monomer.

Figure 4b compares the conformation of ↓F-L-E-K with ↓F(NO_2)-E-A-Nle-S. The rms deviation for the carbon alpha atoms is 0.4Å. The carbon alpha atoms at the P2' position are only 0.2 Å apart. The peptides overlap less well at the other residues: the carbon alphas of P3' are 0.4 Å apart, and the carbon alphas at P1' are 1.3 Å apart. A similar trend holds true for the overlap

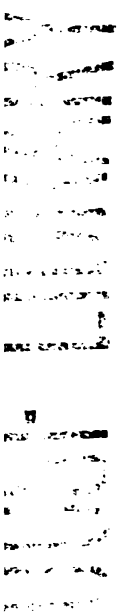
with the peptidomimetic inhibitor SKF107457. The N-terminal nitrogen atoms of the peptide products are 1.6Å apart. Water 401 is shifted towards the aspartic acids in the ↓F(NO_2)-E-A-Nle-S structure by about 0.5Å, and the flaps are slightly more closed by about 0.6Å. According to an estimation of the errors in the atomic coordinates as a function of B-factor (Stroud & Fauman, 1995), the atomic coordinates for the backbone atoms of ↓F-L-E-K, and ↓F(NO_2)-E-A-Nle-S at residues P1'-P3' are accurate to within 0.2Å. Therefore changes of the order of 0.3Å between ↓F-L-E-K, and ↓F(NO_2)-E-A-Nle-S are significant.

Model of substrate:

The substrate was modeled by superimposing the SIV protease structure bound to ↓F-L-E-K onto the HIV-1 protease structure bound to Ac-S-L-N-F↓. The distance between the scissile carbon of the P-product Ac-S-L-N-F↓, and nitrogen of the Q-products ↓F-L-E-K or ↓F(NO_2)-E-A-Nle-S, was 2.2 Å or 1.3Å, respectively. An idealized peptide bond is 1.45 Å and planar. Figure 6 shows the two positions of the carbonyl after removing each of the carboxyl oxygens in turn and regularizing the geometry. Removal of OT1, near Asp 25', led to much less movement of the carbonyl than removal of OT2, near Asp 25. The black line connecting these positions represents conformations of the peptide bond of the same minimal energy; beginning regularization with the oxygen at any of these positions did not shift the position of the oxygen. Varying the atoms fixed during regularization did not affect the outcome of regularization: fixing P2 to P4 and P2' to P4' gave the same results as allowing all residues the freedom to move.

Reforming the peptide bond was achieved with minimal movements of atoms: 0.7Å and 0.8Å for the scissile carbon and nitrogen respectively, 0.3

and 0.5Å for the carbon alpha atoms of P1 and P1' respectively, and less than 0.2Å for the backbone atoms more distant from the scissile bond. The distance between the carbon alpha atoms of the P1 and P1' residues in the products was 4.6Å, and in the final substrate was 3.9Å. The phi and psi angles rotated 15° or less. The phi and psi angles for the resulting peptide were within the allowed region on a Ramachandran plot.



DISCUSSION

Product Complexes are Reaction Intermediates:

Although product complexes of various proteases have been studied crystallographically, we report the first structures of the non-covalent product complexes of aspartyl proteases. Two peptide substrates and their proteolyzed products were co-crystallized with HIV-1 and SIV proteases. The substrate Ac-S-L-N-F/P-I-V-NH_2 is a modification of the p17/p24 cleavage site of the viral polyprotein (Rich et al., 1990). The second substrate, R-V-Nle/F(NO_2)-E-A-Nle-S, is a non-natural chromogenic substrate used in spectrophotometric assays. An additional product, ↓F-L-E-K, was studied, derived from the reverse transcriptase/integrase cleavage site for SIV protease: R-Q-V-L↓F-L-E-K (Grant et al., 1991). The structures of two amino-terminal products (the "P"-product, Ac-S-L-N-F↓ and P-I-V-NH_2↓) with HIV-1 protease and two carboxy-terminal products (the "Q" products, ↓F(NO_2)-E-A-Nle-S and ↓F-L-E-K) with SIV protease showed well ordered density for the peptides. No structures with both products bound simultaneously were attained.

HIV-1 and SIV proteases are structurally and functionally similar. Their sequences are 50% identical. Their structures differ by an rms deviation of 1.2 Å between the carbon alpha atoms when the dimers are superimposed (Rose et al., 1993). SIV protease is capable of cleaving the HIV-1 polyprotein cleavage sites with varying efficiencies (Debouck, 1991, Grant et al., 1991). The product complex structures with each of these proteins are therefore comparable.

Kinetic studies have demonstrated the formation of product complexes. Competitive and non-competitive inhibition patterns for product inhibitors of HIV-1 protease with millimolar inhibition constants have been

determined (Hyland et al., 1991b). Our experience with the peptide /P-I-V-NH_2 indicates that short peptide products may bind the protease in either orientation, complicating the interpretation of these inhibition studies. Amidation contributed to the misorientation of this peptide with respect to the substrate. It oriented the C-terminus perpendicular to that of the carboxyl group of Ac-S-L-N-F↓, similar to the orientation of the hydroxyethylene hydroxyl group of U85548-E. Since the sequences studied by Hyland, *et al* were all acetylated and amidated, the assumption that the products all bind in one orientation is probably incorrect. Despite this caveat, these inhibition studies indicated random release of products (Meek et al., 1994). We were able to attain structures with P- and Q- products, consistent with a random release mechanism.

The P-product Ac-S-L-N-F↓ adopts a conformation essentially identical to the "P" portion of the peptidomimetic inhibitor JG365, except for the scissile carbonyl carbon. The re-orientation of this carbon may reflect both the flexibility required to cleave substrates with a P1' proline and the extra bond of the hydroxyethylamine isostere of JG365. The "P" portion of U-85548E also superimposes closely with Ac-S-L-N-F↓, despite the difference in sequence. The P-product Ac-S-Q-N-Y↓ has been shown to bind to HIV-1 protease in an enzymatically competent conformation in solution (Hyland et al., 1991b). The protease exchanged ^{18}O from $H_2{}^{18}O$ into the C-terminal carboxyl group, similar to the exchange of ^{18}O into substrate. We assume that the C-terminal carboxyl group of the bound product Ac-S-L-N-F↓ reflects this conformation.

The carboxy terminus of the P-product, Ac-S-L-N-F↓, forms a "triangular" interaction with the two catalytic aspartic acids as shown in Figure 2. Each carboxyl group is protonated to form the three hydrogen bonds of the triangle. In order to form hydrogen bonds at pH 5.4, the pH of the

crystals, the pKa's of the three carboxyl groups are raised from the solution value of about 3 in the P-product complex. The pKa of the proton shared between the two catalytic aspartic acids has been measured to be between 5.5 and 6.8 (Hyland et al., 1991b, Ido et al., 1991). Raising the pKa of the other protons of the triad is consistent with the P-product binding to the protease more tightly at low pH, where these bonds can be maintained, than at high pH, where the hydrogen bonds would be broken. The peptide Ac-S-Q-N-Y↓ binds 5 times tighter at pH 3.5 (K_i=0.2 millimolar) than at pH 6 (K_i=1.0 millimolar) (Hyland et al., 1991b). A deuterium isotope effect which decreased between pH 6 and pH 3.2 was also observed (Hyland et al., 1991a). This was interpreted as a non-catalytic step becoming rate-limiting at low pH, possibly involving release of the P-product.

Unlike the P-product, the Q-product's mainchain atoms adopt a sequence-dependent conformation when bound to the protease, except at P2' (Figure 4b). The more rigidly held P2' sidechain is often a glutamate or glutamine in naturally occurring substrates and the site is less forgiving of mutations than are other subsites (Poorman et al., 1991, Tomasselli & Heinrikson, 1994, Wlodawer & Erickson, 1993). The glutamate in the P2' site of ↓F(NO_2)-E-A-Nle-S is hydrogen bonded to the backbone nitrogens of Asp 29 and Asp 30, as in HIV-2 protease (Griffiths & al, 1992, Tong et al., 1993), providing some of the constraints at this sidechain.

The positions of the N-terminal nitrogens of ↓F-L-E-K and ↓F(NO_2)-E-A-Nle-S differ by 1.6Å. The para-nitro-phenylalanine residue of ↓F(NO_2)-E-A-Nle-S may cause the amino terminus to be closer to both aspartic acids. The peptide ↓F-L-E-K, with the smaller P1' Phe residue, hydrogen bonds with only one aspartic acid.

Substrate binds without strain:

The idea that substrate strain can assist in bond cleavage has been discussed in the context of aspartyl proteases for some time (Davies, 1990, Pearl & Blundell, 1984). Fruton proposed that the strain in the substrate upon binding to pepsin could contribute to bond cleavage (Fruton, 1976). Monomeric aspartyl proteases and HIV-1 protease complexed with peptidomimetic inhibitors with tetrahedral geometry at the scissile carbon have been interpreted in terms of strain of the planar peptide bond of a substrate(Jaskólski et al., 1991, Suguna et al., 1987b). Modeling the substrate from the product complex structures indicates the absence of strain in the reconnected peptide bond.

In modeling the substrate from the products, we assume that the products bind to the protease in the most energetically favorable conformation for that sequence. Because the product complex structures contain only one of the products, their positioning is independent of the repulsive forces that would occur in the newly cleaved substrate. A similar analysis has been carried out with dihydrofolate reductase in which a ternary complex was modeled by overlapping two binary complexes (Davies et al., 1990).

The scissile carbonyl of the modeled substrate can occupy a range of positions of equal energy without strain, as shown in Figure 6. These conformations are attained by minimal movements of the P1 and P1' residues, indicating that a substrate bound to the protease utilizing similar interactions to the products would not deform the scissile peptide bond. The movements in the P1 and P1' residues are small enough to be accommodated

by the protein without much energetic cost. This argues against the development of strain as a result of substrate binding.

A planar peptide bond has been modeled from a hydroxyethylene peptide isostere using a similar regularization procedure (Jaskólski et al., 1991). The carbonyl oxygen of the hydroxyethylene isostere is located between the carboxyl oxygens of Ac-S-L-N-F↓. Positioning of the scissile carbonyl was sensitive to the atoms held fixed while the geometry was regularized, unlike our model. This lead to the conclusion that strain developed in the scissile bond upon substrate binding, deforming the planar peptide bond towards the tetrahedral transition state. This strain occurred as a result of rigidly fixing the P1 and P1' atoms (except for the P1 carbonyl oxygen) by the protease to their conformation in the peptidomimetic inhibitor. Based on our structural results, we contend that the protease accommodates the products readily and these residues are not so rigidly held that they would strain the scissile bond.

Differences between the monomers of HIV-1 protease bound to Ac-S-L-N-F↓ may reflect conformational changes and increased thermal stability of the protein accompanying ligand binding. Some of these differences may also result from different crystal contacts made by each monomer. The "liganded" monomer is more compact than the "unliganded" monomer, reflecting contacts made between the protease and the peptide. The flaps, which close in response to ligand binding, differ conformationally by up to 1.8 Å in the liganded and unliganded monomers. Differences also occur in the base of the binding pocket and the S1 subsite. The B-factors of the unliganded monomer are 5 to 10 $Å^2$ greater than for the liganded monomer. The flaps have similar B-factors in the two monomers.

The proteolytic mechanism:

Biochemical (Hyland et al., 1991a, Hyland et al., 1991b) and structural (Jaskólski et al., 1991, Silva et al., 1996) studies of HIV protease have led to proposals for its mechanism based on the mechanism of pepsin and other monomeric aspartyl proteases. We propose structures for the major steps of proteolysis derived from the product complexes: the Michaelis complex and gemdiol intermediate. Figure 7 shows a chemical mechanism based on these structures.

There is general agreement that the ordered water molecule hydrogen bonded between the aspartic acids in the unliganded structure becomes the catalytic nucleophile (James et al., 1992, Jaskólski et al., 1991, Suguna et al., 1987b). In Figure 7 the water is depicted forming bifurcated hydrogen bonds to the two catalytic aspartic acids. This water is displaced as a result of substrate binding and formation of a hydrogen bond between the scissile carbonyl oxygen and Asp 25'.

The position of the nucleophilic water during catalysis is usually modeled from the position of an oxygen of the putative transition state or intermediate structure, for instance from pepstatin (Bott et al., 1982, James & Sielecki, 1985) or difluoroketone (James et al., 1992, Parris et al., 1992, Veerapandian et al., 1992). We model the water near one carboxy terminal oxygen of the P-product. Assuming that the nuclei do not move significantly during the time of the reaction, the oxygen of the attacking water will be positioned similarly before and after hydrolysis. We also assume that the carboxy terminus of the P-product Ac-S-L-N-F↓ has not shifted significantly after hydrolysis. The positions of the carbonyl and catalytic water are similar to those proposed by Suguna, *et al* (Suguna et al., 1987b) for the reduced isostere inhibitor and by the difluoroketone inhibitors except that the second proton of the water is too far from the aspartic acids to hydrogen bond. The

alternative to repositioning the water with substrate binding is to rotate the peptide carbonyl to attain a good angle of attack, as proposed by Jaskólski, *et al* (Jaskólski et al., 1991).

Figure 8 depicts the catalytic water modeled with the substrate. The carbonyl adopts the position within hydrogen bonding distance of Asp 25' (the primed monomer associated with the Q-product), close to the position of one C-terminal carboxyl oxygen of the product Ac-S-L-N-F↓. The peptide carbonyl displaces the catalytic water, which forms a bifurcated hydrogen bond with Asp 25, 2.1Å from OD1 and 3.3Å from OD2. In this position, close to the position of the other C-terminal carboxyl oxygen of the product Ac-S-L-N-F↓, the water is hydrogen bonded to only one of the aspartic acids and is 1.6Å from the carbon of the scissile bond. The proximity of the water to the carbon may mean that nucleophilic attack begins as the substrate binds. The water molecule is depicted attacking the carbonyl stereospecifically at an 80° angle. This angle is close to the optimal 107° angle for nucleophilic attack of a carbonyl (Burgi et al., 1973). The direction of attack is the same as predicted by Jaskólski, *et al* (Jaskólski et al., 1991).

Crystal structures with pepstatin (James & Sielecki, 1985) reduced peptide isosteres (Suguna et al., 1987b), hydroxyethylene isosteres (Veerapandian et al., 1990), and most recently diflouroketone (James et al., 1992, Parris et al., 1992, Silva et al., 1996, Veerapandian et al., 1992) have been used as models of the gemdiol intermediate. We propose that the conformation of the gemdiol intermediate is similar to that of the carboxy terminus of Ac-S-L-N-F↓, with each hydroxyl interacting with one of the catalytic aspartic acids. This is in contrast to the conformation of the difluoroketone gemdiol and the concerted mechanism proposed by Jaskólski, *et al* (Jaskólski et al., 1991).

In the substrate model, the scissile bond nitrogen is 4Å from Asp 25, too far for efficient donation of a proton. The source of this proton is depicted in Figure 7 as one of the gemdiol hydroxyls. Silva, *et al* have proposed an anti to gauche transition of the gemdiol modeled from an HIV-1-difluoroketone structure by rotation around the C-N bond (Silva et al., 1996). This rotation would bring the nitrogen closer to the catalytic aspartic acids, although the associated conformational change of the backbone and sidechains is not evident in our model of the substrate from the products.

These modeled structures provide a framework for the proteolytic mechanism. The major differences between the mechanism proposed by these structures and that proposed by Hyland, *et al* are the positioning of the catalytic water and the symmetrical disposition of the gemdiol intermediate relative to the catalytic aspartic acids. Our model of the substrate suggests the two aspartic acids are not equivalent: Asp 25 (associated with the P-product) acts as the general base extracting a proton from the water. Asp 25' acts as the general acid, donating a proton to the carbonyl of the scissile bond 2.6Å away. This would echo the inequivalence of the catalytic aspartic acids in the monomeric aspartyl proteases: Asp 213 is thought to act as the general base in penicillopepsin, while Asp 33 acts as the general acid (James et al., 1992).

CONCLUSIONS

Clearly much can be learned regarding the mechanism of catalysis from detailed analyses of protease-product complexes. Subsequent structural studies will address factors other than substrate strain contributing to catalysis. Particularly the role of conformational changes in the protein will be investigated. Ultimately these studies will contribute to an explanation of the complex specificity patterns displayed by many aspartyl proteases.

ACKNOWLEDGMENTS

We thank Earl Rutenber for advice on structure determination and interpretation. Earl Rutenber, Carleton Sage, Peter Sayre for useful discussions.

REFERENCES

Antonov, V. K., Ginodman, L. M., Kapitannikov, Y. V., Barshevskaya, T. N., Gurova, A. G., & Rumsh, L. D. (1978) *FEBS Lett. 88*, 87-90.

Antonov, V. K., Ginodman, L. M., Rumsh, L. D., Kapitannikov, Y. V., Barshevskaya, T. N., Yavashev, L. P., Gurova, A. G., & Volkova, L. I. (1981) *Eur. J. Biochem. 117*, 195-200.

Bott, R., Subramanian, E., & Davies, D. R. (1982) *Biochemistry 21*, 6956-6962.

Brünger, A. T. (1992) *Nature 355*, 472-474.

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458-460.

Burgi, H. B., Dunitz, J. D., & Shefter, E. (1973) *J. Am. Chem. Soc. 95*, 5065-5067.

Davies, D. R. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 189-215.

Davies, J. F., Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H., & Kraut, J. (1990) *Biochemistry 29*, 9467-9479.

Debouck, C. (1991) *Adv. Exp. Med. Biol. 306*, 407-15.

Fruton, J. S. (1976) in *Methods Enzymol.* (Meister, A., Ed.) pp 1 35, Siley, John and sons, New York.

Grant, S. K., Deckman, I. C., Minnich, M. D., Culp, J., Franklin, S., Dreyer, G. B., Tomaszek, T. A., Jr., Debouck, C., & Meek, T. D. (1991) *Biochemistry 30*, 8424-8434.

Griffiths, J. T., & al, e. (1992) *Biochemistry 31*, 5193-5200.

Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M., & Oroszlan, S. (1988) *Human Retrovirus, Cancer and AIDS: Approaches to Prevention and Therapy*, Liss, New York.

Hyland, L. J., Tomaszek, T. J., & Meek, T. D. (1991a) *Biochemistry 30*, 8454-8463.

Hyland, L. J., Tomaszek, T. J., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., Culp, J. S., & et, a. l. (1991b) *Biochemistry 30*, 8441-8453.

Ido, E., Han, H. P., Kezdy, F. J., & Tang, J. (1991) *J Biol Chem 266*, 24359-24366.

James, M. N., & Sielecki, A. R. (1985) *Biochemistry 24*, 3701-3713.

James, M. N., Sielecki, A. R., Hayakawa, K., & Gelb, M. H. (1992) *Biochemistry 31*, 3872-3886.

James, M. N. G., Hsu, I.-N., & Delbaere, L. T. J. (1977) *Nature 267*, 808-813.

James, M. N. G., Sielecki, A., Salituro, F., Rich, D., & Hofmann, T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6137-6141.

Jaskólski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B., & Wlodawer, A. (1991) *Biochemistry 30*, 1600-9.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4686-4690.

McKeever, B. M., Navia, M. A., Fitzgerald, P. M., Springer, J. P., Leu, C. T., Heimbach, J. C., Herbert, W. K., Sigal, I. S., & Darke, P. L. (1989) *J. Biol. Chem. 264*, 1919-1921.

McQuade, T. J., Tomasselli, A. G., Liu, L., Karacostas, V., Moss, B., Sawyer, T. K., Heinrikson, R. L., & Tarpley, W. G. (1990) *Science 247*, 454-456.

Meek, T. D., Rodriguez, E. J., & Angeles, T. S. (1994) *Methods Enzymol. 241*, 127-156.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B., & Wlodawer, A. (1989) *Science 246*, 1149-1152.

Otwinowski, Z. (1993) in *CCP4 Study weekend: Data collection and processing, 29-30 Jan 1993* (Sawyer, L., Isaacs, N., & Bailey, S., Eds.) pp 56-62, SERC Daresbury Laboratory, England.

Parris, K. D., Hoover, D. J., Damon, D. B., & Davies, D. R. (1992) *Biochemistry 31*, 8125-8141.

Pearl, L., & Blundell, T. (1984) *FEBS Lett. 174*, 96-101.

Polgár, L. (1987) *FEBS Lett. 219*, 1-4.

Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L., & Kézdy, F. J. (1991) *J. Biol. Chem. 266*, 14554-14561.

Rich, D. H., Green, J., Toth, M. V., Marshall, G. R., & Kent, S. B. H. (1990) *J. Med. Chem. 33*, 1285-1288.

Richards, A. D., Phylip, L. H., Farmerie, W. G., Scarborough, P. E., Alvarez, A., Dunn, B. M., Hirel, P.-H., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V., & Kay, J. (1990) *J. Biol. Chem. 265*, 7733-7736.

Rosé, J. R., Salto, R., & Craik, C. S. (1993) *J. Biol. Chem. 268*, 11,939-11945.

Rose, R. B., Rosé, J. R., Salto, R., Craik, C. S., & Stroud, R. M. (1993) *Biochemistry 32*, 12498-12507.

Rutenber, E., Fauman, E. B., Keenan, R. J., Fong, S., Furth, P. S., Ortiz de Montellano, P. R., Meng, E., Kuntz, I. D., DeCamp, D. L., & Salto, R. (1993) *J. Biol. Chem. 268*, 15343-15346.

Sack, J. (1988) *J. Mol. Graphics 6*, 225.

Schecter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157-162.

Seelmeier, S., Schmidt, H., Turk, B., & von der Helm, K. (1988) *Proc. Nat. Acad. Sci. U.S.A. 85*, 6612-6616.

Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., & James, M. N. G. (1990) *J. Mol. Biol. 214*, 143-170.

Silva, A. M., Cachau, R. E., Sham, H. L., & Erickson, J. W. (1996) *J. Mol. Biol. 255*, 321-346.

Stroud, R. M., & Fauman, E. B. (1995) *Protein Sci. 4*, 2392-2404.

Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H., & Davies, D. R. (1987a) *J. Mol. Biol. 196*, 877-900.

Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D., & Davies, D. R. (1987b) *Proc. Natl. Acad. Sc.i U.S.A. 84*, 7009-7013.

Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B., & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805-8809.

Tomasselli, A. G., & Heinrikson, R. L. (1994) *Methods Enzymol. 241*, 279-301.

Tong, L., Pav, S., Pargellis, C., Do, F., Lamarre, D., & Anderson, P. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8387-8391.

Veerapandian, B., Cooper, J. B., Sali, A., & Blundell, T. L. (1990) *J. Mol. Biol. 216*, 1017-1029.

Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., & Hoover, D. J. (1992) *Protein Sci. 1*, 322-328.

Wlodawer, A., & Erickson, J. W. (1993) *Annu. Rev. Biochem. 62*, 543-585.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616-621.

Zhao, B., Winborne, E., Minnich, M. D., Culp, J. S., Debouck, C., & Abdel-Meguid, S. S. (1993) *Biochemistry 32*, 13054-13060.

Table Ia: Statistics for HIV-1 Protease Structures:

peptide in structure[a]	Ac-SLNF↓	PIV-NH_2↓	ambiguous	Ac-SLNF↓
peptide crystallized[b]	Ac-SLNF/	Ac-SLNF/PIV-NH_2	/PIV	Ac-SLNF/ + /PIV
resolution (Å)	2.2	2.3	2.5	2.3
no. of reflections ($I \geq 1\sigma(I)$)	29,000	42,000	15,000	14,000
no. unique reflections	9,700	8,700	6,400	7,800
completeness (%)	96	97	93	89
completeness in highest res. bin (%)	93	97	97	90
R_{sym} (%)	8	7	9	8
Average B factor ($Å^2$)	21	22	14	17
R-factor (from 7Å)	19.8	17.9	17.5	17.0
R-free (from 7Å)	28.3	28.5	28.0	26.2
sidechains without density[c]	7, 41, 43, 55, 70, 7',14',18',21',41', 43',55',65',70'	7, 8, 14, 41, 55, 70 7',41',43',55'	7, 18, 41, 43, 55 7',14',41',70',80'	7, 41, 43, 55, 70 7',14',41',43',55'
no. waters	31	42	19	36
rmsd from ideal:				
bond length (Å)	0.014	0.015	0.018	0.014
bond angles (°)	2.9	3.1	3.1	3.0
dihedral angles (°)	27.0	27.3	27.0	27.0

Table Ib: Statistics for SIV Protease Structures:

peptide in structure[a]	↓FLEK	↓F(NO_2)EAA	AAA↓	AAA↓
peptide crystallized[b]	/FLEK	/F(NO_2)EANleS	RVNle/	RVNle/ F(NO_2)EANleS
resolution (Å)	2.2	2.5	2.6	2.2
no. of reflections ($I \geq 1\sigma(I)$)	21,000	13,000	7,000	15,000
no. unique reflections	4,900	3,500	4,300	7,750
completeness (%)	93	98	85	80.5
completeness in highest res. bin (%)	83	100	86	72
R_{sym} (%)	5.3	7.9	5.1	7.4
Average B factor ($Å^2$)	29	14	20	21
R-factor (from 7Å)	21.1	17.4	18.2	17.0
R-free (from 7Å)	29.5	28.5	28.0	26.0
sidechains without density[c]	7,37,70,72	37	7, 70, 72, 79 7',61',70',72'	7, 37, 70, 72 7',13',37',70',72'
no. waters	19	15	24	42
rmsd from ideal:				
bond length (Å)	0.016	0.014	0.008	0.016
bond angles (°)	3.1	3.1	1.55	3.1
dihedral angles (°)	27.0	26.7	26.6	27.5

[a]Arrow denotes peptide residue closest to the catalytic aspartic acids in the structure. An arrow to the right of the sequence denotes a P-product, an arrow to the left denotes a Q-product.

[b]Slash denotes location of the scissile bond in the substrate or, for the products, the substrate from which the product was derived.

[c]These residues were built as alanines in the structure. For structures with a dimer in the asymmetric unit, unprimed residues belong to the monomer more closely associated with the P-product; primed residues belong to the monomer more closely associated with the Q-product.

Table IIa: Dimensions of HIV-1 Protease Unit Cells:

peptide space group	Ac-SLNF/ $P2_12_12_1$	Ac-SLNF/PIV-NH_2 $P2_12_12_1$	/PIV $P2_12_12_1$	Ac-SLNF/ +/PIV $P2_12_12_1$
a	51.1	51.3	51.2	51.5
b	59.3	59.1	58.9	59.3
c	61.4	62.3	58.9	62.4

Table IIb: Dimensions of SIV Protease Unit Cells:

peptide	/FLEK	/F(NO_2)EANleS	RVNle/	RVNle/ + F(NO_2)EANleS
space group	$C222_1$	$C222_1$	C2	C2
a	63.7	62.3	62.8	62.6
b	32.5	32.1	32.0	32.0
c	97.7	96.3	96.4	96.3
β			90°	90°

Table III: Protease Sidechains Forming the Peptide Binding Subsites:

subsite	protease	[a]peptide	[b]protease contacts
S4	HIV-1	Ac-**S**-L-N-F↓	**D29, D30, K45, I47**
S3	HIV-1	Ac-S-**L**-N-F↓	**G48, D29**, R8'
S2	HIV-1	Ac-S-L-**N**-F↓	**G27, A28, D29, D30, I47, G48, I84**, I50'
S1	HIV-1	Ac-S-L-N-**F**↓	**G49, I50**, L23', P81', V82', I84'
S1'	SIV	↓**F**-L-E-K	**G27'**, L23, I82, D25, I84
S2'	SIV	↓F-**L**-E-K	**A28', D30', I32', I84', G48'**, I50
S3'	SIV	↓F-L-**E**-K	**D29'**, R8
S4'	SIV	↓F-L-E-**K**	**I46', D30', E58', K45', M76'**

[a]In bold and underlined are the residues of the peptide occupying the respective subsites

[b]In bold are the residues of the monomer which makes the backbone hydrogen bonds to the peptide product. The unprimed monomer is hydrogen bonded to the P-product Ac-S-L-N-F↓. The primed monomer is hydrogen bonded to the Q-product ↓F-L-E-K.

FIGURE LEGENDS

Figure 1: Final $2|F_o|-|F_c|$ density (contoured at 1σ) in the binding pocket of HIV-1 protease bound to the peptide products a) Ac-S-L-N-F↓ and b) P-I-V-NH_2↓. Density is also shown for the active site aspartic acids and water 401. Density for residues 46-53 and 46′-53′ (the flaps), and residues 26-30 and 26′-30′ (the base of the active site) is not displayed. a) The peptide and three water molecules have been built into the binding pocket. These waters do not explain all of the density. b) The peptide is seen to bind "backwards" from the sequence in the substrate Ac-S-L-N-F↓P-I-V-NH_2, with the valine in the P1 position. The unliganded half of the binding pocket contains density for three water molecules: waters 521 and 522 (displayed) and water 520 (not displayed because it is directly in front of the C-terminus). (cross-eyed stereo)

Figure 2: Diagram of the "triangular interaction" among the three carboxyl groups: the active site aspartic acid residues and the C-terminus of the phenylalanine residue of the peptide product Ac-S-L-N-F↓. The distances between carboxyl oxygens and the angles between carbon-oxygen-oxygen atoms are displayed. (cross-eyed stereo)

Figure 3: Tautomers of the N- and C-termini of products interacting with the catalytic aspartic acids. a) Two tautomeric forms of the carboxy terminus of the P-product Ac-S-L-N-F↓. The left hand tautomer is geometrically favored. b) Two tautomers of the amino terminus of the Q-product ↓F-L-E-K. The amino terminus forms a hydrogen bond with one of the aspartic acids. At pH 5.4 the amino terminus would be protonated. c) Two tautomers of the amino terminus of the Q-product ↓F(NO_2)-E-A-Nle-S. The amino terminus forms hydrogen bonds with both aspartic acids.

Figure 4: The conformation of the bound product peptides are compared with those of peptidomimetic inhibitors. The peptides were overlapped by superimposing the carbon alpha atoms of the proteases. a) The P-product Ac-S-L-N-F↓ (black) is overlapped with the inhibitor JG365 (gray), of the same N-terminal sequence. The peptides differ the most at the P1 carbonyl of Ac-S-L-N-F↓. The third peptide, U-85548E (white), adopts the same backbone conformation. b) The Q-products ↓F(NO_2)-E-A-Nle-S (black) and ↓F-L-E-K (gray) overlap closely at the P2' carbon alpha, but less well at P1' and P3'. The same is true of the peptidomimetic inhibitor SKF107457 (white). The P2' residue is usually a glutamate or glutamine in substrates. (cross-eyed stereo)

Figure 5: Final $2|F_o|-|F_c|$ density in the binding pocket of SIV protease bound with the peptide products a) ↓F-L-E-K (contoured at 0.9σ) and b) ↓F(NO_2)-E-A-Nle-S (contoured at 0.8σ). The monomers of the dimer are related by a crystallographic 2-fold axis; density is shown only for one of the peptides in the active site. Density is also shown for the active site aspartic acids and for water 401, between the flaps. Density for residues 46-53 and 46'-53' (the flaps), and residues 26-30 and 26'-30' (the base of the active site) is not displayed. b) Density between the aspartic acids is unaccounted for. (cross-eyed stereo)

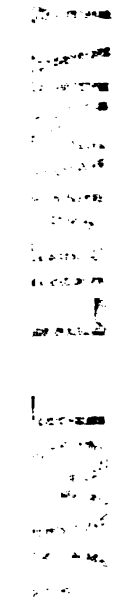


Figure 6: Model of the substrate (gray) overlapped with the two products (white): Ac-S-L-N-F↓ and ↓F-L-E-K. The scissile carbonyl of the substrate adopts a range of equal energetic conformations (black line) without strain: the two extreme conformations are displayed. The position closest to OT1 of the C-terminal carboxyl group of Ac-S-L-N-F↓ is within hydrogen bonding distance (dotted line) of Asp 25' (black). Hydrogen bonds are also shown to water 401. (cross-eyed stereo)

Figure 7: A proposed chemical mechanism with the four primary reaction intermediates represented. The unliganded protease shows the catalytic water hydrogen bonded between the two aspartic acids. Rotations around the C_β-C_α bond are represented by an arched arrow. The structure of the Michaelis complex shows the catalytic water displaced by the scissile carbonyl, in position for nucleophilic attack. The hydroxyls of the gemdiol intermediate hydrogen bond with both catalytic aspartic acids equally. The nitrogen of the scissile bond is shown being protonated by one of the gemdiol hydrogens. Following cleavage of the peptide bond, the products are released randomly (double arrows). The amino group of the Q-product becomes protonated. The carboxyl group of the P-product can form the triangular interaction with the two aspartic acids.

Figure 8: Model of the Michaelis complex. The scissile carbonyl of the substrate (white) is hydrogen bonded (dotted lines) to Asp 25' (gray). The catalytic water (black) forms a bifurcated hydrogen bond to the two oxygens of Asp 25. Hydrogen bonds to water 401 are shown, which also forms hydrogen bonds with the flaps (black). (cross-eyed stereo)

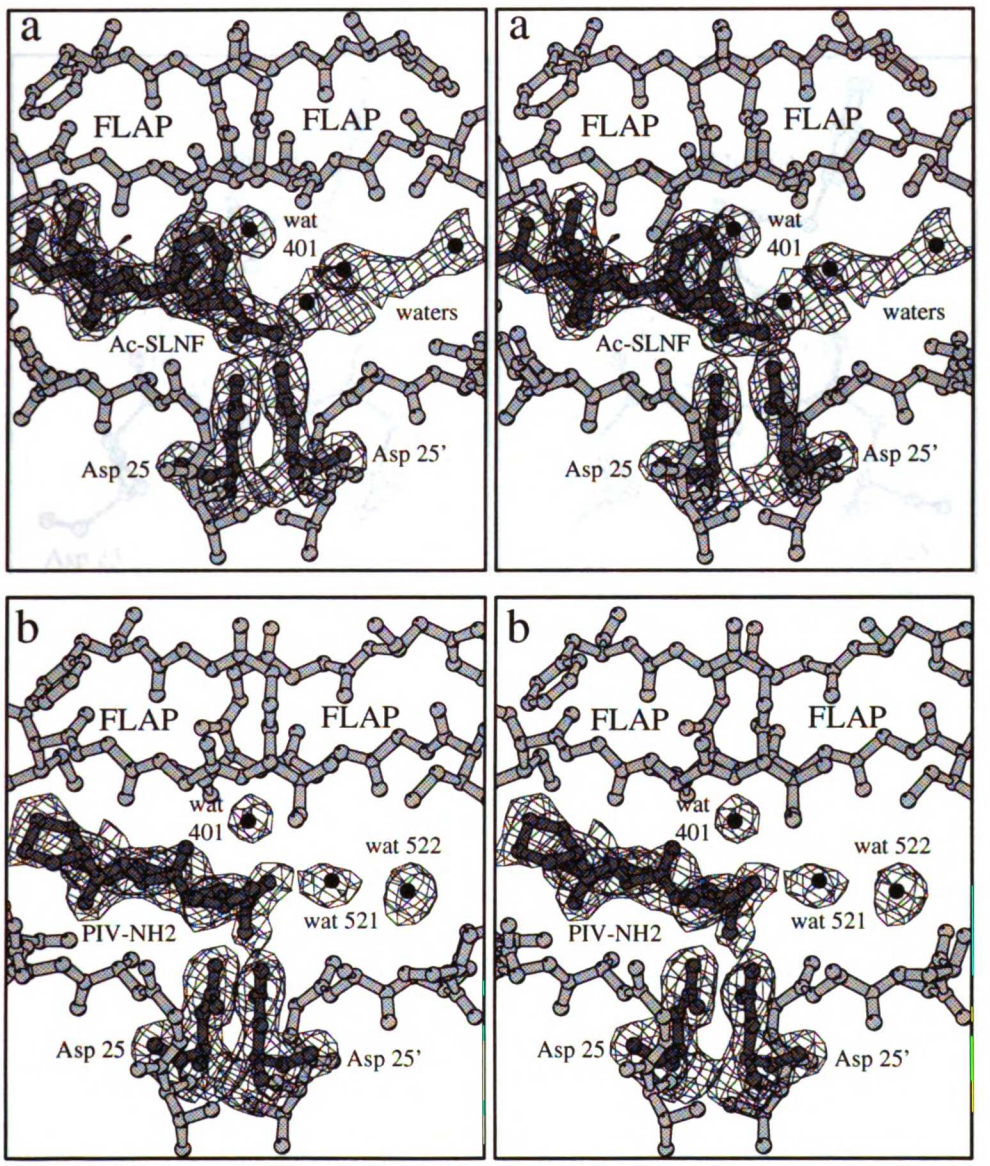

a
FLAP
FLAP
wat
401
waters
Ac-SLNF
Asp 25
Asp 25'
a
FLAP
FLAP
wat
401
waters
Ac-SLNF
Asp 25
Asp 25'
b
FLAP
FLAP
wat
401
wat 522
wat 521
PIV-NH2
Asp 25
Asp 25'
b
FLAP
FLAP
wat
401
wat 522
wat 521
PIV-NH2
Asp 25
Asp 25'

Figure 1

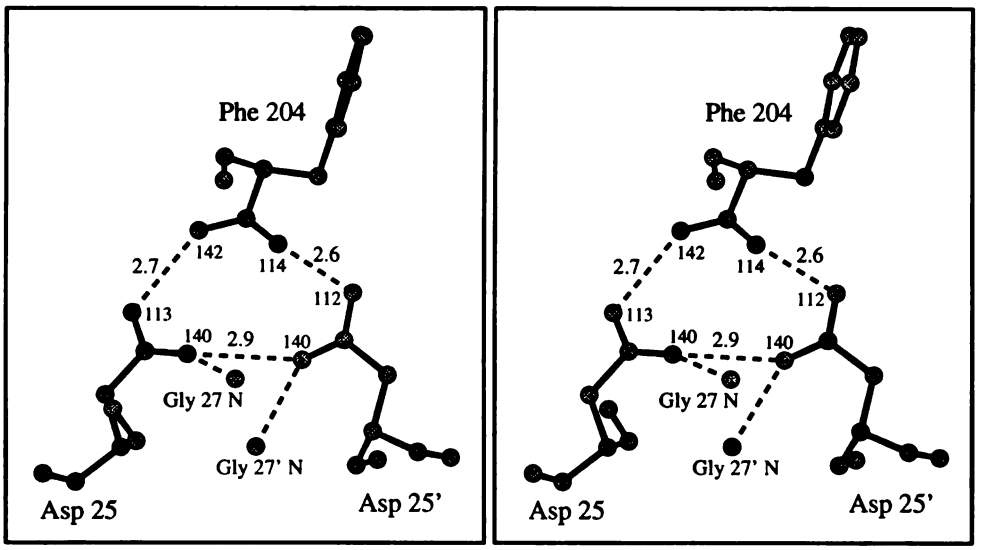
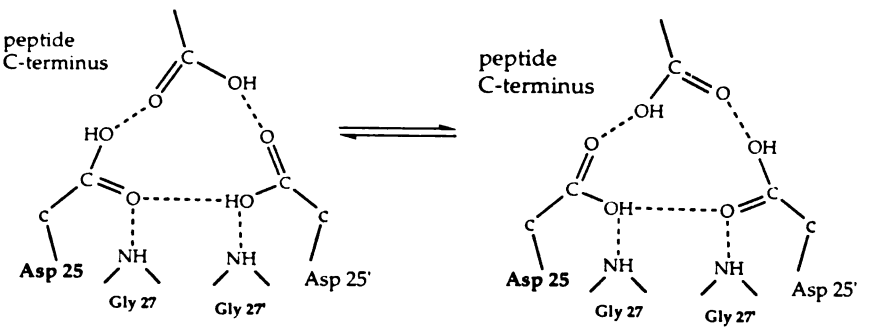
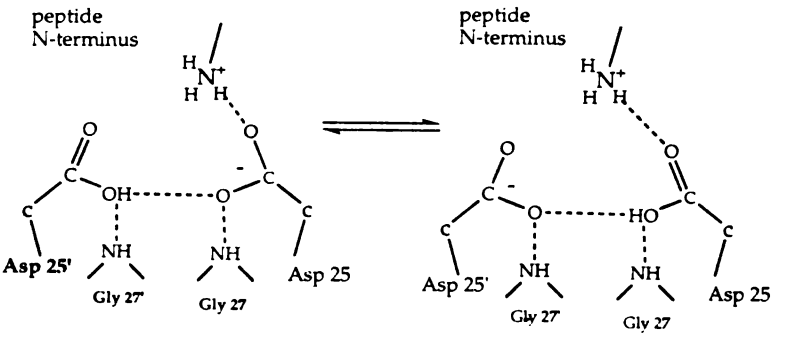

Phe 204
2.7
142
114
2.6
112
113
140
2.9
140
Gly 27 N
Gly 27' N
Asp 25
Asp 25'
Phe 204
2.7
142
114
2.6
112
113
140
2.9
140
Gly 27 N
Gly 27' N
Asp 25
Asp 25'

Figure 2

a. Ac-S-L-N-F



b. F-L-E-K



c. F(NO_2)-E-A-Nle-S

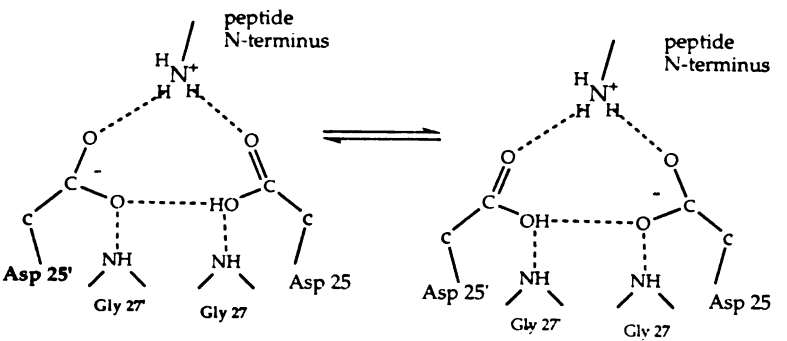


Figure 3

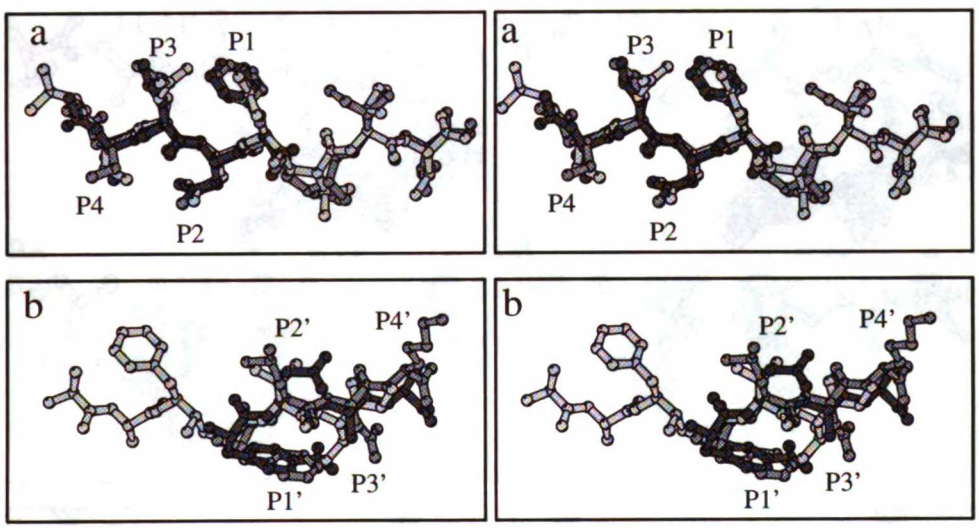


Figure 4

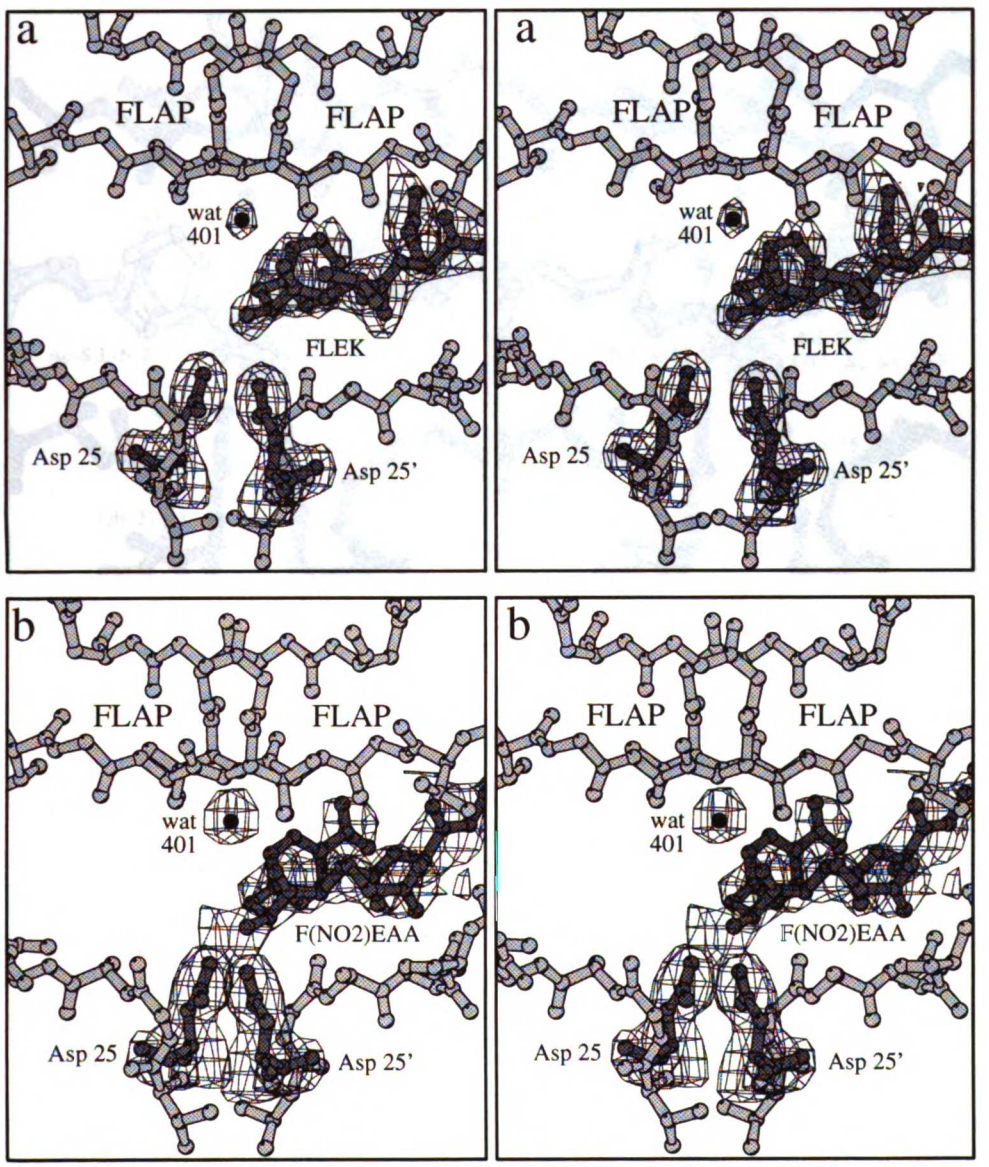


Figure 5

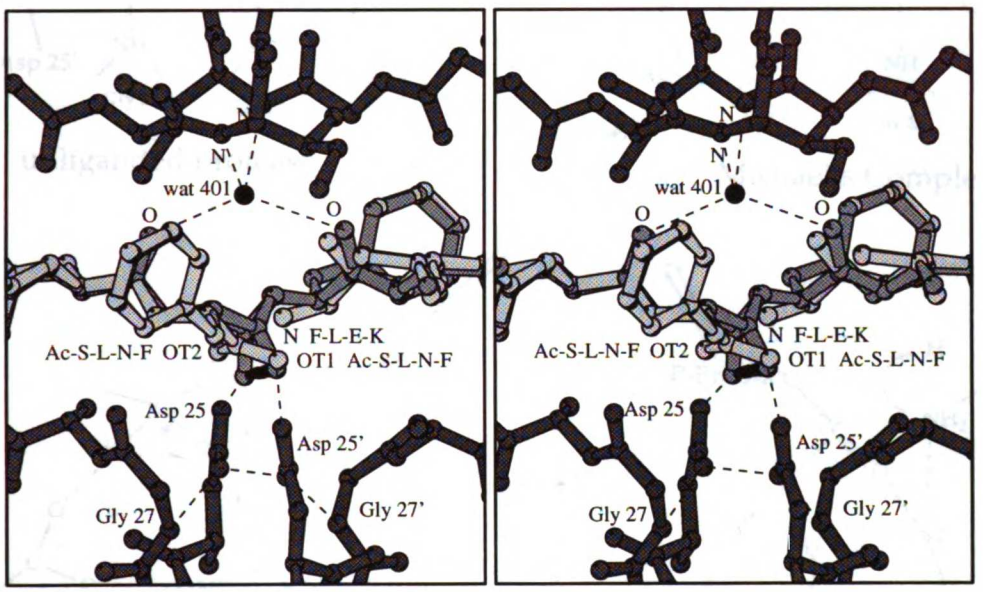

N
N
wat 401
O
O
N F-L-E-K
Ac-S-L-N-F OT2
OT1 Ac-S-L-N-F
Asp 25
Asp 25'
Gly 27
Gly 27'
N
N
wat 401
O
O
N F-L-E-K
Ac-S-L-N-F OT2
OT1 Ac-S-L-N-F
Asp 25
Asp 25'
Gly 27
Gly 27'

Figure 6

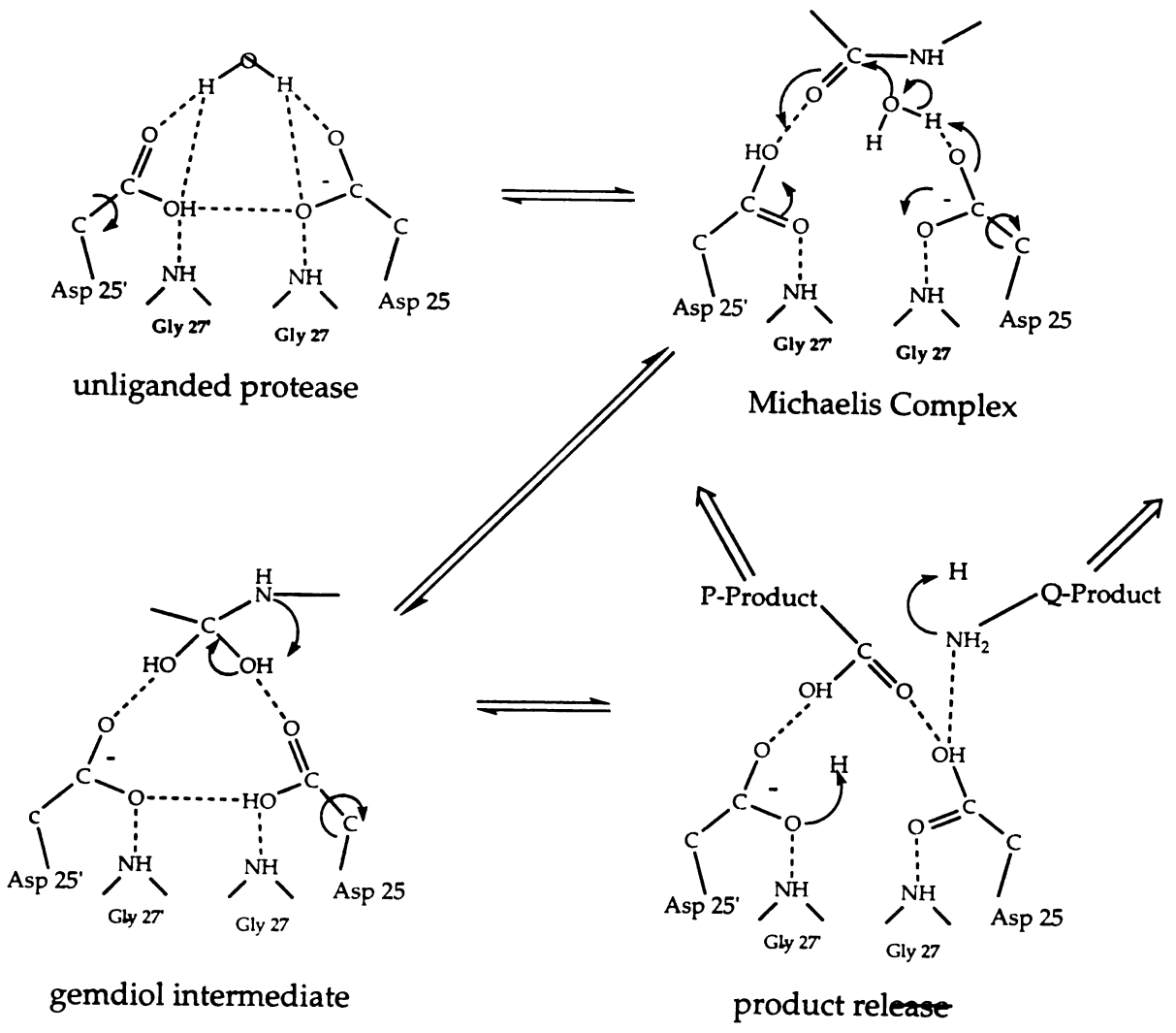

unliganded protease
Michaelis Complex
gemdiol intermediate
product release
P-Product
Q-Product
Asp 25'
Gly 27'
Gly 27
Asp 25

Figure 7

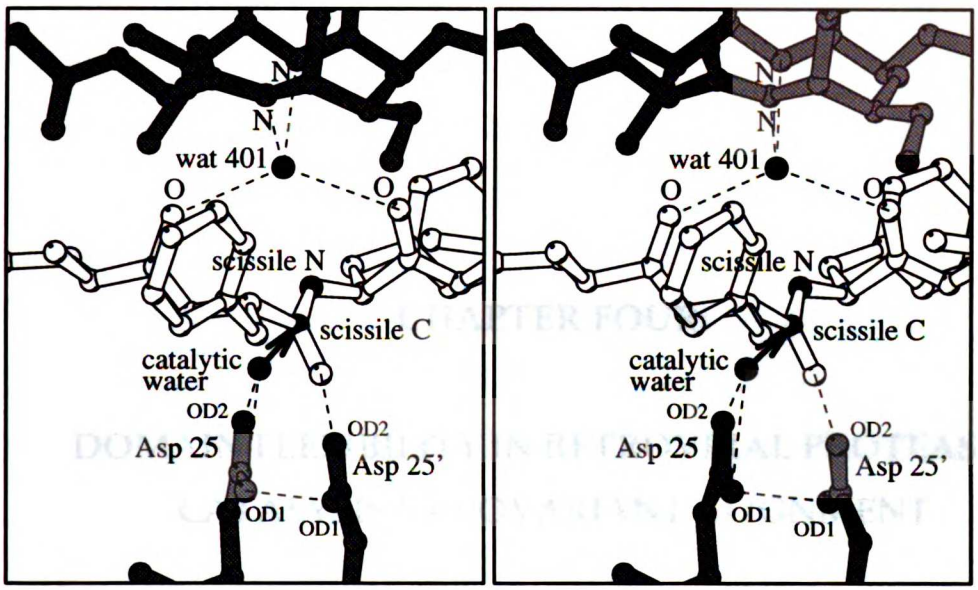

N
N
wat 401
O
O
scissile N
scissile C
catalytic
water
OD2
Asp 25
OD2
Asp 25'
OD1
OD1
N
N
wat 401
O
O
scissile N
scissile C
catalytic
water
OD2
Asp 25
OD2
Asp 25'
OD1
OD1

Figure 8

CHAPTER FOUR:

DOMAIN FLEXIBILITY IN RETROVIRAL PROTEASES: CATALYSIS BY COVARIANT ALIGNMENT

Submitted for publication to *Proteins*

Domain Flexibility in Retroviral Proteases:

Catalysis by Covariant Alignment

Robert B. Rose[1], Charles S. Craik[1,2], Robert M. Stroud[1,2*]

[1]Graduate Group in Biophysics, and [2]Departments of Pharmaceutical Chemistry and Biochemistry and Biophysics, The University of California, San Francisco, San Francisco, California 94143

*Correspondence should be addressed to Dr. Robert M. Stroud, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143-0448.

phone: (415) 476-4224

fax: (415) 476-1902

short title: Domain Flexibility in Retroviral Proteases

key words: HIV, SIV, AIDS, proteinase, aspartyl, flaps, apo, resistance, crystallography, X-ray

Abbreviations: EPNP, 1,2-epoxy-3-(*p*-nitrophenoxy)propane; HIV, Human Immunodeficiency Virus; rms, root-mean-square; rmsd, root-mean-square deviation; SIV, Simian Immunodeficiency Virus; amino acids of proteins designated by the three-letter code; amino acids of peptide ligands designated by the one-letter code.

ABSTRACT

Domain motions associated with substrate binding are identified for retroviral proteases, in addition to closing of the flaps. A crystal structure of unliganded SIV protease (with the point mutation Ser 4 His) was determined in a new space group, $P3_221$, to 2.0Å resolution. The structure is in an "open" conformation, with 6 residues of the flaps disordered. Comparing the structure with liganded structures by difference distance matrices identified five rigid body domains of the protease dimer: one terminal domain encompassing the N- and C-terminal beta sheet of the dimer, two core domains containing the two catalytic aspartic acids, and two flap domains. The two core domains rotate together during substrate binding, positioning the aspartic acids and reshaping the binding pocket for catalysis. A mechanism of catalytic alignment is proposed which employs the substrate binding energy to stabilize the enzyme and substrate in their catalytically active conformations.

INTRODUCTION

Liganded and unliganded structures of Human Immunodeficiency Virus (HIV) protease show significant conformational changes that accompany substrate binding [1]. The movement has been described as a "hinge motion" between monomers of the protease homodimer. This motion narrows the binding pocket slightly, but the major conformational rearrangement observed was of two beta strands, the flaps, that encircle the bound substrate. Flexibility of the flaps is necessary for binding the protein substrate or intact gag and gag-pol polyproteins of the virus. Once "closed" the flaps form the "ceiling" of the protease binding pocket. The catalytic aspartic acids, one contributed from each monomer of the protease, are located at the "floor" of the binding pocket. The flaps in the "open" conformation of unliganded HIV-1 protease crystal structures are not sufficiently separated to allow a polyprotein substrate to enter the active site [2-5]. The unliganded structure of a related retroviral protease from Rous Sarcoma Virus shows nine residues of the flaps to be disordered, indicating multiple conformations of the flaps [6]. Rapid flap motions have been measured by NMR [7] and fluorescence [8]. Larger motions of the flaps have been modeled by molecular dynamics [9-11], including motions of residues bordering the flaps described as the 'cantilever' [10, 11].

The monomeric proteases are bilobal instead of dimeric, with a pseudo-two-fold symmetry axis relating the two lobes. The retroviral proteases are related evolutionarily to the monomeric aspartyl proteases, as indicated by the conserved catalytic Asp-Thr-Gly sequence and the structural homology [12-14]. One catalytic aspartic acid is positioned on each lobe of the monomeric proteases. Only one of the lobes has retained a flap, which also closes over a

bound substrate. Conformational changes in the flap of the monomeric proteases associated with ligand binding have also been observed [15]. But comparing the liganded and unliganded structures has identified more global rigid body rotations of the two lobes [16]. We describe a similar rigid body rotation of domains of the retroviral proteases.

Conformational changes of the pepsin-like aspartyl proteases have been invoked to explain kinetic observations such as the broad substrate specificity which is a feature of many aspartyl proteases, the increase in k_{cat} with substrate length, biphasic substrate binding, and the rate enhancement of k_{cat} in the presence of small non-cleaved peptides [17, 18]. The movements implied by the crystal structures have reinforced these explanations of the kinetic data [16, 19].

We have previously reported liganded structures of SIV protease with their flaps closed [20, 21]. We report an unliganded structure of the Simian Immunodeficiency Virus (SIV) protease. A point mutation Ser 4 His has been engineered into the protein to retard autoproteolysis [21]. The structure is more "open" than described in previous unliganded HIV protease structures. Indeed an unliganded SIV protease structure has been reported [22] in which the flaps were in the closed conformation. As a result, the authors proposed that the open HIV-1 protease structures were artifacts of crystal packing. Our unliganded SIV protease structure indicates that the open conformation occurs readily, as required by the protease function.

The unliganded protease structure is compared with liganded structures to identify rigid body motions of domains, as found in the pepsin-like aspartyl proteases. The conformational changes of the protein can be discussed in terms of "induced fit" [23]. The rigid body motions suggest one mechanism by which HIV protease is capable of mutating to select against

anti-protease drugs - by altering the equilibrium between the open and closed conformations of the protease.

MATERIALS AND METHODS

Recombinant SIV protease, with the point mutation Ser 4 His, was purified and frozen as described previously [21]. The protein was thawed and concentrated for crystallization using an 8ml. Collodion apparatus (Schleicher and Schuell), following the same protocol as previous crystallizations [21]. A few hundred microliters of 1N sodium hydroxide was added to the protein after thawing to prevent precipitation. The protein solution, now at pH 10, was transferred to the Collodion membrane and dialyzed against the sodium acetate buffer at pH 5.4.

Crystals were grown by the hanging drop vapor diffusion method. Crystals grew to a size of 500 x 200 x 100 microns. The crystal morphology was pyramidal, different than the plates we described previously that were in space group $C222_1$ [21]. Data was collected on an R-axis II image plate detector using X-rays generated from a Rigaku 18kW generator. Data collection statistics are listed in Table I. The reflection intensities were integrated and scaled using the programs DENZO and SCALEPACK [24]. The rotational symmetry was determined with the program XPREP (Copyright 1991, Siemens Analytical X-ray Instruments). The orientation of the molecule in the new space group was determined by Molecular Replacement using the program XPLOR [25]. XPLOR was also used to refine the structure. Manual building was carried out using the program CHAIN [26].

Two inhibitors, curcumin and biliverdin, were also crystallized with SIV protease in the same space group as the unliganded structure. Curcumin

reversibly inhibits HIV-1 and HIV-2 proteases with IC_{50}'s of 100mM and 250mM, respectively [27]. Biliverdin has been shown to have anti-viral effects in cell culture [28]. It inhibits SIV protease with an IC_{50} of 8μM [29]. A range of inhibitor concentrations were tested for co-crystallization with SIV protease. Crystals grew at about 200μM concentration in both cases, well below the IC_{50} of curcumin. The final protein to inhibitor ratio in the crystallization drop was about one to one. Higher inhibitor concentrations interfered with crystal growth.

Structure comparisons

The unliganded SIV protease structure was compared with two liganded structures: with the bound product peptide F-L-E-K [20] and with a bound peptidomimetic inhibitor (1siv.pdb) [30]. A further comparison was made between unliganded and liganded HIV-1 protease structures. Three unliganded HIV-1 structures have been deposited in the Brookhaven Protein Data Bank (PDB) [2, 4, 5], and these were compared with the liganded protease structure determined previously [31] and 18 liganded structures from the PDB (Table II). The structures were superimposed by a least squares minimization of the carbon alpha positions implemented in the program GEM (Eric Fauman, personal communication).

Rigid body domains were identified using a difference distance matrix of the carbon alpha atoms of the liganded and unliganded structures [32]. Domains composed of spatially contacting residues were identified with a difference distance of less than 0.5Å using the program NewDome [33], The domains were then compared visually to determine the most logical endpoints between the domains. For example, the one short helix of the structure was assigned completely to one domain. Individual residues were inspected which were associated structurally with a particular domain but deviated by more than 0.5Å from the difference distance of the rest of the domain.

Rigid body motion was described in terms of a screw axis which results in the best superposition of one structure onto the other [34]. The program GEM calculates the screw axis resulting from a superposition. In comparing the domain motions between the unliganded and liganded structures, the

"terminal" domains were first superimposed, then the axis was determined which best superimposed the "core" domains.

RESULTS

Structure determination of unliganded SIV protease:

The data was reduced in the space group $P3_121$ or $P3_221$ with unit cell dimensions of 45.4Å, 45.4Å, 87.9Å(R_{symm}=8.6%). The crystal density of 2.4Å^3/dalton indicated the crystal contained a protease monomer in the asymmetric unit [35]. The protease dimer axis was therefore coincident with the crystallographic 2-fold axis. The positioning of the molecule was determined by Molecular Replacement. The initial search model consisted of an SIV protease monomer as determined in our laboratory, excluding the ligand and the flap residues 48-52 [21]. After Patterson Correlation refinement, only one solution gave a correlation coefficient greater than 0.1, equal to 0.12. The translation search identified the enantiomorph, and the screw axis, as $P3_221$ instead of $P3_121$. The R-factor after the translational search was 48%.

Extra parameters were included based on use of the free R-factor [36]. After rigid body and positional refinement, the density for the flap residues 48-52 and 48'-52' was discontinuous and weak. 39 water molecules were added to the structure. Simulated annealing reduced the R-factor by 1% and R-free by 1.4%. Manual rebuilding and positional refinement led to a final R-factor of 20.4% and R-free of 26.6%. Table I indicates final refinement statistics.

Description of unliganded SIV protease structure

There was discontinuous density for flap residues 48-53. In a 2|Fo|-|Fc| map there was some density for the mainchain of residue 53, but no density for the Phe 53 sidechain. Residues 54-55 and 46-47 were in good density, except for the Ile 54 sidechain. These residues were from 1-1.5Å retracted from the dimer 2-fold axis relative to the liganded structure. The B-factors of these residues increased to about 60Å^2, from an average of 20Å^2 for the protein. The gap between the ordered flap residues is 20Å wide, large enough to allow a substrate to enter.

Subsequently, we obtained liganded SIV protease structures in space group $C222_1$ [21]. The unliganded SIV protease has also been crystallized in the "closed" conformation in this crystal form [22]. The closed conformation grew from 100mM sodium acetate, 0.2M NaCl, at pH 5.5. The conditions in which the open conformation of the unliganded protease was crystallized were similar except for the pH (100mM sodium cacodylate, pH 6.5).

The catalytic aspartic acids are hydrogen bonded to a water molecule, as observed in the unliganded HIV-1 structures [2]. The water molecule is within 2.7Å of the inner carboxyl oxygen of the aspartic acid sidechain, and 3.1Å of the outer oxygen. No 2|Fo|-|Fc| density in the map greater than 1σ is visible in the active site. Particularly, the water normally hydrogen bonded to the amide nitrogens of Ile 50 of both flaps when substrate analogs are bound is not visible, as was true of the unliganded HIV-1 protease structures.

Domains of SIV protease

The monomers of the "closed" (liganded) and "open" (unliganded) forms of SIV protease superimpose more closely than do the dimers. Comparing the unliganded structure to that of SIV protease bound to the product peptide sequence F-L-E-K [20], the rms deviation of the carbon alphas for the dimers was 1.5Å, while the rms deviation for the monomers was 0.7Å. A "core" domain of the monomer encompassing the catalytic aspartic acids was identified using a difference distance matrix. A second "flap" domain included the beta strands of the flap. A third "terminal" domain consisted of the N-and C-terminal residues and the helices from both monomers. Figure 1 is a ribbon diagram of the SIV protease structure in the closed flap conformation [21]. The domains are identified. Features of the protease that will be referred to in the text are labeled here for reference.

Figure 2 shows the dimer of SIV protease bound to the peptide F-L-E-K superimposed on the respective domains of the unliganded SIV protease structure. Each of these domains superimposed with an rms deviation for the carbon alpha atoms of 0.4Å. The error in the atomic coordinates estimated from the average B-factor of the carbon alpha atoms of each domain was: for the unliganded structure, 0.2Å for each domain, for the structure bound to F-L-E-K, 0.3Å for the terminal and core domains, and 0.4Å for the flap domain [37]. Therefore the rms deviation expected between the two structures is about 0.4Å for each domain. The deviation of the three domains superimposed are within the expected random error from the measurements. The deviations of the monomer or dimer superpositions are significantly greater than the expected random error.

Similar conclusions resulted from a comparison of the unliganded SIV protease and the protease bound to the inhibitor SKF107457, 1siv.pdb (from

Brookhaven data bank) [30] (Table II). The first two residues of the terminal beta domain have a different conformation in this structure and have been excluded from the calculations of rms deviation.

The terminal domain encompasses the four-stranded beta sheets, residues 1-3 and 95-99 of each monomer, the turn, residues 4-9, and the two helices, residues 86-94, of each monomer. The backbone hydrogen bonds of these residues are either solvated or satisfied within the domain. Arg 8 is the only residue in this domain in direct contact with a peptide ligand. In the absence of ligand the sidechain repositions to hydrogen bond with solvent.

The core domain consists of six primarily beta-strand structures. The domain sequence, residues 10-32 and 63-85, is disrupted by the flap residues. The catalytic aspartic acid (residue 25) is part of a conserved turn structure at the interface of the core domains of the two monomers, forming the "fireman's grip" with the symmetry-related turn. The largest deviations between these domains are 0.7Å and occur in the loop 77-82: residues 77-79 contact residues 55-57 of the flap domain, and residues 81 and 82 form part of the S1 subsite. There are also deviations this large in two external loops of 1siv.pdb: residues 62-63 and residue 19.

The interface between the core and terminal domains is composed primarily of small hydrophobic residues. Figure 3a shows 2|Fo|-|Fc| density for the interface. The helix of the terminal domain packs against several beta-strands of the core domain. The helical residues are completely hydrophobic except for the N-terminal residues: Arg 87 and Asn 88.

The flap domain has been assigned to residues that shift significantly when the core domains of the liganded and unliganded structures are superimposed. This includes residues preceding and following the flap residues that are disordered in the unliganded structure (residues 48-52). The

preceding residues, residues 33-46, include a solvent exposed loop, residues 38-44, plus five residues that penetrate into the core domain, residues 33-37. Residues 55-58 extend the beta-strand of the flap, hydrogen bonding to residues 43-45. Residues 57-62 form a turn that is mostly solvent exposed.

Comparison with HIV-1 protease

Three unliganded HIV-1 structures have been deposited in the Brookhaven protein data bank: 3hvp.pdb [2], 3phv.pdb [5], and 1hhp.pdb [4]. These structures are almost identical: the rms deviation for the carbon alpha atoms of the dimer between 3hvp.pdb and 3phv.pdb or 1hhp.pdb are 0.6Å and 0.5Å respectively. We compared the unliganded structure, 3hvp.pdb, with 2 liganded structures: bound to a product peptide, S-L-N-F [20], and bound to the peptidomimetic inhibitor JG365 [38]. Again the monomers superimposed with a smaller rms deviation (0.8Å for residues 1-45, 55-99) than that of the dimers superimposed (1.2Å for residues 1-45, 55-99 of both monomers). When the core domains as defined for SIV protease were superimposed, the rms deviations were 0.6Å or 0.7Å, depending which monomer of 7hvp.pdb was compared. Excluding the flexible loop, residues 77-81, the rms deviations were both 0.5Å. The terminal domains superimposed with an rms deviation of 0.3Å, and the flap domains superimposed with an rms deviation of 0.6Å.

In order to evaluate whether the orientation of these domains varied depending on the ligand in the active site, we superimposed 18 liganded structures deposited in the Brookhaven protein data base. Table III lists the structures compared along with the type of ligand. Figure 4 shows the carbon alpha traces of the structures superimposed on the dimer of 7hvp.pdb (residues 1-45, 55-99 of both monomers). The variation among the structures

cannot be explained by a difference in domain rotation but instead correlates with the B-factor (see below).

Rigid body movement of the core domain

The screw axes which describe the rigid body movement of the "core" domains relative to the "terminal" domain are displayed in Figure 5 for both SIV and HIV-1 proteases. In all cases the translation along the screw axis is negligible. The rotation is about 7° for each core domain of SIV protease, and about 4° for each core domain of HIV-1 protease. In the liganded structures, the core domains are rotated together, bringing the flaps together and narrowing the binding pocket. The axes are within 5Å from the carbon alpha atoms of the active site aspartic acids in SIV protease and within 2.5Å from the carbon alpha atoms of the aspartic acids in HIV-1 protease and perpendicular to the dimer 2-fold axis. The proximity of the axes to the core of the protein reflects the greater shifts at the exterior of the protein, as great as 4Å between liganded and unliganded structures when the terminal domains are overlapped. Figure 3b compares the interactions at the interface of the terminal and core domains in the open and closed conformations.

The "fireman's grip", residues 25-27, rotates less than predicted by the average rotation of the core domain relative to the terminal domain, suggesting some flexibility between this structure and the rest of the domain. The sidechains of the catalytic aspartic acids are rotated apart, resulting in an increase of the distance between the two catalytic aspartic acids in the unliganded structure. Table IV compares the distance between atoms of the two catalytic aspartic acids of the liganded and unliganded SIV protease structures. The C_α–C_α distances are identical, within error, and the difference between the distances increases along the sidechain. The carboxyl oxygens of

the aspartic acids sidechains are within optimal hydrogen bonding distance with the peptidomimetic inhibitor bound (2.8Å in 1siv.pdb), but are too far apart to form a good hydrogen bond in the unliganded structure (3.2Å).

The relative movements of the catalytic aspartic acids in the liganded and unliganded HIV-1 protease structures was less pronounced (Table IV). The carboxyl oxygens of the catalytic aspartic acids are 2.8Å apart in the liganded structure, and 3.0Å apart in the unliganded structure.

Substrate binding

The substrate binding pocket is dramatically reshaped in the open conformation. Protease structures with peptidomimetic inhibitors indicate that substrates bound to the protease lie in an extended beta sheet conformation [1]. The backbone hydrogen bonds of the P1-P4 positions of the substrate are made to one core domain, while those of the P1'-P4' positions are made to the other core domain (nomenclature of Schecter and Berger) [39]. In the unliganded structure, these domains are rotated apart and therefore cannot form hydrogen bonds to substrate residues on both sides of the scissile bond, at least with the peptide in its extended conformation. Figure 6 shows a superposition of the unliganded and peptide-bound SIV protease structures. The monomer associated with the P1'-P4' residues of the ligand has been used for the superposition. In this figure, the aspartic acids of the unliganded structure can be seen to be rotated apart.

The S1 and S3 subsites are also disrupted in the open structure. When the core domains of one monomer of the liganded and unliganded structures are superimposed, residues Leu 23, Pro 81, Ile 82, and Ile 84 from the other monomer, forming the S1 and S1' subsites, are rotated away. Table V lists the shifts in position of these residues in the liganded versus unliganded

structures. Also Gly 27 from the superimposed monomer reorients in response to ligand binding. In the unliganded structure, the carbonyl of Gly 27 is almost 90° from its position when it is hydrogen bonded to the backbone amide of the P2 residue. The S3 and S3' subsites are formed partly by the sidechains of Arg 8' and Arg 8. The sidechain reorients to interact with the ligand when it binds.

Local Flexibility

The rigid body movements have been damped by the crystal contacts or ligand binding in these structures. Some of the shorter range flexibility of the protein is reflected in the temperature factors. The superposition of 19 liganded HIV-1 protease structures indicates flexible regions of the protein. These structures derive from two space groups. The deviation from the mean for the 11 structures in space group $P2_12_12_1$ and the structures in space group $P6_1$ only differed significantly at residues 35-39, which were more dispersed in the hexagonal space group. The C-terminal 2 residues are in identical positions in the hexagonal space group and in different conformations in the orthorhombic space group. The average distance from the mean position of each carbon alpha atom is plotted in Figure 7. This is compared with the average B-factor at that position (B-factors were normalized to the average B-factor of structure 7hvp.pdb). The trends of these two descriptions are highly correlated, suggesting that the B-factors are a good indicator of the flexible regions of the protein. The "fireman's" grip residues (residues 25-27, which includes the catalytic aspartic acids) are particularly stable. The most mobile residues are residues 37-40 and residues 16-17 which belong to a surface loops, and residues 77-81 which participate in forming the P1 and P1' binding pockets.

DISCUSSION

We report an unliganded SIV protease structure which is more "open" than the unliganded HIV-1 protease structures previously reported. Distance difference matrices were used to determine domains of SIV protease that were relatively rigid in the liganded and unliganded structures. Proper identification of the domains is necessary to characterize the motions accompanying ligand binding. We identify five domains of the protease dimer: two core domains, each containing one of the catalytic aspartic acids, a terminal domain forming the interface between the core domains, and two flap domains. Ligand binding and flap closure is accompanied by a rigid body rotation of the two core domains relative to the terminal domain. The description of these domains as rigid bodies is not meant to imply that they are rigid; they are also deformable. For instance the Asp-Thr-Gly turns of the "fireman's" grip rotate less than the rest of the core domain. Residues belonging to surface loops or forming the P1 binding pocket are flexible as well.

A range of conformations for unliganded retroviral protease structures have now been reported, from the a completely closed conformation of SIV protease [22] to the most open conformation of SIV protease reported here. The unliganded HIV-1 protease structures are intermediate between open and closed as judged by the rotation about the axis between the core and terminal domains (7° in SIV protease and 4° in HIV-1 protease) and the order of the flaps. These structures suggest a model in which many conformations are accessible to the protease in solution, all in equilibrium with the one closed conformation. Each crystal conformation is just a snapshot of the ensemble of

conformations available to the enzyme. The binding energy of the substrate must overcome an entropic barrier by stabilizing the protease in its closed conformation, in addition to a possible enthalpic barrier.

The two catalytic aspartic acids, one from each core domain, are too far apart in the unliganded SIV protease structure to share a hydrogen bond. They are either protonated or hydrogen bonded to the backbone nitrogens of Gly 27 and Gly 27' [40]. The pH of the solution may influence the equilibrium between the closed and open conformation of the dimer. Above pH 6.2, both aspartic acids will be unprotonated and therefore repel each other. This may explain why the "open" conformation was selected for at a higher pH in the unliganded SIV protease structure.

Tight-binding inhibitors modeled on peptides select for the closed conformation of the protease. This is demonstrated by the 19 liganded HIV-1 protease structures. Not all liganded protease crystal structures adopt the closed conformation, for example the Gln 7 Lys mutation of HIV-1 protease co-crystallized with the haloperidol-based inhibitor UCSF8 crystallized in a conformation resembling the open protease structure [31]. The two core domains were rotated by 3.8° and 3.3° relative to the terminal domain, slightly less than the 4° rotation of the open conformation. Inhibitor binding may not select for the closed conformation if the inhibitor is disordered when it binds or binds in several orientations, as with UCSF8, or if the inhibitor binds better to a more open conformation. The rotation of the core domain relative to the terminal domain of SIV protease bound to curcumin showed the protease to be completely open, supporting the evidence that there was no curcumin bound to the active site. This is not surprising since the crystals formed at curcumin concentrations well below the IC_{50}. In contrast, the structure of SIV protease co-crystallized with biliverdin was in the closed

conformation. This was unexpected because the crystals grew in the space group $P3_2$. The other SIV protease crystallized in this space group were open. The density in the active site was too diffuse to distinguish a ligand, or ligands.

Comparable domain rotations of monomeric aspartyl proteases:

The domains of the dimeric retroviral proteases are analogous to those of the monomeric proteases. The two structurally similar N- and C-terminal lobes of pepsin-like proteases, presumed to have arisen by gene duplication from an ancestral dimeric protease [41], correspond to the core domains of the two retroviral protease monomers. A third "central" domain at the interface between the N- and C-terminal lobes is composed of six beta strands and two helices [16, 42]. This corresponds to the four beta-stranded terminal domain identified in the retroviral proteases. Because the domain structure has been conserved evolutionarily, it is presumably important for the catalytic mechanism.

The rigid body of the two lobes of the monomeric aspartyl proteases is similar to the rigid body rotations of the two core domains of the dimeric proteases [16]. Both core domains of the symmetrical dimeric proteases move relative to the terminal beta sheet domain, whereas the central beta sheet domain of the monomeric aspartyl proteases is associated with the N-terminal lobe. In both cases, the catalytic aspartic acids are located near the rotation axes. In both cases the scissile bond is positioned between two domains with the P1-P4 residues associated primarily with one domain (hydrogen bonded to the C-terminal domain in pepsin and one core domain in HIV protease), and the P1′-P4′ residues associated primarily with the other

domain (the N-terminal domain in pepsin and the other core domain in HIV protease).

The interface between the core and terminal domains of HIV and SIV proteases is formed by hydrophobic sidechains, with one hydrogen bond between a backbone carbonyl and a flexible lysine sidechain. The analogous interface in the monomeric proteases consists of a helix from the central domain and a 3_{10} helix from the C-terminal lobe. Sali, et al characterized the rearrangements at this interface that accompany the rigid body motion in terms of the "helix interface sheer mechanism" [16, 43] in which large domain shifts, of several angstroms, are accommodated by small torsion angle shifts of sidechains at a helix-helix interface. The sidechain shifts are small enough not to disrupt the favorable packing interactions at the interface. While the interface between the core and terminal domains of the retroviral proteins are not helices packed together, a similar description applies: small shifts in the hydrophobic packing accommodate the reorientation of the domains.

Covariant Alignment of Substrate and Enzyme

The induced fit mechanism as proposed by Koshland [23] hypothesized that substrate binding caused a conformational change in the enzyme, properly orienting residues from the enzyme for catalysis. While proper oreintation of the catalytic residues is essential for catalysis, the aspartyl protease mechanism is better described as covariant alignment of the substrate and protease. In this model, substrate binding does not cause the conformational changes of the enzyme. Instead the binding energy is utilized to select one of the many accessible conformations of both substrate and enzyme that will be catalytically active. The unliganded SIV protease

structure indicates some of the degrees of freedom available to the protease.

While the open conformation facilitates binding of the polyprotein substrate, the closed conformation facilitates catalysis. In the closed conformation, the flaps hydrogen bond to the substrate and form subsites for the substrate sidechains. The catalytic aspartic acid residues, one from each core domain, can rotate together, which is necessary for them to act as the general acid and general base during catalysis [44, 45]. Rotation of the core domains positions them to hydrogen bond with the substrate in the orientation necessary for catalysis. It also shapes the P1, P1', P3, and P3' binding pockets. These conformational changes in the protein also selects a conformation of the substrate which is suitable for cleavage.

The importance of properly aligning the substrate and enzyme for catalysis contrasts with the hypothesis that the binding energy from the substrate primarily is used to strain the scissile peptide bond towards the tetrahedral transition state [18, 44, 46-48]. Modeling the substrate from products indicates that substrate strain does not contribute to catalysis. Instead, a good substrate stabilizes the closed conformation of the protease, which in turn orients the scissile bond of the substrate properly for catalysis.

The rigid body rotation may explain the inefficiency with which aspartyl proteases cleave short substrates. The monomeric protease pepsin will cleave a dipeptide, though with a much reduced k_{cat} [18]. Short peptides are also not good substrates for the retroviral proteases. These peptides might not have enough binding energy to overcome the entropic cost of stabilizing the protease in the closed conformation, meaning the scissile bond would be rarely aligned for catalysis.

The conformational change may also assist in product release. Substrate binding stabilizes the closed conformation partly by bridging

between the core domains of the two monomers. Upon bond cleavage the products are each associated with one of the monomers, allowing the core domain of each monomer to rotate apart and discouraging the reverse reaction.

The Role of Rigid Body Rotations in Developing Drug Resistance

The rapid rate of mutation of the virus results in resistance to all drugs administered so far. There are at least 23 residues known to mutate during treatment with the protease inhibitors currently in clinical trials or approved for trials (Chong-Hwan Chang, personal communication). These mutated enzymes presumably decrease the binding of the inhibitor to the protease more than binding of the substrates. Some of the resistance mutations are in direct contact with the ligand, and some are far from the substrate binding pocket. Six are in the interface between the core and terminal domains (Val 10, Leu 24, Ile 71, Asn 88, Leu 90, Leu 97 in the SIV sequence), and 4 are in the interface between the core and flap domains (Ile 36, Ile 54, Glu 63, Ile 75 in the SIV sequence). Residues Pro 81, Ile 82, and Ile 84 (in the SIV sequence) are part of the loop forming the P1 pocket that is formed by both core domains. That the mutations appear in the interface between the domains suggests that conformational changes of the protease may be a factor in the development of resistance mutations to anti-protease drugs.

Fersht showed that conformational changes of the enzyme common to all bound ligands cannot enhance specificity since both substrates, or substrate and inhibitor, have to overcome the same energy required for the conformational change [49, 50]. The domain rotations described here are common to all ligands: substrates and inhibitors. Variations in the degree of rigid body rotation may contribute to substrate selectivity for the monomeric

aspartyl proteases [16, 19], but the closed conformations detected in the retroviral proteases were all identical.

In order to enhance specificity, a conformational change in an enzyme has to affect the binding constants of the two ligands differentially [50, 51]. Proteases incorporating resistance mutations in direct contact with ligands have been studied crystallographically [52-54]. Their decreased selectivity for inhibitor versus substrate has been explained by contacts that are more unfavorable to the inhibitor than to substrates, in the closed conformation.

Because the protease surrounds the ligands in the closed conformation, initial binding is to an open conformation. Therefore mutations in the protease affecting ligand binding to an open conformation may affect selectivity also. Selectivity cannot result from affecting something in common to binding of both substrate and inhibitor, like domain closing. But disrupting the equilibrium between the open and closed conformation can select against the inhibitor more than substrate, for instance if substrate binding is not completely an equilibrium process but has a kinetic component. If the open conformation is more accessible to the substrate, since it is bigger and more flexible than an inhibitor, and if once the substrate binds it does not dissociate in the time it takes for the enzyme to close, then substrate binding will exclude inhibitor binding. Stabilizing the open conformation will then affect the binding constant of the inhibitor to a greater extent than the substrate. The mutated protease may look identical to the wild type protease in the closed conformation. There may therefore be a role for less rigid inhibitors, or an inhibitor targeting the open conformation of HIV protease, to compliment the existing drugs which target the closed conformation. Such a combination of protease inhibitors may overcome the rapid mutation rate of the virus.

ACKNOWLEDGEMENTS

We thank Nancy Douglas for purifying protein, and Nick Endres for setting up crystals. We thank Earl Rutenber and Janet Finer-Moore for editorial comments. This work supported by NIH Grant GM 39552, NIH Training Grant CA-09215, and the Biotechnology Resources and Education Program Training Grant.

REFERENCES

1. Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. Science 246:1149-1152, 1989.
2. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J.Kent, S.B.H. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 245:616-621, 1989.
3. Navia, M.A., Fitzgerald, P.M., McKeever, B.M., Leu, C.T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L.Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337:615-20, 1989.
4. Spinelli, S., Liu, Q.Z., Alzari, P.M., Hirel, P.H.Poljak, R.J. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie 73:1391-6, 1991.
5. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danley, D.E., Geoghegan, K.F.et al. X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. Nature 342:299-302, 1989.
6. Jaskólski, M., Miller, M., Mohana Rao, J.K., Leis, J.Wlodawer, A. Structure of the aspartic protease from Rous Scarcoma retrovirus refined at 2-Å resolution. Biochemistry 29:5889-5898, 1990.
7. Nicholson, L.K., Yamazaki, T., Torchia, D.A., Grzesiek, S., Bax, A., Stahl, S.J., Daugman, J.D., Wingfield, P.T., Lam, P.Y.S., Jadhav, P.K., Hodge, C.N., Domaille, P.J.Chang, C. Flexibility and function in HIV-1 protease. Struct. Biol. 2:274-280, 1995.

8. Rodriguez, E.J., Debouck, C., Deckman, I.C., Abu-Soud, H., Raushel, F.M.Meek, T.D. Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry. Biochemistry 32:3557-3563, 1993.
9. Collins, J.R., Burt, S.K.Erickson, J.W. Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics. Nat. Struct. Biol. 2:334-8, 1995.
10. Harte, W.E., Jr., Swaminathan, S., Mansuri, M.M., Martin, J.C., Rosenberg, I.E.Beveridge, D.L. Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. Proc. Natl. Acad. Sci. U.S.A. 87:8864-8868, 1990.
11. Harte, W.E., Jr., Swaminathan, S.Beveridge, D.L. Molecular dynamics of HIV-1 protease. Proteins 13:175-94, 1992.
12. Lin, Y., Fusek, M., Lin, X., Hartsuck, J.A., Kezdy, F.J.Tang, J. pH dependence of kinetic parameters of pepsin, rhizopuspepsin, and their active-site hydrogen bond mutants. J. Biol. Chem. 267:18413-18418, 1992.
13. Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J.Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature 337:576-579, 1989.
14. Pearl, L.H., Taylor, W.R. A structural model for the retroviral proteases. Nature 329:351-354, 1987.
15. James, M.N.G., Sielecki, A., Salituro, F., Rich, D.Hofmann, T. Conformational flexibility in the active sites of aspartyl proteinases revealed by a pepstatin fragment binding to penicillopepsin. Proc. Natl. Acad. Sci. U.S.A. 79:6137-6141, 1982.
16. Sali, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T.Blundell, T.L. Domain flexibility in aspartic proteinases. Proteins 12:158-70, 1992.

17. Fruton, J.S. Fluorescence studies on the active sites of proteinases. Mol. Cell. Biochem. 32:105-114, 1980.
18. Fruton, J.S. The mechanism of the catalytic action of pepsin and related acid proteinases. Methods Enzymol. 44:1-35, 1976.
19. Abad-Zapatero, C., Rydel, T.J.Erickson, J. Revised 2.3 Å structure of procine pepsin: evidence for a flexible subdomain. Proteins 8:62-81, 1990.
20. Bayer, E.A., Ben-Hur, H.Wilchek, M. Isolation and properties of streptavidin. Methods Enzymol. 184:80-89, 1990.
21. Rose, R.B., Rosé, J.R., Salto, R., Craik, C.S.Stroud, R.M. Structure of the Protease from Simian Immunodeficiency Virus: Complex with an Irreversible Nonpeptide Inhibitor. Biochemistry 32:12498-12507, 1993.
22. Wilderspin, A.F. Sugrue, R.J. Alternative native flap conformation revealed by 2.3 A resolution structure of SIV proteinase. J. Mol. Biol. 239:97-103, 1994.
23. Koshland, D.E.J. Application of a theory of enzyme specificity to protein synthesis. Proc. Nat. Acad. Sci. U.S.A. 44:98-104, 1958.
24. Otwinowski, Z. *Oscillation data reduction program.* in *CCP4 Study weekend: Data collection and processing, 29-30 Jan 1993.* 1993. SERC Daresbury Laboratory, England.
25. Brünger, A.T., Kuriyan, J.Karplus, M. Science 235:458-460, 1987.
26. Sack, J. J. Mol. Graphics 6:225, 1988.
27. Sui, Z., Salto, R., Li, J., Craik, C.Ortiz de Montellano, P.R. Inhibition of the HIV-1 and HIV-2 proteases by curcumin and curcumin boron complexes. Bioorg. Med. Chem. 1:415-22, 1993.
28. Mori, H., Otake, T., Morimoto, M., Ueba, N., Kunita, N., Nakagami, T., Yamasaki, N.Taji, S. In vitro anti-human immunodeficiency virus type 1 activity of biliverdin, a bile pigment. Jpn. J. Cancer Res. 82:755-757, 1991.

29. Ochoa, G.-C., *Structure-based inhibition of the Simian Immunodeficiency Virus protease*, in *Biochemistry*. 1994, University of California, San Francisco: San Francisco.

30. Zhao, B., Winborne, E., Minnich, M.D., Culp, J.S., Debouck, C.Abdel-Meguid, S.S. Three-dimensional structure of a simian immunodeficiency virus protease/inhibitor complex. Implications for the design of human immunodeficiency virus type 1 and 2 protease inhibitors. Biochemistry 32:13054-13060, 1993.

31. Rutenber, E., Fauman, E.B., Keenan, R.J., Fong, S., Furth, P.S., Ortiz de Montellano, P.R., Meng, E., Kuntz, I.D., DeCamp, D.L.Salto, R. Structure of a non-peptide inhibitor complexed with HIV-1 protease. J. Biol. Chem. 268:15343-15346, 1993.

32. Richards, F.M. Kundrot, C.E. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. Proteins 3:71-84, 1988.

33. Perry, K.M., Fauman, E.B., Finer-Moore, J.S., Montfort, W.R., Maley, G.F., Maley, F.Stroud, R.M. Plastic adaptation toward mutations in proteins: structural comparison of thymidylate synthases. Proteins 8:315-33, 1990.

34. Goldstein, H., *Classical Mechanics*. 1981, Menlo Park: Addison-Wesley Publishing Co.

35. Matthews, B.W. J. Mol. Biol. 33:491-497, 1968.

36. Brünger, A.T. The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355:472-474, 1992.

37. Stroud, R.M. Fauman, E.B. Significance of structural changes in proteins: expected errors in refined protein structures. Protein Sci. 4:2392-2404, 1995.

38. Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Schneider, J., Kent, S.B.Wlodawer, A. X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. Proc. Natl. Acad. Sci. U.S.A. 87:8805-9, 1990.
39. Schecter, I. Berger, A. Biochem. Biophys. Res. Commun. 27:157-162, 1967.
40. James, M.N., Sielecki, A.R., Hayakawa, K.Gelb, M.H. Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. Biochemistry 31:3872-3886, 1992.
41. Tang, J., James, M.N.G., Hsu, I.N., Jenkins, J.A.Blundell, T.L. Structural evidence for gene duplication in the evolution of the acid proteases. Nature 271:618-621, 1978.
42. Sielecki, A.R., Fedorov, A.A., Boodhoo, A., Andreeva, N.S.James, M.N.G. Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8Å resolution. J. Mol. Biol. 214:143-170, 1990.
43. Chothia, C. Lesk, A.M. Helix movements in proteins. Trends In Biol. Sci. 10:116-120, 1985.
44. Rose, R.B., Craik, C.S., Douglas, N.L.Stroud, R.M. Three-dimensional structures of HIV-1 and SIV protease product complexes. Biochemistry 1996.
45. Polgár, L. The mechanism of action of aspartic proteases involves 'push-pull' catalysis. FEBS Lett. 219:1-4, 1987.
46. Davies, D.R. The structure and function of the aspartic proteinases. Annu. Rev. Biophys. Biophys. Chem. 19:189-215, 1990.
47. Pearl, L. Blundell, T. The active site of aspartic proteinases. FEBS Lett. 174:96-101, 1984.

48. Jaskólski, M., Tomasselli, A.G., Sawyer, T.K., Staples, D.G., Heinrikson, R.L., Schneider, J., Kent, S.B.Wlodawer, A. Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. Biochemistry 30:1600-1609, 1991.

49. Fersht, A., *Enzyme structue and mechanism.* 1985, New York: W. H. Freeman and Co.

50. Herschlag, D. The role of induced fit and conformational changes of enzymes in specificity and catalysis. Bioorg. Chem. 16:62-96, 1988.

51. Ray, J., W. J., Post, C.B.Puvathingal, J.M. Reaction of the isosteric methylenephosphonate analog of alpha-D-glucose 1-phosphate with phosphoglucomutase. Induced-fit specificity revisited. Biochemistry 32:38-47, 1993.

52. Chen, Z., Li, Y., Schock, H.B., Hall, D., Chen, E.Kuo, L.C. Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials. J. Biol. Chem. 270:21433-21436, 1995.

53. Baldwin, E.T., Bhat, T.N., Liu, B., Pattabiraman, N.Erickson, J.W. Structural basis of drug resistance for the V82A mutant of HIV-1 proteinase. Nat. Struct. Biol. 2:244-249, 1995.

54. Ridky, T. Leis, J. Development of drug resistance to HIV-1 protease inhibitors. J. Biol. Chem. 270:29621-29623, 1995.

55. Wlodawer, A. Erickson, J.W. Structure-based inhibitors of HIV-1 protease. Annu. Rev. Biochem. 62:543-85, 1993.

56. Dreyer, G.B., Lambert, D.M., Meek, T.D., Carr, T.J., Tomaszek, T.A., Jr., Fernandez, A.V., Bartus, H., Cacciavillani, E., Hassell, A.M., Minnich, M.et al. Hydroxyethylene isostere inhibitors of human immunodeficiency

virus-1 protease: structure-activity analysis using enzyme kinetics, X-ray crystallography, and infected T-cell assays. Biochemistry 31:6646-59, 1992.

57. Hoog, S.S., Zhao, B., Winborne, E., Fisher, S., Green, D.W., DesJarlais, R.L., Newlander, K.A., Callahan, J.F., Moore, M.L., Huffman, W.F.et al. A check on rational drug design: crystal structure of a complex of human immunodeficiency virus type 1 protease with a novel gamma-turn mimetic inhibitor. J Med Chem 38:3246-52, 1995.
58. Murthy, K.H., Winborne, E.L., Minnich, M.D., Culp, J.S.Debouck, C. The crystal structures at 2.2-A resolution of hydroxyethylene-based inhibitors bound to human immunodeficiency virus type 1 protease show that the inhibitors are present in two distinct orientations. J. Biol. Chem. 267:22770-22778, 1992.
59. Priestle, J.P., Fassler, A., Rosel, J., Tintelnot-Blomley, M., Strop, P.Grutter, M.G. Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 53820, a novel pseudosymmetric inhibitor. Structure 3:381-389, 1995.
60. Thanki, N., Rao, J.K., Foundling, S.I., Howe, W.J., Moon, J.B., Hui, J.O., Tomasselli, A.G., Heinrikson, R.L., Thaisrivongs, S.Wlodawer, A. Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: comparisons with molecular modeling. Protein Sci. 1:1061-1072, 1992.
61. Abdel-Meguid, S.S., Zhao, B., Murthy, K.H., Winborne, E., Choi, J.K., DesJarlais, R.L., Minnich, M.D., Culp, J.S., Debouck, C., Tomaszek, T.A., Jr.et al. Inhibition of human immunodeficiency virus-1 protease by a C2-symmetric phosphinate. Synthesis and crystallographic analysis. Biochemistry 32:7972-7980, 1993.

62. Thompson, S.K., Murthy, K.H., Zhao, B., Winborne, E., Green, D.W., Fisher, S.M., DesJarlais, R.L., Tomaszek, T.A., Jr., Meek, T.D., Gleason, J.G.et al. Rational design, synthesis, and crystallographic analysis of a hydroxyethylene-based HIV-1 protease inhibitor containing a heterocyclic P1'--P2' amide bond isostere. J. Med. Chem. 37:3100-3107, 1994.

63. Kim, E.E., Baker, C.T., Dwyer, M.D., Murcko, M.A.al., e. Crystal structure of hiv-1 protease in complex with vx-478, a potent and orally bioavailable inhibitor of the enzyme. J. Am. Chem. Soc. 117:1181-1182, 1995.

64. Jhoti, H., Singh, O.M., Weir, M.P., Cooke, R., Murray-Rust, P.Wonacott, A. X-ray crystallographic studies of a series of penicillin-derived asymmetric inhibitors of HIV-1 protease. Biochemistry 33:8417-8427, 1994.

65. Hosur, M.V., Bhat, T.N., Kempf, D.J., Baldwin, E.T.al., e. Influence of stereochemistry on activity and binding modes for C(2) symmetry-based diol inhibitors of hiv-1 protease. J. Am. Chem. Soc. 116: 1994.

66. Lam, P.Y., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., Chang, C., Weber, P.C., Jackson, D.A., Sharpe, T.R.Erickson-Viitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science 263:380-384, 1994.

67. Abdel-Meguid, S.S., Metcalf, B.W., Carr, T.J., Demarsh, P., DesJarlais, R.L., Fisher, S., Green, D.W., Ivanoff, L., Lambert, D.M., Murthy, K.H.et al. An orally bioavailable HIV-1 protease inhibitor containing an imidazole-derived peptide bond replacement: crystallographic and pharmacokinetic analysis. Biochemistry 33:11671-11677, 1994.

68. Bone, R., Vacca, J.P., Anderson, P.S.Holloway, M.K. X-ray crystal structure of the hiv protease complex with l-700,417, an inhibitor with pseudo C2 symmetry. J. Am. Chem. Soc. 113:9382-9384, 1991.

69. Fitzgerald, P.M., McKeever, B.M., VanMiddlesworth, J.F., Springer, J.P., Heimbach, J.C., Leu, C.T., Herber, W.K., Dixon, R.A.Darke, P.L. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution. J Biol Chem 265:14209-14219, 1990.

70. Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N.et al. Design, activity, and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. Science 249:527-533, 1990.

FIGURE LEGENDS

Figure 1: Ribbon diagram of SIV protease with the flaps in the "closed" conformation. The rigid body domains are distinguished by differed shading: the terminal domain (residues 1-9 and 86-99 or both monomers), colored light gray; the core domain (residues 10-32 and 63-85 or a single monomer), colored black; and the flaps (residues 33 to 62 of a single monomer), colored medium gray. The beta strands and alpha helix of the left monomer of the dimer are labeled as in the pepsin-like proteases [55]. Residues along the monomer on the right are labeled, as well as the catalytic aspartic acids a several features of the protease.

Figure 2: Carbon alpha traces of the unliganded (bold lines) and peptide product (F-L-E-K) liganded SIV protease structures after superposition of the carbon alpha atoms. a) dimer superposition: residues 1-45 and 55-99 of both monomers of the dimer. b) monomer superposition: residues 1-45 and 55-99 of one monomer. c) terminal domain superpostion: residues 1-9 and 86-99 of both monomers of the dimer. d) core domain superpostion: residues 10-32 and 63-85 of one monomer. The flap residues 48-53 of the unliganded structure are not visible in the density.

Figure 3: Residues at the interface between the terminal and core domains. a) the helix of the terminal domain (on the right) is colored medium gray in 2|Fo|-|Fc| density. The b′ and c′ beta-strands of the core domain, colored black, is on the left. 2|Fo|-|Fc| density is shown for residues 69 to 73. The rest of the beta strand, residues 64 to 68, are colored light gray. b) Reorientation of the residues at the interface as a result of the rigid body rotation. The terminal domains of the unliganded SIV protease, colored gray, and the protease bound to the product peptide F-L-E-K, colored black, were superimposed.

Figure 4: Carbon alpha traces of 17 HIV-1 protease structures after superimposition of the carbon alpha atoms onto 7hvp.

Figure 5: Axes determined for the rotation of the core and flap domains. a) Carbon alpha trace of the liganded SIV protease structure (1siv) viewed down the dimer 2-fold axis towards the flaps. The two axes about which the core domains rotate relative to the terminal domain are marked by dotted lines. X marks the position of the catalytic aspartic acids. b) Carbon alpha trace of the liganded HIV-1 protease structure (7hvp) viewed down the dimer 2-fold axis towards the flaps. The axes about which the core domains rotate are marked by dotted lines. X marks position of the catalytic aspartic acids. c) Carbon alpha trace of the unliganded SIV protease structure with the shoulder domain of the liganded structure (1siv) superimposed. The axis about which the shoulder domain rotates with respect to the core domain is marked by a dotted line.

Figure 6: Ball-and-stick representation of the binding pocket of the unliganded (in gray) and liganded (1siv, in black) SIV protease structures with one core domain superimposed. The ligand (in white) is displayed in the active site, within hydrogen bonding distance of the superimposed monomer of the unliganded structure but too far to hydrogen bond to the unsuperimposed monomer. The carbon alpha atoms of the primed core domains were superimposed. The orientation of Asp 25 of the unliganded structure is seen to be rotated with respect to Asp 25 of the liganded structure.

Figure 7: Comparison of the average deviation of the carbon alpha atoms from 18 superimposed HIV-1 protease structures (continuous line) with the average B-factors for the same atoms (dotted line). The B-factors have been scaled by a factor of 0.2 to plot on the same graph. The deviation of the coordinates follow the B-factors.

Table I: Data Collection and Refinement Statistics

highest resolution (Å)	2.0
no. of unique reflections[a]	7,300
redundancy	5-fold
completeness of 2.06-2.00Å resolution shell (%)	94
R_{symm}[b] (%)	8.6
final R-factor (7-2Å) (%)	20.4
final R-free (7-2Å)(%)	26.6
Average B-factor (Å^2)	20.1
rms deviation of bond lengths (Å)	0.014
rms deviation of bond angles (°)	3.0
rms deviation of dihedral angles (°)	27.6

[a]no sigma cutoff applied

[b]$R_{symm}=\Sigma_i\Sigma_j \mid <I_i>-I_{ij} \mid / \Sigma_i\Sigma_j I_{ij}$

Table II: HIV-1 Protease Structures From Brookhaven Protein Data Bank

Brookhaven code	ligand	space group	[a]rmsd. (Å) from 7hvp	references
3hvp	none	$P4_12_12$		[2]
3phv	none	$P4_12_12$		[5]
1hhp	none	$P4_12_12$		[4]
	[b]UCSF8	$P2_12_12_1$	0.58	[31]
1aaq	peptidomimetic	$P6_1$	0.70	[56]
1hbv	SB203238	$P6_1$	0.59	[57]
1hef	SKF 108738	$P6_122$	0.55	[58]
1hih	CGP 53820	$P2_12_12_1$	0.62	[59]
1hiv	U75875	$P2_12_12_1$	0.54	[60]
1hos	SB204144	$P6_1$	0.57	[61]
1hps	SB206343	$P6_1$	0.60	[62]
1hpv	VX-478	$P6_1$	0.61	[63]
1htg	GR137615	$P2_12_12_1$	0.64	[64]
1hvi	A77003	$P2_12_12_1$	0.53	[65]
1hvr	XK263	$P6_1$	0.59	[66]
1sbg	SB203386	$P6_1$	0.55	[67]
4hvp	MVT-101	$P2_12_12_1$	0.37	[1]
4phv	L-700,417	$P2_12_12_1$	0.63	[68]
5hvp	acetyl-pepstatin	$P2_12_12_1$	0.69	[69]
7hvp	JG-365	$P2_12_12_1$	0.0	[38]
8hvp	U-85548E	$P2_12_12_1$	0.66	[48]
9hvp	A-74704	$P6_1$	0.64	[70]

[a]rmsd is calculated by a least square superposition of the carbon alpha atoms of the structure onto 7hvp and calculating the root mean square deviation of the carbon alpha atoms from 7hvp.
[b]crystallized with a point mutant of HIV-1 protease, Gln 7 Lys

Table III: Root mean square deviations of superpositions of liganded and unliganded SIV and HIV-1 protease structures.

	[b]1siv	[c]FLEK	[d]7hvp	[e]SLNF
protease	SIV	SIV	HIV-1	HIV-1
apo structure	this paper	this paper	[f]3hvp	[f]3hvp
[a]rms dimer	1.6	1.5	1.2	1.2
[a]rms monomers	0.8	0.7	0.8	0.7
[a]rms terminal domain	0.4	0.4	0.4	0.3
[a]rms core domain	0.4	0.4	[g]0.5 or 0.6	[g]0.5 or 0.6
[a]rms shoulder domain	0.4	0.4	0.6	0.5

[a]Root-mean-square deviations after superposition of carbon alpha atoms of the specified subsets of the liganded and respective unliganded structures.
[b]SIV protease bound to a peptidomimetic inhibitor, referred to by the Brookhaven Protein Data Bank code (ref)
[c]SIV protease bound to peptide product of sequence F-L-E-K (ref)
[d]HIV-1 protease bound to a peptidomimetic inhibitor, referred to by the Brookhaven Protein Data Bank code (ref)
[e]HIV-1 protease bound to peptide product of sequence S-L-N-F (ref)
[f]Unliganded HIV-1 protease structure, referred to by the Brookhaven Protein Data Bank code (ref)
[g]Residues 77-81 are excluded from this calculation since their position varied significantly from the rest of the domain. The two results listed are for the two monomers. The monomers of the dimer are not symmetrical in this structure.

Table IV: Distance between atoms of the two catalytic aspartic acids of the dimer

SIV protease:

atoms	unliganded	1siv	FLEK
C_α–C_α	6.8 Å	6.6 Å	6.6 Å
C_β–C_β	7.7 Å	7.2 Å	7.1 Å
C_γ–C_γ	5.6 Å	5.0 Å	4.9 Å
$O_{\delta 1}$–$O_{\delta 1}$	3.2 Å	2.8 Å	3.0 Å

HIV-1 protease:

atoms	unliganded	7hvp	SLNF
C_α–C_α	6.7 Å	6.4 Å	6.4
C_β–C_β	7.6 Å	7.0 Å	7.2
C_γ–C_γ	5.3 Å	5.0 Å	5.0
$O_{\delta 1}$–$O_{\delta 1}$	3.0 Å	2.8 Å	2.9

Table V: Movements of atoms forming the S1, S1' and S3 subsites between the "open" to the "closed" conformation of SIV protease

Subsite	Residue	[a]C_α-C_α distance (Å)
S1'	Leu 23	1.6
S1'	Ile 81	3.1
S1'	Pro 82	2.5
S1'	Ile 84	3.9
S1	Gly 27	1.6
[b]S3	Arg 8'	0.8

[a]The carbon alpha atoms of the core domains (primed monomer) of the liganded and unliganded (1siv) SIV protease structures were superimposed, and the distances between the residues lining the S1', S1 and S3 subsites from the unsuperimposed (unprimed) monomers were measured.
[b]The Arg from the terminal domain of the primed monomer reaches into the S3 subsite.

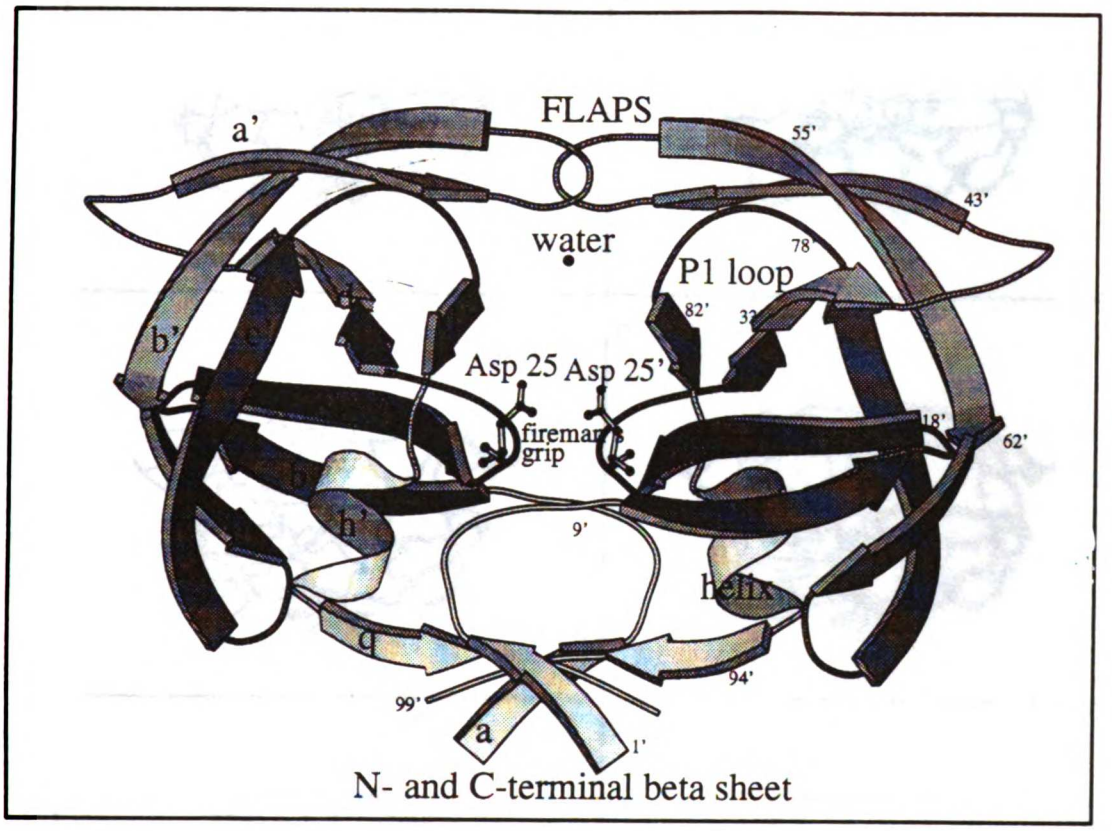

FLAPS
a'
55'
43'
water
78'
P1 loop
82'
b'
Asp 25
Asp 25'
fireman's
grip
18'
62'
h'
9'
helix
94'
99'
a
1'
N- and C-terminal beta sheet

Figure 1

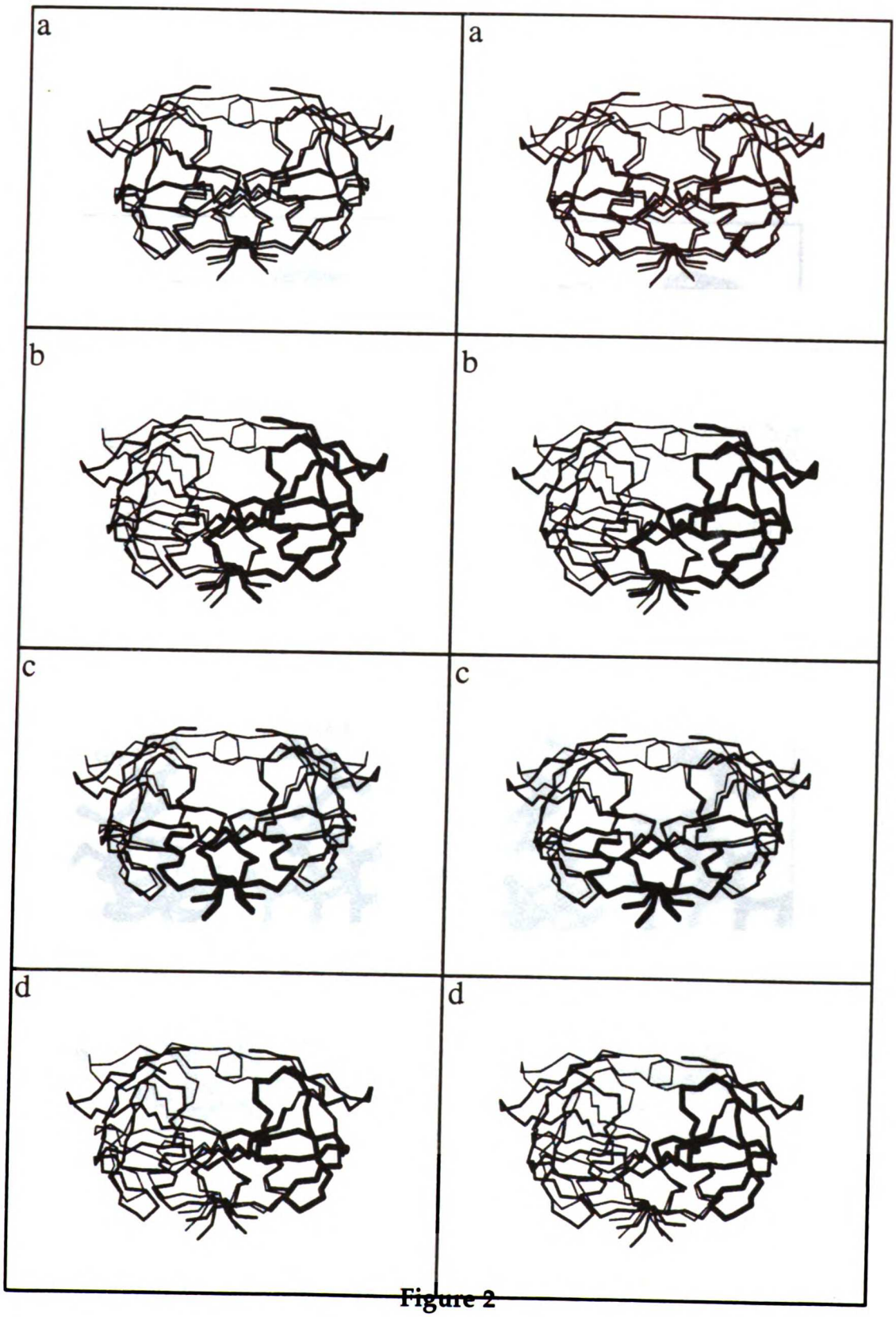


Figure 2

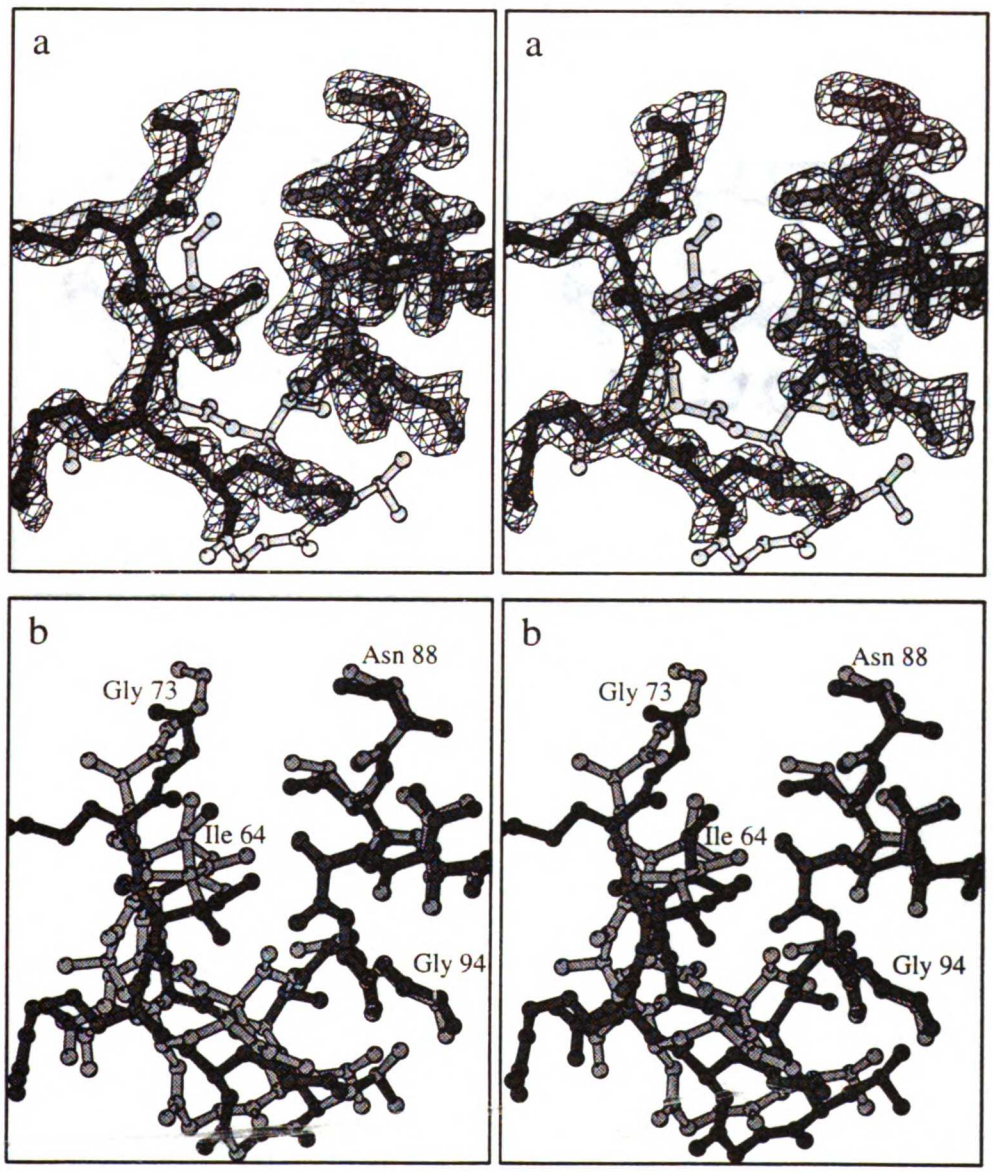


Figure 3

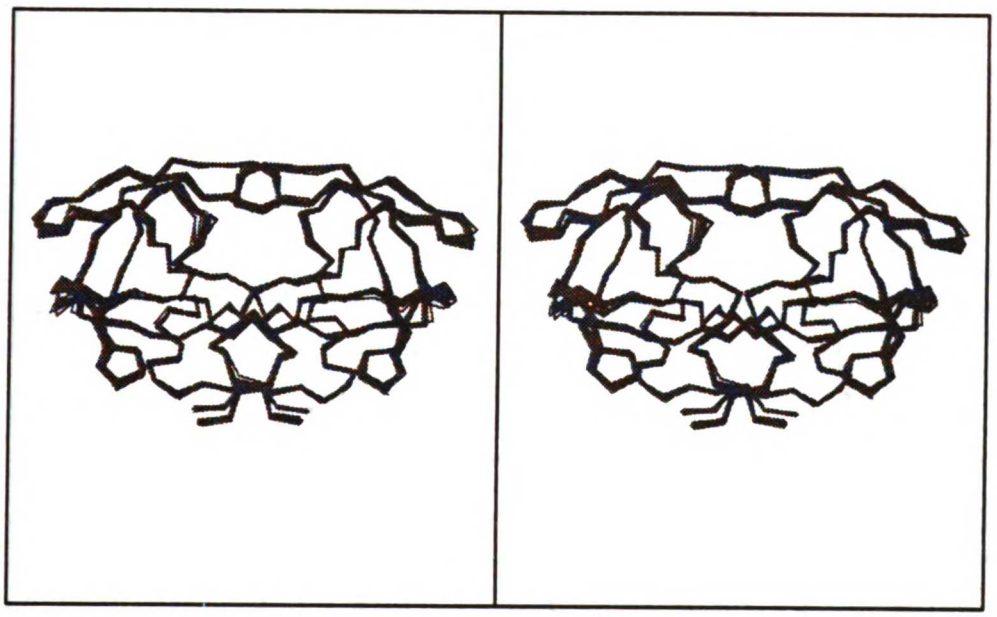


Figure 4

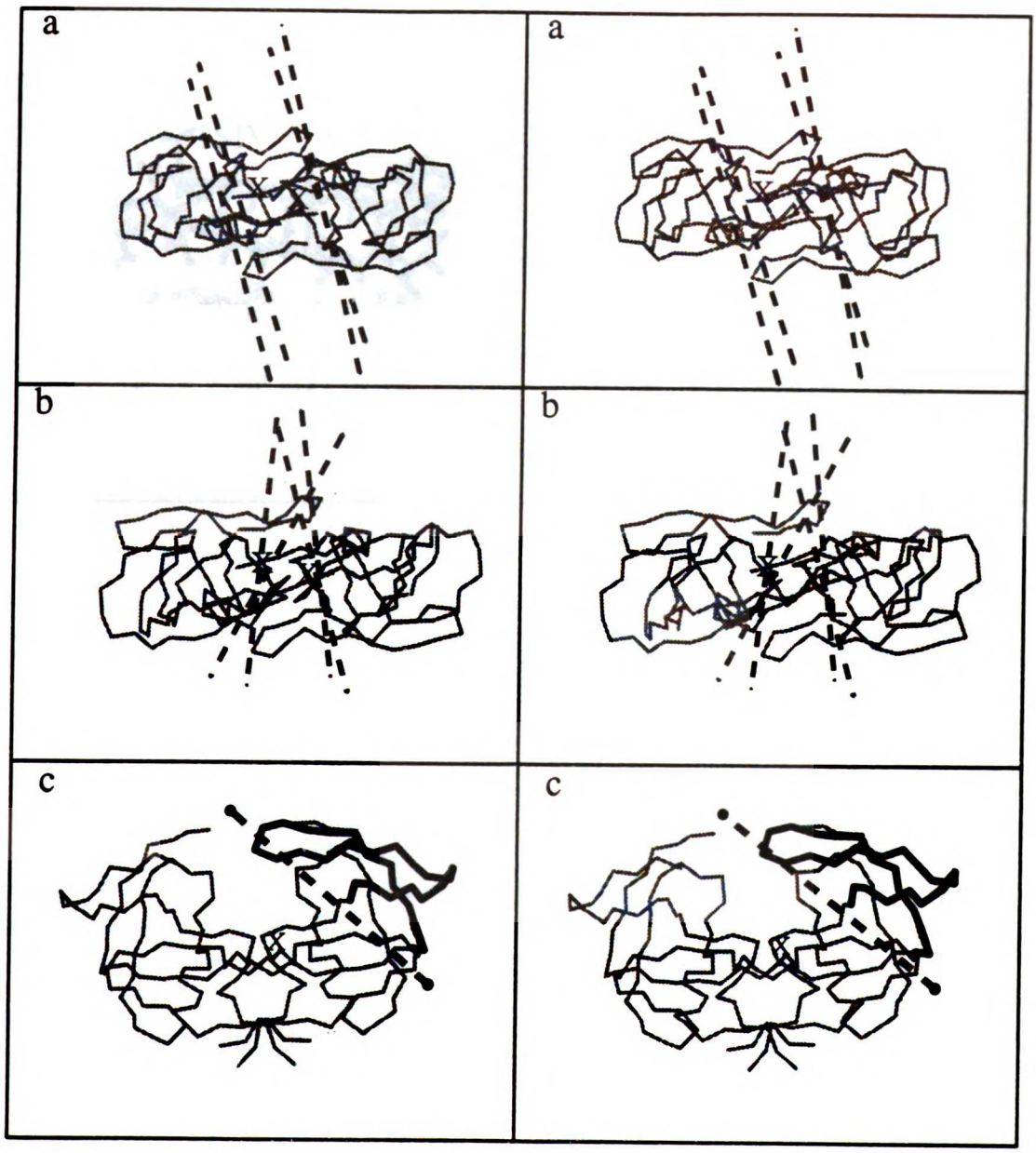


Figure 5

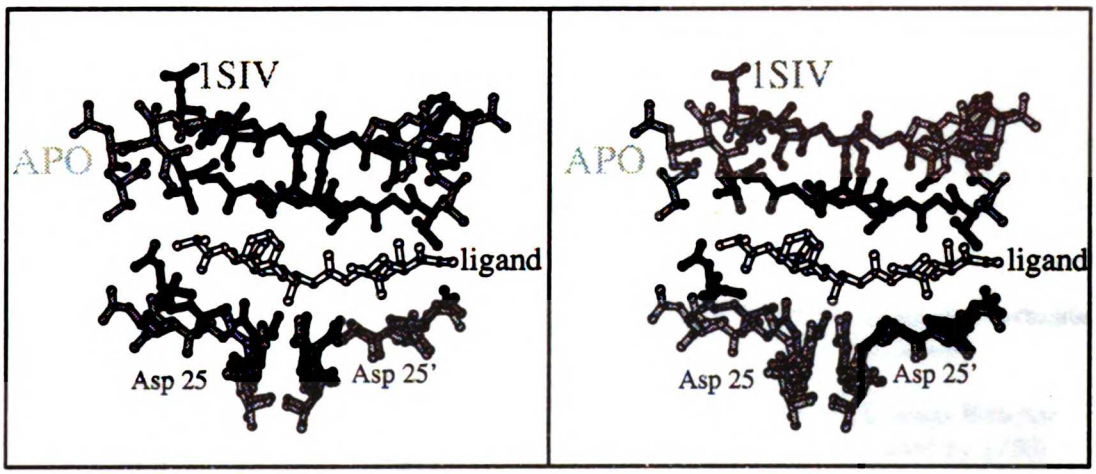


Figure 6

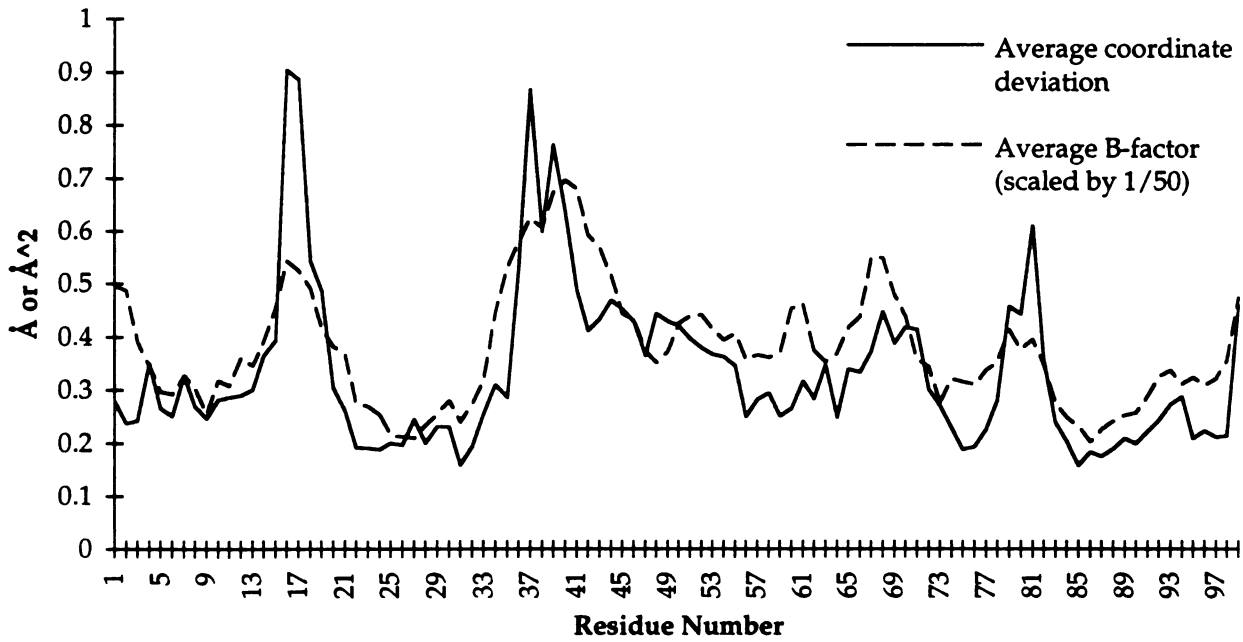

Average coordinate deviation
Average B-factor (scaled by 1/50)
Å or Å^2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
77
81
85
89
93
97
Residue Number

Figure 7

CHAPTER FIVE:

HIV INTEGRASE:
SOLUBILITY, CRYSTALLIZATION TRIALS, AND DNA BINDING

Introduction:

Two unique retroviral enzymes allow the virus's RNA to be incorporated into the host cell's genome: reverse transcriptase and integrase. Both are associated in a nucleo-protein complex when the virus infects the cell (Farnet & Haseltine, 1991). The association with other proteins may explain the insolubility of these enzymes in isolation. While the reverse transcriptase synthesizes the DNA copy of the viral genome, the integrase splices the DNA copy into the chromosome. Reverse transcriptase adds the Long Terminal Repeat (LTR) sequences at the termini of the DNA copy of the viral genome (Goff, 1990). In HIV, the 3' and 5' sequences are distinct (referred to as U3 and U5). These sequences are necessary for proper integration of the viral genome (Craigie et al., 1991).

An *in vitro* assay reproduces the main functions of integrase: cleavage of two basis on the 3' strand of the LTR sequences and ligation of this sequence non-specifically into a second oligonucleotide (Craigie et al., 1990). The site of integration is not random but the mechanism for discriminating integration sites is not understood. DNA binding assays for integrase have not demonstrated sequence-specific binding to the LTR's. Therefore it is not known whether the LTR and the chromosomal binding sites of integrase are identical or whether the protein binds as a monomer, dimer, or even tetramer.

Purification:

The major obstacle we confronted with purification was insolubility of integrase. The final purification scheme is describe in Leavitt, *et al* (Leavitt et al., 1992) (see appendix). The yeast expression system worked well and yielded on the order of 20mg of purified protein per liter of cells. The cells were lysed

in a low salt buffer and spun at 12,000 x g. The integrase associated with the pellet. Andrew Leavett tested many conditions to solubilize the integrase from the pellet, including many detergents, and found the sterol-based detergent CHAPS to be most effective (personal communication). The remainder of the purification was done in the presence of 7mM CHAPS. Integrase was purified to greater than 95% purity, as measured by a coomasie-stained SDS gel. Two columns were used routinely: an Aphigel-Heparin ion exchange column followed by a phenyl-Sepharose column. A size exclusion step was added to increase the purity. Minor contaminating bands were still visible on the SDS gel. In anticipation of further purification steps, it was found that integrase bound to phosphocellulose, hydroxyapitite, and CM-50 matrices. Integrase did not bind to a CM Memsep column at 50mM (low) NaCl.

Association of integrase with the pellet after the cells were lysed may be a result of association with the lipid membrane or the DNA. If the pellet is solubilized by 1M NaCl and 8mM LDAO, a fraction of the integrase is found in the supernatant. DNA is also present in the supernatant, as visualized by ethidium bromide staining of an agarose gel. DNA was not present in the supernatant when CHAPS was used to solubilize the integrase. Integrase may be soluble in the LDAO condition because it is associated with DNA. Solubilization with CHAPS required the detergent concentration go be greater than its critical micellar concentration (cmc) (the cmc for CHAPS is approximately 7mM), suggesting that the integrase was solubilized in CHAPS micelles. Because low salt conditions were found to precipitate the protein, the cells were lysed in a high salt, 1M NaCl, buffer to determine if the low salt condition destabilizes the protein. The resulting protein was no more soluble than from the low salt lysis protocol.

The purity of the protein after a size exclusion step was tested by amino acid analysis. Five amino acids (Val, Met, Ile, Leu, Phe) deviated by more than 5% from the expected values (as much as 35%), suggesting significant quantities of contaminating proteins. An amino acid analysis of the protein after an additional CM-50 purification step showed even greater deviations from the expected values. It is possible that the discrepancy resulted from precipitation of integrase in the sample so that the contaminating proteins were preferentially measured in this assay. The integrase samples were sent by mail to Stanford for amino acid analysis.

The purity of the protein was also tested by mass spectrometry. The protein, after size exclusion chromatography, was TCA precipitated. The pellet was washed 3 times with water to extract the CHAPS and solubilized in 100% formic acid. This wash could have extracted contaminating soluble proteins. The result of electrospray mass spectrometry was a very clean spectrum with a mass of 32,228$\pm$6. The calculated molecular weight was 32,211.

The protein eluted from the phenyl-Sepharose column at up to 9mg/ml at the peak. The protein could then be frozen without loss of activity in the presence of 10% glycerol. Some protein loss resulted from freezing: the protein concentration was always at least 10% less than the pre-frozen concentration.

Protein was typically concentrated in an Amicon stirred cell or a Centricon for crystallization. Many of the solubility experiments described below, and all of the crystal trials, utilized protein concentrated in this manner. Only towards the end of these experiments did we realize that CHAPS was being concentrated along with the protein. To measure the CHAPS concentration, concentrated integrase samples were blotted onto a

TLC (thin layer chromatography) plate. The solvent system used was 95% methanol, 5% ammonium hydroxide. Upon staining with iodine, the CHAPS concentrations were determined to be as high as 30mM. This may explain some of the inconsistencies observed while testing solubility. In addition, a faster migrating species appeared in each integrase lane. This could have been an integrase-CHAPS complex. When glycerol was present, this species migrated slightly more slowly.

Integrase ran on an SDS PAGE gel close to a 31,000 molecular weight protein marker. It was found that not boiling the samples before loading them onto the gel minimized smearing of the integrase band. DTT also helped minimize smearing. A second band sometimes appeared at about the size of a dimer. On a Western blot, this band also reacted with a polyclonal antibody to integrase. This band was especially pronounced after the protein had precipitated. It is not known if this "dimerization" is related to solubility.

Solubility:

Integrase is soluble at 3-5 mgs/ml in greater than 7mM CHAPS, at 4° C, pH 7.5 and 1M NaCl for longer than one week. (We later realized that CHAPS was also concentrated in the Amicon concentrator, so the CHAPS concentration was much greater than 7mM in these trials). Crystal trials were initiated starting from these conditions. The fact that no membrane protein had been crystallized in the presence of CHAPS influenced much of our effort to find non-CHAPS conditions that would solubilize the protein. CHAPS concentrations at near the cmc interferes with the *in vitro* activity assay, also suggesting that CHAPS may not be a good condition for crystallization.

solubility methods:

Several approaches were used to test solubility conditions, including exchange on columns, dialysis, and dilution. Buffer exchanges were tested on the Heparin, phenyl-Sepharose and sizing columns. Exchange on the sizing column diluted the protein significantly, making it difficult to assess the solubility without further concentrating the sample. Detergent exchange on the phenyl-Sepharose column modified how well the protein bound to the column, reducing the resulting purity. For some conditions tested, the integrase did not elute from the phenyl or heparin columns, presumably precipitating on the column.

20ml dialysis buttons (Cambridge Repetition Engineering) were used to test many conditions with a small volume of protein. Initially the appearance of flocculant was used as a measure of solubility. This measure worked for some conditions, which produced obvious and rapid precipitation, but it was inadequate for other conditions. Many conditions which did not visibly

precipitate the protein, or caused small amounts of visible precipitation, surprisingly had very little soluble protein. When the samples were retrieved from the buttons and centrifuged, little protein was detected in the supernatant by Bradford or on an SDS PAGE gel. It appeared that the protein could precipitate or stick to the surface of the button or dialysis tubing without forming a visible floculant. The protein's ability to stick to the surfaces seemed to vary with the buffer condition. When some of these conditions were tested in larger quantities in dialysis bag, significant precipitation was visible. An additional difficulty with dialysis is that the rate of diffusion of CHAPS through the dialysis membrane is dependent on the condition. The cmc of CHAPS decreases as the salt concentration increases, stabilizing the micelles at lower CHAPS concentrations. If solubility of the protein is dependent on the presence of micelles, this will also maintain the solubility of integrase for a longer time.

A further complication emerged upon investigation of the button dialysis technique. When the integrase sample was dialyzed into the same buffer that the protein was stored in, the dialyzed sample precipitated more rapidly than a similar sample not dialyzed. This observation was repeatable. The presence of the dialysis tubing in the not-dialyzed sample did not promote precipitation. Either the presence of the button promoted precipitation or the integrase is associated with some unknown factor which is dialyzed away. This association would have to survive the purification protocol. This observation was not further resolved.

An additional technique for testing solubility was by dilution into various buffer conditions followed by light scattering using a wavelength scan on a spectrophotometer between 200 and 600nm. While this technique was

fast and used relatively small amounts of material, the dilution (at least 10-fold) meant that the protein was being tested at a low concentration.

solubility results and discussion:

Reducing the CHAPS concentration to less than 7mM precipitated the protein at high (greater than 1mg/ml) protein concentrations. Reducing the CHAPS concentration slowly, by changing the CHAPS concentration in the dialysate over time, did not enhance the protein's solubility. At 7mM CHAPS, integrase was soluble longer at 4° C than room temperature. This observation was consistent with precipitation resulting from hydrophobic aggregation. High ionic strength also enhanced solubility; low salt concentrations precipitated the protein immediately. There were indications that integrase precipitated more readily at pH 5 than at pH 7 or 9. 10% glycerol enhanced the solubility of the protein, though it could not substitute for CHAPS to maintain solubility. 10mM spermine could also not replace CHAPS.

Much effort was expended in trying to exchange the CHAPS detergent for b-octylglucoside (OG) because OG has been used to crystallize several membrane proteins (Michel, 1991). Integrase was soluble in 2% OG, but at concentrations much lower than 1mg/ml. Even at low concentrations, the protein would precipitate over time. A buffer exchange on a heparin column resulted in the protein not eluting from the column. Conditions tested concurrently with buffer exchanges into OG were pH (from 6-9), NaCl concentration (50-500mM), and temperature (4° versus room temperature).

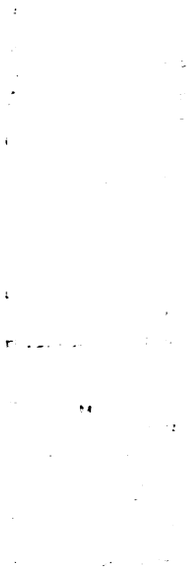
Exchange of the CHAPS for high salt concentrations, either $AmSO_4$, NaCl or KCl, caused immediate precipitation of the protein. $NaPO_4$ also did not improve the solubility. The CHAPS could also not be exchange for up to

20% ethanol. Additives such as higher DTT concentrations (to 20mM), cholate, ATP, $CaCl_2$, $MgCl_2$ or $LiCl_2$ did not improve solubility.

$MnCl_2$ is essential for the *in vitro* integration assay (Leavitt et al., 1992). 15mM $MnCl_2$, the concentration used in the activity assay, did not improve the protein's solubility in any of the conditions tested. Dialysis experiments with and without $MnCl_2$ indicated that the cation promoted precipitation. Dilutions experiments to 0.1mg/ml integrase, 200mM NaCl also showed light scattering in the presence of $MnCl_2$, but not in its absence.

Dialysis into 2M urea and high salt concentrations (1M NaCl or 0.8M $AmSO_4$), in 12,000-14,000 molecular weight cutoff dialysis membranes, did keep the protein soluble at 1-2mg/ml for 5 days. The CHAPS could be exchanged for 2M urea, 1M NaCl on a sizing column, but the resulting sample was very dilute. When the eluted protein was concentrated in a centricon to 0.2mg/ml, it showed no activity in an *in vitro* activity assay. 1M urea, 0.5M NaCl also maintained the solubility at these low concentrations of integrase.

Dilution experiments were used to exchange detergent environment quickly. Aliquots of 1mg/ml integrase were diluted 10-fold. The CHAPS concentrations in these samples are not known accurately (see below). Some of the variation with protein concentration may be a result of uncontrolled CHAPS concentrations. In any case, the 10-fold dilution assures the CHAPS concentration is below the cmc. (All of these experiments were performed at pH 7.5, 10mM Hepes buffer). At greater than 30mM NaCl, the sample did not scatter light. At 20mM NaCl, light scatter was observed. This suggests that detergent concentration and ionic strength contribute to solubility independently. If the protein concentration were increased to 0.4mg/ml, then greater than 400mM NaCl is required to eliminated scatter. Detergents were



added to the dilution buffer at conditions that otherwise resulted in scatter. In this way it was determined that 20mM cholic acid, 10mM deoxycholate, and 0.1% SDS also maintain the solubility of the protein at 0.4mg/ml integrase. These detergent concentrations are above their cmc. Detergents that did not solubilize the protein even at these low concentrations were: 50mM OG, 8mM LDAO, 18mM NG, 2.5% Brij, 0.2M Zwittergent 3-16, and 1mM DDM.

A mutant protein, W235E, was also tested for solubility. This mutation was generated to its affect on activity. The point mutation did not obviate the need for CHAPS to maintain solubility.

Analytical Ultracentrifugation:

Another assay considered to investigate conditions that would prevent aggregation was equilibrium analytical ultracentrifugation. This technique could also be used to determine the multimeric state of the protein. Samples were tested in 7mM CHAPS, 50mM NaCl, Hepes buffer, pH 7.5, at an integrase concentration of 1.5 mg/ml. When the radial distribution of the absorbance was fit to an exponential curve, the residuals were somewhat parabolic in behavior instead of random. Fitting the data to the sum of exponentials did not improve the residual fit. A single exponential was therefore used. Comparing the molecular weights determined in H_2O and D_2O, the partial specific volume was determined to be 0.75, approximately equal to the to the 0.73 for the average protein. With this value the protein molecular weight was determined to be 32,200, close to the monomer molecular weight. A third sample with an integrase concentration of 0.3mg/ml also yielded a molecular weight of 32,000. This molecular weight suggested that the protein was not associated with CHAPS molecules. The experiment was repeated with a different integrase sample, at a protein concentration of 0.3mg/ml, with

15mM $MnCl_2$ or 15mM $MgCl_2$. The radial distributions of the absorbance fit a single exponential with random residuals, but the molecular weight was determined to be 53,000 daltons even in the absence of divalent cation. We cannot explain the discrepancy in molecular weight, but sometimes dimers were seen in SDS PAGE gels as well. There was a slow loss of area over time with these samples, suggesting aggregation and pelleting of the protein.

Oligonucleotides:

We considered it possible that binding to DNA would increase the solubility of integrase. Because no binding assay for DNA was available, nothing was known about the conditions necessary for binding. The closest assay available was the *in vitro* activity assay, which measured binding plus activity. From this assay it was known that low CHAPS concentrations, low salt concentrations (less than 150mM NaCl), and $MnCl_2$ (15mM was used) were necessary for the reaction. These conditions were assumed to be necessary for DNA binding as well. Because only small amounts of DNA was available, the solubility of integrase in the presence of DNA was tested by mixing the integrase with oligonucleotides in dialysis buttons then dialyzing out the CHAPS and salt.

The first indication that DNA might enhance the solubility of integrase was using non-specific DNA. Solubility was tested with sonicated salmon sperm DNA: 25ml of 10mg/ml DNA was added to 25ml of 3mg/ml of integrase. Assuming integrase occupies 10 bases when bound to DNA, the DNA binding site to integrase ratio was about 10:1. 2000 molecular weight cutoff dialysis membranes were used to prevent diffusion of the DNA. The salmon sperm DNA maintained the solubility of about half of the protein for 4 days, even at low salt concentrations. The same conditions without DNA

precipitated the integrase after 1.5 hours of dialysis. A specific (U5 of the LTR) 20-mer double-stranded oligonucleotide at a 6:1 DNA to protein ratio, and a random sequenced single-stranded DNA at a 6:1 DNA: protein ratio, did not maintain the solubility of integrase.

We decided to initiate crystal trials with a y-shaped oligonucleotide which mimicked the reaction intermediate of the LTR sequence ligated to the target DNA (Chow et al., 1992). The 3' cytosine was modified to a dideoxy-cytosine to prevent disintegration of the y-mer (suggested by Sam Chow, personal communication). It is still not known whether integrase has a separate binding domain for the specific LTR sequences than for non-specific DNA. We reasoned that having both sequences present would enhance the binding to integrase and that integrase would be more likely to bind in a specific orientation to this sequence. At 20:1 y-mer to integrase concentration, with integrase at 1mg/ml, we were able to dialyze out the CHAPS and reduce the salt concentration to 50mM and maintain half of the integrase in solution for 10 days. Light precipitate was visible in the dialysis buttons. Again without y-mer present, all of the integrase precipitated. This experiment was repeated with 2.5mg/ml integrase, with similar results observationally. This time the samples were retrieved after 2 days, spun in a microfuge and loaded onto an SDS PAGE gel. Despite the lack of precipitate in the y-mer buttons, most of the integrase was found in the pellet. A similar aggregation was noticed by members of Pat Brown's laboratory in the *in vitro* activity assay. They observed that most of the integrase was pelletable, and the pelletable integrase was active (personal communication). Only when 10% glycerol was present did much of the protein remain in the supernatant.

Crystallization trials:

The crystal trials without oligonucleotides were set up in the presence of CHAPS. The trials with oligonucleotides were dialyzed for 2-3 days to reduce the CHAPS concentration. Longer dialysis times produced visible precipitation. The conditions for the oligonucleotide samples were modeled after the conditions used in the *in vitro* activity assay. Because the CHAPS was concentrated along with the integrase in the Amicon or centricon, the exact concentration is unknown. The protein concentration varied between 2mg/ml. up to 10mg/ml. Most of the conditions tested below precipitated the protein, unless otherwise stated. Crystal trials were set up at 4°C unless otherwise specified, because the protein was more soluble at that temperature than at room temperature.

(1) incomplete factorial screne **(ref)**: at 4°C and room temperature.

(2) 0-1.5M ammonium sulfate at pH 4.6-9.6: room temperature or 4°C all precipitated.

(3) 0-1.5M ammonium sulfate/ 0-0.4M sodium citrate at pH 7, 7.5, 8: lower ammonium sulfate concentrations precipitated first. Greater than 4.3 total ionic strength caused phase separation.

(4) 3-26% ammonium sulfate and glycerol at pH 7.5: without glycerol all drops precipitated, with 10% glycerol they were clear.

(5) 13-15% ammonium sulfate with 1% PEG 400: drops appeared clear for a month.

(6) concentration of drops by 0-1.5M ammonium sulfate in the wells, no ammonium sulfate in the drops, at pH 7, 7.5, 8: The pH 7.5 and 8 drops precipitated. The pH 7 drops remained visibly clear for 1 month.

(7) 0-4.5M ammonium chloride at pH 5.6, 7.5, 8.5, 9.7: Drops were clear without phase separation. With 1% PEG 400 or PEG 4000 added, the drops were still clear.

(8) 0-4M ammonium phosphate or 0-1.2M sodium tartrate: all drops precipitated.

(9) 0.5-4M ammonium chloride at pH 4.6-9.6: The pH 4.6-6.5 drops precipitated. At pH 7.5-9.6 and greater than 0.5-1.5M ammonium chloride the drops remained clear.

(10) 10-25% DMSO at pH 5.6, 7.5, 9.5: all drops precipitated.

(11) 15-30% acetone at pH 5.6, 7.5, 9.5: all drops precipitated.

(12) 0-35% MPD at pH 5.6, 7.5, 9.5: At room temperature the 5-35% MPD drops precipitated. AT 4°C the 1-11% MPD drops remained clear.

(13) 10-35% PEG 400, 5-20% PEG 4000: all drops precipitated.

(14) 15-90% lithium sulfate at pH 4.6, 7.5, 8.5: at pH 4.6, greater than 45% $LiSO_4$ is needed to maintain solubility for over 1 month. At pH 7.5 and 8.5, less than 45% $LiSO_4$ is needed to maintain solubility for over 1 month.

(15) 15-90% sodium formate at pH 4.6, 7.5, 8.5: The pH 4.6 drops precipitated first. The pH 7.5 and 8.5 drops were soluble at less than 60% sodium formate. Phase separation occurred at 90% sodium formate.

(16) With DNA - random sequence double stranded 17-mer, random sequence double stranded 13-mer, or y-mer - with and without $MnCl_2$, using ammonium sulfate, PEG 400, PEG 4000 or MPD as the precipitating agent: all drops precipitated.

(17) Drops were set up with the W235E point mutation of integrase with many conditions above, including oligonucleotides. The results were similar.

DNA binding assays:

Much effort was directed at developing a DNA binding assay for integrase. An *in vitro* activity assay existed, but this did not distinguish binding from activity. The advantage of a binding assay for crystallization trials would be to screen conditions that would favor binding to an oligonucleotide. Oligonucleotides could be screened for sequences and lengths that bind most tightly. And finally, the nature of the specific versus the non-specific interactions could be investigated. Four binding assays were investigated: fluorescence, circular dichroism, gel shifts, and surface plasmon resonance.

Fluorescence:

Ligand binding can sometimes be detected by a shift in the fluorescence emission spectrum. The emission spectra of integrase was compared in the presence and absence of y-mer. Protein at 1mg/ml was diluted 10-fold into 100mM NaCl, 20mM Hepes, pH 7.5, and no CHAPS, conditions at which the protein is active in the *in vitro* activity assay. Experiments were performed with and without $MnCl_2$. Even if the CHAPS concentration in the initial protein sample was much greater than 7mM, the dilution assured that the final concentration was less than the cmc. The oligonucleotide y-mer sequence was tested at a final concentration of 8mM. The excitation wavelength was chosen to be 305nm to minimize the absorbance of the DNA. The y-mer did not shift the maximal wavelength of the emission spectra, but did reduce the intensity slightly. This reduction of intensity was completely explained by the absorbance of the added y-mer. The large volume of the cuvettes for this fluorimeter (2 ml) prevented testing higher concentrations of

oligonucleotide. It is possible that DNA binding does result in a fluorescence shift but that too little of the integrase was bound to DNA to detect it.

Circular dichroism (CD):

The CD spectrum of integrase has been published (Lin et al., 1989). Changes in secondary structure due to ligand binding can be detected by CD. A 3mg/ml (about 100mM) sample was dialyzed to 50mM NaCl for a few hours because the *in vitro* integration assay indicates that salt concentrations greater than 150mM interfere with activity. The CHAPS concentration of this sample was still high. The integrase is soluble in this low salt, high CHAPS condition for a short time. The final concentration of y-mer tested was 60mM. No change in the integrase spectrum was detected in the presence of the y-mer: the sum of the integrase spectrum and the y-mer spectrum alone equaled the spectrum of the integrase and y-mer mixed. It is again possible that the binding constant for DNA was higher than the range tested. CHAPS may have interfered with DNA binding, since high CHAPS concentrations interfere with the activity assay. The necessity for high protein concentration in this assay prevented the dilution of the CHAPS. It is also possible that secondary structure rearrangements do not accompany DNA binding.

Gel shift assay:

A gel shift assay detects bands on an acrylamide gel resulting from DNA protein complexes (Stone et al., 1991). The unbound DNA migrates more quickly and is separated from the bound species. The binding constant can be determined from the ratio of bound and unbound DNA at various concentrations. The DNA was detected by ^{32}P labeling and exposure to a film or phosphoimager for quantitation. In the experiments performed here, the

concentration of the y-mer was held constant and the concentration of the integrase was varied. In that way the total signal in each lane of the gel was constant. Again the initial conditions tested were those used in the *in vitro* activity assay. The protein was diluted 10-fold to dilute the CHAPS below the cmc.

In the initial conditions for the assay, a large fraction of the y-mer remained in the well in the presence of integrase. We interpreted this to mean that the integrase aggregated in the well. Higher concentrations of y-mer resulted in more signal in the well, instead of solubilizing the aggregates. Conditions for the assay were modified to minimize aggregation. Lowering the acrylamide to bisacrylamide ratio did not allow more integrase to enter the well. Neither did glycerol, variations of the salt concentration, addition of BSA (bovine serum albumin), or the use of non-frozen integrase samples. 20mM cholic acid decreased the aggregation but resulted in a lot of smearing of the y-mer signal in the gel lane. A point mutation of integrase, W235E, produced bands in the gel, but also showed some aggregation and smearing. The addition of 10% DMSO into the gel also improved the quality of the bands. The final gel conditions used were 6% polyacrylamide, 1/2 x TBE buffer (90mM Tris borate, 2.5mM EDTA, pH 8.3). Gels were pre-run at 150 volts for one hour. The samples were loaded with the gel running at 200 volts, and run at 300 volts. Even with the W235E mutant protein, there was variability in the amount of aggregation in the well with different purifications of integrase. 1mg/ml BSA added to the sample decreased aggregation. The addition of non-specific DNA eliminated the aggregation in the well.

Two shifted bands were observed at 5.6mM integrase, 0.5nM y-mer. A dissociation constant of 50mM was calculated for one of these bands assuming the formation of a simple Michaelis complex between integrase and DNA.

These bands were detected in the presence or absence of $MnCl_2$. 300mM NaCl did not interfere with the bands, even though this salt concentration would interfere with the *in vitro* activity assay. Competition with cold y-mer, a random single-stranded 20-mer sequence, and dI-dC double-stranded oligonucleotides eliminated the bands at similar competitor concentrations. The upper band was eliminated first, at 400nM concentration, while the lower band was eliminated by 4.3mM. This suggested that the binding was not specific. The results were the same whether the competitor was added before the y-mer probe or after.

Competition with unlabeled y-mer resulted in more aggregation in the well. Lower concentrations of double-stranded dI-dC resulted in some aggregation in the wells, but higher concentrations of the oligonucleotide broke up the aggregation. One interpretation of this is that the y-mer and dI-dC contain more than one binding site for the integrase, allowing the formation of chains of integrase and oligonucleotide.

Non-specific binding was also tested by ^{32}P labeling other oligonucleotides. One gel shift band was observed with a non-specific 22-residue oligonucleotide at an integrase concentration of 20mM. The LTR U5 sequence, both a 17-residue and 29-residue oligonucleotide, caused a band shift at an integrase concentration of 5mM. The upper band was darker with these probes, though a lower band was also visible.

We were unable to demonstrate directly that the gel shifts were the result of integrase binding, instead of a protein contaminant. Polyclonal antibodies at concentrations much greater than used for Western blots did not interfere with the binding. Western blots of gel shifts were performed to detect integrase associated with the shifted bands. Instead it was found that integrase was entering the well, even in the absence of DNA. The pH of the

TBE buffer was measured as 8.5, while the calculated pI for integrase is 9. At this pH the integrase should not enter the gel. I do not know whether it is common for the protein to enter the gel without the DNA in a gel shift experiment. The extent of migration of the integrase into the gel reached at least the upper band observed in the gel shift. Silver staining of a gel shift experiment confirmed the presence of protein in the gel. Protein purified through another column, the CM-50 column, also showed the same gel shifts. If the gel shifts were the result of a contaminating protein, it was consistently purified with the integrase in similar quantities.

These experiments indicated that integrase binds specific and non-specific DNA sequences with similar affinities, with binding constants on the order of 1-50mM. Others have also been unable to distinguish specific binding to the LTR sequences from non-specific binding (LaFemina et al., 1992).

To substantiate that integrase is responsible for the gel shifts, a specific monoclonal antibody that would interfere with the binding should be found. In retrospect, a better way to do the assay would have been to use an excess of oligonucleotide instead of integrase. This would have minimized aggregation of the integrase because of the lower concentrations needed for the experiment. Also the DNA might prevent aggregation of the integrase. The multiple bands observed may have been a result of multimerization of the integrase at higher protein concentration, which would not have been observed at lower integrase concentrations.

Surface Plasmon Resonance:

The BIAcore machine by Pharmacia Biosensor was used to measure binding by immobilizing an oligonucleotide to the dextran chip and flowing integrase over the surface. The BIAcore measures on-rates and off-rates by

detecting changes in binding of the flowed substituent to the immobilized substituent on the surface. Changes in the amount of mass bound to the surface affect the refractive index of the surface which is detected by alterations in the angle exhibiting surface plasmon resonance (Fisher & Fivash, 1994). These changes are measured in units of RU, as defined by the manufacturer.

To immobilize the oligonucleotide, streptavidin was first cross-linked to the surface using the non-specific cross-linking agents N-ethyl-N'-(3-diethylaminopropyl)-carbodiimide (EDC) and N-hydroxy-succinimide (NHS). A biotinylated oligonucleotide then bound to the surface specifically. Two double-stranded oligonucleotides were synthisized: a random sequence and an LTR sequence, both 20 bases long with a two base overhang to simulate the LTR terminus after the two 3'-bases are removed. In this way integrase will not modify the LTR sequence. The biotinylated base was at the opposite terminus of the oligonucleotide. The biotinylated oligonucleotides were gel purified before use in the BIAcore.

The first difficulty to overcome was that the integrase bound to the negatively charged dextran surface in the absence of oligonucleotide. This is consistent with integrase being positively charged at pH 7.5. Only a small fraction of the dextran surface is cross-linked with each treatment of NHS/EDC. It was found that repeated NHS/EDC treatments, followed by crosslinking of ethanolamine reduced the non-specific binding of integrase considerably. Every new surface prepared in this way was tested by flowing over integrase after the ethanolamine treatments and prior to the immobilization of the oligonucleotide.

Regeneration of the surface involves dissociating the bound integrase without damaging the immobilized oligonucleotide. This is important

because on-rate measurements require repeated use of the same surface. High salt and extremes of pH did not regenerate the surface. 0.5% SDS and 250mM NaCl did dissociate the integrase. The salt seemed to help in the regeneration, perhaps to solubilize the integrase. Streptavidin is stable in the presence of this concentration of SDS (Bayer et al., 1990).

Following is the protocol used for preparing the chip:

flow: 5 ml/minute

(1) inject 50ml of 50% 0.2M EDC, 50% 0.05M NHS

(2) inject 50ml of 1M ethanolamine

(3) inject 50ml of guanidinium-HCl

(4)-(6) repeat steps (1)-(3)

(7) repeat step (1)

(8) inject 50ml of 0.2 mg/ml streptavidin, in 10mM NaAcetate, pH 4.5

(9) repeat step (2)

(10) inject 50ml of 3 mg/ml oligonucleotide in HBS (10mM Hepes, pH 7.4, 150mM NaCl, 3.4mM EDTA, 0.05% Tween 20)

The ratio of streptavidin to DNA on the surface of the chip was calculated according to the change in the RU values from immobilizing each of these to the surface. Using the conversion of 1 ng/mm^2 = 800 RU for DNA (Pharmacia Biosensor, personal communication), and 1 ng/mm^2 = 1000RU for proteins (Pharmacia Biosensor literature), the DNA bound to the chip in a ratio of 1:1 with the streptavidin.

Integrase was then flowed over the surface at different concentrations in 50mM NaCl, 7mM CHAPS, 10mM Hepes, pH 7.5, with and without $MnCl_2$. All experiments were carried out in the presence of 7mM CHAPS to maintain the solubility of integrase. This may have influenced integrase binding to DNA.

The on-rate is determined by plotting ln(dRU/dt) versus time and is dependent on the concentration of the flowed substituent. Time points were taken starting 50 seconds after the integrase flow began to allow time for the refractive index change due to the integrase buffer. The plots were biphasic, with a more rapid initial phase. The origin of this behavior is unclear. It could be an artifact of the background binding of the integrase to the dextran. More interestingly, it could represent a fast and a slow rate of integrase binding to the oligonucleotide. The second slower rate may be a rate of dimerization of the integrase after a significant population of singly-bound oligonucleotides has developed. Immobilization of both specific and non-specific oligonucleotide sequences resulted in the same behavior for binding of integrase.

The off-rate is determined by plotting $\ln(RU_o/RU)$ (RU_o=initial RU value) versus time and should be concentration independent. Again the plots were biphasic with a faster initial off-rate. The initial faster off-rate was concentration dependent. If this biphasic nature is not an artifact, it may represent dissociation of integrase dimers, which would be concentration dependent. Comparing specific and non-specific oligonucleotide binding, the specific sequence resulted in a longer fast phase for the on-rate, though the rates of dissociation were similar. The specific sequence might be better at promoting dimerization than the non-specific sequence. The binding behavior was similar in the presence or absence of $MnCl_2$.

The above results are preliminary. The nature of the biphasic association and dissociation must be investigated further. Particularly, the binding of integrase to the dextran chip may be the source of artifacts. Immobilizing integrase to the surface and flowing the oligonucleotide past would eliminate this source of artifacts. It may also allow the elimination of CHAPS from the buffer, since the immobilized integrase will be unable to aggregate. It is not known if the immobilized integrase would be stable at low CHAPS concentrations, but CHAPS is diluted to low concentrations, along with the integrase, for the *in vitro* activity assay. A second unanswered question is whether integrase is responsible for the observed binding. Interference with antibodies specific to integrase might answer this question.

Discussion:

Integrase must interact with the specific LTR sequences as well as non-specific chromosomal sequences. It is not known whether the two binding sites are distinct. In the simplest cases, recognition for a specific DNA sequence by a DNA-binding protein is reflected in tighter binding affinity for that sequence. Though recognition implies favorable interactions with the targeted sequence, the favorable free energy does not always appear as binding energy. EcoRV is one example in which the binding affinities to specific and non-specific DNA sequences are comparable but cleavage of the specific site is highly selective (Vipond & Halford, 1993). In both the gel shift assay and the BIAcore assay, integrase bound to specific and non-specific DNA with similar affinities. Though insolubility remains the major obstacle to crystallization, the lack of specificity of DNA binding may make the formation of a single species for crystallization difficult. With more work, one of these assays may be perfected to allow screening of conditions necessary for DNA binding.

Acknowledgements:

The integrase work was started as a collaboration with the Varmus laboratory, and particularly Andy Leavitt, a post-doc in the Varmus lab at the time. The work would not have been possible without their efforts. I particularly want to thank Andy Leavitt for his continued support and optimism in the face of a very difficult problem.

The Arris company was very generous in allowing me to use their BIAcore machine when none was available at UCSF. Particularly, Bob Cass instructed me on the use of the machine.

References:

Bayer, E. A., Ben-Hur, H., and Wilchek, M. Isolation and properties of streptavidin. (1990) *Methods Enzymol. 184*, 80-89.

Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. (1992) *Science 255*, 723-6.

Craigie, R., Fujiwara, T., and Bushman, F. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. (1990) *Cell 62*, 829-37.

Craigie, R., Mizuuchi, K., Bushman, F. D., and Engelman, A. A rapid in vitro assay for HIV DNA integration. (1991) *Nucleic Acids Res. 19*, 2729-34.

Farnet, C. M., and Haseltine, W. A. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. (1991) *J. Virol. 65*, 1910-5.

Fisher, R. J., and Fivash, M. Surface plasmon resonance based methods for measuring the kinetics and binding affinities of biomolecular interactions. (1994) *Curr. Opin. Biotechnol. 5*, 389-95.

Goff, S. P. Retroviral reverse transcriptase: synthesis, structure, and function. (1990) *J. Acquir. Immune Defic. Syndr. 3*, 817-31.

LaFemina, R. L., Schneider, C. L., Robbins, H. L., Callahan, P. L., LeGrow, K., Roth, E., Schleif, W. A., and Emini, E. A. Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. (1992) *J. Virol. 66*, 7414-9.

Leavitt, A. D., Rose, R. B., and Varmus, H. E. Both substrate and target oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in Saccharomyces cerevisiae. (1992) *J. Virol. 66*, 2359-68.

Lin, T. H., Quinn, T. P., Grandgenett, D., and Walsh, M. T. Secondary structural analysis of retrovirus integrase: characterization by circular dichroism and empirical prediction methods. (1989) *Proteins 5*, 156-65.

Michel, H. (1991) , CRC Press, Boca Raton.

Stone, S. R., Hughes, M. J., and Jost, J. Qualitative and quantitative studies of protein-DNA interactions by gel mobility-shift assay. (1991) *BioMethods 5*, 163-183.

Vipond, I. B_4, and Halford, S. E. Structure-function correlation for the EcoRV restriction enzyme: from non-specific binding to specific DNA cleavage. (1993) *Mol. Microbiol. 9*, 225-31.

APPENDIX:

BOTH SUBSTRATE AND TARGET OLIGONUCLEOTIDE SEQUENCES AFFECT IN VITRO INTEGRATION MEDIATED BY HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 INTEGRASE PROTEIN PRODUCED IN *SACCHAROMYCES CEREVISIAE*

Reprinted from (1992) *J. Virology 66*, 2359-2368.

Preface

My contribution to this manuscript was developing the purification of integrase in collaboration with Andy Leavitt. Andy discovered that CHAPS was necessary to solubilize the integrase. Prior to this purification, a complicated purification involving high urea concentrations had been used.

JOURNAL OF VIROLOGY, Apr. 1992, p. 2359–2368
0022-538X/92/042359-10$02.00/0

Vol. 66, No. 4

Both Substrate and Target Oligonucleotide Sequences Affect In Vitro Integration Mediated by Human Immunodeficiency Virus Type 1 Integrase Protein Produced in *Saccharomyces cerevisiae*

ANDREW D. LEAVITT,[1,2]* ROBERT B. ROSE,[1] AND HAROLD E. VARMUS[1,3]

Department of Microbiology and Immunology,[1] Cancer Research Institute,[2] and Department of Biochemistry and Biophysics,[3] The University of California, San Francisco, California 94143

Received 19 November 1991 Accepted 17 January 1992

Integration of retroviral DNA into the host cell genome requires the interaction of retroviral integrase (IN) protein with the outer ends of both viral long terminal repeats (LTRs) to remove two nucleotides from the 3′ ends (3′ processing) and to join the 3′ ends to newly created 5′ ends in target DNA (strand transfer). We have purified the IN protein of human immunodeficiency virus type 1 (HIV-1) after production in *Saccharomyces cerevisiae* and found it to have many of the properties described for retroviral IN proteins. The protein performs both 3′ processing and strand transfer reactions by using HIV-1 or HIV-2 attachment (*att*) site oligonucleotides. A highly conserved CA dinucleotide adjacent to the 3′ processing site of HIV-1 is important for both the 3′ processing and strand transfer reactions; however, it is not sufficient for full IN activity, since alteration of nucleotide sequences internal to the HIV-1 U5 CA also impairs IN function, and Moloney murine leukemia virus *att* site oligonucleotides are poor substrates for HIV-1 IN. When HIV-1 *att* sequences are positioned internally in an LTR-LTR circle junction substrate, HIV-1 IN fails to cleave the substrate preferentially at positions coinciding with correct 3′ processing, implying a requirement for positioning *att* sites near DNA ends. The 2 bp normally located beyond the 3′ CA in linear DNA are not essential for in vitro integration, since mutant oligonucleotides with single-stranded 3′ or 5′ extensions or with no residues beyond the CA dinucleotide are efficiently used. Selection of target sites is nonrandom when *att* site oligonucleotides are joined to each other in vitro. We modified an in vitro assay to distinguish oligonucleotides serving as the substrate for 3′ processing and as the target for strand transfer. The modified assay demonstrates that nonrandom usage of target sites is dependent on the target oligonucleotide sequence and independent of the oligonucleotide used as the substrate for 3′ processing.

The retroviral life cycle is characterized by the formation of a provirus, a DNA copy of the viral genome integrated into a host cell chromosome. Genetic studies have demonstrated that two viral components are essential for retroviral integration: (i) integrase (IN), the protein encoded by the 3′ end of *pol* (9, 10, 21, 22, 24), and (ii) attachment (*att*) sites, DNA sequences located at the ends of the viral long terminal repeats (LTRs) (6, 7, 13, 20, 23). IN first acts in the cytoplasm in a 3′ processing reaction which removes the terminal two nucleotides from the 3′ end of each strand of linear viral DNA (4, 14, 23) so that the 3′ ends of the viral DNA terminate with the dinucleotide CA. A nucleoprotein complex containing the processed viral DNA then enters the nucleus (2), where IN cleaves the chromosomal target DNA in a staggered fashion and ligates the processed 3′ ends of viral DNA to the 5′ ends of the newly cut target DNA in a strand transfer reaction (3, 11, 12). The product of the reaction is a gapped intermediate, which is thought to be repaired by host-derived DNA synthetic machinery, thereby generating the virus-specific short direct repeats that flank the provirus.

Recent advances in understanding IN activity have come from newly developed in vitro assays. The first in vitro assay for retroviral integration used nucleoprotein complexes from murine leukemia virus (MLV) infected cells to provide viral DNA and integration activity (3). The assay was then modified by substituting a heterologous DNA substrate with synthetic *att* sites to function in place of the viral DNA synthesized during infection (13). Subsequently, an in vitro integration assay which requires only purified retroviral IN protein and oligonucleotides representing the ends (*att* sites) of viral DNA was developed (8, 15, 16, 25). The *att* site oligonucleotides serve as both donor and target DNA in the reaction and can undergo both the 3′ processing and strand transfer reactions. As conventionally used, the assay does not require the coordinate strand transfer of two ends of viral DNA at a single target cleavage site, as occurs during integration in vivo; in this sense, it is not a complete integration reaction. Nonetheless, this simplified assay has helped to demonstrate that IN from MLV (8), avian leukosis virus (15), and human immunodeficiency virus type 1 (HIV-1) (5, 17, 29) can mediate all but the final step in the integration events that establish a provirus in an infected cell.

To better understand HIV-1 integration, we have produced HIV-1 IN in *Saccharomyces cerevisiae* and used the purified protein in the *att* site oligonucleotide-based assay to evaluate the DNA substrate requirements for HIV-1 IN. We have also modified the assay so that different oligonucleotides are monitored as the substrates for 3′ processing and as the targets for strand transfer. This has allowed us to demonstrate that in vitro integration occurs nonrandomly in a manner dependent upon the target rather than the donor oligonucleotide sequence.

* Corresponding author.

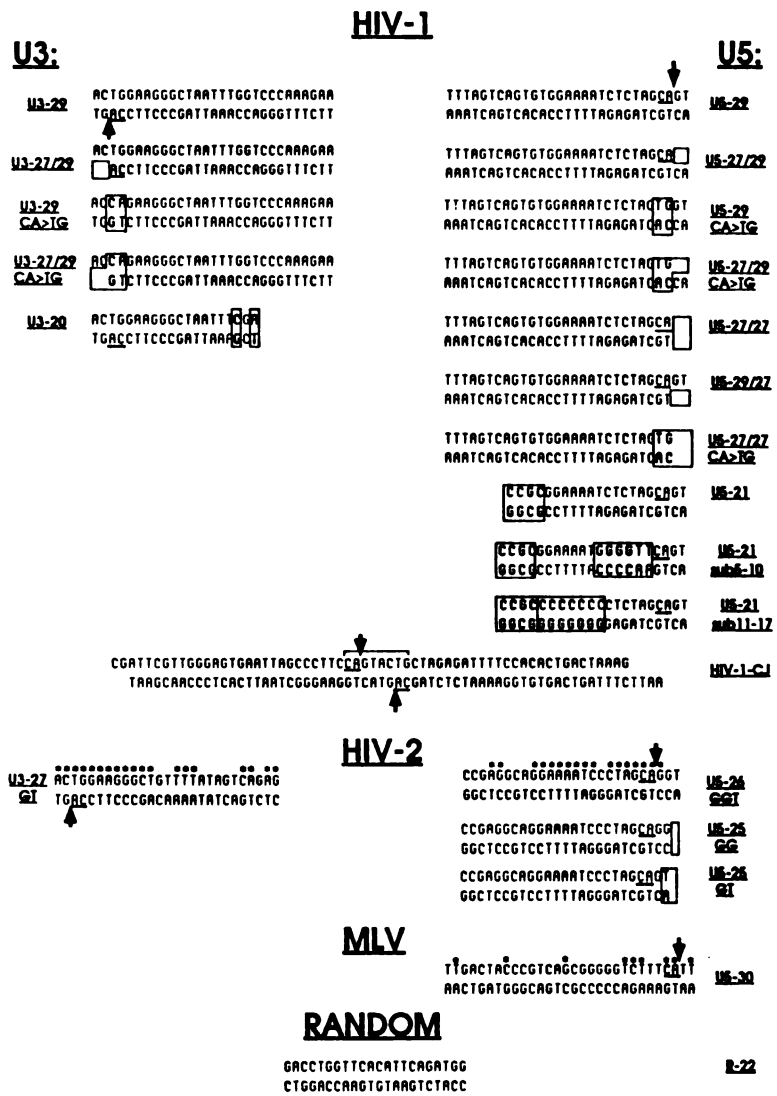


FIG. 1. Oligonucleotides used in assessment of HIV-1 IN activity. Synthetic DNA oligonucleotides were purified and annealed as described in Materials and Methods to produce the duplex oligonucleotides depicted. The conserved 3' CA is underlined. Vertical arrows indicate the 3' processing site. The numbers given after a hyphen refer to the length of each nucleotide strand. A single number after a hyphen implies both strands are a single length. Boxed, boldface base pairs indicate alterations from the wild-type sequence. HIV-1-CJ is the HIV-1 circle junction sequence. The bracketed sequence in HIV-1-CJ represents the inverted repeats. An asterisk over a base pair in the HIV-2 or MLV sequence indicates identity to the corresponding residue in the HIV-1 sequence. The R-22 sequence (RANDOM) comes from the N-II exon of mouse c-*src* and has no overt homology with HIV-1 *att* sites. HIV-1 sequences are from Myers et al. (19).

MATERIALS AND METHODS

Oligonucleotides. Oligonucleotides were obtained from the University of California, San Francisco, Biomolecular Resource Center, except for the HIV-2 oligonucleotides, which were a gift from Steve Hughes, National Cancer Institute, Frederick, Md. Oligonucleotide names and sequences are provided in Fig. 1. T4 polynucleotide kinase (Pharmacia) was used to label the 5' end of oligonucleotides with [γ-^{32}P]ATP (3,000 Ci/mmol; Amersham). Klenow fragment (Pharmacia) was used to label the 3' end of oligonucleotides with [α-^{32}P]dCTP (3,000 Ci/mmol; Amersham).

Yeast strain. *S. cerevisiae* AB116 (*MATα leu2 trp1 ura3-52 prB1-1122 pep4-3 prC1-407* [cir^0]) was propagated under conditions of leucine selection before liquid culture in enriched medium as described elsewhere (1).

Protein purification. The IN-coding sequence from HIV-1 (SF2) was cloned into the yeast expression vector pBS24Ub as described elsewhere (1). The construct creates a fusion gene consisting of the yeast ubiquitin-coding sequence followed in frame by the HIV-1 IN-coding sequence. The wild-type IN sequence, including the amino-terminal phenylalanine, is maintained, and the fusion gene product is accurately cleaved in vivo to yield HIV-1 IN (1). The construct places HIV-1 IN under the transcriptional control of the glucose-regulatable alcohol dehydrogenase 2/glyceraldehyde 3-phosphate dehydrogenase hybrid promoter (1).

After growth in enriched medium for 48 h, cells were pelleted and either lysed immediately or stored at −70°C. Cells were lysed (>90% lysis as determined by phase microscopy) in 50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid; pH 7.5)–5 mM EDTA–1 mM dithiothreitol (DTT) in a Dyno-Mill cell mill (Willy A. Bachofen, Basel, Switzerland), with a 1:2:1.5 ratio of pelleted cells (grams) to glass beads (grams) to buffer (milliliters). Glass beads were removed by filtration over a plastic screen. The lysate was then spun for 20 min at 12,000 × *g* at 4°C. After the pellet was washed once with lysis buffer, it was resuspended in 50 mM HEPES (pH 7.5)–1 mM EDTA–1 mM DTT–1 M NaCl–10 mM CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}. The resuspended pellet was stirred for 2 h at 4°C and then spun for 1 h at 100,000 × *g* at 4°C. The resulting supernatant was either frozen at −70°C in 20% glycerol or used for further purification. Whether used directly or after thawing from −70°C, the supernatant was diluted to give a final composition of 25 mM HEPES (pH 7.5), 1 mM EDTA, 1 mM DTT, and 7 mM CHAPS (buffer A) plus 150 mM NaCl and loaded onto an Aphigel-Heparin (Bio-Rad) column equilibrated with the same buffer. The column was washed with buffer A plus 200 mM NaCl, and adherent proteins were eluted with buffer A plus 350 mM NaCl. Peak elution fractions containing IN were pooled, and the buffer was changed to buffer A plus 100 mM NaCl and 1 M ammonium sulfate (buffer B). The pooled fractions were loaded onto a phenyl-Sepharose column (Pharmacia) equilibrated with buffer B; the column was then washed with buffer B and eluted with buffer A plus 100 mM NaCl. Peak fractions containing IN were pooled and dialyzed into buffer A plus 400 mM NaCl. When this sample was placed on a TSK (Pharmacia) high-performance liquid chromatography sizing column, IN eluted as a well-isolated protein of roughly 32 kDa.

Western blots (immunoblots). Western blots were performed with nitrocellulose filters for transfer and 1% yeast extract (Difco) in phosphate-buffered saline (PBS) as a blocking agent. The primary antiserum was a polyclonal anti-HIV-1 serum provided by Kathelyn S. Steimer (Chiron Corporation, Emeryville, Calif.). The antiserum was raised in rabbits against an HIV-1 IN–human superoxide dismutase fusion protein from *Escherichia coli* (26). The secondary antiserum was a goat anti-rabbit serum conjugated to alkaline phosphatase (Promega). Primary and secondary antisera were incubated at 4°C in PBS plus 1% yeast extract (Difco). The Western blots were developed with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Sigma).

Assay for IN activity using radiolabeled oligonucleotides. Single-oligonucleotide reactions were performed in 50 mM MOPS (morpholinepropanesulfonic acid; pH 7.0)–1 mM DTT–15 mM $MnCl_2$–10% glycerol–100 μg of bovine serum albumin per ml–roughly 20 mM NaCl (from protein storage buffer) with roughly 5 pmol of IN and 0.5 pmol of radiolabeled *att* site oligonucleotide in a total volume of 10 μl. Two-oligonucleotide reactions differed only by the addition of roughly equal molar amounts of unlabeled *att* site oligonucleotide and 3′ radiolabeled target oligonucleotide. The reaction mixtures were incubated at 30°C for 1 h, and the reactions were stopped by adding 10 μl of loading buffer (95% formamide, 20 mM EDTA, 0.05% bromphenol blue, 0.05% xylene cyanol) and heating to 100°C for 5 min. The reaction products were then analyzed by electrophoresis on 20% polyacrylamide gels with 7 M urea in Tris-borate-EDTA (TBE). Autoradiography was performed with either Amersham Hyperfilm-MP or Kodak X-Omat film. Schematic representations of the single-oligonucleotide assay have been published elsewhere (8, 15). Figure 7 shows a schematic representation of our modification of the assay with separate donor and target oligonucleotides.

RESULTS

Isolation of HIV-1 IN expressed in *S. cerevisiae*. We expressed HIV-1 IN in *S. cerevisiae*, using the ubiquitin fusion construct described in Materials and Methods. The fusion protein is cleaved in vivo to yield IN with the wild-type amino acid sequence, including the amino-terminal phenylalanine, in an amount readily detected on a Coomassie blue-stained sodium dodecyl sulfate (SDS)-polyacrylamide gel of a total cell lysate (Fig. 2, lane 2). After centrifugation at 12,000 × *g*, most of the viral protein is located in the pellet of the IN-containing lysate (Fig. 2, lanes 3 and 4). After resuspension in lysis buffer plus 1 M NaCl and 10 mM CHAPS and centrifugation at 100,000 × *g*, IN remains predominantly in the supernatant (Fig. 2, lane 5). Sequential column chromatography on heparin sulfate and phenyl-Sepharose, as outlined in Materials and Methods, yields 1.5 to 2 mg of highly purified IN (>95% pure) from 1 liter of yeast culture (Fig. 2, lane 6). Western blot analysis of the total cell lysate (Fig. 2, lane 8) and of the material eluted from the phenyl-Sepharose column (Fig. 2, lane 9) demonstrated an immunoreactive protein at the expected molecular mass of roughly 32,000 Da. The protein was not detected in the total cell lysate of an isogenic yeast strain that lacks the IN-coding sequence (Fig. 2, lanes 1 and 7). All reactions described below were performed with IN eluted from the phenyl-Sepharose column.

HIV-1 IN from *S. cerevisiae* has specific 3′ nuclease and strand transfer activity in vitro that is impaired by alterations to the HIV-1 *att* site oligonucleotide sequence. We first tested our yeast-derived HIV-1 IN for its ability to mediate 3′ processing and strand transfer by using wild-type HIV-1 *att* site oligonucleotides 29 bp in length (U5-29 and U3-29). The predicted 27-nucleotide primary cleavage product generates the predominant signal after the addition of IN to the 29-bp U3 or U5 *att* site oligonucleotide (Fig. 3A, lanes 1 to 4). The lengths of the primary cleavage products were confirmed by DNA markers (data not shown). Like Bushman and Craigie (5) and LaFemina et al. (17), we found that HIV-1 IN acts more efficiently on U5 than U3 *att* site oligonucleotides. In contrast, Vink et al. (29) observed equal activities with U5 and U3 *att* site oligonucleotides. We also found minimal 3′ processing of an oligonucleotide representing the U5 end of MLV (Fig. 3A, lanes 9 and 10). As previously noted by others with HIV-1 IN produced in insect cells (5) or *E. coli* (17, 29), HIV-1 IN-mediated reactions showed a preference

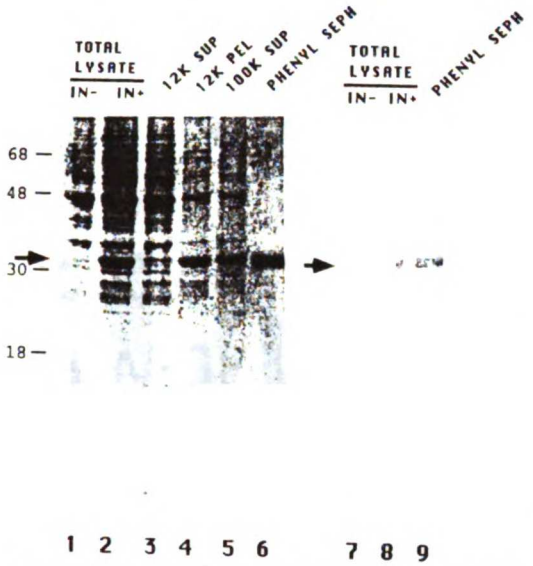


FIG. 2. Production and purification of HIV-1 IN expressed in *S. cerevisiae*. Yeast cells (strain AB116) containing an inducible expression plasmid without (pBS24Ub) or with (pBS24Ub-IN) the HIV-1 IN-coding sequence were grown as described in Materials and Methods. Samples from the purification steps outlined in Materials and Methods were analyzed by SDS–12.5% polyacrylamide gel electrophoresis. Lanes 1 to 6 were stained with Coomassie blue. Lanes 7 to 9 underwent Western blot analysis. Lanes 1 and 7, AB116(pBS24Ub) total cell lysate; lanes 2 and 8, AB116(pBS24Ub-IN) total cell lysate; lane 3, AB116(pBS24Ub-IN) supernatant from centrifugation at 12,000 × *g*; lane 4, AB116(pBS24Ub-IN) pellet from centrifugation at 12,000 × *g*, resuspended in 1 M NaCl–10 mM CHAPS; lane 5, AB116(pBS24Ub-IN) supernatant from centrifugation at 100,000 × *g*; lanes 6 and 9, AB116(pBS24Ub-IN) pooled phenyl-Sepharose eluate. The arrow indicates the location of HIV-1 IN. Molecular mass markers (in thousands of daltons) are indicated on the left.

for $MnCl_2$ over $MgCl_2$ and an absolute requirement for a divalent cation (data not shown). The strand transfer step of integration is detected in vitro by the appearance of a series of slowly migrating bands that represent oligonucleotides longer than the starting probe (Fig. 3) (8, 16). To confirm that the more slowly migrating products we observed were the result of IN-mediated strand transfer, they were isolated from gels after electrophoresis and subjected to Maxam-Gilbert sequencing. From assays which employed either U3-29 or U5-29, the sequences were as predicted for IN-mediated strand transfer (data not shown), in agreement with published reports from other groups (5, 8, 16). In an infected cell, 3′ processing occurs in the cytoplasm and strand transfer occurs in the nucleus, implying that the two steps are not coupled. Successful strand transfer with preprocessed oligonucleotides, U5-27/29 and U3-27/29, demonstrates that these two steps are also separable in vitro (Fig. 3A, lanes 5 to 8).

A 3′ CA dinucleotide constitutes the terminal two nucleotides at each 3′ end of all known proviruses. U5 *att* site oligonucleotides with the CA changed to TG significantly impaired both 3′ processing and strand transfer (Fig. 3B, lanes 1 to 4). Preprocessed oligonucleotides with the same CA-to-TG mutation also showed negligible strand transfer (Fig. 3B, lanes 5 and 6). The 3′ CA dinucleotide is therefore independently required for both 3′ processing and strand transfer. Identical requirements for the CA dinucleotide were seen with U3-derived *att* site oligonucleotides (data not shown).

We initially used oligonucleotides containing 29 bp of terminal HIV-1 sequence (Fig. 3A). However, our HIV-1 IN mediates both 3′ processing and strand transfer with oligonucleotides containing only the distal 17 bp from the U5 or U3 end of HIV-1 (Fig. 3C, lanes 1 to 4). To determine the importance of the 13 bp located adjacent to the U5 CA, we changed, as a unit, residues 1 to 6 or 7 to 13 internal to the CA dinucleotide (U5-21 sub5-10 and U5-21 sub11-17 [Fig. 1]). When incubated with HIV-1 IN, both mutants showed impaired 3′ processing and strand transfer (Fig. 3C, lanes 5 to 8), suggesting that IN has important interactions throughout at least 7 and as many as 13 bp adjacent to the U5 CA. The data do not address the importance of individual residues.

The terminal nucleotides removed during integration are not required for in vitro integration. To determine the integration-related function of the two terminal base pairs removed during retroviral integration, a series of mutant U5 *att* site oligonucleotides with altered 3′ and 5′ termini were assayed for 3′ processing and strand transfer (Fig. 4). As also demonstrated above, strand transfer was only slightly greater with wild-type preprocessed oligonucleotides, which contained two unpaired bases at the 5′ terminus, than for wild-type unprocessed oligonucleotides, which contained 2 bp beyond the conserved CA (Fig. 4, lanes 7 and 8 and lanes 1 and 2, respectively). When the starting oligonucleotide terminated with two unpaired bases 3′ to the CA, wild-type levels of 3′ processing and strand transfer were observed (Fig. 4, lanes 3 and 4). When the *att* site oligonucleotide terminated bluntly at the CA, wild-type levels of strand transfer were still observed (Fig. 4, lanes 5 and 6). To confirm that the strand transfers observed in Fig. 4, lanes 4 and 6, occurred by a mechanism similar to that seen with wild-type *att* site oligonucleotides, we tested another mutant oligonucleotide that contains TG in place of the 3′ CA dinucleotide and terminates bluntly after the TG. This mutant failed to undergo strand transfer (Fig. 4, lanes 9 and 10), as would be expected for an IN-mediated reaction. The results suggest that neither the 3′-terminal nor the 5′-terminal nucleotides removed from each end of the viral DNA during the integration reaction are essential in vitro. Furthermore, 3′ processing and strand transfer can occur in the absence of base pairing of the two nucleotides beyond the CA. Possible conflicts between these findings and results obtained with MLV mutants in vivo (6) are addressed in the Discussion.

An LTR-LTR circle junction oligonucleotide is not preferentially cleaved 3′ to the conserved CA sequences. Recent studies have established that MLV uses linear and not circular DNA forms as the precursor to provirus formation (3, 4, 14, 18). However, the LTR-LTR circle junction form contains the same sequences that are found at each end of the linear DNA species. To determine whether events in the in vitro assay with HIV-1 IN conform to in vivo observations with MLV, we radiolabeled the 5′ end of each strand of an oligonucleotide representing the HIV-1 circle junction and annealed each to unlabeled cDNA. DNA fragments of the lengths predicted for the products of nicking 3′ to the conserved CA residues (30 bp for CJ-3 and 32 bp for CJ-4

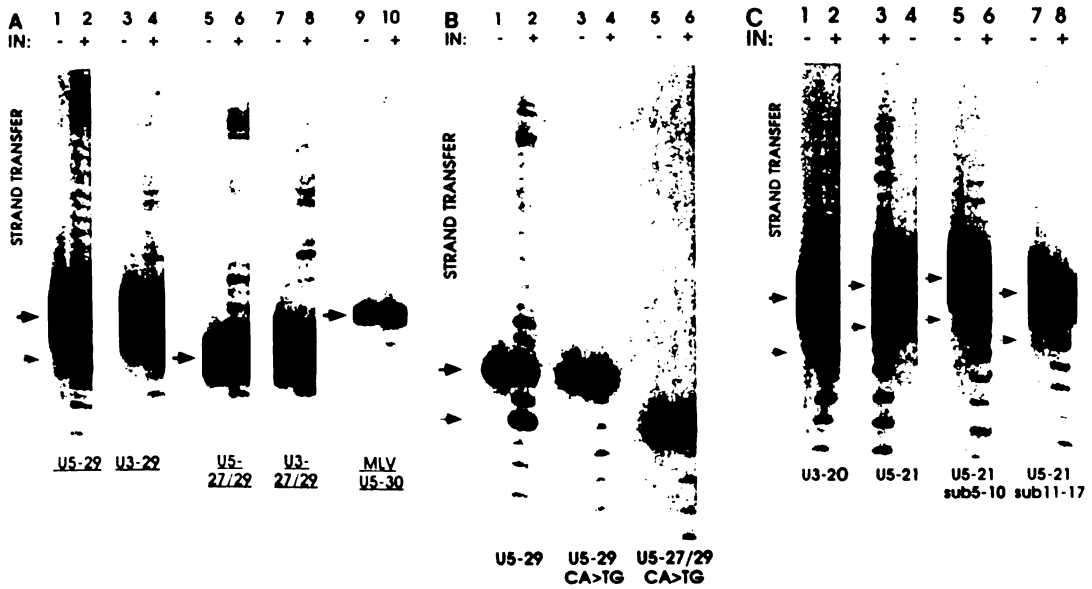


FIG. 3. Yeast-derived HIV-1 IN processes and recombines wild-type HIV-1 *att* site oligonucleotides and is sensitive to *att* site sequence alterations. Assays were performed as described in Materials and Methods. Oligonucleotides are identified below each pair of lanes (Fig. 1). Each oligonucleotide probe was incubated either without (−) or with (+) purified HIV-1 IN. (A) Lanes 1 to 4 contain the wild-type HIV-1 oligonucleotides. Lanes 5 to 8 contain preprocessed HIV-1 oligonucleotides. Lanes 9 and 10 contain the wild-type U5 end of MLV. The solid arrows represent the starting oligonucleotide probe, and the hatched arrow represents the product of 3′ processing. Strand transfer products are indicated. (B) Lanes 1 to 6 contain a series of mutants with CA mutations in the HIV-1 U5 *att* site. The solid arrow represents the starting oligonucleotide probe in lanes 1 to 4; the hatched arrow represents the product of 3′ processing in lane 2 and the labeled, preprocessed probe in lanes 5 and 6. (C) Lanes 1 to 4 contain wild-type HIV-1 oligonucleotides which contain 17 bp from the HIV-1 U3 or U5 *att* site. Lanes 5 to 8 assess the effect of altering groups of base pairs within 13 residues of the conserved CA in the U5 *att* site. The solid arrows represent starting oligonucleotide probes, and the hatched arrows represent the products of 3′ processing.

[Fig. 5B]) were not generated when oligonucleotide HIV-1-CJ was incubated with HIV-1 IN (Fig. 5A). Since 17 bp from either the U3 or U5 end is sufficient for 3′ processing (Fig. 3C, lanes 1 to 4), the lack of 3′ processing of CJ-3 or CJ-4 is not a consequence of an inadequate *att* site sequence but instead suggests a requirement that functional *att* sequences be near an end of the oligonucleotide. The several bands corresponding to oligonucleotides shorter than the starting probe (Fig. 5A, lanes 2 and 4) may be the result of contaminating nucleases, or they may represent the products of *att* site-independent endonuclease activity of IN. The locations of scissions that generated the prominent signals are indicated with solid arrows in Fig. 5B.

HIV-2 ends are functional substrates for HIV-1 IN in vitro. The HIV-1 and HIV-2 genomes have highly similar nucleotide sequences (19). For instance, at the left (U3) end of viral DNA, 15 of 17 nucleotides are identical, and at the right (U5) end, 18 of 23 are identical. However, unique among known retroviruses, unintegrated HIV-2 has 3 bp beyond the conserved 3′ CA at the U5 end, while the conventional 2 bp are present beyond the conserved 3′ CA at the U3 end (30). In contrast, HIV-1 has 2 bp distal to the 3′ CA at both the U5 and the U3 ends (28, 31).

To determine whether HIV-1 IN can mediate 3′ processing on the termini of HIV-2, a series of wild-type and mutant HIV-2 *att* site oligonucleotides were analyzed for 3′ processing and strand transfer by HIV-1 IN. Figure 6 demonstrates that HIV-1 IN preferentially processed HIV-2 *att* site oligonucleotides immediately 3′ of the conserved CA at both the U3 and U5 ends. At the HIV-2 U5 end, this was the case with the wild-type oligonucleotides (with 3 bp beyond the 3′ CA [Fig. 6, lanes 1 and 2]) and with either of two mutant oligonucleotides, each with 2 bp beyond the 3′ CA (Fig. 6, lanes 3 to 6). However, 3′ processing was consistently more efficient with 2 bp 3′ of the CA than with 3 bp 3′ of the CA. These findings agree with the observations of others that the location of the 3′ cleavage is dependent on the position of the CA dinucleotide and is not defined by a fixed distance from the end of the DNA molecule (5, 29). As we observed with HIV-1 *att* site oligonucleotides, the patterns of strand transfer products with the HIV-2 U5 and U3 oligonucleotides differ from each other, and IN-mediated reactions appear to be more efficient with U5 than with U3 *att* site oligonucleotides (Fig. 6, lanes 1 and 2 and lanes 7 and 8, respectively).

The target oligonucleotide influences the strand transfer pattern. There is no known sequence specificity for the target site of retroviral integration. Additionally, there are no known structural requirements for target DNA that are likely to be maintained in the oligonucleotide-based integration assay. Consequently, it is reasonable to assume that all sites within an oligonucleotide may serve equally as target sites for strand transfer. However, the strand transfer patterns show that all target sites are not used equally (Fig. 3A), as also observed by others studying IN from MLV (8) and

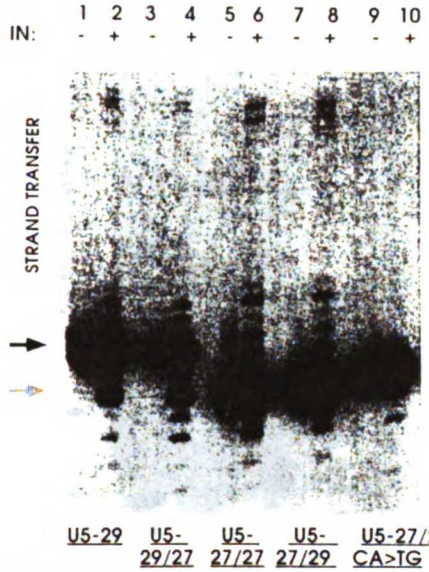


FIG. 4. The terminal nucleotides removed during provirus formation are not required for HIV-1 IN-mediated 3′ processing or strand transfer. Assays were performed as described in Materials and Methods. Oligonucleotides are identified below each pair of lanes (Fig. 1). Each oligonucleotide probe was incubated without (−) or with (+) purified HIV-1 IN. Strand transfer products migrated more slowly than the starting probe. The solid arrow represents the starting oligonucleotide probe in lanes 1 to 4; the hatched arrow represents the products of 3′ processing in lanes 2 and 4 and the starting probe in lanes 5 to 10.

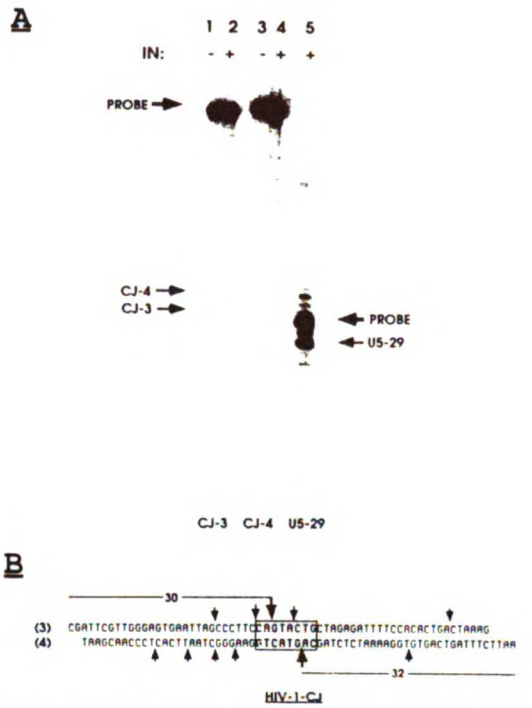


FIG. 5. HIV-1 IN does not cut LTR-LTR circle junction oligonucleotides at the 3′ processing site. Assays were performed as described in Materials and Methods. (A) Oligonucleotides are identified below each pair of lanes. CJ-3 and CJ-4 refer to HIV-1-CJ with either the upper or lower strand (as shown in panel B) 5′ end labeled. Paired lanes with a given oligonucleotide probe are either without (−) or with (+) purified HIV-1 IN. The solid arrows represent the starting probe. The hatched arrows indicate the expected location of the labeled product after 3′ processing. (B) The HIV-1-CJ oligonucleotide sequence is shown. The boldface, boxed sequence depicts the perfect, inverted repeat of the LTR-LTR circle junction. The hatched arrows are located 3′ to the conserved CA in each strand, the usual site of 3′ processing. The solid arrows indicate the location of cleavage corresponding to the prominent bands in lanes 2 and 4 in panel A. The lengths of the labeled products expected for 3′ processing of CJ-3 and CJ-4 are indicated as 30 and 32 nucleotides, respectively.

HIV-1 (5). Not only did we find nonrandom use of target sites, but different patterns were produced with HIV-1 U3 and U5 oligonucleotides. Such nonrandom use of potential target sites could be a function of the substrate or the target.

To examine the separate roles of the substrate and target oligonucleotides, we altered the standard reaction to include two different oligonucleotides, only one of which has the *att* site sequence that permits it to serve as the substrate for 3′ processing (Fig. 7). Although both oligonucleotides can serve as targets of integration, the HIV-1 *att* site oligonucleotide is not radiolabeled and therefore cannot be detected when it acts as a target in a strand transfer reaction. The second oligonucleotide, which does not contain HIV-1 *att* site sequences and therefore cannot serve as a substrate for 3′ processing, is 3′ radiolabeled and must be joined to the unlabeled oligonucleotide to generate the slowly migrating products characteristic of strand transfer. The modified oligonucleotide-based integration assay thus detects strand transfer products but not the products of 3′ processing.

We first used 3′ radiolabeled MLV U5-30 as the target oligonucleotide, because we had shown that it undergoes negligible 3′ processing when mixed with HIV-1 IN (Fig. 3A, lanes 9 and 10). Neither the 3′-radiolabeled target alone nor the target plus unlabeled MLV U5-30 yielded significant amounts of strand transfer products in the presence of IN (Fig. 8A, lanes 1 to 3). However, when unlabeled HIV U5-29 or unlabeled HIV U3-29 was mixed with 3′-radiolabeled MLV U5-30 DNA and IN, strand transfer products were readily seen (Fig. 8A, lanes 4 and 5). Identical strand transfer patterns were observed with the HIV-1 U3 and U5 *att* site oligonucleotides; recall that, in contrast, different patterns were seen with the two HIV-1 *att* sites in the standard assay (Fig. 3, lanes 1 to 4). Therefore, target site selection appears to be independent of the oligonucleotide that undergoes 3′ processing.

Although the use of MLV U5-30 as a radiolabeled target produced identical strand transfer patterns for unlabeled HIV-1 U5 and U3 *att* site oligonucleotide substrates, it did not produce strand transfer patterns consistent with random use of potential target sites. We wondered if this lack of randomness could be due to unsuspected interactions between HIV-1 IN and MLV U5-30 that might not be expected with a target unrelated to known retroviral sequences. We

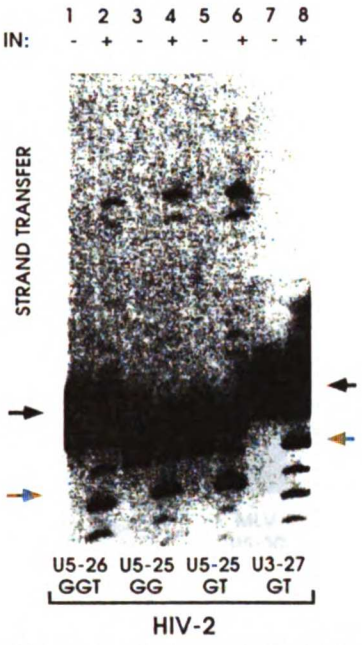


FIG. 6. HIV-2 *att* site oligonucleotides are correctly processed and recombined by HIV-1 IN. Assays were performed as described in Materials and Methods. Oligonucleotides are identified below each pair of lanes (Fig. 1). Each oligonucleotide probe was incubated either without (−) or with (+) purified HIV-1 IN. Solid arrows represent the starting oligonucleotide probe, and hatched arrows represent the products of 3′ processing (removal of three nucleotides in lane 2; removal of two nucleotides in lanes 4, 6, and 8). Strand transfer products are indicated.

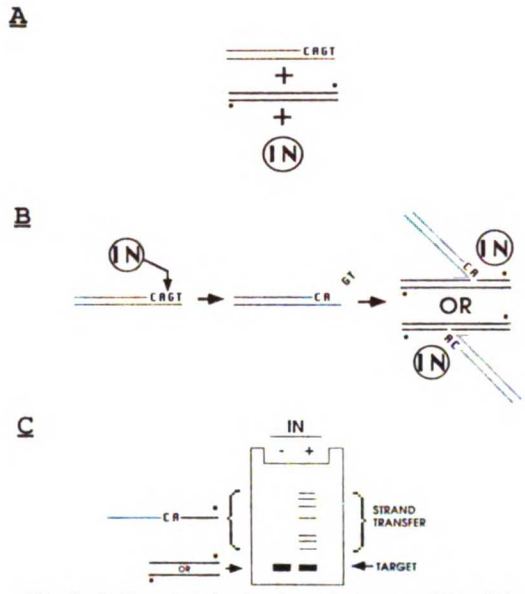


FIG. 7. Outline of an in vitro integration assay which distinguishes the effects of target and substrate oligonucleotides. (A) HIV-1 IN is incubated with two different oligonucleotides; one is a substrate for 3′ processing, and the other is a target for detectable strand transfer events. The substrate for 3′ processing is not radiolabeled and is depicted with broken lines; the target for insertion of 3′ processed DNA is radiolabeled at its 3′ ends and depicted with solid lines. •, the 3′ radiolabel on the target oligonucleotides. (B) The product of 3′ processing is unlabeled and therefore not detectable by autoradiography. Some of the 3′ processed unlabeled *att* site oligonucleotides will be joined to either unlabeled *att* site oligonucleotides (dashed lines) or oligonucleotides radiolabeled at their 3′ ends (solid lines). OR indicates that strand transfer of the *att* site oligonucleotide can occur into either of the target strands. (C) After heating and electrophoresis on a denaturing 20% acrylamide gel, the products of strand transfer into the 3′ radiolabeled target oligonucleotide are detected by autoradiography.

therefore used the R-22 oligonucleotide, which is based on the sequence from a c-*src* intron, as a radiolabeled target. When HIV U5-29 or HIV U3-29 was mixed with the 3′-radiolabeled R-22 in the presence of IN, strand transfer products were readily formed (Fig. 8B). Again, in contrast to the standard assay results (Fig. 3A), the strand transfer patterns were identical when unlabeled HIV U5-29 or HIV U3-29 was used. However, R-22 is also used nonrandomly as a target, suggesting that HIV-1 IN prefers certain target sequences for integration in vitro.

DISCUSSION

Retroviral integration, an essential step in the virus life cycle, involves a series of DNA cutting and joining events. To better understand the role of HIV-1 IN in these events, we expressed the protein in *S. cerevisiae* and developed a procedure to obtain highly purified IN for in vitro analysis. Using a now standard oligonucleotide-based in vitro assay for retroviral integration (8, 15), we have characterized the functional activities of the purified IN and assessed the sequence and structural requirements for oligonucleotide substrates. In addition, by modifying the assay (Fig. 7), we have shown that nonrandom use of target sites by yeast-derived HIV-1 IN in vitro is a function of target and not *att* site DNA sequence.

The HIV-1 IN isolated from *S. cerevisiae* has the native amino acid sequence of HIV-1 IN, and we can recover 1.5 to 2 mg of highly purified (>95%) protein per liter of cell culture (Fig. 2). The yeast-derived IN performs the 3′ processing and strand transfer reactions involved in retroviral integration, showing a strong preference for cognate *att* site sequences (Fig. 3). Our results corroborate the findings of others, using MLV (8), avian leukosis virus (15), and HIV-1 (5, 17, 29) IN proteins, that IN alone is sufficient for these two important steps in the integration of retroviral DNA.

To address the requirement for a functional *att* site for HIV-1 IN, a series of *att* site oligonucleotides were assayed for their effects on 3′ processing and strand transfer. Using blunt-ended and preprocessed U5 and U3 *att* site oligonucleotides, we showed that alteration of the conserved 3′ CA to TG has an independent and marked impairment of both 3′ processing and strand transfer (Fig. 3B). These findings are in agreement with those of LaFemina et al. (17), obtained with HIV-1 U5 *att* site oligonucleotides. Others have also

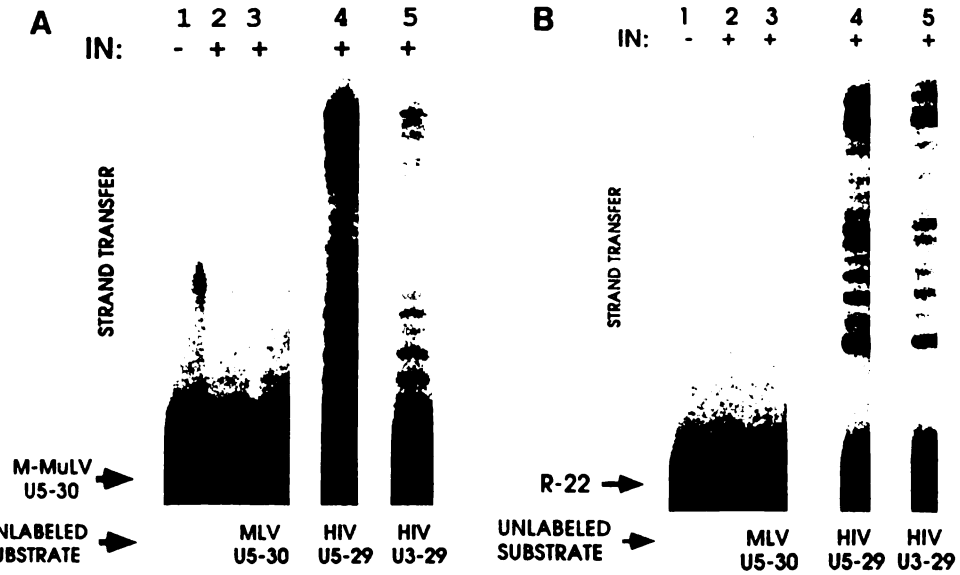


FIG. 8. Target site selection is dependent on the target and independent of the substrate oligonucleotides. The modified oligonucleotide assay shown in Fig. 7 was performed with unlabeled substrate oligonucleotides for 3' processing, as indicated below each lane. The 3' radiolabeled target DNA was MLV U5-30 (A) or R-22 (B). − and + indicate the absence and presence of HIV-1 IN, respectively. Strand transfer products are indicated. See Fig. 7 and its legend for a detailed description of the assay. M-MuLV, Moloney MLV.

demonstrated the importance of the U3 (5) and U5 (29) CA on 3' processing, but preprocessed mutants were not tested, precluding conclusions about independent effects on strand transfer. In addition, two groups have shown that both the C residue and the A residue must be altered to achieve nearly complete ablation of 3' processing (17, 29).

While the 3' CA is necessary for both efficient 3' processing and strand transfer, it is not sufficient. For example, the MLV U5 *att* site oligonucleotide is a poor *att* site substrate for HIV-1 IN (Fig. 3A), yet it possesses an appropriately placed CA dinucleotide. Furthermore, nucleotides internal to the conserved CA are important for IN function: alteration of the 6 bp immediately adjacent to the conserved CA (U5-21 sub5-10) or those located 7 to 13 bp from the CA (U5-21 sub11-17) impaired 3' processing and associated strand transfer (Fig. 3C, lanes 5 to 8). Notably, changes located 7 to 13 bp from the conserved CA resulted in a greater reduction of IN activity than did changes located 1 to 6 bp from the conserved CA. Similar findings have been recently reported by Vink et al. (29); however, Bushman and Craigie (5) did not see significant impairment of IN activity with a grouped alteration of base pairs located 6 to 11 residues from the U3 CA. These differences may reflect the choice of new sequences as well as the loss of wild-type sequences. In general, mutations that change only one to three residues are unlikely to impair IN-mediated functions significantly unless they include the 3' CA (17, 29). Thus, while no single base pair outside the HIV-1 3' CA has been shown to have a strong inhibitory effect on 3' processing or strand transfer, nucleotides up to 13 positions from the CA may influence IN activity in vitro if enough residues are altered. Definition of the terminal sequence requirements for integration of HIV-1 DNA in vivo has not yet been achieved.

We also investigated the role of the two terminal base pairs which are removed during provirus formation (Fig. 4). These 2 bp appear dispensable for either IN-mediated reaction. The terminal 2 bp can be removed completely (Fig. 4, lanes 5 and 6), or unpaired nucleotides can be present as 3' or 5' extensions (Fig. 4, lanes 2 and 3 and lanes 7 and 8, respectively), yet 3' processing (U5-29/27) and strand transfer (U5-29/27, U5-27/29, and U5-27/27) occur as with wild-type (U5-29/29) *att* site oligonucleotides. We have confirmed that the strand transfer reaction seen with oligonucleotides that end bluntly at the conserved CA is dependent on the CA dinucleotide, as would be expected of an IN-mediated reaction (Fig. 4, lanes 9 and 10). In contrast, an MLV mutant which lacks the 2 bp beyond the U5 CA is unable to establish a provirus in vivo, despite synthesis of normal amounts of viral DNA (6). This apparent conflict between our in vitro and published in vivo results could represent either an unexpected difference between HIV-1 and MLV IN proteins or a difference between the demands of the in vivo and in vitro integration reactions. This question could be resolved by appropriate comparisons of MLV and HIV *att* site mutants in vivo and in vitro, tests that might challenge the verisimilitude of the in vitro assay.

After reverse transcription of the incoming retroviral RNA, linear viral DNA enters the nucleus, where some DNA molecules remain linear and others become one- or two-LTR circles (for a review, see reference 27). Recent studies with MLV strongly argue that linear DNA is the precursor to proviral DNA (3, 4, 14, 18). In agreement with these findings, linear (Fig. 3A) but not two-LTR circle junction (Fig. 5) forms of HIV-1 *att* site oligonucleotides serve as substrates for 3' processing and strand transfer for HIV-1 in vitro. Others have shown a progressive decline of

3' processing by HIV-1 IN as the number of base pairs 3' to the conserved CA is increased from 2 to 5 (29). Taken together, these results suggest that correct 3' processing by HIV-1 IN occurs only when the *att* sequence is very near the end of a DNA molecule. However, the location of the 3' processing site for HIV-1 IN is not a fixed distance from the end, but is determined by the position of the CA dinucleotide (Fig. 6) (5, 29). In contrast, avian sarcoma virus IN preferentially nicks avian sarcoma virus DNA 3' to the conserved CA dinucleotide in circle junction oligonucleotides (16) or in larger fragments containing a circle junction (14a). However, the activity observed with circle junction oligonucleotides was less than that seen with terminal *att* sites (16). These findings may be due to intrinsic differences between HIV-1 and avian sarcoma virus IN proteins or differences in reaction conditions.

The strand transfer patterns seen with our yeast-derived HIV-1 IN in the now standard oligonucleotide assay demonstrate nonrandom use of target sites for integration (Fig. 3A), an observation also made by others (5, 8, 15). In addition, we noticed that the usage patterns of U3 and U5 *att* site oligonucleotides differ (Fig. 3A). To distinguish the influence of substrate and target oligonucleotides on target site usage, we altered the assay so that target and substrate oligonucleotides can be independently assessed. In the modified assay, a 3'-radiolabeled oligonucleotide serves as a target for another unlabeled oligonucleotide. Only the unlabeled oligonucleotide has an HIV-1 *att* site and can serve as a donor in a strand transfer reaction, while both oligonucleotides can serve as targets for integration. However, only integration into the radiolabeled target can be detected. When a common oligonucleotide was used as a target, HIV-1 U5 and U3 *att* site oligonucleotide substrates produced identical strand transfer patterns (Fig. 8). This contrasts with the different strand transfer patterns seen when HIV-1 U5 or U3 *att* site oligonucleotides served as both substrate and target in the standard in vitro integration reaction (Fig. 3A, lanes 1 to 4). However, regardless of the target used, target site usage is nonrandom, and the patterns are different with *src*-derived or MLV-derived oligonucleotide targets. These results suggest that the strand transfer pattern is a function of target sequence and not a function of the donor DNA sequence undergoing 3' processing.

The influence of the nucleotide sequence on the selection of the integration site appears not to be confined to the oligonucleotide-based integration assay described here. We have also observed that target nucleotide sequences influence the location of integration sites when minichromosomes or naked circular plasmids are used as targets in in vitro integration reactions, with purified IN or viral nucleoprotein complexes as a source of integration activity targets (21a). The significance of such site selectivity during integration in vivo is under investigation.

ACKNOWLEDGMENTS

We thank Phil Barr for assistance with protein expression, Paul Feucht and Ian Bathurst for assistance with the Dyno-Mill, Kathelyn Steimer for HIV-1 IN antisera, and Steve Hughes for donating HIV-2 *att* site oligonucleotides. We also thank Paul Bates, Pat Brown, Peter Pryciak, Lily Shiue, and John Young for critical reading of the manuscript.

A.D.L. is supported by a Burroughs Wellcome Research Foundation Award. H.E.V. is an American Cancer Society Research Professor. This work was supported by grants from the National Institutes of Health.

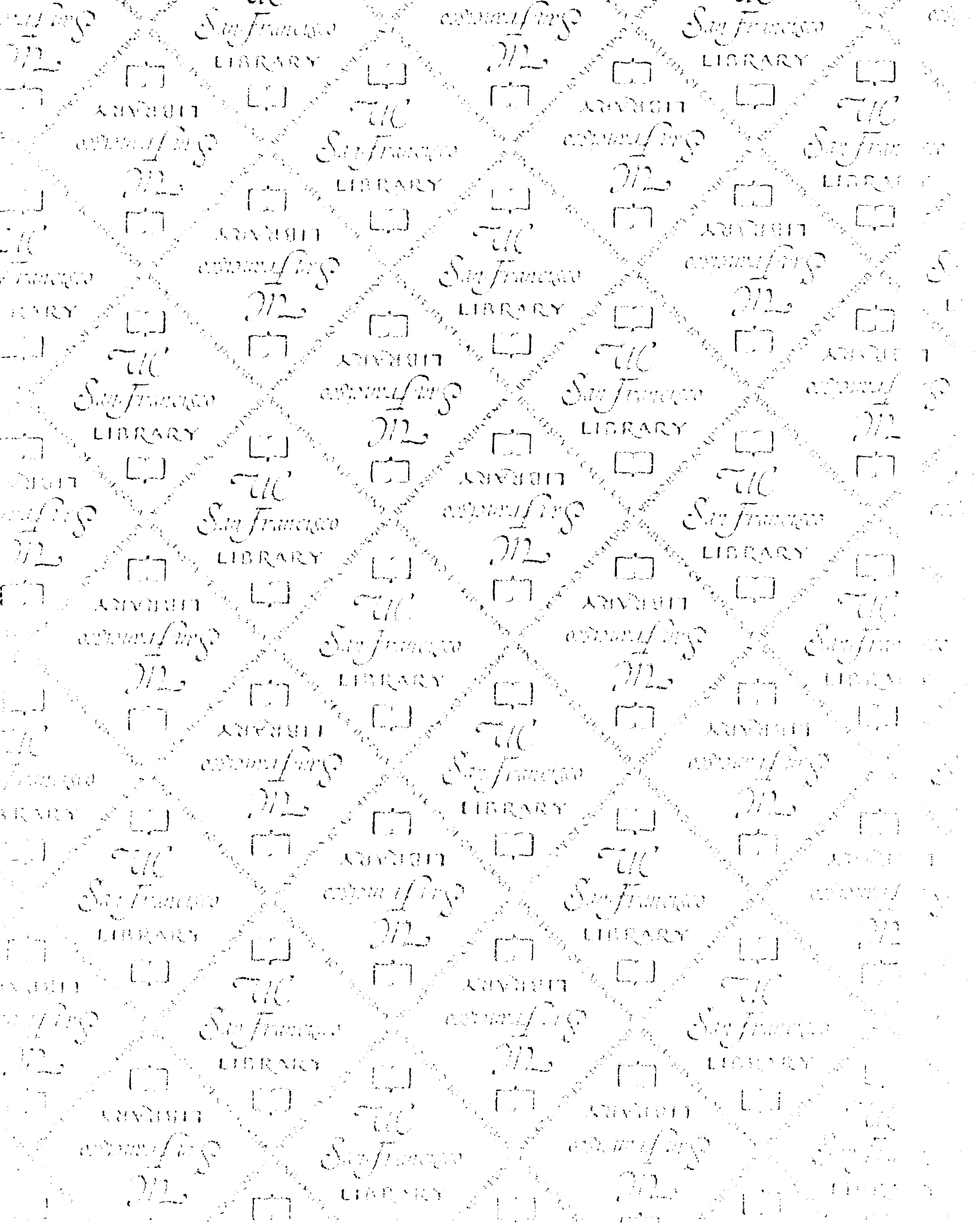
REFERENCES

1. **Barr, P. J., E. A. Sabin, C. T. L. Ng, K. E. Landsberg, K. S. Steimer, I. C. Bathurst, and J. R. Shuster.** 1991. Ubiquitin fusion approach to heterologous gene expression in yeast: high level production of amino-terminal authentic proteins. *In* R. T. Hatch, C. Goochee, T. Moreira, and Y. Alroy (ed.), Expression systems and processes for recombinant DNA products. American Chemical Society, Washington, D.C.
2. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3**:469–478.
3. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA *in vitro*. Cell **49**:347–356.
4. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA **86**:2525–2529.
5. **Bushman, F. D., and R. Craigie.** 1991. Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. Proc. Natl. Acad. Sci. USA **88**:1339–1343.
6. **Colicelli, J., and S. P. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. Cell **42**:573–580.
7. **Colicelli, J., and S. P. Goff.** 1988. Sequence and spacing requirements of a retrovirus integration site. J. Mol. Biol. **199**:47–59.
8. **Craigie, R., T. Fujiwara, and R. Bushman.** 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. Cell **62**:829–837.
9. **Donehower, L. A.** 1988. Analysis of mutant murine leukemia viruses containing linker insertion mutations in the 3' region of *pol*. J. Virol. **62**:3958–3964.
10. **Donehower, L. A., and H. E. Varmus.** 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA **81**:6461–6465.
11. **Ellison, V., H. Abrams, T. Roe, J. Lifson, and P. Brown.** 1990. Human immunodeficiency virus integration in a cell-free system. J. Virol. **64**:2711–2715.
12. **Farnet, C. M., and W. A. Haseltine.** 1990. Integration of human immunodeficiency virus type 1 DNA *in vitro*. Proc. Natl. Acad. Sci. USA **87**:4164–4168.
13. **Fujiwara, T., and R. Craigie.** 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. Proc. Natl. Acad. Sci. USA **86**:3065–3069.
14. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.

14a. **Grandgenett, D. P., and A. C. Vora.** 1985. Site specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. Nucleic Acids Res. **13**:6205–6221.

15. **Katz, R. A., G. Merkel, J. Kulkosky, J. Leis, and A. M. Skalka.** 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. Cell **63**:87–95.
16. **Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis.** 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. **63**:5319–5327.
17. **LaFemina, R. L., P. A. Callahan, and M. G. Cordingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. J. Virol. **65**:5624–5630.
18. **Lobel, L. I., J. E. Murphy, and S. P. Goff.** 1989. The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. J. Virol. **63**:2629–2637.
19. **Myers, G., S. F. Josephs, and J. A. Berzofsky (ed.).** 1989. Human retroviruses and AIDS 1989: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
20. **Panganiban, A. T., and H. M. Temin.** 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. Nature (London) **306**:155–160.

21. **Panganiban, A. T., and H. M. Temin.** 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. Proc. Natl. Acad. Sci. USA **81**:7885–7889.
21a. **Pryciak, P. M., and A. D. Leavitt.** Unpublished data.
22. **Quinn, T. P., and D. P. Grandgenett.** 1988. Genetic evidence that the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. J. Virol. **62**:2307–2312.
23. **Roth, M. J., P. L. Schwartzberg, and S. P. Goff.** 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: Dependence on IN function and terminal DNA sequence. Cell **58**:47–54.
24. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**:1043–1052.
25. **Sherman, P. A., and J. A. Fyfe.** 1990. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. Proc. Natl. Acad. Sci. USA **87**:5119–5123.
26. **Steimer, K. S., K. W. Higgins, M. A. Powers, J. C. Stephans, A. Gyenes, C. George-Nascimento, P. A. Luciw, P. J. Barr, R. A. Hallewell, and R. Sanchez-Pescador.** 1986. Recombinant polypeptide from the endonuclease region of the acquired immune deficiency syndrome retrovirus polymerase (*pol*) gene detects serum antibodies in most infected individuals. J. Virol. **58**:9–16.
27. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
28. **Vink, C., M. Groenink, Y. Elgersma, R. A. M. Fouchier, M. Tersmette, and R. H. A. Plasterk.** 1990. Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA. J. Virol. **64**:5626–5627.
29. **Vink, C., D. C. van Gent, Y. Elgersma, and R. H. A. Plasterk.** 1991. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. J. Virol. **65**:4636–4644.
30. **Whitcomb, J. M., and S. H. Hughes.** 1991. The sequence of human immunodeficiency virus type 2 circle junction suggests that integration protein cleaves the ends of linear DNA asymmetrically. J. Virol. **65**:3906–3910.
31. **Whitcomb, J. M., R. Kumar, and S. H. Hughes.** 1990. Sequence of the circle junction of human immunodeficiency virus type 1: implications for reverse transcription and integration. J. Virol. **64**:4903–4906.


The San Francisco LIBRARY