THE UNIVERSITY OF CHICAGO

PROTEIN DYNAMICS AND FUNCTION ARE CORRELATED IN
THE CHEMICAL MECHANISM OF HIV-1 PROTEASE CATALYSIS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY
VLADIMIR Y. TORBEEV

CHICAGO, ILLINOIS
JUNE 2009
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF PUBLICATIONS</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF SCHEMES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>xii</td>
</tr>
</tbody>
</table>

## CHAPTER

1. Introduction – chemistry of enzyme catalysis and HIV-1 protease.........................1

2. Convergent chemical synthesis and crystal structure of a 203 amino acid ‘covalent
dimer’ HIV-1 protease enzyme molecule........................................................................19

3. Crystal structure of chemically synthesized HIV-1 protease and a ketomethylene
isostere inhibitor based on the p2/NC cleavage site ......................................................31

4. Dynamics of ‘flap’ structures in three HIV-1 protease/inhibitor complexes probed by
total chemical synthesis and pulse-EPR spectroscopy...............................................41

5. Protein dynamics and function are correlated in the chemical mechanism of HIV-1
protease catalysis ...............................................................................................................49

6. The true ionization state of the catalytic dyad Asp25/25’ in unliganded HIV-1 protease:
$^{13}$C NMR studies of HIV-1 protease and several chemical analogues, unliganded and
complexed with inhibitors.................................................................................................74

7. Hydrogen bonds at the protein-inhibitor interface in the HIV-1 protease / inhibitor
complexes probed by total chemical synthesis and X-ray crystallography.....................85
8. Short, strong H-bonds in complexes of the HIV-1 protease with substrate-derived inhibitors detected by NMR spectroscopy .................................................................100

9. Significance and future perspectives ........................................................................127

Appendix A List of chemically-synthesized HIV-1 protease and its analogues ..........138
Appendix B Supporting information for chapter 2 .......................................................140
Appendix C Supporting information for chapter 3 .......................................................148
Appendix D Supporting information for chapter 4 .......................................................153
Appendix E Supporting information for chapter 5 .......................................................169
Appendix F Supporting information for chapter 6 .......................................................192
Appendix G Supporting information for chapter 7 .......................................................194
Appendix H Supporting information for chapter 8 .......................................................204
Appendix I Single-molecule fluorescence assays of proteolytic activity of HIV-1 protease ..................................................................................................................207
Appendix J Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme ...........................................................................................................231
LIST OF PUBLICATIONS

Chapter 2

Chapter 3

Chapter 4

Chapter 5.

Chapters 6 – 8.
V.Yu.Torbeev, S.B.H.Kent, to be submitted.

Appendix J
LIST OF FIGURES

Figure
1.1. Catalytic mechanism in which covalent tetrahedral enzyme-substrate intermediate forms ................................................................................................7
1.2. General acid-general base mechanism ...........................................................................8
1.3. Northrop’s chemical and kinetic ‘isomechanism’ for an aspartic protease .........9
2.1. Analytical HPLC traces (λ = 214 nm) of kinetically controlled ligation of (A1-A40)-\(\alpha\)COSC\(\text{C}_6\text{H}_4\text{CH}_2\text{COOH}\) and Cys-(A42-A99)-\(\alpha\)COSCH\(\text{CH}_2\text{CH}_2\text{Arg}_4\) ...........................................22
2.2. Analytical HPLC traces (λ = 214 nm) corresponding to the final step in the convergent synthesis of the HIV-1 ‘covalent dimer’ construct .............................................24
2.3. FT-ICR ESI LC-MS spectrum of ‘covalent dimer’ HIV-1 protease ..........................25
2.4. X-ray crystallographic structure of the ‘covalent dimer’ HIV-1 protease complexed with MVT-101 inhibitor .................................................................26
3.1. Total chemical synthesis of wild-type HIV-1 PR and the [D25N]HIV-1 PR analogue ..............................................................................................................33
3.2. Active site region in complexes of inactive D25N chemical analogue and active HIV-1 protease with KVS-1 inhibitor .................................................................38
3.3. Superposition of active site regions of structure of wild-type HIV-1 protease complexed with KVS-1 inhibitor ........................................................................39
4.1. Hypothetical reaction coordinate diagram for peptide bond hydrolysis catalyzed by HIV-1 protease .......................................................................................43
4.2. Snapshots of the various conformational states of the HIV-1 protease ‘flaps’ ..................................................................................................................44
5.1. Structural features of HIV-1 protease ..............................................................................51
5.2. Dynamic properties of chemically synthesized Gly51 analogues of HIV-1 protease ................................................................................................................57
5.3. Active site structures revealed by MD simulations, and a scheme for the mechanism of HIV-1 protease catalysis taking into account the dynamics of the flaps ........................................................................................................61
6.1. 226 MHz \(\text{\textsuperscript{13}C-\{\text{\textsuperscript{1}H}\}}\) NMR spectra for unliganded HIV-1 protease and three chemical analogue enzymes ........................................................................................................80
6.2. 226 MHz \(\text{\textsuperscript{13}C-\{\text{\textsuperscript{1}H}\}}\) NMR spectra for [L-Ala51/51′]HIV-1 protease with inhibitors ..................................................................................................................81
6.3. 226 MHz \(\text{\textsuperscript{13}C-\{\text{\textsuperscript{1}H}\}}\) NMR spectra for complex of [L-Ala51/51′]HIV-1 protease with the reduced isostere inhibitor MVT-101 ..............................................................................82
7.1. Flap structures (residues 47–54 and 47′-54′), catalytic aspartates (25 and 25′) in crystal structures of chemical analogues of HIV-1 protease complexed with the reduced isostere inhibitor MVT-101: (a) wild-type [Gly51/51′]HIV-1 protease, (b) [L-Ala51/51′]HIV-1 protease, (c) [D-Ala51/51′]HIV-1 protease and (d) [Aib51/51′]HIV-1 protease ..................................................................................................................91
7.2. Flap structures (residues 47–54 and 47′-54′), catalytic aspartates (25 and 25′) in crystal structures of chemical analogues of HIV-1 protease complexed with the reduced isostere inhibitor MVT-101: (a) [L-Ala51/Gly51′]HIV-1 protease, (b)
[Gly51/D-Ala51']HIV-1 protease, (c) [L-Ala51/D-Ala51']HIV-1 protease and (d) [Aib51/Gly51']HIV-1 protease. .................................................................92

7.3. Flap structures (residues 47–54 and 47’–54’), catalytic aspartates (25 and 25’) in crystal structures of chemical analogues of HIV-1 protease complexed with the reduced isostere inhibitor MVT-101: (a) [O-Ile50/50’]HIV-1 protease, (b) [Ile50/O-Ile50’]HIV-1 protease, (c) [O-Gly51/51’]HIV-1 protease and (d) [alloIle50/50’]HIV-1 protease. .................................................................93

7.4. Flap structures (residues 47–54 and 47’–54’), catalytic aspartates (25 and 25’) in crystal structures of chemical analogues of HIV-1 protease complexed with the ketomethylene isostere inhibitor KVS-1.................................................................94

7.5. Flap structures (residues 47–54 and 47’–54’), catalytic aspartates (25 and 25’) in crystal structures of chemical analogues of HIV-1 proteases complexed with inhibitors .............................................................................................................95

7.6. Selected interatomic distances at the HIV-1 protease flaps / inhibitor MVT-101 interface.................................................................................................................97

8.1. 1H NMR spectra acquired with binominal 1331 water suppression for HIV-1 protease preparations complexed with MVT101 .........................................................103

8.2. 1H NMR spectrum for complex of [L-Ala51/51’]HIV-1 protease with the ketomethyleneisostere inhibitor KVS-1, acquired with watergate 3919 water suppression.................................................................................................................109

8.3. 900 MHz 1H spectra acquired with 1331 binominal water suppression with maximum excitation set at ~16 ppm for wild-type HIV-1 protease complex with MVT-101 inhibitor at six different pH values ......................................................110

8.4. Signal intensities of deshielded protons in HIV-1 protease complex with MVT-101 inhibitor plotted as a function of the H2O content in the H2O/D2O mixture.................................................................................................................111

8.5. 600 MHz 1H spectra for [L-Ala51/51’]HIV-1 protease complexed with MVT-101 inhibitor acquired with 1331 binominal water suppression with maximum excitation set at ~16 ppm at different pH values ......................................................112

8.6. 600 MHz 1H spectra for [D-Ala51/51’]HIV-1 protease complexed with MVT-101 acquired with 1331 binominal water suppression with maximum excitation set at ~16 ppm at different pH values......................................................113

8.7. 600 MHz 1H spectra for [allo-Ile50/50’]HIV-1 protease complexed with MVT-101 acquired with 1331 binominal water suppression with maximum excitation set at 16 ppm at six different pH values ......................................................................114

8.8. 600 MHz 1H spectra for [O-Ile50/50’]HIV-1 protease complexed with MVT-101 acquired with 1331 binominal water suppression with maximum excitation set at 16 ppm at seven different pH values ..............................................115

8.9. Low-field resonances acquired with 1331 binominal pulse sequence for complex of wild-type HIV-1 protease complexed with [CH215NH]MVT-101 inhibitor at pH 3.1 and two temperatures.................................................................116

8.10. 15N decoupler offset was varied in order to determine approximate chemical shift of 15N nucleus coupled to resonance at 13.3 ppm......................................................117

8.11. Resonance at ~13.3 ppm undergoes shift 0.15 ppm downfield upon increase of D2O content in solution of HIV-1 protease / MVT-101 complex ..............................117
8.12. Compare the ratio of resonances at 16.3 ppm and 15.4 ppm with nearly the same pH value taken for (a) complex of wild-type HIV-1 protease with [15N]MVT-101 inhibitor and (b) complex of wild-type HIV-1 protease and non-labeled MVT-101.................................118

9.1. Potential energy surfaces for solvent fluctuations and proton transfer as a function of driving force ................................................................................134

9.2. Some transition state mimicking isosteres to replace the scissile amide bond in aspartic protease inhibitors ........................................................................135

B.1. Analytical HPLC traces (\(\lambda = 214\) nm) of transthioesterification reaction ......142
B.2. Analytical HPLC traces (\(\lambda = 214\) nm) of native chemical ligation of Thz\(^{201}\)-Gly\(_4\)-(B1-B40)-\(^{23}\)COSCH\(_2\)CH\(_2\)Arg\(_4\),amide and (Cys\(_{B41}\)-B99)-COOH..........143
B.3. Enzymatic activity of the ‘covalent dimer’ HIV-1 PR .....................................144
B.4. Experimental (A) and simulated (B) FT-ICR data for the most abundant charge state 23H\(^+\)..........................................................................................................145

C.1. Analytical HPLC traces (\(\lambda = 214\) nm) for two-segment native chemical ligation en route to the synthetic wild-type HIV-1 protease........................................150
C.2. Analytical HPLC profiles (\(\lambda = 214\) nm) together with ESI-MS data (insert) corresponding to each diastereomer.................................................................151
D.1. Illustrates experimental steps in the total chemical synthesis of spin-labeled wild-type HIV-1 protease and its [D25N] inactive analogue .............................155
D.2. Analytical HPLC (\(\lambda = 214\) nm) and ESI mass-spectra of spin-labeled constructs of (a) active HIV-1 protease, and (b) its [D25N] inactive analogue...............156
D.3. Data analysis for nitroxide spin-labeled inhibitor-free inactive [D25] chemical analogue .................................................................159
D.4. Data analysis for nitroxide spin-labeled inhibitor-free active HIV-1 protease .................................................................................................160
D.5. Data analysis for nitroxide spin-labeled [D25N] analogue of HIV-1 protease with MVT-101 inhibitor .................................................................161
D.6. Data analysis for nitroxide spin-labeled active HIV-1 protease with MVT-101 inhibitor ......................................................................................162
D.7. Data analysis for nitroxide spin-labeled [D25N] analogue of HIV-1 protease with KVS-1 inhibitor (keto-form) ..........................................................163
D.8. Data analysis for nitroxide spin-labeled active HIV-1 protease with KVS-1 inhibitor (hydrated form) ........................................................................164
D.10. Data analysis for nitroxide spin-labeled active of HIV-1 protease with JG-365 inhibitor ....................................................................................166
E.1. Analytical HPLC (\(\lambda = 214\) nm) and FT-ICR ESI-MS characterization of chemically synthesized 99-residue HIV-1 protease monomers .........................................172
E.2. Analytical HPLC (\(\lambda = 214\) nm) and FT-ICR ESI-MS characterization of chemically synthesized 203-residues covalent dimers ...........................................173
E.3. Analytical HPLC (\(\lambda = 214\) nm) and FT-ICR ESI-MS characterization of chemically synthesized 99-residue [D25N]HIV-1 protease monomers labelled with MTSSL-nitroxide spin-label for EPR studies ..............................................174
E.4. Analytical HPLC ($\lambda = 214$ nm) and FT-ICR ESI-MS characterization of chemically synthesized 99-residue $^{15}$N- and $^{13}$C-labelled HIV-1 protease monomers for NMR studies .................................................. 175

E.5. Flap structures (residues 47–54 and 47’–54’), catalytic aspartates (25 and 25’) in crystal structures of [L-Ala51; D-Ala51’] covalent dimer HIV-1 protease with (a) MVT-101 inhibitor, and (b,c) JG-365 inhibitor .............................................. 177

E.6. Plots of crystallographic inhibitor, and (b,c) JG-365 inhibitor .............................................. 177

E.6. Plots of crystallographic temperature $B$-factors for amino acid sequence of two ‘flap’ structures (residues 37-61 and 37’-61’) .............................................. 178

E.7. $^1$H-detected CPMG $^{15}$N relaxation dispersion data for $^{15}$N-labelled residues in [L-Ala51/51’] HIV-1 protease (left column), [D-Ala51/51’] HIV-1 protease (middle column), and [Aib51/51’] HIV-1 protease (right column) .............................................. 184

E.8. (continuation of Figure E.7). $^1$H-detected CPMG $^{15}$N relaxation dispersion data for $^{15}$N-labelled residues in [L-Ala51/51’] HIV-1 protease (left column), [D-Ala51/51’] HIV-1 protease (middle column), and [Aib51/51’] HIV-1 protease (right column) .............................................. 185

E.9. Pulse-EPR measurements data. Background subtracted dipolar echo evolution curves (left), dipolar spectra after Fourier transformation (center) and $L$-curves for Tikhonov regularization with optimal regularization parameter $\lambda = 10$ in all cases (right) .......................................................................................................................... 187

E.10. Nitroxide-to-nitroxide’ distance distribution profile for nitroxide-labelled homodimeric L-Ala51/51’ (top) and D-Ala51/51’ (bottom) unliganded HIV-1 protease molecules obtained in pulse-EPR measurements .............................................. 188

E.11. Active-site titration of chemically synthesized enzymes by surface-plasmon resonance .................................................................................................................. 189

E.12. Sequence of substrate derived reduced isostere inhibitor with hexapeptide binding site for HIV-1 protease and analogues depicted in red and polyethylene glycol and biotin moieties in blue ................................................................................................................. 190

F.1. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of $^{13}$C-labeled (a) wild-type [1-99] HIV-1 protease, and (b) its L-Ala51 analogue .............................................. 192

F.2. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of $^{13}$C-labeled (a) D-Ala51 [1-99] HIV-1 protease, and (b) Aib51 [1-99] HIV-1 protease .............. 193

G.1. Active-site titration of chemically synthesized enzymes by surface-plasmon resonance .................................................................................................................. 199

G.2. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of (a) alloIle50 [1-99] HIV-1 protease, and (b) O-Ile50 [1-99] HIV-1 protease .... 201

G.3. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of (a) O-Gly51 [1-99] HIV-1 protease, and (b) O-Gly52 [1-99] HIV-1 protease ....... 202

G.4. (a) Analytical HPLC ($\lambda = 214$ nm) and (b) FT-ICR ESI mass-spectra of [Ile50;O-Ile50’] HIV-1 protease .............................................. 203

H.1. Temperature dependence of 16.35 ppm and 15.45 ppm resonances .................. 204

H.2. Arrhenius plots of the effect of temperature on the observed transverse relaxation rates ($1/T_{2obs} = \pi \cdot$ line width) of the 16.35 ppm and 15.45 ppm peaks ................................................................................................................. 205
H.3. 900 MHz $^1$H-$^1$H NOESY with 1–1 water suppression was performed for [L-Ala51/51']HIV-1 protease with MVT-101 inhibitor in an attempt to correlate low-field resonances .................................................................206

I.1. Two approaches for single-molecule detection of proteolytic activity of HIV-1 protease .............................................................................................................212

I.2. In the total chemical synthesis of 203-amino acid ‘covalent dimer’ HIV-1 protease molecule, last alkylation step is replaced by the reaction with maleimide-containing PEGylated biotin-tag, which is further used to tether such HIV-1 protease construct to the quartz surface employing biotin-neutravidin method .............................................................................................................213

I.3. Monitoring the cleavage reaction of Cy5-substrate-Cy3-biotin with TIR microscopy ...........................................................................................................214

I.4. Average donor (green) and acceptor (red) signal intensities plotted versus time ..............................................................................................................215

I.5. Examples of single-molecule traces for proteolysis of Cy5-substrate-Cy3-biotin ..............................................................................................................216

I.6. Monitoring the binding of HIV-1 protease to Cy5-inhibitor-Cy3-biotin with TIR microscopy .......................................................................................217

I.7. Single-molecule fluorescence traces for HIV-1 protease binding to FRET-labelled inhibitor .................................................................................................218

I.8. TIR-monitoring of proteolytic reaction of HIV-1 protease tethered to quartz slide .............................................................................................................219

I.9. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides ..................................................................225

I.10. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides. (a) Cy3-substrate-Cy5-Btn; (b) Cy3-inhibitor-Cy5-Btn .............................................................................................................226

I.11. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides. (a) Alexa647-substrate-Alexa555-Btn; (b) Alexa555-substrate-Alexa647-Btn .............................................................................................................227

I.12. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides. (a) Cy3B-substrate-Cy5-Btn; (b) Cy3-substrate-Cy5 .............................................................................................................228
LIST OF SCHEMES

Scheme
1.1. Computational enzyme design protocol for a multistep reaction ........................................2
2.1. Convergent synthesis of ‘covalent dimer’ HIV-1 protease ....................................................21
2.2. Kinetically controlled ligation (KCL) of peptide-($\alpha$thioarylester) and peptide- ($\alpha$thioalkylester) ..................................................................................................................23
8.1. Conformational isomerisations of the ‘flaps’ taking place in the HIV-1 protease complex with the reduced isostere inhibitor MVT-101 ...............................................................123
E.1. Total chemical synthesis approach for preparation of chemical analogues of HIV-1 protease has allowed to site-specifically $^{15}$N-label residues of interest, namely, selected residues in flaps, residue Gly40 in elbows and residues Asp25 and Gly27 in catalytic site ..................................................................................................180
E.2. Final fitting for flap residues in $[L$-Ala51/51’] and $[D$-Ala51/51’] chemical analogues was performed according to 3-site exchange model. 3-site exchange model was invoked based on interflap distance measurements by pulse-EPR and previous MD simulations, where three principal conformers were identified ................................................................................................................186
I.1. In the synthesis of substrates and inhibitors for single-molecule spectroscopic observation of proteolytic activity of HIV-1 protease, the polypeptide synthesized with the help of Boc- SPPS chemistry contained one amino-group functionality (either peptide N-terminus or amino-group of lysine) and one sulfohydryl-group of cysteine ........................................................................................................221
I.2. Structures of fluorophore-labeled peptide substrates and reduced isostere inhibitors (p2/NC cleavage site in red) ........................................................................................................222
I.3. Alexa647 / Alexa555 dyes were tested to improve photostability and make substrates more hydrophilic ..................................................................................................................223
I.4. Cy3B dye was introduced as a more photostable variant of Cy3-dye (in a). In (b) structure of Cy3 / Cy5 labelled substrate tested in experiment with HIV-1 protease construct tethered to the surface ......................................................................224
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.</td>
<td>Crystal parameters, data collection and refinement statistics ........................................37</td>
</tr>
<tr>
<td>5.1.</td>
<td>Steady-state kinetics of proteolysis by chemically synthesized HIV-1 protease and Gly51 analogues .................................................................53</td>
</tr>
<tr>
<td>7.1.</td>
<td>Selected interatomic distances (in Å) at the flaps / inhibitor MVT-101 interface. For clarification see Figure 7.6 .................................................................96</td>
</tr>
<tr>
<td>7.2.</td>
<td>Steady-state kinetics of proteolysis by chemically synthesized analogues of HIV-1 protease (see Table 5.1 in Chapter 5) ....................................................98</td>
</tr>
<tr>
<td>8.1.</td>
<td>$^1$H NMR chemical shifts observed in low-field region of spectra for chemical analogues of HIV-1 protease with reduced isostere MVT-101 inhibitor ........104</td>
</tr>
<tr>
<td>A.1.</td>
<td>Chemically-synthesized HIV-1 protease (HIVPR) and its analogues ..........138</td>
</tr>
<tr>
<td>B.1.</td>
<td>Data collection and refinement statistics (highest resolution is in parenthesis) .................................................................147</td>
</tr>
<tr>
<td>D.1.</td>
<td>Summary of the data for EPR distance measurements ........................................167</td>
</tr>
<tr>
<td>E.1.</td>
<td>Data collection and refinement statistics for complexes of [L-Ala51;D-Ala51']HIV-1 protease with MVT-101 and JG-365 inhibitors (highest resolution shell is in parenthesis) ..................................................179</td>
</tr>
<tr>
<td>E.2.</td>
<td>Amide chemical shifts for $^{15}$N-labelled chemical analogues of HIV-1 protease measured on 600 MHz spectrometer and referenced to DSS-d$_6$ ................181</td>
</tr>
<tr>
<td>E.3.</td>
<td>Amide chemical shifts for $^{15}$N-labelled chemical analogues of HIV-1 protease acquired on 900 MHz spectrometer referenced to DSS-d$_6$ ..................181</td>
</tr>
<tr>
<td>E.4.</td>
<td>T$_1$, T$_2$ and $^1$H-$^{15}$N NOE values measured for $^{15}$N-labelled chemical analogues of HIV-1 protease on 600 MHz spectrometer. Residues located at the tips of the flaps are highlighted in magenta .........................................182</td>
</tr>
<tr>
<td>E.5.</td>
<td>Generalized $S^2$ order parameters derived with model-free analysis using anisotropic diffusion tensor and T$_1$, T$_2$ and $^1$H-$^{15}$N NOE parameters. Residues located at the tips of the flaps are highlighted in magenta .........................................183</td>
</tr>
<tr>
<td>E.6.</td>
<td>Initially all residues were fitted with 2-site exchange model without constrains on populations (only $k_{ex}$, s$^{-1}$ are listed, population of minor component varied in the range 1–3%) .........................................................186</td>
</tr>
<tr>
<td>E.7.</td>
<td>Active-site titration of chemically synthesized enzymes by surface-plasmon resonance .................................................................................................................190</td>
</tr>
<tr>
<td>G.1.</td>
<td>Data collection and refinement statistics for chemical analogues of HIV-1 protease (highest resolution is in parenthesis) .........................................................194</td>
</tr>
<tr>
<td>G.2.</td>
<td>Active-site titration of chemically synthesized enzymes by surface-plasmon resonance .................................................................................................................200</td>
</tr>
<tr>
<td>H.1.</td>
<td>Parameters extracted from fitting to equation (1) ........................................................................................................205</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I express my deepest gratitude to my mentor Prof. Stephen B. H. Kent for proposed line of research, guidance, fruitful discussions, endless support and encouragement. I acknowledge all other Kent Lab members for being friendly and cooperative colleagues. Especially, I thank Dr. Kalyan Mandal, with whom I collaborated in the synthesis of KVS-1 inhibitor and Dr. Joanna Wojnar for critical reading of the selected chapters of this thesis. I am grateful to Dr. H. Raghuraman for great help with EPR measurements, his friendship and continuous enthusiasm and acknowledge Prof. Eduardo Perozo for providing access to the pulsed-EPR spectrometer and his interest in the project. Special thanks are to the staff members of the National Magnetic Resonance Facility at Madison, Dr. Marco Tonelli and Dr. William M. Westler, for their assistance with various measurements and fruitful discussions. I also thank Dr. Valentina Tereshko for teaching me the basics of protein crystallography. Molecular dynamics simulations were performed in collaboration with Prof. Donald Hamelberg (Georgia State University) and Prof. Sanjib Senapati (Indian Institute of Technology, Madras), whose help I sincerely appreciate. Single-molecule fluorescence spectroscopy experiments were done in collaboration with Dr. Sua Myong and Prof. Teakjip Ha (University of Illinios at Urbana-Champaign). I acknowledge both of them for their willingness to collaborate and their help with experiments.

Finally, I acknowledge all my teachers from my undergraduate years, Prof. Rem G. Kostyanovsky (Institute of Chemical Physics, Russian Academy of Science, Moscow), Dr. Konstantin Lyssenko (Institute of Organoelement Compounds, Russian Academy of Science, Moscow), Prof. Meir Lahav and Prof. Leslie Leiserowitz
(Weizmann Institute of Science, Israel). Knowledge and experimental skills learnt under their careful mentorship helped a lot in the efficient realization of ideas when in the graduate school at the University of Chicago.
Chapter 1. Introduction – chemistry of enzyme catalysis and HIV-1 protease.

Enzymes play fundamental roles in almost all life processes. They accelerate a great variety of metabolic reactions and they control signaling, energy transduction, and transcription and translation of genetic information. Their ability to catalyze reactions by many orders of magnitude allows cells to carry out reactions that otherwise would not occur on biologically useful time scales. There is, therefore, broad interest in understanding the origin of this catalytic power on a molecular level.

Many proposals have been put forward to rationalize the catalytic power of enzymes. Proximity and orientation, induced-fit, specialized microenvironments (e.g. desolvation) for catalytic groups, strain imposed on the substrate, near attack conformations and electrostatic stabilization of the transition state in the reaction mechanism have been brought up to explain the enormous rate accelerations by more than 20 orders of magnitude in some cases compared to the rates of the uncatalyzed reactions in water. However, as discussed previously, some of these proposals are problematic or difficult to analyze quantitatively. The ultimate test for our understanding of enzyme catalysis would be the de novo design of a highly proficient enzyme, however, as will be shown below, we are not quite there yet.

De novo design of enzymes. Baker and co-workers recently published two seminal manuscripts describing successful computational designs of de novo enzymes catalyzing chemical transformations that do not exist in biological systems. Their first report focused on a ‘retro-aldol’ enzyme, catalyzing the breaking of a carbon-carbon bond in a non-natural substrate. The second report described the design of a protein
catalyst for a Kemp elimination, a model reaction for proton transfer from carbon, and also a reaction not catalyzed by naturally-occurring biocatalysts\textsuperscript{12}. The design protocol that Baker et al. used is outlined in Scheme 1.1.

\textbf{Scheme 1.1} Computational enzyme design protocol for a multistep reaction (reproduced with permission from ref.\textsuperscript{11}).

The first step in this design protocol was to choose a catalytic mechanism for the reaction to be catalyzed, and then to use quantum mechanics to perform calculations of key intermediates and transition states (TSs) in the envisioned reaction pathway, in the context of a specific catalytic motif composed of protein functional groups (\textit{theozyme}\textsuperscript{13}). These models were then superimposed to create an initial composite active site description. Using advanced search algorithms, the authors then searched from
a library of protein scaffolds for those with backbone positions capable of supporting these idealized active sites (typical number of possible realizations exceeded ~100,000). For each of these ‘matches’, a gradient-based energy minimization was used to optimize the rigid body orientation of the transition state and the torsional degrees of freedoms of the catalytic side chain to best satisfy all catalytic geometrical constrains. Subsequently, residues at neighboring positions were redesigned both to maximize the stability of the active site conformation, and the affinity for the transition state. Designs were further ranked on the basis of the catalytic geometry and the computed TS-binding energy; selected designed proteins were recombinantly expressed and tested for catalytic activity.

The strategy turned to be highly successful. In the case of the retro-aldol enzyme, out of 72 designed proteins 31 proteins (43 %) displayed the desired catalytic activity, with catalytic enhancements \(k_{cat}/k_{uncat}\) approaching \(2 \times 10^4\) for the most active proteins. In the case of Kemp elimination enzyme design, 8 proteins out of 59 (14 %) were found to be catalysts for the Kemp reaction, with \(k_{cat}/k_{uncat}\) up to \(10^5\). Further application of in vitro directed evolution to these designed protein enzyme ‘leads’ gave improved catalytic potencies, with enhancement \(k_{cat}/k_{uncat} > 10^6\). Natural enzymes are, however, much better catalysts with \(k_{cat}/k_{uncat}\) approaching \(10^{17} – 10^{20}\) for the fastest enzymes and \(~10^5\) for the slowest\(^9,14,15\). Hence, current computational methodology (even coupled with directed evolution) is capable only of producing catalytic proteins with catalytic efficiencies of the poorest biocatalysts evolved in nature.

Baker and colleagues\(^11,12\) specify weak binding of the substrate as one potential shortcoming of the many of the designs. Although the computational design methodology has the advantage of being able to explicitly place key catalytic residues, this may come
at a cost of overall substrate and TS affinity, since the design process does not explicitly model configurational entropy changes both in the substrate and the protein molecule, and also does not model the effects of dynamics of the protein molecule. Other factors contributing to the rather poor performance of the designed catalysts are complications in accurately computing free energies of buried polar interactions, and sampling problems of side-chain identity and conformational combinatorics. These last two problems are technical / computational, and can be solved by employing higher level computational facilities, by designing more advanced search algorithms, and by applying higher level atomic force-fields.

Protein dynamics. Protein dynamics is, however, the essential missing conceptual component in the design scheme used by Baker and co-workers: the principles by which intrinsic protein motions at various time-scales contribute to chemical mechanisms of enzyme catalysis only started to be unraveled\textsuperscript{16-18}. Motions that aid catalysis may be broadly divided into three categories\textsuperscript{19}: (A) protein domain hinge and shear motions that facilitate binding and release of substrate and product; (B) protein vibrational modes that are coupled to the reaction coordinate as the enzyme-substrate (ES) complex proceeds from reactant to product; and, (C) conformational flexibility of catalytic residues and coupled protein fragments, which contributes to frequencies of appearances and lifetimes of existence of preorganized geometries for particular TS, and also contributes to conformational reorganization of the protein molecule as the ES complex proceeds from one TS to another.

Hinge and shear motions of domains do not directly influence the chemical barriers of the catalyzed reaction, but are essential for substrate binding and product
release, and they can be rate-limiting. Such motions occur on the order of milliseconds to microseconds. The second category, protein vibrational modes, are the so-called co-catalytic motions that facilitate the progress of the reaction from one stationary point to another one as it proceeds from reactant complex to products within the ES complex. A subtype of motions proposed to be coupled to catalysis are the very fast, low amplitude atomic vibrations within the residues in and around the active site. Their timescales are on the order of lifetime of atomic vibrations, picoseconds to 100 femtoseconds. Conformational flexibility, the third type of motions, can reorient the enzyme active site residues along the reaction coordinate, and thus can connect either the ground-state immediately preceding a chemical transformation to the TS of that step, or connect different chemical steps that must have different transition states. In the former case, the motions of the protein molecule essentially promote bond formation/breaking and the reorganization of active site dipoles to enhance TS stabilization. Protein conformational motions can also effectively reorient protein functional groups to catalyze a chemical reaction at a different position on the substrate, but no bonds in the substrate(s) are formed or broken during the reorganization. Examples of such reorganizational motions include the compressing motion on a hydride donor–acceptor pair that brings them from the Michaelis complex to the TS for hydride transfer\textsuperscript{20}. This compressing motion has been suggested to occur in liver alcohol dehydrogenase\textsuperscript{21}, cyclophillin A\textsuperscript{22}, and dihydrofolate reductase\textsuperscript{23}. Multi-step reorganization was also suggested to play a role for triose phosphate isomerase\textsuperscript{17,24}.

Two challenges are present: how to design and carry out experiments that will enable us to decipher which protein molecular motions are playing a key role(s) in
enzyme catalysis; this is particularly important, because most inferences concerning
the role of protein dynamics in enzyme catalysis are still coming from theoretical work.
Secondly, how to implement such knowledge into novel concepts for proficient enzyme
design?

The first goal of this thesis work is to understand how protein dynamics is linked
to the chemical catalytic mechanism of the enzyme HIV-1 protease.

HIV-1 protease. The HIV-1 protease is an aspartic protease\(^25\), and the enzyme
molecule is formed from two 99-residue monomers and containing a single active site.
Each of the monomeric polypeptide chains contributes one of two catalytic aspartates at
the dimer interface\(^26\). HIV-1 protease is a key target in AIDS chemotherapy\(^27\); in HIV
infected individuals, inhibition of the enzyme by potent small-molecule drugs, leads to
prevention of AIDS. Current combination therapies for HIV-1 infection\(^28,29\) include
inhibiting the viral enzymes, protease, reverse transcriptase (as well as integrase); but the
effectiveness of these enzyme inhibitors can be diminished by the occurrence of
resistance mutations in the virus. New inhibitors of the HIV-1 protease are thus needed
which would inhibit not only wild-type but also mutated forms of the enzyme.

Three distinctly different chemical mechanisms have been proposed for catalysis
of peptide bond cleavage by aspartic proteases\(^30,31,34\). In the first one, a nucleophilic
aspartic acid side chain carboxylate attacks the carbonyl-group of the scissile peptide
bond, forming a covalent enzyme-substrate tetrahedral intermediate, followed by
expulsion of the amine component with consequent formation of a peptide-enzyme
anhydride complex which is subsequently hydrolyzed to release the other half of the
peptide substrate and regenerate the enzyme molecule (Figure 1.1)\(^30\).
The second mechanism proposed for aspartic proteases involves general acid-general base catalysis, where one catalytic aspartate side chain carboxylate (COO⁻) acts as a general base to remove a proton from the water molecule nucleophile, while another aspartic acid side chain carboxyl (COOH) general acid donates a proton to the carbonyl oxygen atom of the scissile peptide bond (Figure 1.2)₃¹-₃₃.

In the third chemical and kinetic ‘isomechanism’, a 10-membered cyclic structure is formed, involving the two catalytic aspartic acid side chain carboxylate groups (COO⁻), with a proton between their distal oxygens, and the water molecule nucleophile situated between the proximal oxygens; this mechanism allows for energy-inexpensive proton shuffling within the cyclic structure along reaction coordinate (Figure 1.3)₃⁴. The last two mechanisms also may invoke a low-barrier hydrogen bond (LBHB); in the
general acid/general base mechanism, the LBHB would stabilize the transition state\textsuperscript{35,36}, while in the kinetic isomechanism it enables hydrogen tunneling\textsuperscript{37}.

*The second goal of this thesis is to establish, which chemical mechanism is actually operating in proteolytic catalysis by HIV-1 protease.*

**Figure 1.2.** In the general acid-general base mechanism, aspartate B functions as a base to abstract a proton from water, while another aspartate A donates a hydrogen bond to the tetrahedral intermediate resulting from the addition of water to the carbonyl-group of the scissile amide bond of the peptide. Now protonated, aspartate B donates a proton to the leaving amino group of the tetrahedral intermediate (see ref.\textsuperscript{31}).
The involvement of low barrier hydrogen’s bonds (LBHBs) in enzyme catalysis is a controversial concept in biochemistry and enzymology. When originally proposed by Cleland, Kreevoy, Frey and colleagues\textsuperscript{38,39}, it provided a new paradigm for enzyme mechanisms. “A weak hydrogen bond in the ground state becomes a low-barrier hydrogen bond [very strong (10 – 20 kcal/mol), short (less than 2.5 Å)] in the transition state or in a transient intermediate. The energy released in forming the LBHB is used to help the reaction that forms it, thus lowering the activation barrier for the reaction”\textsuperscript{40}. Such rationale, for example, has provided a thermodynamic explanation for how enzymes find it so easy to abstract protons form carbons next to carboxylate groups (to form carbanion species), where the p$K_a$ of such protons may exceed 30 and p$K_a$ of a basic group in the protein is only 5.7 in the free enzyme (see ref.\textsuperscript{41}). This difference of ~24 pH units would be equivalent to ~ 30 kcal/mol and would be compensated by formation of LBHB in the transition state.
The main argument against the original LBHB concept is the proposed very high strength (up to 30 kcal/mol) of the low barrier hydrogen bond in transition state(s) within ES complex. The idea originates from the large formation energy of \([F\ldots H\ldots F]^-\) in the gas phase (~ 40 kcal/mol)\(^{42}\) and \(F\ldots F\) short bond length (2.26 Å)\(^{43}\) in the solid state. Physico-chemical parameters that have been proposed to identify LBHBs include:\(^{40}\) (A) extreme low-field \(^1\)H nuclear magnetic resonance (NMR) chemical shifts (\(\delta > 15\) ppm); (B) deuterium/tritium\(^{44}\) isotope effects on low-field \(^1\)H resonances; (C) low (<1.0) isotope fractionation factors; and (D) deuterium isotope effects on IR or Raman frequencies. Many small-molecule systems (phthalate and maleate monoanions, proton sponges, and cyclic diamines) possess the necessary characteristics and may serve as model-systems to derive the energy of such an H-bond. Hence, a linear correlation between H-bond energy and \(\Delta pK_a\) between donor and acceptor has been established, from which the energy of an H-bond possessing the above-mentioned physico-chemical properties did not exceed ~5 kcal/mol.\(^{45}\) Bachovchin et al.\(^{46}\) reinvestigated the catalytic triad of serine proteases (the enzymes studied in the original paper of Perry Frey\(^{39}\)) and found that although the Asp102...H...His57 proton exhibits anomalous spectroscopic features (such as chemical shift, \(\delta\) at 17 – 18 ppm and, in addition, low isotope fractionation factor), it is not equally shared between the side chains of the Asp and His residues, but rather is localized > 86 % on N\(^{\delta1}\) nitrogen of His (contrary to LBHB properties); he also found that the energy of the Asp...His hydrogen bond does not exceed ~5 kcal/mol. The harshest criticism of the original LBHB concept came from Warshel et al.\(^{47,48}\), who termed ‘anticatalytic’ a LBHB transiently emerging in the transition state, on the basis that formation of a LBHB leads to delocalization of charge,
whereas TS is charged; thus better stabilization of the TS would instead be achieved with localized charges\(^8\).

Until very recently, there was no direct crystallographic observation of a LBHB in a biomolecule. Kataoka et al.\(^49\) succeeded for the first time to localize the position of the hydrogen in a LBHB by studying photoactive yellow protein (PYP) with high-resolution neutron and X-ray diffraction methods. They found two short, strong hydrogen bonds (SSHBs) in the vicinity of \(p\)-coumaric acid, the chromophore of PYP. One of them, between Glu46 and phenolyc oxygen of the chromophore, at a donor-acceptor distance of 2.51 Å is a LBHB with deuterium nuclear density peak distanced by 1.21 Å and 1.37 Å from the oxygen atoms of Glu46 and phenolyc oxygen of \(p\)-coumaric acid, respectively.

The authors suggested a biophysical role for the LBHB in PYP: in the ground state, the quasi-covalent character of the LBHB stabilizes the negative charge on the chromophore, which is in hydrophobic interior of the protein, and in excited state it is disrupted and mediates fast proton transfer. Clearly, there is a lot yet to be understood in the role of LBHBs in protein structure and function.

LBHBs have been suggested to be of mechanistic importance in aspartic proteases\(^50\). In the complexes of endothiapepsin with inhibitors, rather short distances (down to 2.41 Å) were found between donor and acceptor oxygen’s on the catalytic aspartates\(^51\); protonation states were determined with the help of neutron diffraction\(^36\), and low-field resonance NMR peaks were observed, although they were left unassigned\(^50\). For HIV-1 protease, very short 2.3 Å ‘O…O’ (Asp25…Asp25’ side chain) hydrogen bond distance has been observed in an HIV-1 protease in situ products complex, suggesting the presence of a LBHB\(^52\). Calculations on the unliganded form of
HIV-1 protease concluded on 2.5 Å inter-aspartate donor-acceptor H-bond distance\textsuperscript{53}, and eventually led Northrop to propose his mechanism\textsuperscript{34}. A combined high resolution neutron and X-ray diffraction study has been performed for HIV-1 protease complex with KNI-272 inhibitor; the localized protons/deuteriums in the active site, however, did reveal any LBHBs\textsuperscript{54}.

The third goal of this thesis work is to verify whether any LBHBs are present at all in the active site of HIV-1 protease and to determine their mechanistic relevance.

**Thesis summary.** The key experimental technique for the work presented in this thesis is total chemical synthesis of proteins, which enables the atom-by-atom alteration of protein structure with surgical precision, the incorporation of residues with unnatural moieties, such as isotope labels, the control of electrostatics and, moreover, the systematic variation of conformational propensities in selected sites in both the protein backbone and side-chains\textsuperscript{55,56}.

Chapter 2 describes the elaboration of a robust and practical synthetic strategy for the preparation of a 203 amino-acid residue ‘covalent dimer’ HIV-1 protease enzyme molecule, which is currently the largest polypeptide chain assembled by chemical means. Such methodology enabled the chemical synthesis of a series of unique asymmetric chemical analogues of HIV-1 protease.

Chapter 3 aims to test the nucleophilic Asp side chain/covalent intermediate mechanistic hypothesis by synthesis and crystallographic study of a substrate-derived ‘ketomethylene isostere’ inhibitor KVS-1 complexed with the HIV-1 protease. No covalent intermediate was observed in our studies. Instead, we have established that the ketone function of the KVS-1 inhibitor is hydrated into-gem diol by HIV-1 protease.
catalytic machinery and not in solution, thus positioning KVS-1 inhibitor as a truly mechanism-based inhibitor and potential molecular scaffold to elaborate new drug leads.

**Chapter 4** presents a study of the conformational heterogeneity of the protein flaps in HIV-1 protease complexes with three different inhibitors by pulse-EPR method. We found that enzyme is more rigidified with inhibitor mimicking early transition state of the proteolytic reaction and becomes much more flexible with inhibitor mimicking later transition state inhibitor. In addition, we refined previous observations for unliganded HIV-1 protease and determined that there are indeed three principle conformers with respect to flaps (closed/closed, semiopen/semiopen and open/open) versus previously reported broad Gaussian-like distribution of conformers.

**Chapter 5** describes the systematic correlation of protein structural dynamics and catalytic function in the HIV-1 protease. Correlation of the dynamic properties of the ‘flap’ structures and the dynamics of the catalytic aspartates was demonstrated on sub-ns and µs-ms time scales by use of NMR and pulse-EPR methods. In addition, we found that pre-organization of the enzyme molecule is critical for catalysis and, moreover, that predisposed conformers are asymmetric despite C2 symmetry of the unliganded dimeric protein at the polypeptide level. In essence, our results contradict the Northrop mechanism and support the general acid – general base mechanism.

**Chapter 6** describes $^{13}$C NMR studies of wild-type HIV-1 protease and selected chemical protein analogues, as well as complexes with inhibitors. It corrects previously published data for unliganded HIV-1 protease.

**Chapter 7** and **Chapter 8** describe high resolution X-ray and NMR studies, respectively, of active site hydrogen bonds (including short, strong hydrogen bonds)
observed in HIV-1 protease complexes with inhibitors. We provided evidence for
delocalization of protons in the active site of the enzyme/inhibitor complexes, and
demonstrated how variation of structural dynamics of the flaps in various chemical
analogues of HIV-1 protease affects proton transfer.

Chapter 9 summarizes the findings of this thesis research. Conclusions and
significance of the work are discussed.

Appendix A provides with the table listing all chemically synthesized analogues of
HIV-1 protease and measurements performed with them mentioned in this thesis work.

Appendices B-H contain supplementary experimental data for chapters 2 – 8,
respectively.

Appendix I describes single-molecule fluorescence studies of proteolytic cleavage
by HIV-1 protease.

Appendix J describes the total chemical synthesis and high-resolution X-ray
structure of human lysozyme.

References:


**Chapter 2.** Convergent chemical synthesis and crystal structure of a 203 amino acid ‘covalent dimer’ HIV-1 protease enzyme molecule.

Total chemical synthesis of proteins of size larger than about 15 kDa is still a challenging task, even utilizing modern methods for the ligation of unprotected peptides. The most effective ligation chemistry is thioester-mediated amide-forming reaction at Cys residues (‘native chemical ligation’) and typically peptides are ligated sequentially in the C-to-N direction. As a consequence of handling and other losses, synthesis by sequential reaction is inefficient (even in the case of ‘one pot’ ligations) and consequently the yield of final polypeptide is low. Recent advances in convergent methodology for protein total chemical synthesis have been proposed to improve the situation.

In our recently reported ‘kinetically controlled ligation’ strategy, the peptide1-(α-thioarylester) selectively reacts with a Cys-peptide2-(α-thioalkylester) – in the absence of added thiol – to form the peptide1-peptide2-(α-thioalkylester) in high yield, because of the higher intrinsic reactivity of α-thioarylesters. This simple concept has resulted in two important implications. First, synthesis (including sequential ligation) from N-terminal segment towards C-terminal segment became possible. Second, two large polypeptides can be assembled in this way, one having a thioester moiety on the C-terminus and the other one having Cys on the N-terminus. Native chemical ligation of these two large polypeptides at the final stage of the synthesis constitutes a fully convergent approach to the total synthesis of proteins.

We are undertaking detailed studies of the enzymatic mechanism of HIV-1 protease, one of the targets in therapeutic treatment against AIDS. In its native form, the
HIV-1 protease enzyme molecule is a homodimer of two 99 amino acid residue polypeptide chains and with a single active site formed at the dimer interface. The chemical analogues we are constructing to investigate the catalytic mechanism will incorporate different functionalities in the two monomer polypeptide chains. To enable nonsymmetric incorporation of functionalities (or labels), the two 99-residue monomers have to be covalently joined via a short linker. Previous approaches to covalent linking have included recombinant expression of ~210 residue polypeptides, or have employed a synthetic strategy involving directed formation of a disulfide bond between the two chains. Although enzymes made in this way has led to insights about the catalytic mechanism, the overall synthesis was inefficient and, thus, a more robust synthetic route was required for further work. Herein we report the convergent chemical synthesis of a polypeptide chain with 203-amino acids from four peptide segments. We demonstrate the full catalytic activity of the resulting enzyme molecule and report its high resolution X-ray structure.

The first step in the convergent synthesis of the target polypeptide (Scheme 2.1) is the kinetically controlled ligation of the two peptide segments (A1-A40)-($\alpha$thioarylester) (1) and Cys-(A42-A99)-($\alpha$thioalkylester) (2). Segment 1 was obtained after transthioesterification of (A1-A40)-($\alpha$thioalkylester) with excess of 4-mercaptophenylacetic acid. Ligation was performed at pH 6.3 in order to slow down all the reactions and thus get better overall control. Two main by-products were present in the reaction mixture (Figure 2.1, Scheme 2.2). One is the branched thioester, formed by reaction of the ligation product (A1-CysA41-A99)-($\alpha$thioalkylester) (3) with 1. The second is the internal thiolactone, formed from intramolecular transthioesterification of
the ligation product. After an empirically determined optimal reaction time of 1 hour, excess of 4-mercaptophenyl acetic acid was added to a total concentration of 200 mM at pH 6.0; this leads to breakdown of the branched thioester 7, thereby releasing more of the ligation product 3 and regenerating starting peptide 1, which can further ligate with any remaining 2. Moreover, both the internal thiolactone 8 and the ligation product 3 undergo transthioesterification to form the desired ligation product 4. The sulfhydryl functionality of Cys\textsuperscript{A41} was subsequently capped with 2-bromoacetamide to form ψ-Gln\textsuperscript{A41} at the ligation site.

Scheme 2.1. Convergent synthesis of ‘covalent dimer’ HIV-1 protease. Designations ‘-S-aryl’ = 4-mercaptophenylacetic acid thioester, ‘-S-alkyl’ = 3-mercaptopropionic acid tetra-arginine amide, KCL = kinetically controlled ligation, NCL = native chemical ligation, Thz = thiazolidine. For sequence see ref.\textsuperscript{10} and for detailed experimental procedures, see Experimental Section and Appendix B.

The segment Cys-Gly\textsubscript{4}-B1-B99 was synthesized by conventional native chemical ligation\textsuperscript{2}. Two peptides Thz-Gly\textsubscript{4}-(B1-B40)- ("thioalkylester) and Cys-(B42-B99) were ligated at pH 7.0 using 4-mercaptophenylacetic acid as catalyst\textsuperscript{11}. Residue Cys\textsuperscript{B41} at the ligation site was then alkylated with 2-bromoacetamide and the ligation product treated with MeONH\textsubscript{2}·HCl to convert the N-terminal thiazolidine to Cys residue.
Figure 2.1. Analytical HPLC traces ($\lambda = 214$ nm) of kinetically controlled ligation of (A1-A40)$^{a}$COSC$_6$H$_4$CH$_2$COOH (1) and Cys-(A42-A99)$^{a}$COSCH$_2$CH$_2$Arg$_4$ (2). A) $t < 1$ min, pH 5.0; 5 indicates the product (A1-A40)-CO$_2$H arising from hydrolysis (see Scheme 2.2), 3 is the product of ligation (A1-A99)$^{a}$COSCH$_2$CH$_2$Arg$_4$. B) $t = 1$ h, pH 6.3; 7 + 2 indicates coeluting (A1-A99)-(A1-A40)$^{a}$COSCH$_2$CH$_2$Arg$_4$ and recovered (A41-A99)$^{a}$COSCH$_2$CH$_2$Arg$_4$, 6 is cyclic (Cys$^{A41}$-A99) and 8 (right shoulder) is the internal thiolactone. C) After addition of 200 mM 4-mercaptophenylacetic acid at pH 6.0, for 3 h. D) After S-alkylation with 2-bromoacetamide. E) After purification by HPLC. The asterisk indicates 4-mercaptophenylacetic acid (in A-C) or S-alkylated 4-mercaptophenylacetic acid and buffer components (in D).
Scheme 2.2. Kinetically-controlled ligation (KCL) of peptide-(α-thioarylester) $^1$ and peptide-(α-thioalkylester) $^2$. Desired ligation peptide $^3$ and by-products $^7$ and $^8$ were transformed to a single product $^4$ by treatment with 4-mercaptophenylacetic acid.

The purified segments (A1-A99)-(α-thioarylester) and Cys-Gly$_4$-(B1-B99) were then joined together by native chemical ligation to form a final polypeptide chain consisting of 203 amino acids (Figure 2.2). The Cys residue at the final ligation site was converted into ψ-Gln by treatment with 2-bromoacetamide. After removal of formyl protecting groups from tryptophans$^{12}$, the product was purified by reversed-phase HPLC (RP-HPLC; 6.7% overall yield based on limiting peptide segment). The 203-residue synthetic polypeptide was characterized by LC-MS, and analyzed by Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) (Figure 2.3). Within the limits of experimental certainty, the product had the expected mass (observed 21869.8 ± 0.4 Da; calcd 21869.8 Da, average isotope composition).
Figure 2.2. Analytical HPLC traces ($\lambda = 214$ nm) corresponding to the final step in the convergent synthesis of the HIV-1 ‘covalent dimer’ construct. A) $t < 1$ min, B) $t = 5$ h, pH 7.0. C) Product after S-alkylation with 2-bromoacetamide, removal of the formyl protecting groups, and HPLC purification. The asterisks indicate guanidine·HCl (Gn·HCl) and tris(2-carboxyethyl)phosphine (TCEP) in the case of (A) and added 50 mM 4-mercaptophenylacetic acid in (B).
Figure 2.3. FT-ICR ESI LC-MS spectrum of ‘covalent dimer’ HIV-1 protease (see Appendix B for more details).

The synthetic polypeptide was folded by two step dialysis against acetate buffer pH 5.6 (29% yield). A standard fluorogenic assay of the enzymatic activity was performed in 50 mM NaOAc, 0.2 M NaCl at pH 5.6 and 37 ºC with Abz-Thr-Ile-Nle-Phe(p-NO₂)-Gln-Argamide (Abz = 2-aminobenzoyl)\textsuperscript{13}. The $k_{\text{cat}}$ and $K_m$ values of 23.4 ± 0.2 s$^{-1}$ and 27 ± 1.4 µM, respectively (see Appendix B), are in agreement with previously reported ones for similar assay conditions\textsuperscript{14}. As a control, chemically synthesized homodimeric HIV-1 protease (that is 2 × 99 residues) was assayed under the same conditions ($k_{\text{cat}}$ 23.4 ± 0.4 s$^{-1}$, $K_m$ 25.1 ± 1.2 µM).
Figure 2.4. X-ray crystallographic structure of the ‘covalent dimer’ HIV-1 protease complexed with MVT-101 inhibitor (the linker region \(\psi\)-Gln\(^{201}\)GlyGlyGlyGly\(^{205}\) is shown in red).

X-ray structural analysis was performed to verify if the synthetic protein had the correct three-dimensional fold of the HIV-1 protease covalent dimer enzyme. Crystals grown in the presence of the inhibitor MVT-101 (Ac-Thr-Ile-Nle-\(\psi\)-(CH\(_2\)NH)-Nle-Gln-Argamide) were isomorphous to previously reported crystals of synthetic and recombinantly expressed HIV-1 proteases, and diffracted to 1.65 Å resolution. The X-ray structure of the protein molecule with 203 amino acids (Figure 2.4) was found to be essentially identical to the previously reported structures of homodimeric HIV-1 protease\(^{15}\), as well as to the structures of recombinantly expressed tethered dimer of HIV-1 protease\(^{16}\), with the linker region being partially disordered.

This 21870 Da protein with full enzymatic activity and correct three-dimensional structure is, to the best of our knowledge, the largest linear polypeptide chain prepared to date by chemical synthesis. Total synthesis of a protein of this size, in straightforward
fashion, demonstrates the great potential of recently developed methods for fully convergent chemical protein synthesis. Facile synthetic access to the 203-residue ‘covalent dimer’ HIV-1 protease will enable the preparation of a wide range of unique chemical analogues to systematically dissect the molecular basis of function of this important enzyme.

**Experimental section.**

In a kinetically-controlled ligation, \((A1-A40)\)\(^{-}\text{OC}S\text{C}_6\text{H}_4\text{CH}_2\text{COOH}\) (8.2 mg, 1.8 µmol) and \((\text{Cys}^{A41}-A99)\)\(^{-}\text{OCOSCH}_2\text{CH}_2\text{Arg}_4\) (13.5 mg, 1.9 µmol) were dissolved in aqueous buffer (1.46 mL) containing 6 M Gn·HCl, 0.2 M Na\(_2\)HPO\(_4\) and 19 mM TCEP at pH 6.3. After 1 hour 4-mercaptophenylacetic acid was added to give a total concentration of about 200 mM and the pH value was adjusted to 6.0. After 3h, 2-bromoacetamide (52 mg, 0.377 mmol) was added and the pH value adjusted to 6.7. After 15 min, 4-mercaptophenylacetic acid (51 mg, 0.304 mmol) was added to neutralize excess of 2-bromoacetamide. The product was purified by RP-HPLC with a shallow gradient of water/acetonitrile with 0.1% trifluoroacetic acid (TFA). LC-MS: found: 10956 ± 0.8 Da, calcd: 10955.9 Da (Figure 2.1). Isolated yield 5.7 mg (0.52 µmol, 29%). For the synthesis of Cys-Gly\(_4\)-(B1-B99) see Appendix B.

In the final native chemical ligation, \((A1-A99)\)\(^{-}\text{OCSC}_6\text{H}_4\text{CH}_2\text{COOH}\) (4.8 mg, 0.44 µmol) and Cys-Gly\(_4\)-(B1-B99) (5.5 mg, 0.49 µmol) were dissolved in buffer (1.6 mL) containing 8 M Gn·HCl, 0.1 M Na\(_2\)HPO\(_4\), and 25 mM TCEP. 4-mercaptophenylacetic acid was added to give a concentration of 50 mM and the pH value was adjusted to 7.0. After 12 h, the reaction mixture was deluted with buffer (1 mL), and
2-bromoacetamide (100 mg, 0.72 mmol) was added at pH 6.7. After 15 min, the reaction was quenched with an excess of 4-mercaptophenylacetic acid. Deformylation was performed by treatment with a mixture of 2-mercaptoethanol and piperidine (1:1 (v/v), 3.6 mL) on ice for 15 min, and then neutralizing with HCl. The reaction mixture was diluted twofold with buffer (6 M Gn·HCl, 0.1 M Na₂HPO₄) and purified by RP-HPLC. LC-MS: found: 21869.8 ± 0.4 Da, calcd: 21869.8 Da. Yield of isolated product 2.2 mg (0.1 µmol, 23%). The overall yield based on the limiting peptide segment is 6.7 %. For more experimental details see Appendix B.

References:


10. Sequence of tethered construct of HIV-1 protease (from N-terminus to C-terminus):

\[ \text{PQITL} \text{WKRPL}^{A10} \text{VTIRIGQQLK}^{A20} \text{EALLDTGADD}^{A30} \text{TVIEEN} \text{NleNLPG}^{A40} \psi- \text{Gln} \text{WKPKNleIGGI}^{A50} \text{GGFIKVRQYD}^{A60} \text{QIPVEI} \text{AbuGHK}^{A70} \text{AIGTVLVGPT}^{A80} \]
\[ \text{PVNIIGRNLL}^{A90} \text{TQIG} \text{AbuTLNF}^{A99} \psi-\text{Gln}^{201} \text{GGGG}^{205} \text{PQITLWKRPL}^{B10} \]
\[ \text{VTIRIGQQLK}^{B20} \text{EALLDTGADD}^{B30} \text{TVIEEN} \text{NleNLPG}^{B40} \psi-\text{Gln} \text{WKPKNleIGGI}^{B50} \]
\[ \text{GGFIKVRQYD}^{B60} \text{QIPVEI} \text{AbuGHK}^{B70} \text{AIGTVLVGPT}^{B80} \text{PVNIIGRNLL}^{B90} \]
\[ \text{TQIG} \text{AbuTLNF}^{B99}. \text{Unnatural amino acids are in italics in three-letter code. Nle} = \text{norleucine, Abu} = \alpha\text{-aminobutyric acid, } \psi\text{-Gln} = \text{pseudo-homoglutamine. Residues from} \]
N-terminal 99-residue part (part A) are specified by letter ‘A’ placed before the number of the residue, correspondingly C-terminal 99-residue part (part B) has the letter ‘B’ before the number. The five amino acid linker region is numbered from 201 to 205. Ligation sites are underlined.


12. Four tryptophans are present: $W^A_6$, $W^A_{42}$, $W^B_6$, $W^B_{42}$.


Chapter 3. Crystal structure of chemically synthesized HIV-1 protease and a ketomethylene isostere inhibitor based on the p2/NC cleavage site.

Abstract. Here we report the X-ray structures of chemically synthesized HIV-1 protease and the inactive [D25N]HIV-1 protease complexed with the ketomethylene isostere inhibitor Ac-Thr-Ile-Nleψ[CO-CH₂]Nle-Gln-Argamide at 1.4 Å and 1.8 Å resolution, respectively. In complex with the active enzyme, the keto-group was found to be converted into the hydrated gem-diol, while the structure of the complex with the inactive D25N enzyme revealed an intact keto-group. These data support the general acid–general base mechanism for HIV-1 protease catalysis.

Three distinctly different chemical mechanisms have been proposed for catalysis of peptide bond cleavage by aspartic proteases. In the first one, a nucleophilic aspartic acid side chain carboxylate attacks the carbonyl-group of the peptide bond, forming a covalent enzyme-substrate tetrahedral intermediate, followed by expulsion of the amine component\(^1\). Second, is the general acid-general base mechanism, where one catalytic aspartate side chain carboxylate (COO\(^–\)) acts as a general base to remove a proton from the water molecule nucleophile, while another aspartic acid side chain carboxyl (COOH) general acid donates a proton to the carbonyl oxygen atom of the scissile peptide bond\(^2,3\). In the third ‘kinetic isomechanism’, a 10-membered cyclic structure is formed, involving the two aspartic acid side chain carboxyl groups (COO\(^–\)), with a proton between them, and the water molecule nucleophile; this mechanism allows for energy-inexpensive proton shuffling within the cyclic structure along reaction coordinate\(^4\). The last two mechanisms also invoke a low-barrier hydrogen bond (LBHB); in the general
acid/general base mechanism the LBHB would stabilize the transition state\textsuperscript{5,6}, while in the kinetic isomechanism it allows for hydrogen tunneling\textsuperscript{4}.

X-ray structures of aspartic protease enzymes co-crystallized with inhibitors that are available in Protein Data Bank do not contain the lytic water molecule. Thus, the hypothesis that the lytic water is not initially present in the active site and that catalysis occurs via a covalent enzyme-substrate tetrahedral intermediate cannot be ruled out. A ketomethylene isostere, in which the scissile peptide bond $[C(O)\text{NH}]$ is substituted by a $\psi[C(O)\text{CH}_2]$ linkage, is the most suitable substrate surrogate to test the ‘covalent intermediate’ hypothesis. If a water molecule is initially present in the active site, then the active enzyme should catalyze the hydration of the keto group to form the gem-diol; on the other hand, direct nucleophilic attack by an ionized aspartic acid side chain would lead to a covalent adduct with the inhibitor that would be readily observed by X-ray crystallography.

Chemical synthesis of substrate-derived ketomethylene inhibitors and their tight binding to HIV-1 protease was reported previously ($IC_{50}$ down to 4.6 nM)\textsuperscript{7}. We have reproduced the chemical synthesis for the selected ketomethylene isostere ($IC_{50}$ 6.3 nM\textsuperscript{7}, mimicking p2/NC cleavage site) and co-crystallized the resulting substrate-derived inhibitor with wild-type HIV-1 protease (based on the SF2 isolate) and with the inactive [D25N]HIV-1 protease analogue.

Total chemical synthesis of the HIV-1 protease\textsuperscript{8} was based on a two segment native chemical ligation (Figure 3.1 a)\textsuperscript{9}. Both segments were prepared by \textit{in situ} neutralization Boc chemistry SPPS\textsuperscript{10}, and after ligation the Cys41 residue at the ligation site was alkylated with 2-bromoacetamide to form a $\psi$-\textit{homo}-Gln41 residue.
Figure 3.1. (a) Total chemical synthesis of wild-type HIV-1 PR and the [D25N]HIV-1 PR analogue. Native chemical ligation of two unprotected synthetic peptides was followed by alkylation with 2-bromoacetamide to convert Cys41 to ψ-homo-Gln41. After deformylation of Trp6 and Trp42 and reverse phase HPLC purification, the (1-99)-polypeptide was folded by two-step dialysis against 10 mM NaOAc, pH 5.6. MPaArg4 = 3-mercaptopropionate tetraarginineamide (i.e. the thioester leaving group). Analytical HPLC traces (λ=214 nm) and ESI-MS of purified wild-type HIV-1 protease (b) and the [D25N]HIV-1 PR analogue (c). In (d), the structure of ketomethylene isostere inhibitor is shown. After the synthesis it exists as a mixture of diastereomers (with respect to the carbon atom corresponding to the α-carbon of Nle4), which were separated by reverse phase HPLC. (e) Analytical HPLC traces (λ=214 nm) and ESI-MS of the (R)-‘Nle4’ diastereomer (KVS-1) and the (S)-‘Nle4’ diastereomer (KVS-2). See Appendix C for more information.
After removal of the formyl groups from Trp6 and Trp42, the (1-99)-polypeptide was purified by reverse-phase HPLC and folded by two-step dialysis to form fully active enzyme ($k_{\text{cat}}$ 23.4 ± 0.4 s$^{-1}$, $K_m$ 25.1 ± 1.2 µM). The inactive [D25N]HIV-1 protease analogue was synthesized according to the same strategy.

Chemical synthesis of the ketomethylene isostere Ac-Thr-Ile-Nleψ[CO-CH$_2$]Nle-Gln-Arg.amide (Figure 3.1 d) gave two diastereomers (abbreviated KVS-1 and KVS-2), differing in their stereochemistry at the tertiary carbon corresponding to the $\alpha$-carbon of the Nle4 residue. The diastereomers could be readily separated by reverse phase HPLC on C18, 10 × 250 mm column (see Figure 3.1 e; and Appendix C), and their stereochemical identity was inferred on the basis of their co-crystallization behavior with HIV-1 protease. For the KVS-1 diastereomer, crystals could be obtained within 1-2 days using standard crystallization conditions$^{11}$ developed for HIV-1 protease/inhibitor complexes.

Use of a ketone functionality in aspartic protease inhibitors was originally suggested for ketone analogues of statines (statones) to mimic the presumed tetrahedral intermediate$^{12,13}$. It was further discovered that statones are weaker binders than their hydroxy counterparts; this was attributed to the unfavorable equilibrium for hydration of the ketone function$^{14}$. The [C(O)CH$_2$] moiety was thus substituted by 2,2-difluorostatone moiety [C(O)CF$_2$] which is readily hydrated in water to give the gem-diol$^{15}$. The 2,2-difluorostatones were found to be 50-1000 fold better inhibitors than non-fluorinated analogues$^{16}$. In contrast to statones, peptide substrate-derived non-fluorinated ketomethylene isosteres are 5-10 times more potent inhibitors of HIV-1 protease than their hydroxyethyl counterparts$^7$. 
In the work reported here, we have determined high resolution X-ray structures for a substrate-derived ketomethylene isostere complexed with fully active (‘wild type’) HIV-1 protease, and for the complex of the same ketomethylene inhibitor with inactive [D25N]HIV-1 protease (Table 3.1). In the complex with inactive [D25N]HIV-1 protease, the ketone group was found to be intact indicating that the ketone exists predominantly in non-hydrated form in water solution (see Figure 3.2 a and c), which is in agreement with previous studies of the variety of carbonyl compounds\textsuperscript{17}. The complex with the wild type HIV-1 protease, however, clearly demonstrated the tetrahedral geometry for hydrated ketone group (Figure 3.2 b and d). This result unambiguously demonstrates that hydration of the ketomethylene isostere is due to the first step of the proteolytic catalytic mechanism of HIV-1 protease.

Recently, Kumar et al.\textsuperscript{18} have been able to directly view by X-ray crystallography an actual tetrahedral intermediate by soaking peptide substrate into crystal of apo-HIV-1 protease. In the crystal, the enzyme is locked in a ‘closed’ conformation and, presumably, this inflexibility prevents dissociation of the tetrahedral intermediate to release products. With an alternative approach, Kovalevsky et al.\textsuperscript{19} successfully crystallized the tetrahedral intermediate by incorporating Thr at P1’ position, which is known to greatly diminish the ability of protease to hydrolyze the peptide bond. Figure 3.3 depicts superposition of active site fragment in structure of HIV-1 protease and hydrated KVS-1 (this work) and tetrahedral intermediates for wild-type and I54V mutants, determined at 1.46 Å and 1.50 Å resolution, respectively. Although the substrate peptide sequences are not the same, the methylene unit serves as a very natural replacement for NH-moiety. A number of short H-bonds were observed in the complex of wild-type HIV-1 protease with KVS-1
inhibitor (Figure 3.2 d). However, whereas in the structures of actual tetrahedral intermediates very short, strong hydrogen bond has been observed between one of the aspartates and hydroxyl group of gem-diol moiety (2.3 Å in PDB ID:3B7V and 2.5 Å PDB ID:3B80) in structure with KVS-1 the identical H-bond is loosened to 2.7 Å (Figure 3.2 d).

Our present data do not support the ‘covalent intermediate’ mechanism of catalysis in HIV-1 protease; if direct attack had taken place one would expect for this inhibitor (KVS-1) to be covalently bound to one of the aspartates. On the other hand, the structure of the inhibitor mimics a natural tetrahedral intermediate very closely and hence may be of use in advanced studies by X-ray crystallography and neutron diffraction to elucidate the nature of short, strong hydrogen bonds in the active sites of aspartic acid proteases.
Table 3.1. Crystal parameters, data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Complex</th>
<th>[D25N]HIV-1 PR + KVS-1</th>
<th>wt HIV-1 PR + KVS-1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>23ID-D</td>
<td>23ID-B</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>0.97934</td>
<td>1.03320</td>
</tr>
<tr>
<td>Space group</td>
<td>$P\overline{2}_12_12_1$</td>
<td>$P2_12_12_1$</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$ (Å)</td>
<td>51.028</td>
<td>51.200</td>
</tr>
<tr>
<td>$b$ (Å)</td>
<td>58.551</td>
<td>58.077</td>
</tr>
<tr>
<td>$c$ (Å)</td>
<td>61.584</td>
<td>61.658</td>
</tr>
<tr>
<td>$\alpha = \beta = \gamma$ (º)</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.00 – 1.80</td>
<td>50.00 - 1.40</td>
</tr>
<tr>
<td>$R_{merge}$</td>
<td>0.089 (0.590)$^a$</td>
<td>0.084 (0.384)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>31.3 (3.9)</td>
<td>23.9 (6.3)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.8</td>
<td>6.8</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20.00-1.80</td>
<td>20.00 – 1.40</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.61</td>
<td>99.66</td>
</tr>
<tr>
<td>No. reflections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work/free set</td>
<td>16491 / 882</td>
<td>35053 / 1846</td>
</tr>
<tr>
<td>$R_{work}/R_{free}$</td>
<td>0.194 / 0.223</td>
<td>0.204 / 0.215</td>
</tr>
<tr>
<td>No. protein atoms</td>
<td>1518</td>
<td>1518</td>
</tr>
<tr>
<td>No. inhibitor atoms</td>
<td>55</td>
<td>56</td>
</tr>
<tr>
<td>No. water atoms</td>
<td>83</td>
<td>113</td>
</tr>
<tr>
<td>Average B-factor ($\text{Å}^2$)</td>
<td>24.05</td>
<td>11.995</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.017</td>
<td>0.017</td>
</tr>
<tr>
<td>Bond angles (º)</td>
<td>1.595</td>
<td>1.684</td>
</tr>
<tr>
<td><strong>PDB ID</strong></td>
<td>3DCK</td>
<td>3DCR</td>
</tr>
</tbody>
</table>

$^a$ Highest resolution is in parenthesis.
Figure 3.2. Active site region in complexes of: (a) inactive D25N chemical analogue of HIV-1 PR and intact keto-form of KVS-1 inhibitor; and, (b) wild type HIV-1 PR and hydrated ‘gem-diol’ KVS-1 inhibitor. (c) Residues Asn25 and Asn25’ hydrogen-bonded to the ketomethylene isostere in the inactive complex; and, (d) catalytic Asp25 and Asp25’ hydrogen-bonded to the gem-diol, in the wild type complex. The $|2F_o-F_c|$ electron density map for selected residues contoured at 2.0σ (in a and b) and 1.8σ (in c and d).
Figure 3.3. Superposition of active site regions of structure of wild-type HIV-1 protease complexed with KVS-1 inhibitor (gray) onto structures of tetrahedral peptide intermediates of wild-type HIV-1 protease (PDB ID: 3B7V, in cyan) and I54V mutant (PDB ID: 3B80, in magenta).

References:


8. PQITLWKRPL\textsuperscript{10}VTIRIGGQLK\textsuperscript{20}EALLXaaTGADD\textsuperscript{30}TVIEENleNLPG\textsuperscript{40}$\psi$-GlnWKPKNleGGI\textsuperscript{50}GGFIKVRQYD\textsuperscript{60}QIPVELAbuGHK\textsuperscript{70}AIGTVLVGPT\textsuperscript{80}PVNIIGRNLL\textsuperscript{90}TQIGAbuTLNF\textsuperscript{99}. Unnatural amino acids are in italics in three-letter code. Nle = norleucine, Abu = $\alpha$-aminobutyric acid, $\psi$-Gln = pseudo-homoglutamine, Xaa = either Asp or Asn. Ligation site is underlined.


11. For crystallization conditions and structure solution and refinement see Appendix C.


Chapter 4. Dynamics of ‘flap’ structures in three HIV-1 protease / inhibitor complexes probed by total chemical synthesis and pulse-EPR spectroscopy.

Abstract. The unliganded form of nitroxide spin-labeled HIV-1 protease and three different complexes with inhibitors were studied by pulse-EPR spectroscopy in order to determine ‘inter-flap’ distance distributions in solution. In the unliganded enzyme, we observed a rather broad distribution with three maxima corresponding to three flap conformers; the principal form is a ‘semi-open/semi-open’ conformer. In the complexes with inhibitors, the dominant conformer is an asymmetric ‘closed/semi-open’ form. Moreover, the distance distribution profile is significantly varied among the different inhibitors, which mimic different species on the reaction coordinate for enzyme catalyzed proteolysis.

Heterogeneity of the mobile ‘flap’ structures in HIV-1 protease (residues 37-61 in each domain of the homodimeric protein) has recently attracted much attention. Molecular dynamics simulations suggested an equilibrium between closed, semiopen, and fully open ensembles in the unliganded HIV-1 protease, with semiopen conformers being predominant; this is also in agreement with experimental NMR solution studies. This equilibrium is believed to play a role in catalysis by ‘fine-tuning’ substrate binding and product release. Also, it was shown that H-bonds donated by the amide –NH– moieties of Ile50 and Ile50’ to carbonyls of substrate or inhibitor through a well-ordered conserved water molecule (‘water 301’) are vital for the enzyme to be active. If both Ile50 and Ile50’ amides were replaced by esters, thus knocking out the possibility to form amide –NH– H-bonds, the enzyme had reduced proteolytic activity.
An experimental approach that is very suitable for elucidating conformational states of the flap structures in solution is distance measurements by pulse-EPR spectroscopy in spin-labeled constructs, and was realized for the first time by Fanucci and co-workers\textsuperscript{5}. They found a very broad distance distribution for the unliganded form of HIV-1 protease, and a much more narrow distribution for the complex with the inhibitor Ritonavir. Here, we report interflap distances determined by the Double Electron Electron Resonance (DEER) method\textsuperscript{6} for complexes of HIV-1 protease with three different peptidomimetic inhibitors. Our data also suggest a refinement of the results previously obtained for the unliganded HIV-1 protease.

Both active HIV-1 protease and its inactive [D25N] analogue were prepared in nitroxide spin-labeled form by total chemical synthesis\textsuperscript{7}. Active protease, despite having MTSSL-spin label at positions 55 and 55', retained native-like enzymatic activity suggesting that such substitution does not significantly affect the conformational equilibrium in spin-labeled analogues. The three inhibitors studied here represent three different principal states on the reaction coordinate of enzyme-catalyzed peptide bond hydrolysis (Figure 4.1 a).\textsuperscript{4} The MVT-101 inhibitor ($K_d$ 0.83 µM) is structurally similar to an ‘earlier’ transition state, whereas the KVS-1 inhibitor (IC$_{50}$ 6.3 nM) in hydrated gem-diol form is a fully isosteric, nonhydrolyzable replacement for the tetrahedral intermediate in the reaction.\textsuperscript{7} The JG-365 inhibitor ($K_d$ ~2 nM) has an additional methylene unit (–CH$_2$–) and thus mimicks a ‘later’ transition state, where the substrate is undergoing fission at gem-diol (–C(OH)$_2$–) and –NH$_2$+– junction (Figure 4.1 b).
Figure 4.1. (a) Hypothetical reaction coordinate diagram for peptide bond hydrolysis catalyzed by HIV-1 protease adopted from ref. 4. E.I = enzyme.intermediate. (b) The inhibitors studied were designed to mimick different species on the reaction coordinate.

Solutions of the spin-labeled enzymes complexed with inhibitors for DEER measurements were prepared at 40-50 µM concentration for HIV-1 protease, and inhibitors were added at 30-100–fold molar excess to ensure saturation of the enzymes. Measurements were performed at 55 K with flash-frozen solutions using 4-pulse DEER, and data acquisitions took 24-48 hours depending on spin-echo modulation intensity for different complexes to achieve the best possible S/N ratio. Distance distribution profiles were determined using an optimal Tikhonov regularization parameter when fitting the spin-echo evolution curves.
Figure 4.2. (a) Snapshots of the various conformational states of the HIV-1 protease ‘flaps’. Mean nitroxide-to-nitroxide O…O distance is depicted for each state. (b-d) DEER distance profiles extracted using DeerAnalysis2008 software\(^8\) for complexes with inhibitors and (e) for inhibitor-free HIV-1 protease. Red line = wild-type HIV-1 protease; blue line = [D25N] analogue. See Appendix D for more details.
Very sharp distance distributions were observed for the complexes with MVT-101 inhibitor, and only slightly wider distributions for the complex of active HIV-1 protease with KVS-1, indicating that the flaps are quite rigid in these two complexes. In contrast, when complexed with the JG-365 inhibitor the flaps gave a very broad distance distribution, but as in the two former complexes three well-developed maxima were observed (Figure 4.2 a-c). It should be noted that in the case of KVS-1 the results were different for the complex of active HIV-1 protease versus the inactive [D25N] chemical analogue. We recently demonstrated that active HIV-1 protease enzyme converts the keto-group of KVS-1 into a gem-diol in analogy with the first step of the proteolytic reaction, whereas with the [D25N] inactive enzyme analogue the keto-group remains intact\(^7\). Interestingly, this difference in the chemical structure of the active site region appeared to be reflected in the interflap distance profiles.

Previously, a very broad interflap distance distribution obtained by the DEER method has been reported for unliganded HIV-1 protease; the data were fitted with a single broad Gaussian\(^5\). With two spin-labels introduced at the same positions we also observed a rather broad distance distribution by the DEER method for unliganded HIV-1 protease; however, we observed three distinct maxima (Figure 4.2 d). We were able to reproduce this result three times for active HIV-1 protease, and found an essentially identical profile for the inactive [D25N] chemical analogue of the enzyme, indicating that products of autoproteolysis (up to 5-10 % by reverse-phase HPLC analysis) that appear when folding and handling samples of the active enzyme do not interfere with the measurements. Remarkably, recent full-atom explicit water MD simulations for the unliganded ‘apo’ form of HIV-1 protease reproduced our experimental results, showing
three maxima. It is possible that insufficient S/N ratio in the previous measurements hindered the authors in fitting their data for the unliganded enzyme with a profile containing three major conformers; however, this does not undermine the importance of their innovative work.

It is widely assumed that in complexes with inhibitors, the HIV-1 protease flaps adopt a ‘closed/closed’ conformation. This belief has its origin in the numerous crystal structures determined for a variety of HIV-1 protease/inhibitor complexes, in which the flaps are observed to be closed over inhibitor molecules. Our data, however, do not support such a view for the enzyme complexed with inhibitors when in solution. Using flash frozen samples that represent snapshots of the solution state, we observed three distinct conformations of the flaps. The major peak (at ~32.6 Å) does not fit well to the ‘closed/closed’ model for the two flaps, where the average distance is much shorter (~25-29 Å). Indeed, such a 25-29 Å conformer was observed in all three complexes with inhibitors, albeit sparsely populated, with the third (minor) peak at >40 Å corresponding to an ensemble of ‘open/open’ conformers. We thus interpret the major peak at ~32.6 Å in all three inhibitor complexes as an asymmetric ‘closed/semiopen’ conformational state. In the unliganded form of the HIV-1 protease, the major peak is at a distance of ~34 Å and corresponds to a symmetric ‘semi-open/semiopen’ conformational ensemble.

The intrinsic asymmetry of the two domains of HIV-1 protease complexed with the peptidomimetic inhibitor KNI-272 has been observed previously by NMR; moreover, the two domains were found to have different dynamic properties. However, these studies were aimed at the dynamics on sub-ns time scale, which is not indicative of large-scale motions such as opening–closing of the flaps. In the unliganded enzyme, large scale
flap opening–closing events occurring on the µs–ms time scale were experimentally
elucidated by NMR$^2$; such a long time scale explains why such events are rarely observed
in explicit water MD simulations. When the HIV-1 protease is complexed with inhibitors,
the large-scale dynamics should attain even slower rates, which makes it even more
difficult to study with MD methods in explicit solvent. Interestingly, however, in a recent
MD simulation of the HIV-1 protease complexed with a urea-based inhibitor (i.e. lacking
water 301), the authors have observed a dynamic ensemble of ‘closed/semiopen’
conformers, where on average only one flap is H-bonded to the inhibitor at any given
moment$^{12}$.

Overall, these results suggest that at different stages of the catalytic reaction the
flaps are likely to adopt different dynamic properties. In the ‘earlier’ transition state and
the tetrahedral intermediate, a rigid conformation with stable hydrogen-bond(s) is
preferred, while in the ‘later’ transition state the flaps become more flexible (similar to
when without inhibitor), possibly assisting in product release. A dominant
‘closed/semiopen’ conformer is in agreement with a previously suggested hypothesis and
associated experimental data, where a mechanistic role of flap-substrate interactions is to
stabilize a transition state via formation of hydrogen bonds and where interaction with the
backbone –NH– of Ile50/50’ of one flap only is needed to retain wild-type enzyme
catalytic activity$^{13}$. 
References:


Chapter 5. Protein dynamics and function are correlated in the chemical mechanism of HIV-1 protease catalysis

Abstract. Increasing evidence suggests that internal protein motions at a wide range of time scales play a decisive role in promoting protein functions such as enzyme catalysis\(^1\text{-}\text{4}. We have designed and experimentally realized dynamics-function correlations for the HIV-1 protease, an enzyme which has received considerable attention as a target for drug design for the treatment of AIDS\(^5\). Chemical protein synthesis was used to substitute, in both symmetric and asymmetric fashion, the pivotal amino acid residues Gly51/51' (Figure 5.1) at the tip of each flexible ‘flap’ structure (residues 37-61 in each monomer of the homodimeric protein molecule) with \(L\)-Ala, \(D\)-Ala, or Aib (\(\alpha\)-aminoisobutyric acid). Most of these analogues - although containing substitutions in regions distant from the catalytic aspartates - showed substantially reduced catalytic activity. Exceptionally, an asymmetric covalent dimer HIV-1 protease molecule with \(L\)-Ala51 in one flap and \(D\)-Ala51’ in the other flap retained full enzyme activity. Chemical protein synthesis was also used to prepare site-specifically \(^{15}\)N-isotope labelled enzymes for NMR dynamics experiments, and nitroxide spin-labelled analogues for pulse-EPR studies of protein dynamics. These spectroscopic studies provided strong evidence for critical differences in dynamic properties of the \(L\)-Ala51-containing flap versus the \(D\)-Ala51’-containing flap, with the latter being more rigid both on the sub-ns and \(\mu\)s-ms time scales. Using NMR relaxation-dispersion measurements we also detected coupling between the dynamics of the flaps and the catalytic aspartate residues of the HIV-1 protease. Molecular dynamics simulations suggest that the asymmetric enzyme
analogue with \textit{L-Ala/D-Ala}-containing flaps, but not the symmetric \textit{L-Ala/L-Ala} or \textit{D-Ala/D-Ala} enzymes, uniquely stabilizes the productive conformational state of catalytic residues Asp25 and Asp25’, with the nucleophilic water molecule being preorganized for catalysis\textsuperscript{6}. Our results suggest that the homodimeric viral HIV-1 protease and the single polypeptide, inherently asymmetric, cell-encoded aspartic proteases share the same asymmetric catalytic mechanism.

The dynamics of the HIV-1 protease molecule have attracted a great deal of attention. NMR studies have identified regions of the protein molecule with enhanced mobility\textsuperscript{7,8}, and attempts have been made with the help of molecular dynamics (MD) simulations to utilize knowledge of the dynamic properties of the HIV-1 protease molecule for drug design, and to rationalize drug resistant mutations\textsuperscript{9-11}. Two highly mobile regions in the HIV-1 protease are the pair of so-called ‘flaps’ (Figure 5.1 a), which close over the substrate or substrate-derived inhibitor in the crystal state. As observed in X-ray structures, the tips of the flaps adopt two different \(\beta\)-turn conformations, type I or type II, in the complex with a peptide substrate or inhibitor (Figure 5.1 b, c)\textsuperscript{12}. The differences between the two flap structures in the enzyme-substrate/inhibitor complex are most pronounced for the highly conserved residues Gly51 and Gly51’ in the middle of the \(\beta\)-turns, where one Gly residue adopts an \textit{L}-amino acid conformation (\(\phi = -102.82^\circ, \psi = -3.18^\circ\)) while the other Gly has a \textit{D}-amino acid conformation (\(\phi = 97.36^\circ, \psi = -13.22^\circ\)). In this paper, we report a series of studies characterizing how asymmetry in the flaps is related to the HIV-1 protease catalytic mechanism. Our observations further reinforce the similarity of the catalytic mechanism.
of this enzyme and the catalytic mechanism of the inherently asymmetric single polypeptide chain eukaryotic aspartic proteases\textsuperscript{13}.

**Figure 5.1.** Structural features of HIV-1 protease. \textbf{a}, Dimeric catalytically active form of HIV-1 protease (2 × 99 amino acids) complexed with the peptidomimetic reduced isostere MVT-101 inhibitor. Proteolytic machinery includes two aspartic acid residues (Asp25 and Asp25\textsuperscript{'}), and catalysis occurs by general acid-general base mechanism. \textbf{b}, The β-turn structures (residues 37-61) from each monomer, known as ‘flaps’, donate hydrogen bonds to the substrate (or inhibitor) through a structural water molecule (in green). \textbf{c}, Flaps with β-turn type I (top left) and β-turn type II (bottom right) conformations depicted separately for clarity. Residue Gly51 has D-amino acid conformation in the β-turn type I structure and L-amino acid conformation in the β-turn type II structure. \textbf{d}, Flap X-ray structures in chemically synthesized [L-Ala51; D-Ala51\textsuperscript{'}] ‘covalent dimer’ HIV-1 protease molecule. 2Fo-Fc electron density contoured in blue at 1.5σ level for the 1.6 Å resolution crystal structure (PDB ID 3FSM; see Appendix E).
Using total chemical protein synthesis, we have prepared a series of protein analogues in which residues Gly51/Gly51’ were substituted with L-Ala, D-Ala, or Aib (α-aminoisobutyric acid) in one or both of the flaps of homodimeric HIV-1 protease molecule. To enable asymmetric incorporation of different α-amino acids in the flaps, the two 99-residue monomers had to be covalently joined through a short linker (5-amino acids). The asymmetric analogues were prepared by a fully convergent total synthesis of the protein, based on a combination of native chemical ligation\(^{14}\) and kinetically-controlled ligation\(^{15}\), giving a 203-amino acid residue ‘covalent dimer’ HIV-1 protease enzyme molecule with full catalytic activity\(^{16}\).

The steady-state kinetics of proteolysis by the Gly51/Gly51’-substituted enzyme analogues were measured with the fluorogenic substrate Abz-NF6\(^*\). The results are shown in Table 5.1. In most cases we observed a significant reduction in the catalytic activity, with only the two asymmetric enzyme molecules \([L-\text{Ala}51/D-\text{Ala}51’]\) and the \([L-\text{Ala}51/Gly51’]\) having native-like catalytic activity. Overall, the observed pattern of kinetic constants strongly supported distinct roles for the two flaps in HIV-1 protease catalysis of peptide bond hydrolysis. High resolution X-ray structures with peptidomimetic inhibitors MVT-101 and JG-365 were obtained for the \([L-\text{Ala}51/D-\text{Ala}51’]\) chemical analogue; analysis of crystallographic temperature \(B\)-factors in this analogue protein suggested different flexibility of the two flaps (see Appendix E).

In order to characterize the dynamic properties of these unique enzyme analogues, we then performed NMR studies using site-specifically \(^{15}\text{N}\)-labelled protein molecules. Key residues in the flaps were site-specifically \(^{15}\text{N}\)-labelled; in addition, Gly40 in the
Table 5.1 Steady-state kinetics of proteolysis by chemically synthesized HIV-1 protease and Gly51 analogues.

<table>
<thead>
<tr>
<th>Flap</th>
<th>Flap'</th>
<th>$k_{cat}$ (s$^{-1}$); mean ± s.d.</th>
<th>$K_m$ (µM); mean ± s.d.</th>
<th>$k_{cat}/K_m$; (s$^{-1}$µM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>Gly51</td>
<td>23.4 ± 0.4</td>
<td>25.1 ± 1.2</td>
<td>0.93</td>
</tr>
<tr>
<td>2†</td>
<td>L-Ala51</td>
<td>17.6 ± 0.3</td>
<td>26.1 ± 1.4</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>L-Ala51</td>
<td>22.2 ± 0.6</td>
<td>47.4 ± 2.9</td>
<td>0.47</td>
</tr>
<tr>
<td>4</td>
<td>Gly51</td>
<td>6.2 ± 0.2</td>
<td>43.8 ± 2.8</td>
<td>0.14</td>
</tr>
<tr>
<td>5</td>
<td>L-Ala51</td>
<td>3.7 ± 0.1</td>
<td>50.1 ± 2.6</td>
<td>0.073</td>
</tr>
<tr>
<td>6</td>
<td>Gly51</td>
<td>4.04</td>
<td>104.6</td>
<td>0.042</td>
</tr>
<tr>
<td>7</td>
<td>D-Ala51</td>
<td>4.9</td>
<td>434.2</td>
<td>0.011</td>
</tr>
<tr>
<td>8</td>
<td>Aib51</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*The ‘wild-type’ enzyme 1 was prepared both as a homodimer and as a ‘covalent dimer’, and the catalytic properties of both were found to be identical within experimental uncertainty.

†In a control experiment, the 99-residue L-Ala51 HIV-1 protease polypeptide was folded by dialysis in the presence of an equimolar amount of the 99-residue D-Ala51 HIV-1 protease polypeptide. We observed $k_{cat}$ 9.1 s$^{-1}$ (one half of the value for analogue 2) and $K_m$ 25.6 µM, which corresponded to the properties of the statistical 0.5-molar ratio of (L-Ala51, D-Ala51) heterodimer enzyme in the mixture precluding that either dimerisation equilibrium or the 5-amino acids interdomain linker inserted in covalent dimers affected the observed enzyme kinetics.
‘elbow’ regions (see Figure 5.1 a), and residues Asp25 and Gly27 in the catalytic site were labelled (see Figure 5.1 a, b and Appendix E). Measurements were performed with symmetrically substituted chemical analogues (homodimers) in order to simplify interpretation (Figure 5.2). Order parameters \( S^2 \) for the protein backbone amides were derived from model-free analysis of measured \( R_1 \) (spin-lattice relaxation rate), \( R_2 \) (spin-spin relaxation rate), and heteronuclear \(^1\)H-\(^{15}\)N NOE (nuclear Overhauser effect) values\(^{17}\). The order parameter \( S^2 \) represents the degree of spatial restriction of internal fluctuation of the amide bond on the sub–ns timescale, ranging from 0 (completely unrestricted motions) to 1 (completely rigid). \( S^2 \) values measured for unliganded enzymes were found to be significantly higher for the \( D \)-Ala51-containing flap (‘D-flap’) and Aib51-containing flap (‘Aib-flap’) than for the \( L \)-Ala51-containing flap (‘L-flap’) (Figure 5.2a, inset). The higher order parameters for the tips of the flaps (residues 48-52) of ‘D-flap’ and ‘Aib-flap’ enzyme molecules indicate greater rigidity on the sub-ns time scale. Remarkably, the ‘\( L \)-flap’ had \( S^2 \) values and hence sub-ns flexibility comparable to those of the corresponding wild-type Gly51-containing \( \beta \)-turn structure\(^{18}\); this is consistent with our observation that of the Gly51-symmetrically substituted enzyme analogues, the \( L \)-Ala51/51’ homodimer had proteolytic activity closest to that of the wild-type enzyme (Table 5.1; entries 5 and 1, respectively).

In order to measure how substitutions at the position Gly51 affect the large-scale motions of the flaps, we incorporated nitroxide spin labels at positions 55/55’ and measured the nitroxide-to-nitroxide distance distribution profiles for symmetric analogues using the pulse-EPR double electron-electron resonance (DEER) method (Figure 5.2 c, d)\(^{19}\). The experiments were performed with ~ 45 µM solutions of
unliganded enzymes, flash-frozen to 55 K, by recording the dipolar spin-echo evolution responses and converting the obtained data into distance ‘snapshots’ of all conformers present in a given sample at the moment of flash-freezing\textsuperscript{20}. We found significant differences in these measurements for the different chemical analogues, especially for the \textit{D}-Ala51/51' symmetric analogue, where the distribution of nitroxide-to-nitroxide distances became much narrower, with the so-called ‘semi-open’ conformer being more populated here than in all other analogues\textsuperscript{20}. Interestingly, this observed predominance of the semi-open conformer for the \textit{D}-Ala51/51’ symmetric analogue is in agreement with the conformations observed for the unliganded form of HIV-1 protease in the crystal structure, where both Gly51 and Gly51’ have a \textit{D}-amino acid backbone conformation and the flaps adopt a semi-open conformation and are symmetrically related via a two-fold axis\textsuperscript{21}.

To elucidate the dynamics of these enzyme analogues in the $\mu$s-ms time regime, $^{15}$N Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion measurements were performed (Figure 5.2 b and Appendix E)\textsuperscript{22}. In the three symmetric analogues containing \textit{L}-flaps’, \textit{D}-flaps’, or \textit{Aib-flaps’}, we observed a systematic decrease in mobility on the $\mu$s-ms time scale; the previously observed higher sub-ns mobility of the \textit{L}-flap’ enzyme analogue (above) was also reflected in faster $\mu$s-ms-regime dynamics, and a greater sub-ns rigidity corresponding to slower mobility on the $\mu$s-ms time scale was observed for the \textit{D}-flap’ and \textit{Aib-flap’} enzyme analogues. In case of the \textit{L}-flap’ and \textit{D}-flap’ enzyme analogues, all labelled residues showed coherently matched chemical exchange rates in these NMR measurements; after fitting the data on a residue-by-residue basis, they were further fitted globally with a three-site exchange model,\textsuperscript{23} yielding $k_{\text{ex}}$ (fast) $2,350 \pm 100$
s^{-1} (mean ± s.d.), \(k_{ex}\) (slow) 38 ± 2 s^{-1} (mean ± s.d.) for the ‘L-flap’ enzyme, and \(k_{ex}\) (fast) 1,380 ± 60 s^{-1} (mean ± s.d.), \(k_{ex}\) (slow) 80 ± 5 s^{-1} (mean ± s.d.) for the ‘D-flap’ enzyme (see Figures 7 and 8, Table 6 and Scheme 2 in Appendix E). In the ‘Aib-flap’ enzyme analogue, however, it was impossible to perform global fitting because of substantial non-coherent variations in exchange rates for different residues, indicating either greater complexity or a shift in the principal isomerisation rate constant for the flaps as global structures. In fact, several residues lost their \(R_2\)-dispersion in the ‘Aib-flap’ enzyme analogue (see Figures E.7 and E.8 in Appendix E).

The most remarkable differences observed for this series of flap analogues of the HIV-1 protease involved the active site residues. Our NMR measurements on flap analogues showed that the catalytic residues Asp25/25’ as well as the nearby Gly27/27’ residues were correlated in their μs-ms dynamics with the dynamics of the flaps. For the ‘L-flap’-containing enzyme analogue, both Gly27 and Asp25 in the catalytic region have μs-ms chemical exchange constants (\(k_{ex}\) 2,325 ± 200 s^{-1}, mean ± s.d.) similar to the fast exchange constants observed for the flap region; for the ‘D-flap’ enzyme analogue, the slower μs-ms dynamics observed in the flap were found to be correlated to the reduced μs-ms mobility of the Asp25 and Gly27 residues (\(k_{ex}\) 420 ± 70 s^{-1}, mean ± s.d.). And, for the ‘Aib-flap’-containing enzyme both the catalytic Asp25 and residue Gly27 were found to be lacking \(R_2\)-dispersion as was the case for some residues in the flaps in this enzyme analogue (Figure 5.2 b, bottom and Figure E.8 in Appendix E), meaning that the time scale window probed with our experimental approach was not suitable for this particular
Figure 5.2. Dynamic properties of chemically synthesized Gly51 analogues of HIV-1 protease. a, 15N-HSQC spectra overlaid for L-Ala51, D-Ala51 and Aib51 homodimers in blue, red and black, respectively. Peaks for the wild-type HIV-1 protease (in green) for corresponding residues were reconstructed from a previous study. Order parameters $S^2$, obtained by model-free analysis of $R_1$, $R_2$ and $^1$H-15N NOE values, versus residue number are depicted as an inset. b, CPMG 15N relaxation dispersion data for catalytic residue Asp25 in three studied HIV-1 protease symmetric homodimers obtained at two magnetic fields (600 MHz and 900 MHz). Remarkably, the chemical exchange rates for the three enzymes are drastically different and are correlated with the dynamic properties in flaps, and with the catalytic rates of proteolysis (Table 5.1; entries 5, 7 and 8). c, ‘Snapshots’ of the conformational states (‘open’, ‘semi-open’ and ‘closed’ overlaid in d) of the HIV-1 protease chemical analogues labelled with nitroxide-spin label, measured at 55 K using pulse-EPR spectroscopy. In the case of the D-Ala51 homodimer, the ‘semi-open’ conformer is overpopulated.
chemical analogue. Thus, the dynamics of the active site residues were correlated with flap dynamics. This suggests an important role for the flaps in the chemical mechanism of catalysis.

In silico experiments were performed to get a better picture of how the dynamics of the flap structures correlate with the dynamics of the catalytic residues. All-atom explicit water molecular dynamics (MD) simulations of symmetric ‘L-flap’- and ‘D-flap’-containing enzyme analogues gave results in agreement with inter-flap distance distributions observed by pulse EPR measurements (see Figure E.10 in Appendix E). In the active site of these two enzymes MD simulations showed a symmetric cyclic structure consisting of two catalytic Asp25 and Asp25’ residues and the nucleophilic water molecule as the most predominant conformer (Figure 5.3 b). Since proton transfer between the two catalytic aspartates Asp25 and Asp25’ cannot be treated adequately by the classical molecular mechanics approach employed in the MD simulation, in the case of asymmetric [L-Ala51; D-Ala51’] heterodimer we performed calculations on two boundary states, with either the ‘L-flap’-containing domain Asp25 or the ‘D-flap’-containing domain Asp25’ ionized and the other aspartate in the same molecule protonated (note: the monoprotonated state for the two catalytic aspartates is invoked for general acid-general base catalysis\(^2\)). In the case of the ‘L-flap’-containing domain Asp25 having a charged carboxylate and ‘D-flap’-containing domain Asp25’ having a protonated side chain, we observed the same cyclic hydrogen-bonded structure that we found as the most populous state in symmetric homodimers. Strikingly, if we reversed protonation states, i.e. with the ‘L-flap’-containing domain Asp25 now being protonated and the ‘D-flap’-containing domain Asp25’ being charged, MD simulations showed an
enhanced population (~100-fold) of the structure composed of Asp25, Asp25’ residues and the nucleophilic water molecule hydrogen-bonded asymmetrically and with a geometry preorganized for catalysis (Figure 5.3 a)\textsuperscript{25}.

The computational prediction of a greater concentration of conformers preorganized for catalysis being present in the unliganded [L-Ala51, D-Ala51’] heterodimeric enzyme molecule was supported by surface plasmon resonance (SPR) binding experiments (see Appnedix E and Figure E.11 in Appendix E). In the SPR measurements on the HIV-1 protease flap analogues, using a reduced isostere inhibitor to mimic the earlier transition state of the enzyme-catalyzed proteolysis, the [L-Ala51; D-Ala51’] asymmetric enzyme analogue showed tighter binding affinity towards inhibitor than any other studied enzyme analogues, including the wild-type HIV-1 protease (K\textsubscript{d} value of 100 nM, 4-times lower than the K\textsubscript{d} value of 420 nM observed for the wild-type enzyme).

The accepted chemical mechanism for the aspartic proteases involves general acid-general base catalysis, where one catalytic aspartate side chain carboxylate (COO\textsuperscript{−}) acts as a general base to remove a proton from the water molecule nucleophile, while another aspartic acid side chain carboxyl (COOH) general acid donates a proton to the carbonyl oxygen atom of the scissile peptide bond\textsuperscript{24}. Based on the theoretical work of Piana and Carloni\textsuperscript{26}, Northrop has proposed a distinctly different chemical and kinetic ‘isomechanism’, where a 10-membered cyclic structure is formed, involving the two aspartic acid side chain carboxyl groups (COO\textsuperscript{−}), with a proton between them, and the water molecule nucleophile. This mechanism invoked low-barrier hydrogen bond(s), which would allow for energy-inexpensive proton shuffling within the cyclic structure
along the reaction coordinate\textsuperscript{27}. Our results do not support Northrop’s mechanism with a symmetric low-barrier hydrogen bond between the Asp25 and Asp25’ side chains, but rather are in agreement with the general acid-general base mechanism, with a nucleophilic water molecule preorganized for catalysis in the asymmetric environment of the two aspartates, one being protonated and one being ionized (Figure 5.3 c). Thus, our data suggest that in the case of the HIV-1 protease, the so-called ‘L-domain’ (which contains the ‘L-flap’) is more mobile and bears the general acid, while the less mobile ‘D-domain’ (containing the ‘D-flap’) is more rigid and bears the general base in the active site.

We recently reported that flap structures in HIV-1 protease – complexed with three different inhibitors which mimick respectively an early transition state, the tetrahedral intermediate, and a late transition state – display significantly different equilibrium populations of conformers; the flaps are least mobile at the earlier transition state of the reaction, and attaining more flexibility in the course of the reaction en route to product release\textsuperscript{20}. The results obtained from the combination of experiments used here show that the glycine residue at position 51 in each monomer serves as a surrogate for both the \textit{L}- and \textit{D}-amino acids required at that position in each domain of the homodimeric HIV-1 protease. The data suggest that catalysis is not rate-limited by opening and closing events of flaps, but rather by emergence of catalytically-preorganized asymmetric β-turn type I / β-turn type II conformers in the Gly51/51’-containing wild-type HIV-1 protease. If substrate binding / product release were rate-limiting steps, one would expect the \([L-\text{Ala51, } D-\text{Ala51’}]\) covalent dimer enzyme to posses catalytic efficiency lower than the more flexible \([L-\text{Ala51, } L-\text{Ala51’}]\) homodimer and greater than more rigid \([D-\text{Ala51, } D-\text{Ala51’}]\)
Ala51] homodimer. However, the \([L\text{-Ala}51, D\text{-Ala}51']\) heterodimer enzyme is \(~10\) -times and \(~100\)-times, respectively, more efficient as a catalyst than the above mentioned two homodimer enzymes (see Table 5.1).

**Figure 5.3.** Active site structures revealed by MD simulations, and a scheme for the mechanism of HIV-1 protease catalysis taking into account the dynamics of the flaps. **a,** An asymmetric structure of catalytic Asp25 and Asp25’ and the nucleophilic water molecule dominated in the MD trajectory for the asymmetric \([L\text{-Ala}51; D\text{-Ala}51']\) chemical analogue of HIV-1 protease with ‘L-domain’ have a protonated Asp25 and ‘D-domain’ having a charged Asp25’ side chain. **b,** Symmetric structure for catalytic residues and nucleophilic water required in Northrop’s mechanism (ref. 26,27) was observed as the most populous state in MD simulation for the symmetric \([L\text{-Ala}51; L\text{-Ala}51']\) and \([D\text{-Ala}51; D\text{-Ala}51']\) chemical analogues. **c,** Scheme for the mechanism of HIV-1 protease catalysis. Asymmetric conformers are preorganized for catalysis. Conformational isomerisations are depicted by blue arrows and electron rearrangements by red arrows. TS1 and TS2 are earlier and later transition states, respectively, and E.I is enzyme complex with tetrahedral intermediate.
These results provide strong experimental support for a chemical mechanism in which the contribution of the flaps to catalysis is critical. First, flap $\beta$-turn isomerisation attenuates the dynamics of the whole protein molecule, resulting in correlation of the dynamics of flaps and the catalytic residues; then hydrogen-bonding interactions of HN-Ile50 and HN-Ile50’ at the tips of the flaps to the carbonyls of the substrate on either side of the scissile bond are mediated by a (non-nucleophilic) water molecule lock substrate in productive catalytic conformation$^{28}$. In contrast to viral aspartyl proteases, the cell-encoded aspartyl protease enzyme molecules have a single polypeptide chain that folds to form two-domains, and a lone flap that forms direct hydrogen bonds to the substrate or inhibitor$^{13}$. We predict that the dynamic properties of the two domains in eukaryotic aspartic protease molecules will be distinct from one another and reminiscent of those reported here for the artificial $[L-\text{Ala}51; D-\text{Ala}51']$ ‘covalent dimer’ HIV-1 protease prepared by total chemical synthesis.

**Experimental section.**

**Sample preparation.** The 99-residue polypeptide chains of symmetric homodimers and the 203-residue polypeptides for asymmetric heterodimers were prepared by total chemical synthesis, folded by dialysis and assayed as described$^{16,30}$. NMR samples were 0.27 – 0.35 mM protein in H$_2$O/D$_2$O (95%/5%), 20 mM sodium phosphate, 100 $\mu$M DSS-d$_6$, pH 5.7. Samples for pulse-EPR measurements were prepared in 10 mM NaOAc, 10% glycerol, pH 5.7 at concentrations 40 – 50 $\mu$M.

**Peptide and protein synthesis.** Peptides were prepared manually by ‘in situ neutralization’ Boc chemistry stepwise solid phase peptide synthesis$^{29}$ and the unprotected peptide
products purified by reverse phase HPLC. Native chemical ligation was used to covalently condense two synthetic peptide segments to prepare 99-residue monomers of HIV-1 protease\textsuperscript{30}. A combination of native chemical ligation and kinetically controlled ligation was used to synthesize asymmetric 203-residue ‘covalent dimers’ of HIV-1 protease\textsuperscript{16}. Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR MS) analysis was used to characterize product synthetic proteins (see Figures E.1-E.4 in Appendix E).

**Folding.** Typical procedure: 10 mg of an HIV-1 protease synthetic polypeptide chain was dissolved in 10 mL of 6 M Gln·HCl, 50 mM NaOAc, pH 6.0 buffer and dialyzed first against 2 M Gln·HCl, 50 mM NaOAc, pH 5.6 and finally against 10 mM NaOAc, pH 5.6. Yields for homodimers were in the 45-60\% range, whereas for covalent dimers yields of 25-35\% were obtained based on final protein concentration determined by absorbance at 280 nm using a calculated extinction coefficient of 25120 M\(^{-1}\)cm\(^{-1}\). After folding, protein solutions were concentrated using Millipore Centricon Centrifugal Devices (3000 MW cutoff).

**Enzyme kinetics.** Steady-state proteolysis kinetics were assayed with the fluorogenic substrate Abz-NF6* [Abz-Thr-Ile-Nle-Phe(p-NO\(_2\))-Gln-Arg.amide] (Abz = 2-aminobenzoyl)\textsuperscript{31}, which was added at 7-10 different concentrations (each particular substrate concentration in triplicate). A Jobin Yvon FluoroMax-3 spectrofluorimeter (cell 1×1 cm) was used to monitor changes in fluorescence intensity. Assays were performed in 50 mM NaOAc, 0.2 M NaCl buffer with 1 \% (v/v) DMSO at pH 5.6 and 37 °C. Enzymes were added at 20 – 40 nM concentration if were measuring rather fast kinetics (Table 5.1; entries 1-5). The values of \(k_{\text{cat}}\) and \(K_m\) were determined by nonlinear least
square fitting of the initial rates to the Michaelis-Menten kinetic model. For slower enzymes, such as \([D-Ala51;D-Ala51']\) homodimer and \([Gly51;Aib51']\) covalent dimer, two substrate concentrations were chosen (triplicate measurements were done for each) and \(k_{\text{cat}}\) and \(K_m\) were determined from Lineweaver-Burk linear fits. For \([Aib51;Aib51']\) homodimer, which was found to be a very slow enzyme, the \(k_{\text{cat}}/K_m\) ratio was estimated from a cleavage experiment performed at \(c(\text{substrate})\) 20 µM, \(c(\text{enzyme})\) 1 µM.

Assuming first-order reaction conditions, \(k_{\text{cat}}/K_m = k_{\text{obs}}/E_{\text{total}}\), where \(k_{\text{obs}}\) equals to negative of the slope in the \(\ln(\text{substrate})\) versus time plot.

NMR experiments. Assignment of resonances in three studied enzymes was based on the previous work for the wild-type unliganded HIV-1 protease\(^{18}\). The HNCA experiment\(^{32}\) (to isolate residues having \(^{13}\)C-labels and assign Asp\(^{25}\) residue with its unique \(^{13}\)C \(\alpha\)CH chemical shift at 52.6 ppm as opposite to glycines with \(\alpha\)CH at ~ 45 ppm – note that only Asp\(^{25}\) and glycines are labelled with \(^{13}\)C) as well as 3D gradient enhanced \(^{15}\)N-NOESY-HSQC\(^{33}\) and \(^{15}\)N-TOCSY-HSQC\(^{34}\) experiments were performed to unambiguously confirm the assignments.

All relaxation measurements were performed at 25 °C on Varian Inova 600 spectrometer equipped with a room-temperature triple-resonance probe. Backbone amide \(R_1\) and \(R_2\) values were obtained by \(^1\)H-detected longitudinal and transverse relaxation experiments,\(^{35}\) respectively. \(^{15}\)N \(R_1\)’s were measured using relaxation delays of 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600 and 2000 ms, and \(^{15}\)N \(R_2\)’s 10, 30, 50, 70, 90, 110, 130, 150, 170, 190 ms, respectively. To determine experimental error triplicate measurements were done for at least three points for each experiment and sample. The \(^1\)H-detected \(^1\)H-\(^{15}\)N heteronuclear NOE experiments were performed using a water flip-
back sequence at least twice for each sample and resulting values were averaged.\textsuperscript{35} NOE values were measured by taking the ratio of peak intensities from experiments performed with (NOE experiment) and without (NONOE experiment) \textsuperscript{1}H presaturation prior to the evolution of \textsuperscript{15}N magnetization. In the case of NONOE spectra, a net relaxation delay of 5.5 s was employed, while a relaxation of 3 s prior to a 2.5 s proton presaturation period was used for NOE spectra. NOE values were corrected for incomplete \textsuperscript{1}H and \textsuperscript{15}N magnetization recovery using the procedure elaborated by Freedberg et al.\textsuperscript{18} for the wild-type HIV-1 protease.

\textsuperscript{1}H-detected \textsuperscript{15}N constant-time, relaxation-compensated CPMG dispersion experiments\textsuperscript{36,37} were carried out at 25 °C on Varian Inova 600 and Varian Inova 900 spectrometers equipped with room-temperature triple-resonance probes. The gNcpmgex_NH sequence from the Varian Biopack suite of pulse sequences was employed. For the \textsuperscript{15}N CPMG pulse trains, the \textsuperscript{15}N π/2-pulse widths were 49 μs (57 dB) on Varian Inova 600 and 48 μs (59 dB) Varian Inova 900 spectrometers, respectively, and the constant relaxation time was 50 ms. Amide spectra were acquired, applying 10-15 CPMG field strengths ranging from 50 to 1,000 Hz. Per residue $R_{2,\text{eff}}$ values were calculated according to equation: $R_{2,\text{eff}}(\nu_{\text{CPMG}}) = -1/T \cdot \ln(I(\nu_{\text{CPMG}})/I_0)$. Uncertainties were determined from duplicate measurements. $R_{2,\text{eff}}$ dispersion profiles were fitted either to the 2-site or 3-site chemical exchange equations using software generously provided by Dmitry M. Korzhnev and Lewis E. Kay (University of Toronto).\textsuperscript{23}

\textsuperscript{15}N $R_1$, \textsuperscript{15}N $R_2$ and \textsuperscript{1}H-\textsuperscript{15}N NOE values were used to derive generalized order parameters using the Lipari-Szabo model-free approach\textsuperscript{38,39}. In a previous study\textsuperscript{18}, researchers have performed analysis for the wild-type protein uniformly labelled with \textsuperscript{15}N
isotope and derived components for an anisotropic axially-symmetric diffusion tensor at 25 °C (using structure of unliganded HIV-1 protease PDB ID 3PHV), where τ_{iso} is 12.1 ns, \(D_∥/D_⊥\) is 1.27 and \((\theta, \phi)\) are (13.4, 207.6), respectively. We analyzed our data for labelled regions of the protein assuming the anisotropic model using previously and determined diffusion tensor parameters with the help of Modelfree 4.15 software.\(^{40,41}\)

**Pulse-EPR measurements.** Double electron-electron resonance (DEER) experiments were carried out using a Bruker Elexys E580 X-band pulsed spectrometer operating near 9.7 GHz equipped with a split-ring MS2 resonator at a temperature of 55 K. Samples were transferred to quartz 1.1×1.6×100 mm capillaries and flash-frozen to 55 K in the resonator. All the measurements were performed using a constant-time version of the four-pulse DEER sequence \(\pi/2(v_{obs}) – \tau_1 – \pi(v_{obs}) – t’ – \pi(v_{pump}) – (\tau_1 + \tau_2 – t’) – \pi(v_{obs}) – \tau_2 – \text{echo,}^{19}\) where time \(t’\) is varied. The resonator was overcoupled to \(Q \approx 100 – 200\), the pump frequency \((v_{pump})\) was set to the center of the resonator dip and coincided with the maximum of the nitroxide EPR spectrum, while the observer frequency \((v_{obs})\) was set 65–70 MHz higher and coincided with the low-field local maximum of the spectrum. The pulse lengths for \(\pi/2\) and \(\pi\) were 16 ns and 32 ns, respectively, and pump pulse length was 32 – 40 ns. The pump pulse length was optimized using nutation experiment for all cases. In all experiments, a \(\tau_1\) of 200 ns and \(\tau_2\) of 1800 ns were used. Data were recorded at steps of 8 ns and were processed and analyzed using the program DeerAnalysis2008.\(^{42}\)

Homogeneous 3D background model was used to subtract intermolecular background from raw data. The criterion for choosing background subtraction was the frequency-domain spectrum, where neither a positive spike nor an obvious hole had to be present in the center of the Pake pattern. Computation of the \(L\)-curve was performed to choose the
optimal value of the Tikhonov regularization parameter, which corresponded to the corner of the \( L \)-curve (see Figure E.9 in Appendix E). Validation of the results was performed using validation tools available in the DeerAnalysis2008.

**Molecular dynamics simulations.** All-atom molecular dynamics simulations were carried out using the modified version\(^4\) of the Cornell et al.\(^4\) all-atom force field for proteins using the sander module of the AMBER 9 suite of programs.\(^4\) Molecular dynamics simulations were carried on the symmetric ‘\( L \)-Ala51 flaps’ and ‘\( D \)-Ala51 flaps’ of the HIV-1 protease and the asymmetric [‘\( L \)-Ala 51 flap’ + ‘\( D \)-Ala51 flap’] protease, each for 300 ns. For the asymmetric [\( L \)-Ala51; \( D \)-Ala51’] heterodimer, we carried out two separate simulations with the protonated Asp25 residue on either the ‘\( L \)-flap’-containing domain or the ‘\( D \)-flap’-containing domain and with a charged Asp 25’ on the other flap in the same enzyme molecule. The systems were solvated with approximately 8000 TIP3P water molecules in a truncated octahedron periodic box. Five \( \text{Cl}^- \) counterions were added to attain electrostatic neutrality. The electrostatic interaction was treated using the particle mesh Ewald (PME) method. The SHAKE algorithm was applied to all bonds involving hydrogen atoms, and an integration time step of \( \Delta t = 2.0 \) fs was used for integrating Newton’s equation of motion. Snapshots were collected and saved every 5 ps for further analyses.

**SPR assays.** Surface plasmon resonance measurements were performed on a Biacore 3000 instrument. Biotinylated substrate-derived inhibitor was immobilized onto streptavidin-coated chip and each enzyme was then flowed through the microfluidic cell.
at multiple concentrations at a flow rate of 20 \( \mu \text{L/min} \), with association and
dissociation kinetics studied for 120 seconds.

References:
2. Garcia-Viloca, M, Gao, J., Karplus, M., Truhlar, D. G. How enzymes work: analysis
3. Henzler-Wildman, K., Lei, M., Thai, V., Kerns, S. J., Karplus, M., Kern, D. A
4. Smith, A. J. T., Müller, R., Toscano, M. D., Kast, P., Hellinga, H. W., Hilvert, D.,
Houk, K. N. Structural reorganization and preorganization in enzyme active sites:
comparisons of experimental and theoretically ideal active site geometries in the


Chapter 6. The true ionization state of the catalytic dyad Asp25/25’ in unliganded HIV-1 protease: $^{13}$C NMR studies of HIV-1 protease and several chemical analogues, unliganded and complexed with inhibitors.

Direct experimental determination of the ionization states of the side chain carboxyl’s of catalytic residues Asp25 and Asp25’ in HIV-1 protease is of particular importance for understanding the catalytic mechanism and for designing new inhibitors. All potent inhibitors interact with the catalytic carboxyls of Asp25/25’1. X-ray crystallography is not a suitable technique for determining the ionization states of these side chains because of the very weak electron density for hydrogen atoms. Neutron diffraction crystallography on the other hand provides unambiguous solution for the location of the hydrogens for a macromolecule in the crystal state2. However, the methodology is strongly limited due to difficulties with obtaining rather large single crystals (at least $1\times1\times1$ mm$^3$ required in neutron diffraction measurements). An alternative method is the NMR study of isotope-labelled proteins.

NMR measurements have previously been performed for complexes of HIV-1 protease with the C2-symmetric DMP-323 inhibitor$^3$ and with the asymmetric KNI-272 inhibitor$^4$. In the case of DMP-323, it was found that both Asp25 and Asp25’ side chains are protonated over the pH range extending from 2.2 to 7.0 ($^{13}$C $\delta$, 176.4 ppm), whereas for KNI-272 one aspartate side chain Asp25 is charged ($^{13}$C $\delta$, 177.4 ppm) while the other side chain carboxyl Asp25’ is protonated ($^{13}$C $\delta$, 175.8 ppm, deuterium isotope effect 0.17 ppm). Interestingly, the $^{13}$C-chemical shifts with the inhibitor KNI-272 were found
to be essentially independent of pH in the range 2.5 – 7.0, indicating extreme pK_a
values for the side chain carboxyl’s of Asp25 and Asp25’ in this complex with KNI-272.

In a contemporaneous report, researchers used total chemical synthesis to prepare
site-specifically \(^{13}\text{C}\)\(^\gamma\)-Asp25/25’-labelled enzyme and used \(^{13}\text{C}\)-NMR measurements to
establish the ionization states of the Asp25/25’ side chains in a complex of HIV-1
protease with pepstatin\(^5\). Two \(^{13}\text{C}\)-resonances at 178.8 ppm and 172.4 ppm were detected,
with only one of them (178.8 ppm) demonstrating significant (~0.2 ppm upfield) isotope
effect when H\(_2\)O solvent was replaced by D\(_2\)O. Again, as in the complex with KNI-272,
these resonances did not titrate in the pH range of 2.5 – 6.5. In the same paper, these
authors reported a study of unliganded HIV-1 protease and concluded that both Asp25
and Asp25’ are deprotonated at pH 6.0. Their interpretation was based on the observation
of single \(^{13}\text{C}\) NMR \(^{13}\text{C}\)\(^\gamma\)-carboxylate chemical shift at ~180.2 ppm for HIV-1 protease
bearing \(^{13}\text{C}\)-label only at \(^\gamma\)-atom of both Asp25 and Asp25’. To make an assignment, a
model 11-residue peptide (corresponding to residues 20-30 of the HIV-1 protease
polypeptide) was used, where the \(^{13}\text{C}\)\(^\gamma\) aspartate resonated at ~180.2 ppm when charged
and at much lower chemical shift when protonated (pK_a of that Asp was found to be near
4.0).

However, detailed biochemical mechanistic studies by Meeks and co-workers
indicated a monoprotonated state (i.e. one carboxylate, one carboxyl) for the side chain
carboxyls of asp25/25’ in unliganded HIV-1 protease\(^6\). More recent theoretical studies
also suggested monoprotonated state\(^7\). To resolve this discrepancy, we decided to revisit
the experiments published in 1996-year report\(^5\) using the more sensitive equipment
available at the present day.
Site-specifically labelled (1,4)-$^{13}$C-Asp25/25’ (1-99) wild-type HIV-1 protease and selected Gly51Xaa (Xaa = L-Ala, D-Ala, Aib, where Aib is $\alpha$-aminoisobutyric acid) chemical analogues were prepared by total chemical synthesis$^8,9$. We used a 21.1 Tesla NMR spectrometer (226 MHz for $^{13}$C and 900 MHz for $^1$H) equipped with a super cooled preamplifier for the $^{13}$C-channel, which gave enhanced sensitivity for direct detection of NMR of the $^{13}$C-nucleus (National Magnetic Resonance Facility at Madison, Wisconsin). Spectra were collected for samples of unliganded enzymes and for complexes of [L-Ala51/51’]HIV-1 protease with MVT-101 and KVS-1 inhibitors.

$^1$H-decoupled $^{13}$C-spectra are displayed in Figure 6.1 for the unliganded enzymes at pH 5.7. In the case of the essentially inactive [Aib51]HIV-1 protease (see Table 5.1 in Chapter 5), only two peaks at 177.9 ppm (carboxyl(ate)) and 174.9 ppm (amide) were observed (Figure 6.1 a). In the case of the [D-Ala51]HIV-1 protease, which has only poor proteolytic activity, an additional small peak at 180.2 ppm was noticed (Figure 6.1 b). For the more active [L-Ala51]HIV-1 protease (only ~10-times lower activity than that of the wild-type enzyme), additional peaks at ~180.2 ppm, 176.6 ppm and 176.3 ppm were observed (Figure 6.1 c). For the wild-type HIV-1 protease, these additional peaks dominated in the spectrum (Figure 6.1 d), with the 180.2 ppm peak being the most dominant as in the previously reported data$^5$. It should be noted that peaks at 177.9 ppm and 174.9 ppm had much greater linewidth (2-3 times) than the sharp additional peaks at 180.2 ppm, 176.6 ppm and 176.3 ppm. Recording spectra for the same samples in D$_2$O showed no significant isotope effect (30 ppb for carboxyl(ate)s and 50 ppb for amides) for signals at 177.9 ppm and 174.9 ppm at pH 5.7 (pD 6.1). Experiments with titration
have shown significant effect on chemical shift for resonances at 180.2 ppm (Δδ of 1.3 ppm upfield when pH was reduced from 5.7 to 3.9 and pK_a was determined to be 4.5), but only very slight effect on chemical shifts of 177.9 ppm (Δδ of 0.23 ppm upfield when pH was reduced from 5.7 to 3.9, and pK_a ~ 5).

In the 13C-NMR measurements for the complex of [L-Ala51,51']analogue with reduced isostere inhibitor MVT-101 we found sharp peaks of high intensity at 180.2 ppm, 176.6 ppm and 176.3 ppm, the same as in the 13C-spectra for the wild-type HIV-1 protease and its [L-Ala51]analogue (Figure 6.2 b). In addition, two broader peaks of lower intensity were observed at 178.3 ppm and 175.2 ppm. Additional experiments, where inter-pulse delay was varied from 2 s and 5 s, have shown much better intensity for peaks at 178.3 ppm and 175.2 ppm with a 5 s inter-pulse delay, suggesting a much longer T1-longitudinal (Figure 6.3) relaxation time constant for these NMR resonances. In the complex with ketomethylene isostere inhibitor KVS-1, again we have observed three sharp resonances at 180.2 ppm, 176.6 ppm and 176.3 ppm and, in addition, broader peaks at 179.6 ppm, 175.9 ppm and 174.6 ppm (Figure 6.2 c).

A synthetic model 11-residue peptide (residues 20-30 of HIV-1 protease sequence) has indeed shown 13C-NMR signals at 180.2 ppm and 176.6 ppm; the same resonances have been observed for synthetic precursor peptide (1-40).thioester at pH 5.7 when aspartate Asp25 is deprotonated.

The combined results suggest that the previously detected 180.2 ppm 13C-resonance was the product(s) of autoproteolysis (i.e. the signal came from short unfolded peptide(s)) and did NOT originate from Asp25/25’ in the intact, folded HIV-1 protease. This interpretation is supported by the observation that the intensity of the signal at 180.2
ppm is apparently a function of the catalytic efficiency of a given chemical analogue of HIV-1 protease: for unliganded enzymes, the more active is the enzyme, the more abundantly the resonance was observed at 180.2 ppm. To unambiguously rule out the possibility that different peaks might be coming from different conformational states of HIV-1 protease, we performed $^{13}$C-EXSY (exchange spectroscopy) for unliganded [L-Ala51/51’]HIV-1 protease with mixing times of 0.2 s and 1 s and did not find any exchange cross-peaks. Additional support for the 180.2 ppm peak originating from shorter peptide fragments comes from the observed sharper linewidth of these ‘additional’ resonances (6-18 Hz) in comparison with broader peaks coming from folded enzymes and enzyme/inhibitor complexes (23-43 Hz). This indicates slower $T_2^*$-transverse relaxation for sharper peaks, which is consistent with the unfolded nature of short peptides.

We conclude that the ionization state of the catalytic dyad in the unliganded HIV-1 protease enzyme is more likely to be monoprotonated, as elucidated in biochemical experiments by Meek and co-workers. The ‘true’ chemical shift for $^{13}$C$\gamma$-Asp25/25’ in the unliganded enzyme is $\delta$ 177.9 ppm –which is in between the two extremes for charged (181 ppm) and protonated aspartates (175 ppm) – and is in agreement with a monoprotonated state for the catalytic dyad. Most probably, rapid exchange on the NMR time scale makes it impossible to detect two separate $^{13}$C-resonances for protonated Asp25 and charged Asp25’. Furthermore, we have demonstrated that conformational dynamics of the protein is correlated to the dynamics of the catalytic aspartates (see Chapter 5). Hence, the protonation state may be not uniform for all conformers: different conformers are more likely to have different
protonation states. Such complexity may account for the absence of titration behavior (protein dynamics will be altering with pH changes as well, thus altering the effective \( pK_a \) for aspartates) and any significant deuterium isotope effect. We suggest that Fourier transform infrared spectroscopy may be better suited to unambiguously resolve this conundrum\(^\text{10}\).
Figure 6.1. 226 MHz $^{13}$C-<sup>1</sup>H<sub>1</sub> NMR spectra for unliganded HIV-1 protease and three chemical analogue enzymes. (a) [Aib51/51']HIV-1 protease, (b) [D-Ala51/51']HIV-1 protease, (c) [L-Ala51/51']HIV-1 protease, (d) wild-type HIV-1 protease. Red asterisk indicates peaks originating from peptide autoproteolysis products. All samples were prepared in 18.9 mM Na-phosphate buffer (pH 5.7), containing 5.4 % (v/v) D<sub>2</sub>O and 100 µM DSS-d<sub>6</sub>. Concentrations of protein were 0.29 – 0.41 mM.
**Figure 6.2.** 226 MHz $^{13}$C-$^1$H NMR spectra for [L-Ala51/51']HIV-1 protease: (a) unliganded; (b) complexed with the reduced isostere inhibitor MVT-101; and (c) with the ketomethylene isostere inhibitor KVS-1. Red asterisk indicates peaks originating from unfolded peptidic autoproteolysis products. In theses measurements sample of [L-Ala51/51']HIV-1 protease was folded by dialysis without inhibitors present and by HPLC-analysis products of autoproteolysis did not exceed 5 %. Inhibitors were added after dialysis. In (a) acquisition time was set at 0.5 s and inter-pulse delay was 1 s, while in (b) and (c) inter-pulse delay was extended to 5 s to obtain better S/N ratio spectra as with inhibitors protein becomes rigidified thus attenuating relaxation parameters for its signals.
Figure 6.3. 226 MHz $^{13}$C-$^1$H} NMR spectra for complex of [L-Ala51/51']HIV-1 protease with the reduced isostere inhibitor MVT-101, acquired with inter-pulse delay of 5 s (in a) and 2 s (in b). Red asterisk indicates peaks originating from unfolded peptidic autoproteolysis products. In (b) spectrum is dominated by autoproteolytic product and signals originating from protein are relaxation-filtered by application of the shorter inter-pulse delay.

Experimental details.

$^1$H-decoupled $^{13}$C-NMR spectra were acquired on Varian Unity Inova 900 spectrometer operating at 226.3 MHz for $^{13}$C-nucleus. In typical acquisitions for unliganded samples ~5000 transients were collected with acquisition time 0.5 s and inter-pulse delay 1 s, for samples with inhibitors 5000 – 10000 transients were acquired with
acquisition time 0.5 sand inter-pulse delay varied from 2 s to 5 s (see legends of Figure 6.2 and Figure 6.3). All samples were prepared in 18.9 mM Na.phosphate buffer (pH 5.7), containing 5.4 % (v/v) D2O and 100 µM DSS-d6. Inhibitors (MVT101 and KVS-1) were added in 4-fold molar excess to the solution of [L-Ala51/51’]HIV-1 protease after dialysis. Concentrations of protein were determined by integration of total LC-peak at 280 nm and were 0.34 mM for unliganded [L-Ala51/51’]HIV-1 protease, 0.41 mM [D-Ala51/51’]HIV-1 protease, 0.29 mM [Aib51/51’]HIV-1 protease, 0.4 mM for complex of [L-Ala51/51’]HIV-1 protease with MVT-101 inhibitor and 0.22 mM for complex with KVS-1 inhibitor. 13C chemical shifts were referenced indirectly to DSS-d6 using \( \gamma_C/\gamma_H \) ratio\(^\text{11} \). Temperature was set for all experiments to 3.3 °C to slow down autoproteolysis.

Site-specific 13C-labelling was performed by total chemical protein synthesis (Appendix F).\(^\text{8} \) (1,4)-13C-aspartic acid was purchased from Cambridge Isotopes, and was then side-chain-protected with allyl group\(^\text{12} \) and Boc-protected at the alpha amino group, and incorporated into the corresponding polypeptide as described\(^\text{9} \).

References:


Chapter 7. Hydrogen bonds at the protein-inhibitor interface in the HIV-1 protease / inhibitor complexes probed by total chemical synthesis and X-ray crystallography.

Abstract. Here we report high resolution X-ray crystallographic studies of various chemically-synthesized analogues of HIV-1 protease, where residues in the tips of the flap structures (residues 37-61 in each half of homodimer protein) were substituted with amino-acids with inverted stereochemistry (both at the α-carbon and β-carbon atoms) or backbone amide-to-ester modified. Such substitutions, although distant from catalytic residues, generally reduce catalytic activity of the enzymes. Molecular basis for the observed effects is discussed.

Introduction.

Hydrogen bonds play a pivotal role in the structure and function of proteins\(^1\). We set out to explore hydrogen bonding at the protein - inhibitor interface in the HIV-1 protease / inhibitor complexes and to evaluate the role of these hydrogen bonds in catalytic mechanism of proteolysis. Our studies made use of a combination of total chemical protein synthesis and high resolution X-ray crystallography. Recently, we reported the total chemical synthesis and proteolytic activity of a set of chemical analogues of the homodimeric HIV-1 protease molecule, in which the Gly51/51’ residues in the ‘flap’ structures (residues 37-61) were substituted by L-Ala, D-Ala and Aib (α-aminoisobutyric acid) in a symmetric and asymmetric fashion (see Chapter 5)\(^2\). We found that a ‘covalent dimer’ HIV-1 protease with residue Gly51 in one flap substituted by L-Ala51 and residue Gly51’ in the other flap substituted by D-Ala51’ possesses native-like proteolytic activity. In the case of all other analogues in which Gly51/Gly51’ were
replaced, the catalytic activity was found to be significantly reduced. With the help of NMR dynamics experiments and inter-flap distance measurements by pulse-EPR we have determined that dynamics of D-Ala51-containing flap is rigidified both on sub-ns and µs-ms times scales in comparison with dynamics of L-Ala51-containing flap. The unique combination of D-Ala51 and L-Ala51’-containing flaps in ‘covalent dimer’ [L-Ala51;D-Ala51’]HIV-1 protease leads to higher percentage of asymmetric conformers, which are preorganized for catalysis, whereas in symmetric [L-Ala51/51’] and [D-Ala51/51’]HIV-1 proteases such conformational states are less populated and hence such enzymes are less active.

In the current work, we have obtained high resolution X-ray structures for these analogues of HIV-1 protease complexed with the reduced isostere substrate-derived inhibitor MVT-101 (see Figure 7.1 – Figure 7.2 and ref.3). In addition, we performed total synthesis, enzyme assays and crystallographic studies of several other analogues of HIV-1 protease, where the backbone amide nitrogens of residues 50, 51, 52 were individually replaced by oxygens to give ester linkages, in order to probe the importance of hydrogen bonding at the interface of flap structures and MVT-101 inhibitor (Figure 7.3 a-c)4,5. Finally, we evaluated the role of stereochemistry at the β-carbon atom of residues Ile50/50’ by synthesizing [allo-Ile50/50’]HIV-1 protease (where chirality at the β-carbon atom was inverted) (Figure 7.3 d). The data obtained for the HIV-1 protease / MVT101 complexes were compared with complimentary data for complexes with the corresponding ketomethylene isostere inhibitor KVS-1 (Figure 7.4 and Figure 7.5 a and ref.6), and also for the hydroxyethylamine isostere inhibitor JG-365 (Figure 7.5 c,d and ref.7 also see Appendix E, Figure E.5).
Results and Discussion.

High resolution X-ray structures for chemical analogues of HIV-1 protease complexed with the reduced isostere substrate-derived inhibitor MVT-101 (see Figure 7.1–7.3 and Table G.1 in Appendix G) were all found to be isomorphous (space group $P\ 2_1\ 2_1\ 2_1$). The geometry and network of protein / inhibitor interactions were highly preserved among the different enzyme / inhibitor complexes (see Table 7.1 and Figure 7.6). In most of the structures, the flaps were observed to close over the inhibitor, adopting asymmetric $\beta$ I / $\beta$ II turn conformations at the flap tips. Exceptions are the $[L\text{-}\text{Ala}51/\text{Ala}51']$HIV-1 protease / MVT-101 complex, where both flaps close over the inhibitor in symmetric $\beta$ II / $\beta$ II conformation (Figure 7.1 b), and the $[\text{Gly}51/\text{Ala}51']$HIV-1 protease-MVT-101 complex, where asymmetric $\beta$ I / $\beta$ II orientation is switched to $\beta$ II / $\beta$ I relative to the N-to-C orientation of the MVT-101 substrate-derived inhibitor (Figure 7.2 d). In addition, in the case of $[\text{Ala}51/\text{Ala}51']$HIV-1 protease complex with MVT-101 inhibitor, one flap does not fully close over the inhibitor (Figure 7.1 d and Table 7.1, entry 4).

In the case of $[O\text{-}\text{Ile}50/\text{Ile}50']$ and $[\text{Ile}50/O\text{-}\text{Ile}50']$ chemical analogues of the HIV-1 protease$^{4,5}$, electron density for the non-nucleophilic water molecule (“water 301”) bridging flaps and carbonyls of the inhibitor was found to be much less intense (Figure 7.3 a, b); this was to be expected, since at least one hydrogen-bond-donating –NH–amide is replaced by an ester oxygen atom in these analogues, thus knocking out the possibility for a ‘water301’ H$_2$O…HN-Ile50 hydrogen bond.

In the active site of all the proteins complexed with MVT-101 inhibitor, we found that the side chains of the two catalytic aspartates are almost co-planar, and ‘O…O’
distances for the carboxyl oxygens distal from the inhibitor were in the 2.30 – 2.55 Å range for different chemical analogues (see Figure 7.1–7.3 and Table 7.1).

Crystal structures of HIV-1 protease and analogues complexed with the inhibitors KVS-1 and JG-365 were again all isomorphous (space group $P_2_12_12_1$) and H-bond lengths deviated only slightly at the flaps / inhibitor interface (Figure 7.4 – 7.5). However, it should be mentioned that in the case of $[L$-$\text{Ala51}/D$-$\text{Ala51}]$HIV-1 protease, the previously symmetric $\beta$ II / $\beta$ II configuration of the flaps isomerized to the asymmetric $\beta$ I / $\beta$ II turn configuration (compare Figure 7.1 b and Figure 7.4 b). In the crystal structures with these inhibitors, distance between the distal oxygens of the side chains of the two catalytic aspartates is much longer than with MVT-101 and varies in the 2.54 – 2.90 Å range. This is because KVS-1 and JG-365 have additional oxygen atom(s) (hydroxyls) which participate in the extensive H-bond network in the catalytic site and intrude between the two catalytic aspartates.

In the case of the HIV-1 protease complex with JG-365, two alternative conformations for flap structures ($\beta$ I / $\beta$ II versus $\beta$ II / $\beta$ I) and two orientations (N-to-C and C-to-N) for inhibitor were previously observed at 1 : 1 ratio. In all the structures determined in the present study, only in the case of asymmetric $[L$-$\text{Ala51}/D$-$\text{Ala51}]$HIV-1 protease complex with JG-365 inhibitor, $\beta$ I / $\beta$ II versus $\beta$ II / $\beta$ I conformers of the flaps and inhibitor were found to be not equipopulated but present at the ratio of ~ 65 : 35 (see Appendix E, Figure E.5). This observation allowed us to assign the N-to-C orientation of JG-365 as being opposite with respect to $\beta$ I / $\beta$ II turns configuration of the flaps in comparison to that in complexes with MVT-101 and KVS-1 inhibitors (compare for example Figure 7.5 a,b and Figure 7.5 c,d).
Proteolytic activities for some of the analogues being discussed in this chapter were reported previously in Table 5.1 of Chapter 5. Enzyme kinetics for the remaining newly reported enzymes are summarized in Table 7.2 of this Chapter. Both $[O$-Gly51/51’] and $[O$-Gly52/52’] HIV-1 protease analogues are only about 4-times less active than the wild-type enzyme. Such reduction is most probably caused by the change of the dynamic properties of the flaps similar to that observed when Gly51 was substituted with $L$-Ala51, $D$-Ala51 or Aib51 – the alteration of the dynamics of flaps will cause changes in dynamic properties of the catalytic site residues (see Chapter 5). In the case of $[O$-Ile50/50’] enzyme, catalytic activity is reduced 100-times in comparison to the wild-type HIV-1 protease; the asymmetric [Ile50/O-Ile50’] enzyme is 10-times less active than wild-type enzyme. Interestingly, a symmetric [allo-Ile50/50’] enzyme demonstrated catalytic properties very similar to wild-type HIV-1 protease.

We propose that hydrogen bonding network at the flaps / inhibitor interface mediated by a non-nucleophilic water molecule is part of a protein dynamics feedback mechanism in HIV-1 protease. In the case of the ester $O$-Ile50 analogues, the possibility of the amide hydrogen bond(s) present with native NH-Ile50 is eliminated (as evidenced by diffuse electron density for “water 301” in the X-ray structures, see Figure 7.3 a, b), thus destabilizing the rigid ‘closed’ flaps conformational state, which then would lead to higher flexibility of the corresponding flap structure. Thus, the dynamics of the catalytic site would never adjust to achieve the structure necessary for efficient electrostatic stabilization of the transition state of the catalyzed reaction as it happens in the wild-type enzyme. In the wild-type HIV-1 protease, such stabilization is achieved by hydrogen bonds mediated by “water 301” that rigidify the flap structures, lock the substrate in
productive conformation, and contribute to preorganization of structure of the catalytic site.

Despite identical crystal growth conditions (pH 6.0) for all complexes with MVT-101 inhibitor, ‘O…O’ distances for the carboxyl oxygens distal from the inhibitor were varying in the 2.30 – 2.55 Å range for different chemical analogues (see Figure 7.1–7.3). We suggest that the pKa values of catalytic aspartates are affected by different structural dynamics of the flaps in various chemical analogues of HIV-1 protease leading subsequently to the dissimilar degree of protonation and hence variation of average ‘O…O’ distance for Asp25…Asp25’ catalytic dyad observed in the X-ray structures.

Conclusions.

High-resolution X-ray studies of a series of chemical analogues of HIV-1 protease complement previously discussed hypothesis and associated experimental data of coupling of dynamics of flap structures and catalytic residues. It is clear from this work that hydrogen bonding at the flap-inhibitor interface plays a critical role in attenuating flap dynamics. Furthermore, this dynamics information is transferred to the catalytic site leading to preorganization/reorganization of active site residues in order to electrostatically stabilize transition states in the sequence of the catalytic process.
Figure 7.1. Flap structures (residues 47–54 and 47’–54’), catalytic aspartates (25 and 25’) in crystal structures of chemical analogues of HIV-1 protease complexed with the reduced isostere inhibitor MVT-101: (a) wild-type [Gly51/51’]HIV-1 protease, (b) [L-Ala51/51’]HIV-1 protease, (c) [D-Ala51/51’]HIV-1 protease and (d) [Aib51/51’]HIV-1 protease. Side chains are deleted for clarity. The 2Fo-Fc electron density map is contoured at a level of 1σ (magenta) for residues 50, 51, 50’ and 51’, and at a level of 3σ (green) for residues Asp25 and Asp25’.
Figure 7.2. Flap structures (residues 47–54 and 47’-54’), catalytic aspartates (25 and 25’) in crystal structures of chemical analogues of HIV-1 protease complexed with the reduced isostere inhibitor MVT-101: (a) [L-Ala51/Gly51’]HIV-1 protease, (b) [Gly51/D-Ala51’]HIV-1 protease, (c) [L-Ala51/D-Ala51’]HIV-1 protease (see Chapter 5) and (d) [Aib51/Gly51’]HIV-1 protease. Side chains are deleted for clarity. The $2F_{o}-F_{c}$ electron density map is contoured at a level of 1σ (magenta) for residues 50, 51, 50’ and 51’, and at a level of 3σ (green) for residues Asp25 and Asp25’.
Figure 7.3. Flap structures (residues 47–54 and 47’-54’), catalytic aspartates (25 and 25’) in crystal structures of chemical analogues of HIV-1 protease complexed with the reduced isostere inhibitor MVT-101: (a) $[O$-Ile50/50’]HIV-1 protease, (b) $[Ile50/O$-Ile50’]HIV-1 protease, (c) $[O$-Gly51/51’]HIV-1 protease and (d) $[allo$Ile50/50’]HIV-1 protease. Side chains are deleted for clarity. The $2F_o-F_c$ electron density map is contoured at a level of $1\sigma$ (magenta) for residues 50, 51, 50’ and 51’, and at a level of $3\sigma$ (green) for residues Asp25 and Asp25’.
Figure 7.4. Flap structures (residues 47–54 and 47′-54′), catalytic aspartates (25 and 25′) in crystal structures of chemical analogues of HIV-1 protease complexed with the ketomethylene isostere inhibitor KVS-1: (a) [D-Ala51/51′]HIV-1 protease, (b) [L-Ala51/51′]HIV-1 protease, (c) [Ile50/O-Ile50′]HIV-1 protease and (d) [O-Ile50/50′]HIV-1 protease. Side chains are deleted for clarity. The 2F_\text{O}-F_\text{C} electron density map is contoured at a level of 1σ (magenta) for residues 50, 51, 50′ and 51′, and at a level of 3σ (green) for residues Asp25 and Asp25′ in a-c. In (d), 2F_\text{O}-F_\text{C} is at 0.8σ for residues 50, 51, 50′ and 51′, and at a level of 2.5σ (green) for residues Asp25 and Asp25′.
Figure 7.5. Flap structures (residues 47–54 and 47’–54’), catalytic aspartates (25 and 25’) in crystal structures of chemical analogues of HIV-1 proteases complexed with inhibitors: (a) \([\text{alloIle50/50’}]\text{HIV-1 protease with KVS-1 inhibitor}\), (b) \([\text{Asn25/25’}]\text{HIV-1 protease with MVT-101}\), (c) \([\text{Ile50/O-Ile51’}]\text{HIV-1 protease with hydroxyethylamine JG-365}\), (d) \([\text{O-Gly51/51’}]\text{HIV-1 protease with JG-365}\). Side chains are deleted for clarity. The 2\(F_o-F_c\) electron density map is contoured at a level of 1\(\sigma\) (magenta) for residues 50, 51, 50’ and 51’, and at a level of 3\(\sigma\) (green) for residues Asp25 and Asp25’.
Table 7.1. Selected interatomic distances (in Å) at the flaps / inhibitor MVT-101 interface. For clarification see Figure 7.6.

<table>
<thead>
<tr>
<th>Chemical analogue</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$c_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly51/51’</td>
<td>2.55</td>
<td>2.85</td>
<td>3.01</td>
<td>2.97</td>
<td>2.66</td>
<td>2.67</td>
<td>2.94</td>
</tr>
<tr>
<td>$L$-Ala51/51’</td>
<td>2.53</td>
<td>2.83</td>
<td>3.07</td>
<td>3.17</td>
<td>2.81</td>
<td>2.66</td>
<td>n/a</td>
</tr>
<tr>
<td>$D$-Ala51/51’</td>
<td>2.46</td>
<td>2.83</td>
<td>3.12</td>
<td>3.06</td>
<td>2.73</td>
<td>2.67</td>
<td>2.92</td>
</tr>
<tr>
<td>Aib51/51’</td>
<td>2.46</td>
<td>2.88</td>
<td>4.12</td>
<td>2.92</td>
<td>2.89</td>
<td>2.70</td>
<td>4.22</td>
</tr>
<tr>
<td>Gly51/$D$-Ala51’</td>
<td>2.47</td>
<td>2.86</td>
<td>2.99</td>
<td>3.07</td>
<td>2.68</td>
<td>2.69</td>
<td>3.07</td>
</tr>
<tr>
<td>$L$-Ala51/Gly51</td>
<td>2.50</td>
<td>2.94</td>
<td>2.99</td>
<td>3.07</td>
<td>2.71</td>
<td>2.59</td>
<td>3.23</td>
</tr>
<tr>
<td>$L$-Ala51/$D$-Ala51’</td>
<td>2.37</td>
<td>2.89</td>
<td>3.01</td>
<td>3.08</td>
<td>2.71</td>
<td>2.68</td>
<td>3.17</td>
</tr>
<tr>
<td>Gly51/Aib51’</td>
<td>2.30</td>
<td>2.99</td>
<td>3.24</td>
<td>2.83</td>
<td>2.73</td>
<td>2.67</td>
<td>2.87</td>
</tr>
<tr>
<td>$allo$Ile51/51’</td>
<td>2.43</td>
<td>2.95</td>
<td>2.96</td>
<td>3.00</td>
<td>2.73</td>
<td>2.64</td>
<td>3.00</td>
</tr>
<tr>
<td>$O$-Gly51/51’</td>
<td>2.55</td>
<td>2.90</td>
<td>3.02</td>
<td>2.96</td>
<td>2.70</td>
<td>2.65</td>
<td>3.20</td>
</tr>
<tr>
<td>$O$-Ile50/50’</td>
<td>2.52</td>
<td>2.75</td>
<td>3.01</td>
<td>4.48</td>
<td>2.28</td>
<td>2.59</td>
<td>3.23</td>
</tr>
<tr>
<td>Ile50/$O$-Ile50’</td>
<td>2.48</td>
<td>2.90</td>
<td>3.53</td>
<td>3.13</td>
<td>2.47</td>
<td>2.52</td>
<td>3.16</td>
</tr>
<tr>
<td>Asn25/25’</td>
<td>3.08</td>
<td>3.10</td>
<td>2.95</td>
<td>2.92</td>
<td>2.79</td>
<td>2.70</td>
<td>2.93</td>
</tr>
</tbody>
</table>
Figure 7.6. Selected interatomic distances at the HIV-1 protease flaps / inhibitor MVT-101 interface.

$d_1$ is the ‘O…O’ distance between catalytic Asp25 and Asp25’ oxygen atoms distal from the inhibitor.

$d_2$ is ‘O…N’ distance between the proximal O atom of Asp25 and the NH of reduced isostere [CH$_2$NH] moiety of the inhibitor.

$b_1$ is the ‘O…O’ distance between Ile2(C=O) of MVT-101 and ‘water301’.

$b_2$ is the ‘O…O’ distance between Nle4(C=O) and ‘water301’.

$a_1$ is the ‘O…N’ distance between Ile50'(NH) in β-turn type II flap (Flap B) and ‘water301’.

$a_2$ is the ‘O…N’ distance between Ile50(NH) in β-turn type I flap (Flap A) and ‘water301’.

$c_1$ is the ‘O…N’ distance between Ile50(C=O) in Flap A and Gly51’(NH) in Flap B.
Table 7.2. Steady-state kinetics of proteolysis by chemically synthesized analogues of HIV-1 protease (see Table 5.1 in Chapter 5).

<table>
<thead>
<tr>
<th></th>
<th>Flap</th>
<th>Flap’</th>
<th>$k_{cat}$ (s$^{-1}$); mean ± s.d.</th>
<th>$K_m$ (µM); mean ± s.d.</th>
<th>$k_{cat}/K_m$; (s$^{-1}$µM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>allo-Ile50</td>
<td>allo-Ile50</td>
<td>11.7 ± 0.3</td>
<td>37.1 ± 2.8</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>O-Gly52</td>
<td>O-Gly52</td>
<td>9.5 ± 0.2</td>
<td>36.0 ± 3.7</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>O-Gly51</td>
<td>O-Gly51</td>
<td>13.2 ± 0.4</td>
<td>62.0 ± 4.0</td>
<td>0.21</td>
</tr>
<tr>
<td>4</td>
<td>Ile50</td>
<td>O-Ile50</td>
<td>4.2 ± 0.2</td>
<td>41.4 ± 3.2</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>O-Ile50</td>
<td>O-Ile50</td>
<td>0.94</td>
<td>99.6</td>
<td>0.009</td>
</tr>
<tr>
<td>6</td>
<td>truncated*</td>
<td>Gly51</td>
<td>0.23</td>
<td>16.8</td>
<td>0.014</td>
</tr>
<tr>
<td>7</td>
<td>Asn25</td>
<td>Asn25</td>
<td>inactive</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* For the ‘covalent dimer’ analogue 6, one of the flaps was truncated and had the following sequence Nle$^{46}$Ile$^{47}$Gly$^{48}$GlyGlyLys$^{55}$Val$^{56}$. 
References:


Chapter 8. Short, strong H-bonds in complexes of the HIV-1 protease with substrate-derived inhibitors detected by NMR spectroscopy.

Abstract. In this work we have detected strong, short hydrogen bonds in HIV-1 protease / inhibitor complexes by NMR spectroscopy. For the first time, we have observed low-field $^1$H NMR resonances for the complexes of HIV-1 protease with the substrate-derived inhibitors MVT-101 (reduced isostere) and KVS-1 (ketomethylene isostere). In the case of MVT-101, unambiguous assignment of low-field resonances was performed and the high delocalization nature of the corresponding protons was demonstrated. When studying complexes of MVT-101 (which mimicks an early transition state) with chemical analogues of HIV-1 protease in which the dynamic properties of the flaps were systematically modified, we demonstrated correlation of conformational isomerisations in the protein / inhibitor complex with proton transfer reactions.

Low barrier hydrogen bonds (LBHBs) were originally proposed to play a significant role in enzymatic reactions by Frey, Cleland, Kreevoy and colleagues\textsuperscript{1,2}. Since then, the proposed role of low-barrier hydrogen bonds in enzymatic catalysis has been the subject of intense debate\textsuperscript{1-10}. According to the original hypothesis, formation of a LBHB would supply 10–30 kcal/mol to the stabilization of the transition state\textsuperscript{1,2}, in this way facilitating otherwise very difficult reactions\textsuperscript{11,12}. This suggestion has provoked opposing viewpoints leading to a controversy about the strength and even the existence of LBHBs in biomolecules\textsuperscript{3,6-10}. Very recent combined high-resolution neutron and X-ray diffraction studies of photoactive yellow protein (PYP) for the first time provided strong evidence for the existence of a LBHB in a biological macromolecule\textsuperscript{13}: this finding will further
facilitate additional studies in order to elucidate the nature and function of LBHBs and short, strong hydrogen bonds (SSBHs) in the enzymatic catalysis.

In the case of aspartic proteases, D. B. Northrop has suggested a chemical and kinetic ‘isomechanism’ for enzymatic catalysis in which the LBHB is crucial and is formed between two oxygens of two catalytic aspartates. The mechanistic role of the LBHB is to assist in proton transfer via quantum mechanical tunneling\(^{14}\). The proposal is based on the theoretical work by Piana and Carloni, where a very short hydrogen bond (2.5 Å ‘O…O’ distance) was observed in the \textit{ab initio} molecular dynamics (MD) simulation for unliganded HIV-1 protease\(^{15}\). In a recent experimental X-ray diffraction study authors detected even shorter (2.3 Å) distance between oxygens of the two catalytic aspartates of HIV-1 protease complexed with products of proteolytic reaction in situ, providing additional support for the existence of LBHB in the active site of HIV-1 protease\(^{16}\).

Proton NMR spectroscopy has been a primary method for the detection of LBHBs in biomolecules since the pioneering work on serine proteases\(^{17,18}\). The observation of proton NMR signal(s) with an unusually low-field chemical shift (\(\delta > 15\) ppm), combined with low-fractionation factor (\(\varphi < 0.4\)) and large isotope effects on the chemical shift (upon transferring enzyme from \(^{1}\text{H}_2\text{O}\) to \(^{2}\text{H}_2\text{O}\) or \(^{3}\text{H}_2\text{O}\))\(^{19}\) are considered as strong indications for the presence of LBHBs or SSBHs\(^{20}\).

We applied one dimensional \(^1\text{H}\) NMR spectroscopy to screen for low-field resonances (if any) in chemically synthesized analogues of HIV-1 protease. Measurements were performed for both unliganded enzymes and their complexes with various inhibitors. With the help of the binominal 1331 pulse sequence\(^{21}\) for water
suppression we detected a series of signals in the low-field portion of spectra ($\delta > 11$ ppm) in the case of reduced isostere MVT-101 inhibitor (Figure 8.1 and Table 8.1) and a single resonance at 12.6 ppm for ketomethylene isostere KVS-1 inhibitor (Figure 8.2). We did not observe any low-field signals for complexes with hydroxyethylamine JG-365 inhibitor, nor did we see peaks with chemical shifts below (further downfield) 10.5 ppm for any of the unliganded HIV-1 proteases. Binominal water suppression produces zero-net magnetization at the solvent resonance frequency and, therefore allows detection of protons undergoing fast chemical exchange with water. We found that the low-field resonances could also be detected with other binominal pulse sequences (11, 146641, etc.) as well as with watergate3919 water suppression method22, albeit less efficiently. Water suppression by water resonance presaturation eliminates all of the peaks in the low-field region of the spectra thus proving that these signals corresponded to protons undergoing fast chemical exchange with water solvent.

In the case of MVT-101 inhibitor, we performed pH titration experiments and for the wild-type HIV-1 protease we found that resonances at 16.35 ppm and 15.45 ppm titrate with $pK_a \sim 5.5$, such that the 16.35 ppm peak is predominant at lower pH values and the resonance at 15.45 ppm is predominant at higher pH values (Figure 8.3). Resonances at 13.35 ppm and 11.3 ppm also exhibited pH titration with $pK_a \sim 5.3$. Fractionation factors were determined in solutions, where $D_2O/H_2O$ content was varied (Figure 8.4). For resonance at 16.35 ppm fractionation factor, $\phi$ was found to be 0.33 at pH 3.0 at its maximum intensity and for resonance at 15.45 ppm at pH 6.0 it corresponded to 0.67.
Figure 8.1. $^1$H NMR spectra acquired with binominal 1331 water suppression for HIV-1 protease preparations complexed with MVT101: (a) wild-type HIV-1 protease, (b) [L-Ala51/51']HIV-1 protease, (c) [D-Ala51/51']HIV-1 protease, and (d) [Aib51/51'] chemical analogue in 20 mM Na.phosphate at pH 5.6 and 1 °C temperature. Referencing was done directly to DSS-d$_6$. 
Table 8.1. $^1$H NMR chemical shifts observed in low-field region of spectra for chemical analogues of HIV-1 protease with reduced isostere MVT-101 inhibitor.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$^1$H chemical shifts ($\delta$, ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type HIV-1 protease</td>
<td>16.35; 15.45; 13.3; 11.3</td>
</tr>
<tr>
<td>$[L\text{-Ala51/51'}]$HIV-1 protease</td>
<td>16.2; 15.3; 12.9; 11.4</td>
</tr>
<tr>
<td>$[D\text{-Ala51/51'}]$HIV-1 protease</td>
<td>16.35; 15.4; 13.05; 11.4</td>
</tr>
<tr>
<td>$[\text{Aib51/51'}]$HIV-1 protease</td>
<td>15.85; 12.2</td>
</tr>
<tr>
<td>$[\text{alloIle50/50'}]$HIV-1 protease</td>
<td>16.5; 15.5; 13.5; 11.4</td>
</tr>
<tr>
<td>$[O\text{-Ile50/50'}]$HIV-1 protease</td>
<td>16.2; 12.1</td>
</tr>
<tr>
<td>$[\text{Asn25/25'}]$HIV-1 protease</td>
<td>16.1</td>
</tr>
</tbody>
</table>

The value of $\varphi$ at 0.33 indicates very high preference for protium over deuterium in hydrogen-bond resonating at 16.3 ppm, which may be indicative of low-barrier hydrogen bond$^{20}$.

Remarkably, in the course of the pH titration of $[D\text{-Ala51/51'}]$ and $[L\text{-Ala51/51'}]$ enzymes complexed with MVT101, the $pK_a$ value of the downfield protein resonances was found to be shifted to a higher value of 5.8 (Figure 8.5-8.6) and for the $[\text{allo-Ile50/50'}]$ HIV-1 protease the $pK_a$ was found to be lower, at 4.8 (Figure 8.7). More interestingly, in the case of $[L\text{-Ala51/51'}]$ HIV-1 protease the pH titration was found to be irreversible – after being transferred to pH < 4.5, peak at 16.35 ppm becomes dominant and peak at 15.45 ppm never rebuilds back even after prolonged exposure at higher pH values (several months at room temperature). A similar phenomenon has been observed
for chymotrypsin by Markley and colleagues and was attributed to irreversible conformational isomerization of the protein\textsuperscript{23}. In our case, in the crystal structure of $[L$-\textit{Ala51/51'}]HIV-1 protease with MVT-101 inhibitor two flaps are symmetric and both are in $\beta$ II / $\beta$ II conformations, whereas in the complex of the same protein with KVS-101 inhibitor one flap isomerizes to become $\beta$-turn of type I (see Figure 7.1 b and Figure 7.4 b in Chapter 7). \textbf{These observations suggests coupling of the pH titration of the two low-field $^1$H NMR resonances and flap $\beta$-turn isomerizations.}

In the case of $[O$-Ile50/50']HIV-1 protease, only two low-field peaks at 16.2 ppm and 12.0 ppm have been observed initially at pH 5.7. Titrating sample to lower pH values led to appearance of two additional resonances at 15.5 ppm and 10.8 ppm (Figure 8.8). Clearly, structural dynamics of the flaps modifies $pK_a$ values of the groups participating in the forming hydrogen bonds with protons resonating in the low-field part of the spectra.

Based on the X-ray structures of complexes of the enzymes with MVT-101 inhibitor and very short ‘O…O’ Asp25…Asp25’ distances, the first guess for the origin of the peaks in the low-field region of the spectra would be inter-distal O-atom aspartate proton. However, in the control experiments with the inactive [Asn25/25']HIV-1 protease complexed with MVT-101 inhibitor, where the possibility of inter-aspartate O…H-O hydrogen bond is eliminated, we have observed single resonance at 16.1 ppm. Such result suggested an alternative hypothesis, where secondary amine of the MVT-101 inhibitor forming hydrogen bonds with aspartates would cause $^1$H NMR signals in the low-field of the spectra.
To test this hypothesis, we synthesized MVT-101 inhibitor where nitrogen of secondary amine function was labelled with $^{15}$N-isotope. In one dimensional $^1$H NMR spectra of HIV-1 protease complexed with $[^{15}$N]MVT101 we found that the signal at 13.3 ppm split into a doublet with $^1$$J_{NH} \sim 56$ Hz at 2 °C and $^1$$J_{NH} \sim 62$ Hz at 11 °C (Figure 8.9). Application of $^{15}$N-broadband GARP decoupling$^{24}$ collapsed the doublet into a singlet. We further tested decoupling efficiency as a function of decoupler frequency offset and found that the best decoupling is at ~40 ppm, meaning that $^{15}$N chemical shift of corresponding nitrogen nucleus is at about this value (Figure 8.10). We then assigned the protonation state for this amino group by taking spectra in buffers with different D$_2$O / H$_2$O ratio (Figure 8.11). If secondary amine is protonated, then in solutions with high D$_2$O content one may expect to observe isotope effect on chemical shift of the corresponding ammonium proton due to high concentration of the partially deuterated –N(D)H$^+$– species. Indeed, in 80% D$_2$O buffer resonance at 13.28 ppm shifted downfield to 13.43 ppm and isotope effect was thus found to be 0.15 ppm (at 600 MHz). In solutions with intermediate D$_2$O / H$_2$O ratio two peaks at 13.28 ppm and 13.43 ppm were observed simultaneously corresponding to mixture of –N(D)H$^+$– and –NH$^+_2$– species.

$^{15}$N-MVT-101 inhibitor was further separately studied in organic DMF-$d_7$ and CD$_3$OH solvents in order to slow down exchange of the secondary amine/ammonium protons with solvent. In the case of DMF-$d_7$, at temperature of 2 °C diastereotopic H$_a$ and H$_b$ of –N(H$_a$)H$_b^+$– moiety were spectrally resolved and splitted into doublets with chemical shifts at 9.37 ppm ($^1$$J_{NH} 74.2$ Hz) and 8.78 ppm ($^1$$J_{NH} 74.9$ Hz) and the corresponding $^{15}$N chemical shift at 48.2 ppm. In the CD$_3$OH solvent study was performed at –4 °C with chemical shifts found to be only slightly shifted in comparison to
DMF-\(d_7\), \(\delta, 9.2 \text{ ppm (}^{1}J_{NH} 78 \text{ Hz})\) and 8.22 ppm (\(^{1}J_{NH} 74 \text{ Hz}\)) and \(^{15}\text{N} \text{ chemical shift at} 48.14 \text{ ppm}. \text{In addition, unprotonated secondary amine was detected in} \text{CD}_{3}\text{OH at} ^{1}\text{H} \delta, 7.15 \text{ ppm (}^{1}J_{NH} 91.6 \text{ Hz})\) and \(^{15}\text{N} \delta, 88.9 \text{ ppm.}\)

Comparing data for the \(^{15}\text{N}-\text{MVT-101} \text{ inhibitor obtained in the presence of HIV-1 protease in aqueous solution and protic / aprotic organic solvents, the perturbation of} ^{1}\text{H} \text{ chemical shift for secondary ammonium group upon binding of MVT-101 to HIV-1 protease exceeds 4 ppm. Since} ^{1}J_{NH} \text{ coupling is a direct measure of the covalent character of corresponding chemical bond (see ref.}^{10} \text{ for discussion), the low} ^{1}J_{NH} \text{ value (~60 Hz) in the complex with HIV-1 protease in comparison to ~75 Hz when without HIV-1 protease suggests high delocalization of this particular proton with occupancy (localization) at the nitrogen being 80%.}\)

A more perplexing picture was observed for the resonances at 16.35 ppm and 15.45 ppm when recording \(^{1}\text{H} \text{ NMR spectra for HIV-1 protease complexed with} ^{15}\text{N}-\text{labelled MVT-101 inhibitor. The intensity of these two resonances was strongly attenuated (diminished) in comparison to the data acquired with non-labeled variant of MVT-101 (Figure} 8.9). \text{In the case of} ^{15}\text{N}-\text{labeled MVT-101 inhibitor, the signals did not follow expected pH dependence (see comparison in Figure} 8.12), nor did they follow temperature dependence observed previously for non-labelled MVT-101.} ^{15}\text{N}-\text{decoupling at multiple decoupler offsets did not affect in any way the resonances of interest.}\text{Logically, these resonances should be assigned as proton(s) being bound to the same} ^{15}\text{N}-\text{atom in MVT-101 inhibitor as it is the only modification in comparison with non-labelled MVT-101. Acquiring spectra for non-labeled MVT-101 and varying} \text{D}_{2}\text{O / H}_{2}\text{O ratio in the solution did not show any alterations of chemical shifts, thus indicating that these two} \text{resonances should be associated with the specifically modified nitrogen atom.}\
resonances are non-protonated secondary amine species. Unusual temperature and pH dependence for these resonances in the complex with $^{15}\text{N} \text{MVT-101}$ inhibitor can be explained by very high delocalization of the corresponding protons. If so, scalar $^{1}J(15\text{N}-1\text{H})$–coupling and chemical shifts of both proton and particularly nitrogen (which spans very large chemical shift range) are variable and serve as additional relaxation mechanisms (analogous to chemical shift anisotropy relaxation). The intensity of these resonances at different temperatures and pH values will be a function of these additional relaxation mechanisms influenced by the level of delocalization of a given proton at each particular set of conditions.
Figure 8.2. (a) $^1$H NMR spectrum for complex of $[L$-Ala$51/51']$HIV-1 protease with the ketomethyleneisostere inhibitor KVS-1, acquired with watergate 3919 water suppression. Temperature was set to 3 ºC, referencing was done directly to DSS-$d_6$. (Inset) Magnification of peak at 12.61 ppm. (b) X-ray structure of $[L$-Ala$51/51']$HIV-1 protease complexed with the ketomethylene isostere inhibitor KVS-1 at 1.2 Å resolution. Side chains are deleted for clarity. The $2F_o-F_c$ electron density map is contoured at a level of 1σ (magenta) for residues 50, 51, 50’ and 51’, and at a level of 3σ (green) for residues Asp25 and Asp25’. (Inset) The active site region is enlarged. Short (2.6 Å) ‘O…O’ distances are shown. The $2F_o-F_c$ electron density map is contoured at 1σ level (magenta) and the omit $F_o-F_c$ density is contoured at 1.1 σ level (green). Note that according to this structure ketomethylene isostere is hydrated not to the gem-diol, but to a charged oxyanion.
Figure 8.3. 900 MHz $^1$H spectra acquired with 1331 binominal water suppression with maximum excitation set at ~16 ppm for wild-type HIV-1 protease complex with MVT-101 inhibitor at six different pH values. Resonances at {16.35 ppm, 15.45 ppm} and {13.3 ppm, 11.3 ppm} were phased separately, and are thus depicted separately, because of phase roll caused by application of 1331 binominal pulse sequence. Line broadening (30 Hz) was applied. Temperature was set to 11 °C. Vertical scales for the low-field resonances were normalized using the most intense upfield non-exchangeable methyl resonance.
Figure 8.4. Signal intensities of deshielded protons in HIV-1 protease complex with MVT-101 inhibitor plotted as a function of the H$_2$O content in the H$_2$O/D$_2$O mixture. Peak areas of the low-field resonances were normalized to the peak area of the upfield non-exchangeable methyl resonance. Data were fitted to the equation, $I_x = I_o x[\phi(1-x)+x]$, where $x$ is the mole fraction of H$_2$O, $I_x$ is the signal intensity in H$_2$O/D$_2$O mixture, and $I_o$ is the intensity in H$_2$O. (a) Signal intensity of the resonance at 16.3 ppm at pH 3.0. The fractionation factor, $\phi = \frac{[\text{Enz-D}][\text{H}_2\text{O}]}{[\text{Enz-H}][\text{D}_2\text{O}]}$ is $0.33 \pm 0.05$ (s.d.). (b) Signal intensity of the resonance at 15.4 ppm. Fractionation factor, $\phi$ is $0.67 \pm 0.06$ (s.d.).
Figure 8.5. 600 MHz $^1$H spectra for [L-Ala51/51′]HIV-1 protease complexed with MVT-101 inhibitor acquired with 1331 binominal water suppression with maximum excitation set at ~16 ppm at different pH values (in a). Resonances at 16.2 ppm, 15.3 ppm and 12.9 ppm, 11.4 ppm were phased separately, and are thus depicted separately, because of phase roll caused by application of the 1331 binominal pulse sequence. The same vertical scale was used for both regions. Line broadening (30 Hz) was applied. Temperature was set to 11 °C. Vertical scale was normalized to resonance at ~ 10.4 ppm (presumably NH-indole of Trp). Titration is not reversible after being at pH < 4.5. Upon going from 5.8 to 6.2 and than to 5.0 and back to 6.1 some hysteresis has been observed. $pK_a$ is ~5.8 (in b) for 16.2 ppm and 15.3 ppm signals (0.3 units higher than for wild type enzyme).
Figure 8.6. 600 MHz $^1$H spectra for [D-Ala51/51']HIV-1 protease complexed with MVT-101 acquired with 1331 binominal water suppression with maximum excitation set at ~16 ppm at six different pH values (in a). Resonances at 16.35 ppm, 15.4 ppm and 13.05 ppm, 11.4 ppm were phased and thus depicted separately because of phase roll caused by application of 1331 binominal pulse sequence. The same vertical scale was used for both regions. Line broadening (30 Hz) was applied. Temperature was set to 11 °C. Titration is fully reversible. $pK_a \sim 5.8$ (in b) for 16.35 ppm and 15.4 ppm signals (0.3 units higher than for wild type enzyme). No significant titration for resonances at 13.05 ppm and 11.4 ppm.
Figure 8.7. 600 MHz $^1$H spectra for [allo-Ile50/50']HIV-1 protease complexed with MVT-101 acquired with 1331 binomial water suppression with maximum excitation set at 16 ppm at six different pH values (in a). Resonances at 16.5 ppm, 15.5 ppm and 13.5 ppm, 11.4 ppm were phased and thus depicted separately because of phase roll caused by application of 1331 binomial pulse sequence (Note the same vertical scale for both regions). Line broadening (30 Hz) was applied. Temperature was set to 11 °C. Titration is fully reversible. $pK_a \sim 4.8$ (in b) for 16.5 ppm and 15.5 ppm signals (0.7 units lower than for wild type enzyme). $pK_a \sim 5.5$ for resonances at 13.5 ppm and 11.4 ppm.
Figure 8.8. 600 MHz $^1$H spectra for [O-Ile50/50']HIV-1 protease complexed with MVT-101 acquired with 1331 binominal water suppression with maximum excitation set at 16 ppm at seven different pH values. Resonances at 16.2 ppm, 15.5 ppm and 12 ppm, 10.9 ppm were phased and thus depicted separately because of phase roll caused by application of 1331 binominal pulse sequence (Note the same vertical scale for both regions). Line broadening (30 Hz) was applied. Temperature was set to 11 ºC. Titration is fully reversible. $pK_a \sim 4$ for proton at 16.2-15.5 ppm and $pK_a \sim 4$ for signal at 10.9 ppm.
Figure 8.9. Low-field resonances acquired with 1331 binominal pulse sequence for complex of wild-type HIV-1 protease complexed with \([\text{CH}_2\text{N}^{15}\text{H}]\)MVT-101 inhibitor at pH 3.1 and two temperatures (a) 2 °C, and (b) 11 °C. Signal at 13.3 ppm is split in both spectra, with \(^{1}J_{\text{NH}}\) of 56 Hz in (a), and 62 Hz in (b). Upon application of GARP decoupling on \(^{15}\text{N}\) channel (at 50 ppm) doublet collapses.
Figure 8.10. $^{15}$N decoupler offset was varied in order to determine approximate chemical shift of $^{15}$N nucleus coupled to resonance at 13.3 ppm. The linewidth approached its minimum with decoupler offset at ~40 ppm.

Figure 8.11. Resonance at ~13.3 ppm undergoes shift 0.15 ppm downfield upon increase of D$_2$O content in solution of HIV-1 protease / MVT-101 complex. Presence of isotope effect indicates protonated nature ($-$NH$_2^+$ – species) of secondary amino group.
Figure 8.12. Compare the ratio of resonances at 16.3 ppm and 15.4 ppm with nearly the same pH value taken for (a) complex of wild-type HIV-1 protease with $[^{15}\text{N}]$MVT-101 inhibitor and (b) complex of wild-type HIV-1 protease and non-labeled MVT-101.

Such conclusions are supported by high-resolution X-ray crystal structures for chemical analogues of HIV-1 protease with MVT-101 inhibitor previously discussed, where ‘O…O’ Asp25…Asp25’ distance was found to vary in the 2.30 – 2.55 Å range for different chemical analogues (see Chapter 7). Such very short distance would only be sustainable if proton was shared by the two aspartates. However, despite rather high 1.3 Å resolution obtained in the X-ray diffraction measurements in some of the crystal structures we were unable to localize putative protons when inspecting omit $F_o-F_c$ electron density maps. On the other hand, absence of any electron density for shared
proton can be caused by its high delocalization. In addition, in different chemical analogues as demonstrated by NMR titrations experiments, $pK_a$ values of the titrating groups are altered in different chemical analogues. This different degree of protonation of the two aspartates (or level of localization of protons) can also explain the variation of the inter-aspartate ‘O…O’ distance observed in the crystal structures.

In Scheme 8.1 we summarized our observations in the form of equilibrium between three principal conformational states of the HIV-1 protease / MVT-101 inhibitor complex. We propose that conformational isomerization of flaps in solution triggers conformational isomerization of inhibitor via inversion of secondary amine (ammonium) group. In crystal structures of HIV-1 protease / MVT-101 complexes flaps adopt closed/closed conformation. In such conformation protonated ammonium forms short, strong hydrogen bond with one of the carboxylates, which we assigned to $^1$H NMR resonance of 13.3 ppm (Scheme 8.1 a). Upon opening of the flaps inhibitor adopts a conformation in which ammonium group undergoes inversion and makes even shorter hydrogen bond, which resonates at 16.3 ppm (Scheme 8.1 b). At higher pH values one of the aspartates becomes deprotonated making putative SSHB weaker and shifting resonance to 15.4 ppm (Scheme 8.1 c). Delocalization of protons can be achieved through resonance structures similar to those invoked in the Northrop’s mechanism (Scheme 8.1 right panel).

Previously, LBHBs were studied for the single-chain monomeric aspartic protease endothiapepsin with various inhibitors$^{25-27}$. Short hydrogen bond distances were observed by combined X-ray and neutron diffraction studies$^{26,27}$, although these hydrogen bonds were detected between oxygen atom of the inhibitor and another oxygen atom of the
catalytic aspartate and did not correspond to inter-aspartate hydrogen bond. Moreover, multiple NMR resonances were observed for these complexes, although they were left unassigned. In the crystal structure of \([L-Ala51/51']\)HIV-1 protease with ketomethylene isostere inhibitor KVS-1 at high resolution (1.2 Å) we have observed short distances (2.6 Å) between hydroxyl of hydrated inhibitor and oxygens of Asp25 and Asp25’, similar to complexes of endothiapepsin. Moreover, at this resolution we were able to localize protons based on omit \(F_o-F_c\) map (see Figure 8.2 b, inset). In the complimentary \(^1\)H NMR spectrum we found single resonance at 12.6 ppm (Figure 8.2 a). Such correspondence of chemical shift at 12.6 ppm to hydrogen bond length of 2.6 Å is in a good agreement with systematic correlation of hydrogen bond lengths and chemical shifts done for variety of small molecule model compounds.

How does the data presented here bear on Northrop’s LBHB hypothesis for the catalytic mechanism of the HIV-1 protease? First, the short inter-aspartate side chain carboxyl distal O…O distance in many of the crystal structures with MVT-101 suggests presence of either LBHB or SBHB as in the original proposal by Northrop. It is, however, very difficult to elucidate the exact nature of the hydrogen bond and estimate the barrier for proton shuffling between donor and acceptor atoms, where the most conclusive answer would come from high-resolution neutron diffraction studies.

Second, the observed high delocalization of protons bound to secondary amine (or ammonium) in reduced isostere substrate-derived inhibitor complexes provides support for the mechanistic role of LBHB in assisting proton transfer. Current NMR data provide strong support for proton delocalization. As for the original suggestion that proton tunneling contributes to proton transfer, it is again difficult to verify experimentally. The
unambiguous proof for quantum mechanical tunneling contribution would come from inflated kinetic isotope effects or anomalous (non-Arrhenius) temperature dependence of rate constants as commonly postulated for various chemical reactions of small molecules\textsuperscript{30,31}. As it was discussed before in the literature, the detection of either of the two criteria might be problematic in protein biomolecules due to their complex molecular dynamics\textsuperscript{32-34}. With the current experimental data, we are inclined to a point of view in which the major role of LBHB or SSHB in HIV-1 protease catalysis is proton delocalization (with the possibility of tunneling). Transiently-forming LBHBs or SSHBs in the enzymatic mechanism essentially would help in the reorganization from one transition state to another one.

Northrop’s mechanism is however not correct with respect to all its details, as shown by the data obtained in the present and previous works (see also Chapter 5). In Northrop’s original proposal, the preorganized catalytic site has a symmetric 12-membered H-bonded cyclic structure, comprised of two catalytic aspartates and the nucleophilic water molecule. As we have shown before with the asymmetric chemically synthesized $[L\text{-}\text{Ala51}/D\text{-}\text{Ala51}]$HIV-1 protease – which moreover has native-like catalytic activity – the structure of the preorganized state is asymmetric, where nucleophilic water molecule is hydrogen-bound to catalytic aspartates in a bifurcated, not cyclic, fashion (see Chapter 5). Logically, this makes sense since the peptide substrate is asymmetric as well (it lacks 2-fold symmetry because of the peptide bonds in the backbone). Furthermore, eukaryotic single-chain monomeric aspartic proteases are intrinsically asymmetric\textsuperscript{35}. For eukaryotic aspartic proteases, after gene duplication in the course of evolution, symmetry on the polypeptide level is degraded since it is not
required in the chemical mechanism (and perhaps to enhance sequence specificity). In
the calculations of Piana and Carloni\textsuperscript{15} for the wild-type HIV-1 protease, which are the
keystone of the Northrop’s hypothesis\textsuperscript{14}, in addition to functional asymmetric
conformations, non-productive symmetric ones were also present. In such case
simulations have to be performed for very prolonged periods of time to allow for efficient
conformational sampling.

Therefore, it is more likely that formal chemical mechanism of the enzymatic
catalysis in HIV-1 protease bears most of the features of the original mechanism
proposed by Suguna et al.\textsuperscript{36} (see Figure 1.3 in Chapter 1), with additional contributions
introduced for the role of strong, short hydrogen bonds and proton delocalization.
Overall, we predict that the remaining details of the mechanism can only be filled in by
the application of the advanced QM / MM (quantum mechanics / molecular mechanics)
computations, where we suggest to use asymmetric \([L\text{-}\text{Ala51}/D\text{-}\text{Ala51}']\)HIV-1 protease
as a protein scaffold to ensure that only productive conformational states dominate the
protein dynamics trajectory and in this way to reduce any conformational sampling
problems, which were encountered in the simulations of wild-type HIV-1 protease.
Scheme 8.1. Conformational isomerisations of the ‘flaps’ taking place in the HIV-1 protease complex with the reduced isostere inhibitor MVT-101: (a) In closed-closed conformation (as observed in crystal structures) strong, short hydrogen bond (resonating at 13.3 ppm) is formed between ammonium moiety of inhibitor and one of the aspartates. (b) Isomerizations of flap structures in HIV-1 protease lead to isomerization of the inhibitor via nitrogen inversion of the secondary ammonium group and its subsequent deprotonation leading to delocalization of the corresponding proton (resonating at 16.3 ppm). (c) At higher pH values another proton is abstracted from one of the aspartates, reducing the strength of N-H…O(Asp) bond (resonating at 15.4 ppm). This step is also accompanied by conformational isomerization in flaps, as illustrated by irreversible character of pH titration in [L-Ala51/51’]HIV-1 protease complex with MVT-101.
References:


219 – 245.


27. Coates, L.; Tuan, H.-F.; Tomanicek, S.; Kovalevsky, A.; Mustyakimov, M.; Erskine,


29. This is the first study by combined neutron and X-ray diffraction of HIV-1 protease
complex with KNI-272 inhibitor, where the shortest H-bonds were found to be 2.7 Å.
Shoyama, Y.; Kimura, K.; Matsumura, H.; Sugiyama, S.; Adachi, H.; Takano, K.; Mori,


Total protein synthesis

This thesis work is a demonstration of the application of state-of-the-art total protein synthesis\(^1\) for biophysical studies of an enzyme in order to elucidate the molecular details of its function. Total chemical synthesis was used to prepare a variety of chemical analogues of the HIV-1 protease (see Appendix A), where key residues were replaced by unnatural substituents (with inverted stereocenters, such as \(D\)-amino acids, \(allo\)-isoleucine; stereochemically constrained such as \(\alpha,\alpha'\)-substituted amino acid or backbone-modified, such as \(\alpha\)-hydroxy carboxylic acids), and for site-specific nitroxide-spin- and \(\frac{1}{2}\)-nuclei spin-labeling for EPR and NMR spectroscopies, respectively. In the design of the molecular constructs we undertook syntheses of several 203-amino acid residue ‘covalent dimer’ HIV-1 proteases\(^2\), which are the largest polypeptides chemically synthesized to date. Total synthesis of polypeptides of such length was only possible due to recently developed methods of kinetically-controlled ligation for convergent synthesis of proteins\(^3\). Among experimental approaches available to date, total chemical synthesis is the only methodology which allows robust and efficient preparation of such constructs.

Mechanism of HIV-1 protease catalysis.

HIV-1 protease is an extremely well-studied enzyme with hundreds of crystal structures available in the Protein Data Bank\(^4\). This plethora of structural data was mainly due to the importance of this enzyme as a therapeutic target in the treatment of AIDS\(^5\). Despite the large amount of work done for this enzyme, there is still uncertainty about the chemical mechanism of catalysis in HIV-1 protease. From the work reported in this thesis
it is becoming clear that the correct chemical mechanism should be a fusion of the
general acid–general base mechanism$^6$ and certain parts of the Northrop’s chemical and
kinetic ‘isomechanism’$^7$.

Protein molecular dynamics and catalysis.

Total chemical synthesis adds many more dimensions in order to manipulate the
structure of a protein molecule than is possible with molecular biology methods at the
current state. We explored this opportunity in order to decipher the role of intrinsic
protein dynamics in the chemical mechanism of HIV-1 protease catalysis. By systematic
variation of stereochemistry in the key regions of the protein molecule we derived
dynamics–function correlations, which led us to conclude that protein molecular motions
during the catalysis occur in an asymmetric fashion, despite the two-fold symmetry of
protein molecule on a polypeptide level. Moreover, we detected a correlation of motions
of ‘flap’ structures and catalytic residues, and deduced that catalysis is rate-limited by
preorganization of the active site for optimal electrostatic stabilization of a transition
state$^8$. The preorganized enzyme structure (or ensemble of structures) is asymmetric and
has the nucleophilic water molecule in bifurcated hydrogen-bonded geometry, as
suggested in the general acid-general base mechanism$^6$; this is in contrast with Northrop’s
mechanism$^7$, where predisposed conformers are cyclic and symmetric.

Proton transfer and catalysis.

In the complexes of HIV-1 protease with substrate-derived inhibitors (reduced
isostere MVT-101 and ketomethylene isostere KVS-1) we detected short, strong
hydrogen bonds$^9$ by high resolution X-ray crystallography and NMR spectroscopy. In
addition, properties of low-field $^1$H NMR signals indicated a high level of delocalization
of the putative protons. Thus, experimental data support the hypothesis of D. B. Northrop that proton transfer may be mediated by strong, short hydrogen bonds. The thesis work is thus a prelude for the application of advanced QM/MM (quantum mechanics/molecular mechanics) computations in order to fill in the details in the chemical mechanism of HIV-1 protease enzyme catalysis with atomic resolution and to provide a clear scheme for how protein motions correlate with elementary proton transfer steps leading to cleavage of the peptide bond. *It is clear from this work that such studies should be done with deliberately asymmetric [L-Ala51/D-Ala51]HIV-1 protease as a protein scaffold.*

Experimental validation of proton quantum tunneling contribution to the mechanism.

D. B. Northrop suggested that proton transfer in key steps in catalysis by aspartic proteases occurs by quantum tunneling. Quantum mechanical tunneling is firmly established for biological electron transfer and can take place over distances as large as 25–30 Å. For the hydrogen (i.e. protium) nucleus, the mass is 1840 times larger than that of the electron. Protium can therefore tunnel over a distance of 0.58 Å with the same probability as an electron tunneling over 25 Å. This distance is small, but it is similar in length to a reaction coordinate (i.e. width of a potential energy barrier). This suggests that tunneling might be important in enzymatic proton transfer. However, unlike for biological electron transfer, reports of proton-tunneling in enzymatic reactions have been restricted to a small number of enzyme molecules. Experimental validation of quantum tunneling contribution is difficult. Moreover, kinetic isotope effect (KIE), which serves as the main indicator for significant tunneling contribution in small
molecule reactions, might not attain inflated values (KIE $>$ 7) in enzymatic reactions even if tunneling contribution is significant$^{10-12}$. The possible reduced KIE is caused mainly by the complex molecular dynamics nature of protein structures. Moreover, theories have been developed for vibrationally-assisted (or environmentally-coupled) quantum tunneling$^{12-15}$, where the dynamic fluctuations in the protein molecule are likely to compress (transiently) the width of the potential energy barrier and equalize the vibrational energy levels on the reactant and product sides of the barrier. It was also proposed that hydrogen-transfer can occur by a combination of classical and quantum mechanical behavior$^{16,17}$. Experimental differentiation of these mechanisms will be challenging.

The theory of environmentally-coupled quantum tunneling in biomolecules is primarily based on the theoretical framework for proton transfer in solvent, where the transition state is to be found in the solvent coordinate, and not in the proton transfer coordinate$^{18,19}$. The reaction path involves an initial solvent fluctuation to bring the reactant and product state into resonance subsequent to which the proton tunnels through the adiabatic electronic barrier found in the proton transfer coordinate (nonadiabatic transfer). In the case, where there is no electronic barrier in the proton transfer coordinate, the proton moves across the surface in a vibrational motion not requiring the penetration of an electronic barrier (adiabatic proton transfer). Peters and colleagues$^{20}$ explored this theory experimentally by studying kinetics of proton transfer within variously substituted benzophenones / $N, N$-dimethylaniline triplet contact radical ion pairs and found both a normal and an inverted region in the correlation of the rate constants of electron transfer with driving force, kinetic behavior that is consistent with
Marcus nonadiabatic electron transfer theory (incorporating tunneling)\textsuperscript{21}. Driving force or energetics for proton transfer was varied by employing a variety of substituents (CH$_3$O, CH$_3$, F, Cl) at the $p,p'$-positions of benzophenone or by utilization of a variety of substituted anilines (see Figure 9.1 a-c).

Total synthesis of a series of analogues of HIV-1 protease, where active site residues Asp\textsuperscript{25}Thr\textsuperscript{26}Gly\textsuperscript{27} are replaced by heterocyclic phenolic mimics, would allow such linear free energy correlations to be obtained for this particular enzyme system. By varying X-substituent in the phenolic ring (see Figure 9.1 d), the driving force for proton transfer can be altered. If both normal and inverse regions in the correlation of rate constants are found, this will be unambiguous proof for proton tunneling contribution in this system. One may speculate that such studies might also reveal some additional unexpected effects not observed for small molecule system.

Towards new inhibitors of HIV-1 protease.

Current ‘HAART’ treatments for AIDS include inhibiting the protease and reverse transcriptase of HIV with small molecule drugs, but their effectiveness can be diminished by the occurrence of resistance mutations in the virus. New inhibitors of the HIV protease and reverse transcriptase that will inhibit not only wild-type but also mutated forms of the virus’ proteins are thus needed. Altman et al.\textsuperscript{22} have recently proposed the so-called ‘substrate-envelope hypothesis’ for inhibitors of HIV-1 protease, meaning that inhibitors that stay within a consensus substrate shape should be less likely to induce resistance mutations than those that exceed the envelope and provide handles for escape mutation to lower inhibitor affinity and selectivity. If a drug molecule makes the same or fewer interactions and contacts with the target as the substrate, it could be
difficult for mutations that block inhibitor binding yet maintain substrate recognition to evolve. Ideally, escape mutants should never be selected because they would render the pathogen nonviable.

The most important part in the inhibitor of aspartic proteases is an isosteric replacement moiety for the scissile amide bond (see Figure 9.2). It can have similar topology to native substrate or the tetrahedral intermediate in the proteolytic reaction (such as reduced amine, hydroxyethylene, dihydroxyethylene) or have additional atoms inserted increasing separation between Cα atoms in the scissile bond replacement junction (such as hydroxyethylamine, 2-hydroxyproplyamine, statines and statones). In all the cases mentioned above the shape and atomic composition is quite different from the substrate or tetrahedral intermediate. Therefore, such isosteres have deficiency in their potential to solve the problem of drug-resistance based on ‘substrate-envelope’ hypothesis.

In contrast, in the hydrated form as bound to the enzyme, the substrate-derived inhibitor KVS-1 is fully isosteric with the tetrahedral intermediate of the proteolytic reaction (see Chapter 3, Figure 3.3), which makes it very attractive to design new inhibitors, which would be expected to effectively bind to the various mutants of HIV-1 protease. Originally ketomethylene isosteres were synthesized by Marshall and colleagues, who also determined their high inhibitor binding efficiencies. In this thesis work we have for the first time solved high-resolution X-ray structures for the substrate-derived ketoisostere inhibitor KVS-1 and have shown that the ketomethylene moiety gets converted into a gem-diol (or hydroxyl-oxyanion) only in the active site of the enzyme, and not in solution. Enzyme-catalyzed hydration makes the substrate-derived ketoisostere
a truly mechanism-based inhibitor and may potentially contribute to very high selectivity for inhibition of HIV-1 protease among other aspartic proteases. Further exploration of this motif is thus needed in an attempt to solve the long-standing and very demanding problem of effective inhibition of drug-resistant mutants of HIV-1 protease.
Figure 9.1. (a) Potential energy surfaces for solvent fluctuations and proton transfer as a function of driving force, $\Delta G_{RXN}$: $\lambda_S$ is the solvent reorganization energy; $S_R$ is the solvent structure around reactants; $S_{TS}$ is the solvent structure at the transition state in the solvent coordinate; $S_P$ is the solvent structure around the products. Proton tunneling occurs at solvent configuration $S_{TS}$. (b) Benzophenone / $N,N$-dimethylaniline triplet contact radical ion pair system where proton transfer process was studied by Peters and co-workers. (c) Plots of experimental rate constants for proton transfer versus driving force (kcal/mol) defined as $-\Delta G_{RXN}$ for the solvent tetrahydrofuran. Experimental data for various $p,p'$-substituted benzophenones and $N,N'$-substituted anilines showing normal and inverted regions for rate dependence. (d) Heterocyclic mimic for the conserved Asp$^{25}$Thr$^{26}$Gly$^{27}$ motif in HIV-1 protease, where systematic correlation for driving force for proton transfer can be established by varying the X-substituent on the phenolic fragment. Figures 9.1 a-c are reproduced with permission from ref.20.
Figure 9.2. Some transition state mimicking isosteres to replace the scissile amide bond in aspartic protease inhibitors.
References:


**Appendix A.** List of chemically-synthesized HIV-1 protease (HIV-PR) and its analogues.

**Table A.1.** Chemically-synthesized HIV-1 protease (HIVPR) and its analogues.

<table>
<thead>
<tr>
<th>Chemically-synthesized enzyme</th>
<th>Mass, Da</th>
<th>Activity, $k_{cat}/K_m$, s$^{-1}$µM$^{-1}$</th>
<th>X-ray structures, PDB ID</th>
<th>Biophysical study*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>wild-type [1-99]HIVPR ($\psi$-Gln41/41')</td>
<td>10750.4</td>
<td>10749.7</td>
<td>0.93</td>
</tr>
<tr>
<td>2.</td>
<td>wild-type [1-99]HIVPR (S-Me-Cys41/41')</td>
<td>10706.5</td>
<td>10706.7</td>
<td>0.93</td>
</tr>
<tr>
<td>3.</td>
<td>'covalent dimer' [Gly51;Gly51']HIVPR</td>
<td>21869.8</td>
<td>21869.8</td>
<td>0.93</td>
</tr>
<tr>
<td>4.</td>
<td>[L-Ala51;D-Ala51']HIVPR</td>
<td>21897.9</td>
<td>21897.9</td>
<td>0.67</td>
</tr>
<tr>
<td>5.</td>
<td>[L-Ala51;Gly51']HIVPR</td>
<td>21883.5</td>
<td>21883.8</td>
<td>0.47</td>
</tr>
<tr>
<td>6.</td>
<td>[Gly51;D-Ala51']HIVPR</td>
<td>21884.0</td>
<td>21883.8</td>
<td>0.14</td>
</tr>
<tr>
<td>7.</td>
<td>[L-Ala51;L-Ala51']HIVPR (S-Me-Cys41/41')</td>
<td>10720.9</td>
<td>10720.7</td>
<td>0.073</td>
</tr>
<tr>
<td>8.</td>
<td>[Gly51;Aib51']HIVPR</td>
<td>21897.5</td>
<td>21897.9</td>
<td>0.042</td>
</tr>
<tr>
<td>9.</td>
<td>[D-Ala51;D-Ala51']HIVPR (S-Me-Cys41/41')</td>
<td>10720.5</td>
<td>10720.7</td>
<td>0.011</td>
</tr>
<tr>
<td>10.</td>
<td>[Aib51;Aib51']HIVPR (S-Me-Cys41/41')</td>
<td>10734.9</td>
<td>10734.7</td>
<td>0.001</td>
</tr>
<tr>
<td>11.</td>
<td>$^{15}$N,$^{13}$C-isotope labeled [L-Ala51;L-Ala51']HIVPR</td>
<td>10784.0</td>
<td>10784.2</td>
<td>n/a</td>
</tr>
<tr>
<td>12.</td>
<td>$^{15}$N,$^{13}$C-isotope labeled [D-Ala51;D-Ala51']HIVPR</td>
<td>10784.0</td>
<td>10784.2</td>
<td>n/a</td>
</tr>
<tr>
<td>13.</td>
<td>$^{15}$N,$^{13}$C-isotope labeled [Aib51;Aib51']HIVPR</td>
<td>10792.0</td>
<td>10792.4</td>
<td>n/a</td>
</tr>
<tr>
<td>14.</td>
<td>$^{13}$C-(1,4)-Asp25 wild-type HIVPR</td>
<td>10752.2</td>
<td>10751.7</td>
<td>n/a</td>
</tr>
<tr>
<td>15.</td>
<td>$^{13}$C-(1,4)-Asp25 [L-Ala51;L-Ala51']HIVPR</td>
<td>10765.1</td>
<td>10765.7</td>
<td>n/a</td>
</tr>
</tbody>
</table>
**Table A.1 (continued).** Chemically-synthesized HIV-1 protease (HIVPR) and its analogues.

<table>
<thead>
<tr>
<th>Chemically-synthesized enzyme</th>
<th>Mass, Da</th>
<th>Activity, $k_{cat}/K_m$, $s^{-1} \mu M^{-1}$</th>
<th>X-ray structures, PDB ID</th>
<th>Biophysical study*</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. $^{13}$C-(1,4)-Asp25 $[D\text{-Ala51};D\text{-Ala51}']$HIVPR</td>
<td>10765.8</td>
<td>10765.7</td>
<td>n/a</td>
<td>NMR</td>
</tr>
<tr>
<td>17. $^{13}$C-(1,4)-Asp25 $[\text{Aib51};\text{Aib51}']$HIVPR</td>
<td>10779.8</td>
<td>10779.8</td>
<td>n/a</td>
<td>NMR</td>
</tr>
<tr>
<td>18. wild-type MTSSL55/55' nitroxide-labeled HIVPR</td>
<td>10909.0</td>
<td>10909.0</td>
<td>0.93</td>
<td>EPR</td>
</tr>
<tr>
<td>19. Asn25/25'; MTSSL55/55' nitroxide-labeled HIVPR</td>
<td>10907.6</td>
<td>10908.0</td>
<td>no activity</td>
<td>EPR</td>
</tr>
<tr>
<td>20. Asn25/25'; MTSSL55/55' $[L\text{-Ala51};L\text{-Ala51}']$HIVPR</td>
<td>10921.9</td>
<td>10922.0</td>
<td>no activity</td>
<td>EPR</td>
</tr>
<tr>
<td>21. Asn25/25'; MTSSL55/55' $[D\text{-Ala51};D\text{-Ala51}']$HIVPR</td>
<td>10920.0</td>
<td>10922.0</td>
<td>no activity</td>
<td>EPR</td>
</tr>
<tr>
<td>22. Asn25/25'; MTSSL55/55' $[\text{Aib51};\text{Aib51}']$HIVPR</td>
<td>10936.0</td>
<td>10936.0</td>
<td>no activity</td>
<td>EPR</td>
</tr>
<tr>
<td>23. $[\text{alloIle50};\text{alloIle50}']$HIVPR</td>
<td>10747.8</td>
<td>10749.7</td>
<td>0.32</td>
<td>MVT-101 KVS-1</td>
</tr>
<tr>
<td>24. $[O\text{-Gly51};O\text{-Gly51}']$HIVPR</td>
<td>10751.2</td>
<td>10750.7</td>
<td>0.21</td>
<td>MVT-101 JG-365</td>
</tr>
<tr>
<td>25. $[O\text{-Gly52};O\text{-Gly52}']$HIVPR</td>
<td>10750.8</td>
<td>10750.7</td>
<td>0.26</td>
<td>n/a</td>
</tr>
<tr>
<td>26. $[O\text{-Ile50};O\text{-Ile50}']$HIVPR</td>
<td>10750.7</td>
<td>10750.7</td>
<td>0.009</td>
<td>MVT-101 KVS-1</td>
</tr>
<tr>
<td>27. $[\text{Ile50};O\text{-Ile50}']$HIVPR</td>
<td>21870.5</td>
<td>21870.8</td>
<td>0.1</td>
<td>MVT-101 KVS-1</td>
</tr>
<tr>
<td>28. $[\text{Asn25};\text{Asn25}']$HIVPR</td>
<td>10748.6</td>
<td>10748.7</td>
<td>no activity</td>
<td>3DCR(KVS-1) MVT-101</td>
</tr>
<tr>
<td>29. Flap-truncated HIVPR</td>
<td>21440.3</td>
<td>21439.3</td>
<td>0.014</td>
<td>n/a</td>
</tr>
<tr>
<td>30. Biotin-labeled ‘covalent dimer’ HIVPR</td>
<td>22341</td>
<td>22338.2</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

*NMR = nuclear magnetic resonance; EPR = electron paramagnetic resonance; SPR = surface plasmon resonance; MD = molecular dynamics simulation; SMF = single-molecule fluorescence.
Appendix B. Supporting information for chapter 2.

Reagents. Boc-amino acids, 4-Me-benzylhydrylamine-resins and –OCH₂-Pam-resins (free α-carboxyl peptides) were purchased from Peptides International, Louisville, Kentucky; Boc-L-thiazolidine-4-carboxylic acid from NovaBiochem, San Diego; diethyl ether from Fisher; HF from Matheson Tri-Gas; methoxylamine hydrochloride, p-cresol, sodium 2-mercaptoposulphonate (MESNA), triisopropylsilane from Sigma-Aldrich; 4-mercaptophenylacetic acid was obtained from Sigma-Aldrich and purified by HPLC before use; tris(2-carboxyethyl)phosphine hydrochloride (TCEP) from Fluka; trifluoroacetic acid (TFA) from Halocarbon Products, New Jersey.

Peptide Synthesis. Peptides were prepared manually by “in situ neutralization” Boc-chemistry stepwise solid phase peptide synthesis (SPPS).[1] Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH₃Bzl), Glu(OcHex), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(Br-Z), Trp(CHO), His(Bom). The 1,3-thiazolidine-4-carboxyl (Thz) group was introduced to protect the N-terminal Cys of the (Cys-Gly₄-B₁-B₄₀)-(α-thioalkylester) peptide segment, and Boc-L-thiazolidine-4-carboxylic acid was used for peptide synthesis. After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin support by treatment with anhydrous HF containing 10% (v/v) p-cresol for 1 h at 0 °C. After evaporation of the HF under reduced pressure, crude products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA.

LC and LC-MS analysis. Analytical RP-HPLC was performed on an Agilent 1100 system with in-house packed C-4 and C-18 silica columns (2.1 × 50 mm, Microsorb 300 Å, 3 μm) at flow rate of 0.5 mL/min. Peptides were eluted from the column using a gradient of acetonitrile/0.08% TFA (solvent A) versus water/0.1% TFA (solvent B). Peptide masses were obtained using on-line electrospray MS detection with an Agilent 1100 LC/MSD ion trap. Masses of synthetic peptides were:
(A1-A40)-α-COSCH₂CH₂Arg₄amide, obs. 5125.2 ± 0.3 Da, calc. 5125.0 Da;
(Cys⁴⁴¹-A⁹⁹)-α-COSCH₂CH₂Arg₄amide, obs. 7066.3 ± 0.5 Da, calc. 7065.5 Da;
Thz²⁰¹-Gly₄-(B₁-B₄⁰)-α-COSCH₂CH₂Arg₄amide, obs. 5468.5 ± 0.4 Da, calc. 5468.4 Da;
(Cys⁶⁴¹-B₉⁹)-COOH, obs. 6354.1 ± 0.6 Da, calc. 6353.6 Da. Calculated masses were of average isotope compositions.

Preparative HPLC of crude peptides after SPPS was performed on Agilent 1100 prep system on a Agilent C-3 (250 × 22 mm, 300Å, 7µm) column or on in-house packed columns C4 (250 × 7 mm, Microsorb, 300Å, 5µm) and C8 (250 × 10 mm, Microsorb, 300Å, 10µm). Peptides from SPPS and from ligations were eluted from the column using an appropriate shallow gradient of acetonitrile/0.08% TFA versus water/0.1% TFA. Fractions containing the desired purified peptide product were identified by analytical LC and LC-MS, combined and lyophilized. Segments (Cys⁴⁴¹-A⁹⁹)-α-COSCH₂CH₂Arg₄amide and (Cys⁶⁴¹-B₉⁹)-COOH were purified two times.
Conversion of (A1-A40)-(α-thioalkylester) to (A1-A40)-(α-thioarylester). 31.2 mg (6.1 µmol) of peptide (A1-A40)-αCOSCH₂CH₂Arg₄ amide was dissolved in 5 mL buffer (6 M Gn·HCl, 0.1 M Na₂HPO₄) containing 109 mg (0.65 mmol) 4-mercaptophenylacetic acid and 17 mg (0.06 mmol) TCEP. The pH was adjusted to 5.8 and kept for 4 hours, then purified by RP-HPLC on C8 column. LC-MS: obs. 4563.3 Da, calc. 4563.3 Da. Isolated yield 8.2 mg or 1.8 µmol (29.5 %).

\textbf{Figure B.1.} Analytical HPLC traces ($\lambda = 214 \text{ nm}$) of transthioesterification reaction (C4 column). A) $t < 1 \text{ min}$. B) $t = 4 \text{ h}$. C) After HPLC purification (C18 column). Insert is mass-spectrum of (A1-A40)-αCOSC₆H₄CH₂COOH integrated over elution time of target peak. Asterisk shows elution of 4-mercaptophenylacetic acid and other buffer components.
Native chemical ligation of Thz-Gly₄-(B1-B40)-CO₂CH₂CH₂Arg₄ and (Cys⁸⁴ⁱ-B⁹⁹).

7.6 mg (1.39 µmol) Thz²⁰¹-Gly₄-(A₁-A₄₀)-CO₂CH₂CH₂Arg₄.amide and 8 mg (1.3 µmol) (Cys⁸⁴¹-B⁹⁹).COOHwere mixed in 1 mL of buffer containing 8 M Gln·HCl, 0.2 M Na₂HPO₄ and 22 mM TCEP. The 4-mercaptophenylacetic acid was added at 70 mM concentration. After 3 h, the reaction mixture was deluted to 3.5 mL with buffer and 50 mg (0.36 mmol) of 2-bromoacetamide was added (pH 6.7). After 15 min 62 mg (0.38 mmol) of MESNA were added to quench excess of 2-bromoacetamide. 58.5 mg (0.7 mmol) of MeONH₂·HCl (c 0.2 M) was added to deprotect (Thz)Cys. After 8 h of reaction, the product was purified by RP-HPLC. LC-MS: obs. 11137.5 ± 0.7 Da, calc. 11137.1 Da). Isolated yield 5.5 mg or 0.5 µmol (38%).

Figure B.2. Analytical HPLC traces (λ = 214 nm) of native chemical ligation of Thz²⁰¹-Gly₄-(B1-B40)-CO₂CH₂CH₂Arg₄.amide and (Cys⁸⁴¹-B⁹⁹).COOH. A) t < 1 min. B) t = 3 h. C) After HPLC purification. Insert is mass-spectrum of Cys²⁰¹-Gly₄-(B1-B⁹⁹).COOH (obs. 11137.5 ± 0.7 Da, calc. 11137.06 Da) integrated over elution time of target peak. Asterisk shows elution of 4-mercaptophenylacetic acid and other buffer components.
**Folding.** 0.5 mg of tethered construct of HIV-1 protease was dissolved in 3 mL of 6 M Gn·HCl, 50 mM NaOAc, pH 6.0 buffer and dialyzed first against 2 M Gn·HCl, 50 mM NaOAc, 10% (v/v) glycerol, pH 5.6 and then against 50 mM NaOAc, 10% (v/v) glycerol, pH 5.6. Yield is 29.1% based on final concentration of protein determined by absorbance at 280 nm using a calculated extinction coefficient of 25120 M⁻¹cm⁻¹.

**Enzymology.** Kinetic parameters were determined through assay with fluorogenic substrate Abz-Thr-Ile-Nle-Phe(p-NO₂)-Gln-Arg.amide (Abz = 2-aminobenzoyl),[2] which was added at various concentrations. Fluorimeter Jobin Yvon FluoroMax-3 (cell 1×1 cm) was used. Assay was performed in 50 mM NaOAc, 0.2 M NaCl buffer with 1 % (v/v) DMSO at pH 5.6 and 37 ºC. Values of $k_{cat} 23.4 ± 0.2$ s⁻¹, $K_m 27 ± 1.4$ µM were determined by nonlinear least square fitting of the initial rates to the Michaelis-Menten kinetic model with the help of SigmaPlot 9.0 software.

![Figure B.3](image)

**Figure B.3.** Enzymatic activity of the ‘covalent dimer’ HIV-1 PR. A) Initial rates versus concentration of fluorogenic substrate. B) Lineweaver-Burk plot.
FT-ICR experiments were performed at Chicago Biomedical Consortium proteomics facilities at University of Illinois at Chicago on Thermo Finnigan LTQ-FT LC/MS/MS instrument (Superconducting magnet 7.0 Tesla). Spectra were registered in LC-MS mode. Peak in total ion chromatogram was integrated and analyzed. Spectra were processed with Thermo Xcalibur v. 1.4 software. Simulation of isotopic pattern was done in IsoPro v. 3 and Gaussian fit and mathematic analysis was performed in Origin v. 7.0. Deconvolution to get average molecular weight was done with ProMass v. 2.5 software.

Figure B.4. Experimental (A) and simulated (B) FT-ICR data for the most abundant charge state 23H+. C) Superposition of isotope pattern for charge state 23H⁺ for experimental (solid line) and simulated (dashed line) spectra. Superposition is based on Gaussian best fit of maxima for experimental (red line, circles) and theoretical (blue line, rhombes) peaks.
**Crystallography.** Crystals were grown at 20 °C by the hanging drop vapour diffusion method from a well solution consisting of 0.1 M citrate, 0.2 M NaH₂PO₄, 30% (w/v) saturated ammonium sulfate, 10% (v/v) DMSO, pH 6.0. Protein solution (~0.1 mM) was preincubated with a 30-fold molar excess of MVT-101, and was then mixed in different (v/v) ratios with well solution. Crystals grew within 1-4 days and were frozen in liquid nitrogen using mineral oil as cryoprotectant. Data were collected at the GM/CA-CAT (23ID) beamline at the Advanced Photon Source (Argonne National Laboratory). Crystal structure was solved by molecular replacement with the help of MolRep v. 8.1 and refined by Refmac v.5 (Collaborative Computational Project, Number 4. 1994. **“The CCP4 Suite: Programs for Protein Crystallography.”** Acta Cryst. D 1994, 50, 760-763). Data were deposited to Protein Databank (PDB ID 2O40).
Table B.1. Data collection and refinement statistics (highest resolution is in parenthesis).

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>51.326</td>
</tr>
<tr>
<td>b (Å)</td>
<td>58.306</td>
</tr>
<tr>
<td>c (Å)</td>
<td>61.699</td>
</tr>
<tr>
<td>α = β = γ (°)</td>
<td>90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 – 1.65</td>
</tr>
<tr>
<td>Rmerge</td>
<td>5.5 (74.9)</td>
</tr>
<tr>
<td>I/σI</td>
<td>14.6 (2.67)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.8 (4.6)</td>
</tr>
</tbody>
</table>

| Refinement                         |       |
| Resolution (Å)                     | 20 – 1.65 |
| Completeness (%)                   | 98.7   |
| No. reflections                    |       |
| Work/free set                      | 21383 / 1163 |
| Rwork/Rfree                        | 19.1 / 23.7 |
| No. atoms                          |       |
| Protein                            | 1597  |
| Water                              | 57    |
| B-factors (Å²)                     |       |
| Protein                            | 24.7  |
| Inhibitor                          | 27.9  |
| Water                              | 30.6  |
| R.m.s. deviations                  |       |
| Bond lengths (Å)                   | 0.017  |
| Bond angles (°)                    | 1.599  |

References:


Appendix C. Supporting information for chapter 3.

Reagents. Boc-amino acids, 4-Me-benzylhydrylamine-resin (MBHA-resin) and –OCH2-Pam-resins (free α-carboxyl peptides) were purchased from Peptides International, Louisville, Kentucky; diethyl ether from Fisher; HF from Matheson Tri-Gas; p-cresol, triisopropylsilane, 2-bromoacetamide from Sigma-Aldrich; 4-mercaptophenylacetic acid was obtained from Sigma-Aldrich and purified by HPLC before use; tris(2-carboxyethyl)phosphine hydrochloride (TCEP) from Fluka; trifluoroacetic acid (TFA) from Halocarbon Products, New Jersey, (DIEA) from Applied Biosystems.

Peptide Synthesis. Peptides were prepared manually by “in situ neutralization” Boc-chemistry stepwise solid phase peptide synthesis (SPPS). Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), Glu(OcHex), Lys(2Cl-Z), Thr(Bzl), Tyr(Br-Z), Trp(CHO), His(Bom). After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin support by treatment with anhydrous HF containing 10% (v/v) p-cresol for 1 h at 0 ºC. After evaporation of the HF under reduced pressure, crude products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA. Preparative HPLC of crude peptides after SPPS was performed on an Agilent 1100 prep system using a Agilent C-3 (250 × 22 mm, 300Å, 7µm) column or on in-house packed columns C4 (250 × 7 mm, Microsorb, 300Å, 5µm) and C8 (250 × 10 mm, Microsorb, 300Å, 10µm). Peptide products from SPPS and from ligations were eluted from the column using an appropriate shallow gradient of acetonitrile/0.08% TFA versus water/0.1% TFA. Fractions containing the desired purified peptide product were identified by analytical LC and LC-MS, combined and lyophilized. Segment (Cys41-99)-COOH was purified two times.

Total chemical synthesis of 99-residue HIV-1 protease monomers (see Figure C.1). For native chemical ligation, segment (1-40)-αCOSCH2CH2Arg4 (11.5 mg, 2.2 µmol) and (Cys41-99)-COOH (12.4 mg, 2 µmol) were dissolved in aqueous buffer (2.5 mL)
containing 6 M Gn·HCl, 0.2 M Na₂HPO₄ and 10 mM TCEP and 50 mM 4-mercaptophenylacetic acid at pH 7.0. After three hours, 42 mg (300 µmol) of 2-bromoacetamid were added and reaction mixture was stirred for 15 min at pH 6.7. Deformylation was performed by treatment with a mixture of 2-mercaptopethanol and piperidine (1:1 (v/v), 4 mL) on ice for 15 min, and then neutralizing with HCl. The reaction mixture was diluted two-fold with buffer (6 M Gn·HCl, 0.2 M Na₂HPO₄) and purified by RP-HPLC. Yield of isolated product 7.1 mg (0.66 µmol, 31%).

**Analytical RP-HPLC** was performed on an Agilent 1100 system with in-house packed C-4 and C-18 silica columns (2.1 × 50 mm, Microsorb 300 Å, 3 µm) at flow rate of 0.5 mL/min or Agilent C-3 column (4.6×150 mm, Zorbax, 300Å, 3.5 µm) at flow rate of 1.0 mL/min. Peptides were eluted from the column using a gradient of acetonitrile/0.08% TFA (solvent A) versus water/0.1% TFA (solvent B). Peptide masses were obtained using on-line electrospray MS detection with an Agilent 1100 LC/MSD ion trap.

**Folding.** Typical procedure: 1 mg of HIV-1 protease was dissolved in 3 mL of 6 M Gp·HCl, 50 mM NaOAc, pH 6.0 buffer and dialyzed first against 2 M Gp·HCl, 50 mM NaOAc, pH 5.6 and then against 10 mM NaOAc, pH 5.6. Yields were in 45-60% based on final concentration of protein determined by absorbance at 280 nm using a calculated extinction coefficient of 25120 M⁻¹cm⁻¹. After folding, solutions were concentrated using Centricon Centrifugal Devices (3000 MW cutoff, purchased from Millipore).
**Figure C.1.** Analytical HPLC traces ($\lambda = 214$ nm) for two-segment native chemical ligation en route to the synthetic wild-type HIV-1 protease. a) $t = 0$ min, pH 5.6; 1 indicates (1-40)-$\alpha$COSCH$_2$CH$_2$Arg, 2 indicates (Cys$^{41}$-99)-COOH, 3 is product of hydrolysis (1-40)-COOH. b) $t = 3$ h, pH 7.0, 4 indicates 4-mercaptophenylacetic acid, 5 is product of transthioesterification (1-40)-$\alpha$COSC$_6$H$_4$CH$_2$COOH, 6 is product of ligation (1-99)-COOH.

**Synthesis of Ac-Thr-Ile-Nle$\Psi$[COCH$_2$]Nle-Gln-Arg-NH$_2$.** The peptide was synthesized on MBHA polystyrene resin using manual in situ neutralization Boc-chemistry protocols for stepwise SPPS, on a 0.4 mmol scale. Boc-Nle$\Psi$[COCH$_2$]Nle-OH was prepared following the reported procedure.$^2$ After removal of the N-terminal Boc-group, the peptide was neutralized with 5% DIEA in DMF and acetylated using 0.4-molar acetic anhydride in dichloromethane. The peptide was cleaved from the resin and
simultaneously deprotected by treatment at 0 °C for 1 h with anhydrous HF containing 5% *p*-cresol as scavenger. After removal of HF by evaporation under reduced pressure, the crude peptide was precipitated and washed with diethyl ether, then dissolved in 50% aqueous CH$_3$CN containing 0.1% TFA and lyophilized. The crude lyophilized peptide was obtained in 131 mg amount as a mixture (1:1) of the two diastereomers. The diastereomers were purified by reverse phase HPLC on in house packed Silicycle spherical C-18, 10 × 250 mm column at 40 °C using a gradient of 18%-28% CH$_3$CN over 60 minutes, at a flow rate of 10 mL/min. KVS-2 was purified two times. Fractions containing the desired products were identified by LC-MS, combined and lyophilized to furnish pure diastereomer KVS-1 (MS: found 783.8 Da (M+H$^+$), calc. 783.5 Da) in 32.0 mg and diastereomer KVS-2 (MS: 783.9 Da (M+H$^+$), calc. 783.5 Da) in 20.0 mg, respectively. See Figure C.2.

**Figure C.2.** Analytical HPLC profiles ($\lambda$ = 214 nm) together with ESI-MS data (insert) corresponding to each diastereomer. The chromatographic separations were performed using a linear gradient (5–65%) of buffer B in buffer A over 15 min (buffer A = 0.1% trifluoroacetic acid (TFA) in water; buffer B = 0.08% TFA in acetonitrile) on in house packed C-18, 2.1 × 150 mm column at 40 °C. a) Diastereomer KVS-1; b) diastereomer KVS-2.
Crystallization conditions, structure solution and refinement details for ‘wild-type’ HIV-1 protease and its inactive [D25N] analogue with KVS-1 inhibitor.

Crystals were grown at 20 ºC by the hanging drop vapour diffusion method from a well solution consisting of 0.1 M citrate, 0.2 M NaH₂PO₄, 30% (w/v) saturated ammonium sulfate, 10% (v/v) DMSO, pH 6.0. Protein solution (~0.1 mM) was preincubated with a 30-fold molar excess of KVS-1, and was then mixed in different (v/v) ratios with well solution. Crystals grew within 1-2 days and were frozen in liquid nitrogen using mineral oil as cryoprotectant. Crystal structure was solved by molecular replacement with the help of MolRep v. 8.1 and refined by Refmac v.5 (Collaborative Computational Project, Number 4. 1994. “The CCP4 Suite: Programs for Protein Crystallography.” Acta Cryst. D 1994, 50, 760).

References.
Appendix D. Supporting information for chapter 4.

1. General information about reagents and procedures including molecular modeling and MD simulations.

Reagents, procedures for peptide synthesis, analytical methods, folding protocol, and enzyme assays can be found in recently published articles. Figure 4.2a has been prepared on the basis of molecular modeling. ‘Closed/closed’, ‘semi-open/semi-open’ conformers were built starting with crystal structures for ligand-bound (PDB ID 4HVP, in the Figure 4.2a inhibitor is removed) and the unliganded ‘apo’-form of HIV-1 protease (PDB ID 3HVP); asymmetric ‘semi-open/closed’ conformer was modeled on the basis of crystal structure of inactive D25N HIV-1 protease enzyme with peptidic substrate, where one flap is fully closed and another one is in ‘semi-open’ conformation (PDB ID 1FNS). The ‘open/open’ conformer is a ‘snapshot’ from an explicit water MD simulation for apo-HIV-1 protease. Distances (Cβ(Lys55)…Cβ(Lys55’)) in these four structures are 23.6 Å, 28.3 Å, 29.6 Å and 39.4 Å, respectively. The MTSSL-spin-label has been added to both Cys55 side chains and the optimized dihedral angles for its structure have been chosen on the basis of previous comprehensive simulations and further energy minimized. Molecular dynamics (MD) simulations for ‘closed/closed’, ‘closed/semi-open’ and ‘semi-open/semi-open’ conformers of MTSSL-spin labeled HIV-1 protease were carried out in explicit solvent for 5 ns, where the entire backbone of protease was kept fixed, but the side chain of MTSSL neighbor residues 42-58 were allowed to move freely. All minimization and MD steps were performed using the Amber 9.0 package with Amber force fields. The missing interaction parameters in the MTSSL spin label were introduced using antechamber tools in Amber. The atom-centered point charges on MTSSL were determined via fits to the electrostatic potentials obtained from the calculated wave functions using the CHelpG subroutine of Gaussian 03. The nitroxide-to-nitroxide O…O distances were found to be 29 ± 5 Å, 31 ± 4 Å and 34 ± 5 Å for ‘closed/closed’, ‘semi-open/closed’ and ‘semi-open/semi-open’ conformers, respectively, which is in a good qualitative agreement with experimental distances.
2. Total chemical synthesis of spin-labeled active and inactive [D25N] HIV-1 proteases.

The amino acid sequence for MTSSL-spin-labeled constructs is based on sequence for SF2 isolate:

PQITLWKRPL<sup>10</sup>VTIRIGQQLK<sup>20</sup>EALLXaaTGADD<sup>30</sup>TVIEENleNLPG<sup>40</sup>ψGlnWKPK
NleIGGi<sup>50</sup>GGFiCys(MTSSL)VRQYD<sup>60</sup>QIPVELAbuGHK<sup>70</sup>AIGTVLVGPT<sup>80</sup>PVNIIGRN
LL<sup>90</sup>TQIGAbuTLNF<sup>99</sup>. Unnatural amino acids are in italics in three-letter code. Nle = norleucine, Abu = α-aminobutyric acid, ψ-Gln = pseudo-homoglutamine, Xaa = either Asp or Asn. Cys(MTSSL) = spin-label-attached cysteine. Ligation site is underlined.
P₁-Asn/Asp²⁵-G⁴⁰.thioester  Cys¹⁴-Cys⁵⁵(Acm)-Phe⁹⁹

1) NCL
2) alkylation
3) CHO-removal
4) HPLC prep

P₁-Asn/Asp²⁵-ψGln⁴¹-Cys⁵⁵(Acm)-Phe⁹⁹

1) Hg(OAc)$_₂$
2) HPLC purification
3) Nitroxide (MTSL) labelling
4) HPLC purification

P₁-Asn/Asp²⁵-ψGln⁴¹-Cys⁵⁵-Phe⁹⁹

Figure D.1 illustrates experimental steps in the total chemical synthesis of spin-labeled wild-type HIV-1 protease and its [D25N] inactive analogue. In the solid-phase peptide synthesis of the 41-99 segments, residue Lys⁵⁵ was replaced by Cys(Acm)⁵⁵. Native chemical ligation (NCL) was carried out with the corresponding 1-41.thioester (containing either Asn²⁵ or Asp²⁵) at pH 7.0 in 6 M Gn·HCl, 0.1 mM Na₂HPO₄ and 20 mM 4-mercaptophenylacetic acid; then, S-alkylation of the ligation site Cys⁴¹ was carried out with a 50-fold molar excess of 2-bromoacetamide at pH 6.7, and the reaction product was subjected to deformylation of formyl protecting groups of Trp⁶ and Trp⁴² by treatment with equal volume of 1:1 (v/v) mixture of piperidine and 2-mercaptoethanol. The product polypeptide was then purified by RP-HPLC. The Cys(Acm) group was removed for 0.1 mM solution of polypeptide in 1:1 (v/v) water/AcOH mixture in the presence of 5-7 mol. excess of Hg(OAc)$_₂$ (note: AgOAc and AgOTf were found to be inefficient catalysts for Acm removal even at very high concentrations). After 2 h treatment, the reaction was quenched with an equal volume of 6 M Gn·HCl, 0.2 M DTT, pH 7.0 solution. After subsequent RP-HPLC purification, nitroxide labeling was performed by dissolving polypeptide in 6M Gn·HCl, 0.1 M Na₂HPO₄ buffer at pH 6.7 (c 0.2 mM) and adding 10-mol. excess of 75 mM (1-oxyl-2,2,5,5-tetramethyl-Δ₃-pyrroline-3-methyl) methanethiosulfonate (MTSSL) solution in ethanol. After 15 min labeling was complete and the final product was purified by RP-HPLC. Final products (3-5 mg of lyophilized powder) were characterized by analytical HPLC and ESI mass-spectrometry (see Figure D.2).
Figure D.2. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of spin-labeled constructs of (a) active HIV-1 protease, and (b) its [D25N] inactive analogue.
3. Double electron electron resonance (DEER) experiments were carried out using a Bruker ELEXYS E580 X-band pulsed spectrometer operating near 9.7 GHz equipped with a split-ring MS2 resonator at a temperature of 55 K. Spin-labeled active HIV-1 protease or its [D25N] analogue were dissolved in 6 M GnuHCl, 50 mM NaOAc, pH 6.0 and folded by two-step dialysis against 2 M GnuHCl, 10 mM NaOAc, 10% glycerol, pH 5.6 and then 10 mM NaOAc, 10% glycerol, pH 5.6. Solutions were then concentrated using Millipore centrifugal concentrators and final concentrations were determined by analytical HPLC (integrating total peak area at 280 nm with calibration curve done for reference sample submitted for nitrogen elemental analysis) and were 40-50 µM. Samples were stored frozen at –20 ºC. In the experiment with inhibitors, 30-100 mol. excess of inhibitors were added. Solutions were then transferred to quartz 1.1×1.6×100 mm capillaries and flash-frozen to 55 K in the resonator. All the measurements were performed using a constant-time version of the four-pulse DEER sequence \( \pi/2(v_{\text{obs}}) - \tau_1 - \pi(v_{\text{obs}}) - t' - \pi(v_{\text{pump}}) - (\tau_1 + \tau_2 - t') - \pi(v_{\text{obs}}) - \tau_2 - \text{echo} \), where time \( t' \) is varied. The resonator was overcoupled to \( Q \approx 100 – 200 \), the pump frequency \( (v_{\text{pump}}) \) was set to the center of the resonator dip and coincided with the maximum of the nitroxide EPR spectrum, while the observer frequency \( (v_{\text{obs}}) \) was set 65–70 MHz higher and coincided with the low-field local maximum of the spectrum. The pulse lengths for \( \pi/2 \) and \( \pi \) were 16 ns and 32 ns, respectively, and pump pulse length was 32 – 40 ns. The pump pulse length was optimized using nutation experiment for all cases. In all experiments, a \( \tau_1 \) of 200 ns was used. Data were recorded typically at steps of 8-12 ns. The choice for \( \tau_2 \) parameter was dictated by phase memory time. It has been found that it is important for dipolar modulation to fully decay in order to perform correct background subtraction and being able to find stable solution according to Tikhonov regularization (see Figure D.5).

In the previous pulse-EPR experiments with unliganded apo-form and inhibitor-containing HIV-1 protease samples, the authors seem to have overlooked an additional modulation in the dipolar echo evolution curves and may have underfitted (oversmoothed) the data, particularly for the unliganded apo-form of HIV-1 protease, most probably due to poor signal-to-noise ratio. In our measurements, we clearly observed longer phase memory times for unliganded apo-samples and thus tried to get...
signal-to-noise ratio as high as possible. Accumulation times for the different datasets varied between 24-48 hours (100-250 scans). Experiments for the same samples were performed several times and results were found to be highly reproducible.

4. Analysis of pulse-EPR data (see Figures D.3-D.10).

Data were processed and analyzed using the program DeerAnalysis2008. Homogeneous 3D background model was used to subtract intermolecular background from raw data. The criterion for choosing background subtraction was frequency-domain spectrum, where neither a positive spike nor an obvious hole had to be present in the center of the Pake pattern. There were only few solutions for background correction, which met such conditions and they were producing very similar results at the later stage of the analysis. Computation of \( L \)-curve was performed to choose optimal value for Tikhonov regularization parameter, which corresponded to the corner of the \( L \)-curve. Validation of the results was performed using validation tools available in the DeerAnalysis2008. In all cases (apo-enzymes and with three different inhibitors) we have observed three principal maxima for distance distribution. We performed suppression of less populated two components in order to determine their contribution to the quality of the fitting and in all cases excluding such components produced fits with higher r.m.s.d. values. We then performed validation of the distance distributions against artificially enhanced noise. The noise level has been increased by adding pseudorandom numbers by a factor \( L_{\text{noise}} = 1.5 \). The error estimate indicated that the solution for a distance distribution profile including three-components is indeed stable to such increase of the noise level. The stability of the solution was then probed against background correction parameters, which included computation of the distance distribution profiles with a grid of values for background density and modulation depth. Although some combination of values clearly produced solutions, which were wrong based on their dipolar spectrum (incorrect separation of intra- and intermolecular contributions), the solution again was rather stable against variation of the parameters used in the background correction, with only too high error observed for the highest distance component (\(~ 40 \text{ Å}\) ), where it was not highly populated. However, the existence of such component is justified based on suppressing the
contribution of this component, and when it is suppressed the fits have higher r.m.s.d. values.

**Figure D.3.** Data analysis for nitroxide spin-labeled inhibitor-free inactive [D25] chemical analogue. (a) Raw experimental data; (b) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (c) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (d) computed \( L \)-curve for Tikhonov regularization with optimal regularization parameter marked by red arrow; (e) extracted distance distribution; (f) validation of the solution against artificially enhanced noise level \( L_{\text{noise}} = 1.5 \) with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
Figure D.4. Data analysis for nitroxide spin-labeled inhibitor-free active HIV-1 protease. (a) Raw experimental data; (b) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (c) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (d) computed $L$-curve for Tikhonov regularization with optimal regularization parameter marked by red arrow; (e) extracted distance distribution; (f) validation of the solution against artificially enhanced noise level ($L_{\text{noise}} = 1.5$) with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
Figure D.5. Data analysis for nitroxide spin-labeled [D25N] analogue of HIV-1 protease with MVT-101 inhibitor. (a) Raw experimental data acquired with $\tau_2 = 1900$ ns; (b) computed $L$-curve, which does not show sharp corner (stable solution); (c) Raw experimental data acquired with $\tau_2 = 2500$ ns; (d) computed $L$-curve with stable solution at $\lambda = 1$; (e) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (f) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (g) extracted distance distribution; (h) validation of the solution against grid (10 x 10) of different values for background density and modulation depth with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
Figure D.6. Data analysis for nitroxide spin-labeled active HIV-1 protease with MVT-101 inhibitor. (a) Raw experimental data; (b) computed \( L \)-curve; (c) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (d) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (e) extracted distance distribution; (f) validation of the solution against grid (10 x 10) of different values for background density and modulation depth with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
Figure D.7. Data analysis for nitroxide spin-labeled [D25N] analogue of HIV-1 protease with KVS-1 inhibitor (keto-form). (a) Raw experimental data; (b) computed \(L\)-curve; (c) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (d) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (e) extracted distance distribution; (f) validation of the solution against grid (10 x 10) of different values for background density and modulation depth with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
Figure D.8. Data analysis for nitroxide spin-labeled active HIV-1 protease with KVS-1 inhibitor (hydrated form). (a) Raw experimental data; (b) computed $L$-curve; (c) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (d) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (e) extracted distance distribution; (f) validation of the solution against grid (10 x 10) of different values for background density and modulation depth with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
Figure D.9. Data analysis for nitroxide spin-labeled [D25N] analogue of HIV-1 protease with JG-365 inhibitor. (a) Raw experimental data; (b) computed $L$-curve; (c) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (d) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (e) extracted distance distribution; (f) validation of the solution against grid (10 x 10) of different values for background density and modulation depth with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
Figure D.10. Data analysis for nitroxide spin-labeled active of HIV-1 protease with JG-365 inhibitor. (a) Raw experimental data; (b) computed L-curve; (c) background subtracted dipolar echo evolution curves with best r.m.s.d. fit in red, fit with lower distance component suppressed in green, fit with higher distance component suppressed in blue; (d) dipolar spectrum after Fourier transformation in black and its fitting to a solution in red; (e) extracted distance distribution; (f) validation of the solution against grid (10 x 10) of different values for background density and modulation depth with best r.m.s.d. solution depicted as red curve and lower/higher error estimates corresponding to the mean value of the probability minus/plus two times its standard deviation depicted as dash blue curves.
<table>
<thead>
<tr>
<th>Sample</th>
<th>Low-distance minor component, maximum in Å (width at ½ height)</th>
<th>Most populated component, maximum in Å (width at ½ height)</th>
<th>High-distance minor component, maximum in Å (width at ½ height)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[D25N] HIV-1 protease</td>
<td>28.8 (2.9)</td>
<td>33.9 (4.4)</td>
<td>41.6 (4.2)</td>
</tr>
<tr>
<td>Active HIV-1 protease</td>
<td>28.8 (4.0)</td>
<td>34.0 (4.7)</td>
<td>42.1 (3.6)</td>
</tr>
<tr>
<td>[D25N] HIV-1 protease with MVT-101 inhibitor</td>
<td>27.3 (1.6)</td>
<td>32.5 (2.6)</td>
<td>40.8 (1.9)</td>
</tr>
<tr>
<td>Active HIV-1 protease with MVT-101 inhibitor</td>
<td>27.3 (1.7)</td>
<td>32.7 (2.8)</td>
<td>39.3 (1.9)</td>
</tr>
<tr>
<td>[D25N] HIV-1 protease with KVS-1 inhibitor</td>
<td>25.1 (3.4)</td>
<td>32.7 (3.9)</td>
<td>43.8 (3.2)</td>
</tr>
<tr>
<td>Active HIV-1 protease with KVS-101 inhibitor</td>
<td>27.2 (3.1)</td>
<td>32.6 (2.6)</td>
<td>40.9 (2.2)</td>
</tr>
<tr>
<td>[D25N] HIV-1 protease with JG-365 inhibitor</td>
<td>27.7 (4.1)</td>
<td>32.8 (4.5)</td>
<td>43.8 (3.9)</td>
</tr>
<tr>
<td>Active HIV-1 protease with JG-365 inhibitor</td>
<td>27.3 (3.0)</td>
<td>32.5 (5.1)</td>
<td>42.2 (3.2)</td>
</tr>
</tbody>
</table>
References:


Appendix E. Supporting information for chapter 5.

Reagents. Boc-amino acids, 4-Me-benzylhydrylamine-resins and –OCH2-Pam-resins (free α-carboxyl peptides) were purchased from Peptides International, Louisville, Kentucky; Boc-L-thiazolidine-4-carboxylic acid from NovaBiochem, San Diego; diethyl ether from Fisher; HF from Matheson Tri-Gas; methoxylamine hydrochloride, p-cresol, sodium 2-mercaptosulphonate (MESNA), triisopropylsilane, methyl iodide, 2-bromoacetamide from Sigma-Aldrich; 4-mercaptophenylacetic acid was obtained from Sigma-Aldrich and purified by HPLC before use; tris(2-carboxyethyl)phosphine hydrochloride (TCEP) from Fluka; trifluoroacetic acid (TFA) from Halocarbon Products, New Jersey, (DIEA) from Applied Biosystems. HBTU was purchased from Peptides International and HATU from Anaspec. 15N- and 13C-labelled amino acids were purchased from Cambridge Isotopes and further Boc-protected according to standard protocol. Carboxylate side chain of (U13C4; 15N)-aspartic acid was protected with Allyl-protecting group.

Peptide Synthesis. Peptides were prepared manually by “in situ neutralization” Boc-chemistry stepwise solid phase peptide synthesis (SPPS).1 Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), Glu(OcHex), Lys(2Cl-Z), Thr(Bzl), Tyr(Br-Z), Trp(CHO), His(Bom). The 1,3-thiazolidine-4-carboxyl (Thz) group was introduced to protect the N-terminal Cys of the (Cys201-Gly4-1-40)-(α-thioalkylester) peptide segment, and Boc-L-thiazolidine-4-carboxylic acid was used for peptide synthesis. Aib (α-aminoisobutyric acid) residue was coupled with HATU. After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin support by treatment with anhydrous HF containing 10% (v/v) p-cresol for 1 h at 0 ºC. After evaporation of the HF under reduced pressure, crude products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA. Conversion of (1-40)-(α-thioalkylester) to (1-40)-(α-thioarylester) is described in detail in Supporting Information of Angew. Chem. Int. Ed. 46, 1667-1670 (2007). Preparative
HPLC of crude peptides after SPPS was performed on an Agilent 1100 preparative instrument on Agilent C-3 or C-8 (250 × 22 mm, 300Å, 7µm) columns or on in-house packed columns C4 (250 × 7 mm, Microsorb, 300Å, 5µm) and C8 (250 × 10 mm, Microsorb, 300Å, 10µm). Peptides from SPPS and from ligations were eluted from the column using an appropriate shallow gradient of acetonitrile/0.08% TFA versus water/0.1% TFA. Fractions containing the desired purified peptide product were identified by analytical LC and LC-MS, combined and lyophilized. Segments (Cys\(^{41-99}\)-\(^{4}\)COSCH\(_2\)CH\(_2\)Arg\(_4\).amide and (Cys\(^{41-99}\))-COOH were purified two times.

**Analytical characterization of peptides and proteins.** Analytical RP-HPLC was performed on an Agilent 1100 system with in-house packed C-4 and C-18 silica columns (2.1 × 50 mm, Microsorb 300 Å, 3 µm) at flow rate of 0.5 mL/min or Agilent C-3 and C-8 columns (4.6×150 mm, Zorbax, 300Å, 3.5 µm) at flow rate of 1.0 mL/min. Peptides were eluted from the column using a gradient of acetonitrile/0.08% TFA (solvent A) versus water/0.1% TFA (solvent B). Peptide masses were obtained using on-line electrospray MS detection with an Agilent 1100 LC/MSD ion trap. Fourier transform ion-cyclotron resonance mass spectrometry (FT-ICR MS) analysis for final products was performed at Chicago Biomedical Consortium proteomics facilities at University of Illinois at the Chicago on a Thermo Finnigan LTQ-FT LC/MS/MS instrument (Superconducting magnet 7.0 Tesla). Spectra were registered in LC-MS mode. Peak in total ion chromatogram was integrated and analyzed. Spectra were processed with Thermo Xcalibur v. 2.0 software. Simulation of isotopic patterns was done in IsoPro v. 3 (Dr. M. Senko, Thermo Finnigan) and deconvolution to get average molecular weight was done with ProMass v. 2.5 software (see Figures E.1–E.4).
Amino-acid sequence for the HIV-1 protease monomer (from N-terminus to C-terminus) is based on the SF2 isolate: PQITLWKRL₁⁰ VTIRGGQLK₂⁰ EALLDTGADD₃⁰ TVIEENleNLPG⁴⁰ NmmWKPKNleIGGI₅⁰ XaaGFIKVRQYD₆⁰ QIPVELAbuGHK₇⁰ AIGTVELGQP₈⁰ PVNIIGRNLL₉⁰ TQIGAbuTLNE₉⁹. Unnatural amino acids are in italics in three-letter code. Nle = norleucine, Abu = α-aminobutyric acid, Nmm = S-methyl-cysteine or pseudo-homoglutamine, Xaa = variable amino acid (glycine, L-alanine, D-alanine, α-aminoisobutyric acid). Ligation site is underlined.

Amino-acid sequence for ‘covalent dimer’ HIV-1 proteases (from N-terminus to C-terminus) is based on the SF2 isolate: PQITLWKRL₁⁰ VTIRGGQLK₂⁰ EALLDTGADD₃⁰ TVIEENleNLPG⁴⁰ ψ-GlnWKPKNleIGGI₅⁰ XaaGFIKVRQYD₆⁰ QIPVELAbuGHK₇⁰ AIGTVELGQP₈⁰ PVNIIGRNLL₉⁰ TQIGAbuTLNE₉⁹ ψ-Gln²⁰¹ GGGG²⁰⁵ PQITLWKRLB₁⁰ VTIRGGQLKB₂⁰ EALLDTGADDB₃⁰ TVIEENleNLPGB⁴⁰ ψ-GlnWKPKNleIGGIB₅⁰ YbbGFIKVRQYDB₆⁰ QIPVELAbuGHKB₇⁰ AIGTVELGQPB₈⁰ PVNIIGRNLLB₉⁰ TQIGAbuTLNEB₉⁹. Unnatural amino acids are in italics in three-letter code. Nle = norleucine, Abu = α-aminobutyric acid, ψ-Gln = pseudo-homoglutamine, Xaa and Ybb = variable α-amino acids (glycine, L-alanine, D-alanine or α-aminoisobutyric acid in corresponding analogues). Residues from N-terminal 99-residue part (part A) are specified by letter ‘A’ placed before number of residue, correspondingly C-terminal 99-residue part (part B) has letter ‘B’ before number. 5-amino acid linker region is numbered from 201 to 205. Ligation sites are underlined.
Figure E.1. Analytical HPLC ($\lambda = 214$nm) and FT-ICR ESI-MS characterization of chemically synthesized 99-residue HIV-1 protease monomers: (a) wild-type Gly51-monomer, obs. (mass of average isotope composition) 10706.5 ± 1.0 Da, calc. 10706.7 Da, (b) L-Ala51-analogue, obs. 10720.9 ± 0.6 Da, calc. 10720.7 Da, (c) D-Ala51-analogue, obs. 10720.5 ± 0.5 Da, calc. 10720.7 Da, (d) Aib51-analogue obs. 10734.9 ± 0.7 Da, calc. 10734.7 Da. In isotope patterns for charge state 9H$^+$ red circles depict maxima for experimental measurements; blue rhombs correspond to maxima in simulated spectra. Note: residue 41 is S-methyl-cysteine.
Figure E.2. Analytical HPLC ($\lambda = 214$ nm) and FT-ICR ESI-MS characterization of chemically synthesized 203-residues covalent dimers: (a) $[L$-Ala51;Gly51'] covalent dimer, obs. (mass of average isotope composition) 21883.5 ± 0.6 Da, calc. 21883.8 Da, (b) $[Gly51;D$-Ala51'] covalent dimer, obs. 21884.0 ± 0.7 Da, calc. 21883.8 Da, (c) $[L$-Ala51;D-Ala51'] covalent dimer, obs. 21897.9 ± 0.4 Da, calc. 21897.9 Da, (d) $[Gly51;Aib51']$ covalent dimer, obs. 21897.5 ± 0.6 Da, calc. 21897.9 Da. In isotope patterns for charge state $23H^+$ red circles depict maxima for experimental measurements; blue rhombs correspond to maxima in simulated spectra. Retention time differences in (a,d) and (b,c) are caused by application of different HPLC columns. Note: residue 41/41’ is pseudo-homoglutamine.
Figure E.3. Analytical HPLC (λ = 214 nm) and FT-ICR ESI-MS characterization of chemically synthesized 99-residue [D25N]HIV-1 protease monomers labelled with MTSSL-nitroxide spin-label for EPR studies: (a) L-Ala51-analogue, obs. (mass of average isotope composition) 10921.9 ± 0.7 Da, calc. 10922.0 Da, (b) D-Ala51-analogue, obs. 10920.0 ± 0.6 Da, calc. 10922.0 Da, (c) Aib51-analogue obs. 10936.0 ± 0.5 Da, calc. 10936.0 Da. In isotope patterns for charge state 10H⁺ red circles depict maxima for experimental measurements; blue rhombs correspond to maxima in simulated spectra. Note: residue 41 is pseudo-homoglutamine.
Figure E.4. Analytical HPLC ($\lambda = 214$ nm) and FT-ICR ESI-MS characterization of chemically synthesized 99-residue $^{15}$N- and $^{13}$C-labelled HIV-1 protease monomers for NMR studies: (a) $L$-Ala51-analogue, obs. (mass of average isotope composition) 10784.0 ± 0.5 Da, calc. 10784.18 Da (with 11 $^{15}$N-atoms and 10 $^{13}$C-atoms at 98% enrichment). (b) $D$-Ala51-analogue, obs. 10784.0 ± 0.6 Da, calc. 10784.18 Da (with 11 $^{15}$N-atoms and 10 $^{13}$C-atoms at 98% enrichment), (c) Aib51-analogue, obs. 10792.0 ± 0.5 Da, calc. 10792.36 Da (with 9 $^{15}$N-atoms and 6 $^{13}$C-atoms at 98% enrichment). In isotope patterns for the most abundant charge state $12H^+$, left panel depicts simulated spectrum for non-labelled polypeptide, whereas right panel shows experimental spectrum for $^{15}$N, $^{13}$C-labelled polypeptide. In (a) and (b), mass shift of ~1.75 Da corresponds to 21 incorporated ($^{15}$N, $^{13}$C) heavy isotopes, and in (c), mass shift of ~1.25 Da corresponds to 15 incorporated ($^{15}$N, $^{13}$C) heavy isotopes, as designed in their total chemical synthesis (see Scheme E.1). Note: residue 41 is pseudo-homoglutamine.
**Crystallography.** Crystals were grown at 20 °C by the hanging drop vapour diffusion method from a well solution consisting of 0.1 M citrate, 0.2 M NaH₂PO₄, 30% (w/v) saturated ammonium sulfate, 10% (v/v) DMSO, pH 6.0. Protein solution (~0.1 mM) was preincubated with a 30-fold molar excess of MVT-101 or JG-365 inhibitors, and was then mixed in different (v/v) ratios with well solution. Crystals grew within 1-4 days and were frozen in liquid nitrogen using mineral oil as cryoprotectant. Data collection was performed at 100 K at the Advanced Photon Source, Argonne National Laboratory. Images were processed and scaled with HKL2000.² Crystal structure was solved by molecular replacement with the help of MolRep v. 8.1³ and refined by Refmac v.5.⁴ See Table E.1 for more details.
Figure E.5. Flap structures (residues 47–54 and 47'–54') in crystal structures of [L-Ala51;D-Ala51'] covalent dimer HIV-1 protease with (a) MVT-101 inhibitor, and (b,c) JG-365 inhibitor. Side chains are deleted for clarity, 2F_o-F_c is contoured at a level of 1σ in magenta for residues Ile50, L-Ala51, Ile50' and D-Ala51', and at a level of 3σ in green for residues Asp25 and Asp25'. In the case of MVT-101 inhibitor, only one conformer for flaps and inhibitor was observed, whereas in the case of JG-365 inhibitor, two conformers were found in a ratio of 4 : 1. Note that for major conformer N-to-C orientation of JG-365 inhibitor (~65%) with the respect to flaps (L-Ala51 and D-Ala51') is reversed in comparison to structure with MVT-101 inhibitor. Previously, for the complex of wild-type [Gly51;Gly51'] HIV-1 protease with JG-365 inhibitor, two conformers were found at 1 : 1 ratio with non-lytic water molecule (bound to NH-Ile50 and NH-Ile50' and carbonyls of inhibitors, displayed in (a-c) occupying two positions as well. For [L-Ala51;D-Ala51'] HIV-1 protease, such symmetry is broken and it clearly displays preference to bind JG-365 inhibitor in one orientation.
Figure E.6. Plots of crystallographic temperature $B$-factors for amino acid sequence of two ‘flap’ structures (residues 37-61 and 37’-61’) in (a) crystal structure of $[L$-$Ala51;D$-$Ala51’]$HIV-1 protease complexed with MVT-101 inhibitor (PDB ID 3FSM), and (b) crystal structure of $[L$-$Ala51;D$-$Ala51’]$HIV-1 protease complexed with JG-365 inhibitor (PDB ID 3GI0). In both structures, average $B$-factors are higher for $L$-$Ala51$-containing flap than for $D$-$Ala51’$-containing flap suggesting higher mobility for $L$-$Ala51$-containing flap.
Table E.1. Data collection and refinement statistics for complexes of \([L\text{-Ala51};D\text{-Ala51}']\)HIV-1 protease with MVT-101 and JG-365 inhibitors (highest resolution shell is in parenthesis).

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>MVT-101</th>
<th>JG-365</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>APS 23ID-B</td>
<td>APS 23ID-D</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>1.0332</td>
<td>0.9793</td>
</tr>
<tr>
<td>Space group</td>
<td>(P2_12_12_1)</td>
<td>(P2_12_12_1)</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) (Å)</td>
<td>51.472</td>
<td>51.172</td>
</tr>
<tr>
<td>(b) (Å)</td>
<td>58.286</td>
<td>58.456</td>
</tr>
<tr>
<td>(c) (Å)</td>
<td>61.450</td>
<td>60.950</td>
</tr>
<tr>
<td>(\alpha = \beta = \gamma) (º)</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 – 1.6</td>
<td>50 – 1.8</td>
</tr>
<tr>
<td>(R_{merge})</td>
<td>0.079 (0.677)</td>
<td>0.068 (0.438)</td>
</tr>
<tr>
<td>(I/\sigma I)</td>
<td>22.6 (2.9)</td>
<td>26.2 (3.86)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.9 (4.8)</td>
<td>6.1 (6.1)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20 – 1.6</td>
<td>20 – 1.8</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8</td>
<td>99.9</td>
</tr>
<tr>
<td>No. reflections</td>
<td>25039</td>
<td>16579</td>
</tr>
<tr>
<td>Work/free set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R_{work} / R_{free})</td>
<td>19.9 / 23.5</td>
<td>18.5 / 23.5</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1545</td>
<td>1545</td>
</tr>
<tr>
<td>Inhibitor</td>
<td>54</td>
<td>61</td>
</tr>
<tr>
<td>Water</td>
<td>118</td>
<td>97</td>
</tr>
<tr>
<td>B-factors (Å(^2))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>25.2</td>
<td>20.6</td>
</tr>
<tr>
<td>Inhibitor</td>
<td>29.3</td>
<td>18.5</td>
</tr>
<tr>
<td>Water</td>
<td>32.7</td>
<td>26.0</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.016</td>
<td>0.015</td>
</tr>
<tr>
<td>Bond angles (º)</td>
<td>1.634</td>
<td>1.561</td>
</tr>
<tr>
<td><strong>PDB ID</strong></td>
<td>3FSM</td>
<td>3GI0</td>
</tr>
</tbody>
</table>
NMR experiments.

Scheme E.1. Total chemical synthesis approach for preparation of chemical analogues of HIV-1 protease has allowed to site-specifically $^{15}$N-label residues of interest, namely, selected residues in flaps, residue Gly40 in elbows and residues Asp25 and Gly27 in catalytic site.
Table E.2. Amide chemical shifts for $^{15}$N-labelled chemical analogues of HIV-1 protease measured on 600 MHz spectrometer and referenced to DSS-d$_6$.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Xaa = L-Ala51</th>
<th>Xaa = D-Ala51</th>
<th>Xaa = Aib51</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N15</td>
<td>H1</td>
<td>N15</td>
</tr>
<tr>
<td>Val56</td>
<td>116.6671</td>
<td>8.92141</td>
<td>116.402</td>
</tr>
<tr>
<td>Gly52</td>
<td>107.1483</td>
<td>7.40844</td>
<td>104.8247</td>
</tr>
<tr>
<td>Xaa51</td>
<td>122.9544</td>
<td>8.35415</td>
<td>131.4883</td>
</tr>
<tr>
<td>Ile50</td>
<td>120.3052</td>
<td>8.32826</td>
<td>123.2679</td>
</tr>
<tr>
<td>Gly49</td>
<td>110.8749</td>
<td>7.5924</td>
<td>113.397</td>
</tr>
<tr>
<td>Gly48</td>
<td>110.2479</td>
<td>8.62006</td>
<td>110.6247</td>
</tr>
<tr>
<td>Ile47</td>
<td>119.1612</td>
<td>8.88209</td>
<td>119.5694</td>
</tr>
<tr>
<td>Gly40</td>
<td>106.0162</td>
<td>8.40116</td>
<td>105.9506</td>
</tr>
<tr>
<td>Asp25</td>
<td>127.812</td>
<td>8.68316</td>
<td>127.8846</td>
</tr>
</tbody>
</table>

Table E.3. Amide chemical shifts for $^{15}$N-labelled chemical analogues of HIV-1 protease acquired on 900 MHz spectrometer referenced to DSS-d$_6$.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Xaa = L-Ala51</th>
<th>Xaa = D-Ala51</th>
<th>Xaa = Aib51</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N15</td>
<td>H1</td>
<td>N15</td>
</tr>
<tr>
<td>Val56</td>
<td>116.6649</td>
<td>8.92067</td>
<td>116.3737</td>
</tr>
<tr>
<td>Gly52</td>
<td>107.1411</td>
<td>7.41362</td>
<td>104.8122</td>
</tr>
<tr>
<td>Xaa51</td>
<td>122.8338</td>
<td>8.3485</td>
<td>131.4953</td>
</tr>
<tr>
<td>Ile50</td>
<td>120.3882</td>
<td>8.33636</td>
<td>123.2636</td>
</tr>
<tr>
<td>Gly49</td>
<td>110.8475</td>
<td>7.59478</td>
<td>113.4282</td>
</tr>
<tr>
<td>Gly48</td>
<td>110.2604</td>
<td>8.6204</td>
<td>110.6322</td>
</tr>
<tr>
<td>Ile47</td>
<td>119.1305</td>
<td>8.88483</td>
<td>119.5743</td>
</tr>
<tr>
<td>Gly40</td>
<td>106.0097</td>
<td>8.40244</td>
<td>105.9815</td>
</tr>
<tr>
<td>Gly27</td>
<td>107.6669</td>
<td>8.45443</td>
<td>107.635</td>
</tr>
<tr>
<td>Asp25</td>
<td>127.8084</td>
<td>8.68429</td>
<td>127.8561</td>
</tr>
</tbody>
</table>
Table E.4. $T_1$, $T_2$ and $^{1}H-^{15}N$ NOE values measured for $^{15}N$-labelled chemical analogues of HIV-1 protease on 600 MHz spectrometer. Residues located at the tips of the flaps are highlighted in magenta.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$Xaa = L$-Ala</th>
<th>$Xaa = D$-Ala</th>
<th>$Xaa = Aib$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_1$, s</td>
<td>$T_1$, error, s</td>
<td>$T_1$, s</td>
</tr>
<tr>
<td>Val56</td>
<td>0.885</td>
<td>0.016</td>
<td>0.943</td>
</tr>
<tr>
<td>Ile54</td>
<td>0.941</td>
<td>0.020</td>
<td>1.005</td>
</tr>
<tr>
<td>Gly52</td>
<td>0.865</td>
<td>0.023</td>
<td>0.937</td>
</tr>
<tr>
<td>Xaa51</td>
<td>0.750</td>
<td>0.015</td>
<td>0.954</td>
</tr>
<tr>
<td>Ile50</td>
<td>0.773</td>
<td>0.007</td>
<td>1.006</td>
</tr>
<tr>
<td>Gly49</td>
<td>0.893</td>
<td>0.011</td>
<td>0.982</td>
</tr>
<tr>
<td>Gly48</td>
<td>0.934</td>
<td>0.013</td>
<td>0.994</td>
</tr>
<tr>
<td>Ile47</td>
<td>0.897</td>
<td>0.020</td>
<td>0.986</td>
</tr>
<tr>
<td>Gly40</td>
<td>1.115</td>
<td>0.021</td>
<td>1.115</td>
</tr>
<tr>
<td>Gly27</td>
<td>0.788</td>
<td>0.032</td>
<td>0.845</td>
</tr>
<tr>
<td>Asp25</td>
<td>1.045</td>
<td>0.027</td>
<td>1.139</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue</th>
<th>$T_2$, s</th>
<th>$T_2$, error, s</th>
<th>$T_2$, s</th>
<th>$T_2$, error, s</th>
<th>$T_2$, s</th>
<th>$T_2$, error, s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val56</td>
<td>0.0676</td>
<td>0.0008</td>
<td>0.0591</td>
<td>0.0011</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Ile54</td>
<td>0.0377</td>
<td>0.0006</td>
<td>0.0409</td>
<td>0.0010</td>
<td>0.0489</td>
<td>0.0010</td>
</tr>
<tr>
<td>Gly52</td>
<td>0.1027</td>
<td>0.0049</td>
<td>0.0588</td>
<td>0.0006</td>
<td>0.0767</td>
<td>0.0011</td>
</tr>
<tr>
<td>Xaa51</td>
<td>0.0655</td>
<td>0.0014</td>
<td>0.0600</td>
<td>0.0016</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Ile50</td>
<td>0.0852</td>
<td>0.0020</td>
<td>0.0647</td>
<td>0.0013</td>
<td>0.0534</td>
<td>0.0011</td>
</tr>
<tr>
<td>Gly49</td>
<td>0.0715</td>
<td>0.0005</td>
<td>0.0525</td>
<td>0.0010</td>
<td>0.0395</td>
<td>0.0004</td>
</tr>
<tr>
<td>Gly48</td>
<td>0.0718</td>
<td>0.0007</td>
<td>0.0615</td>
<td>0.0010</td>
<td>0.0669</td>
<td>0.0011</td>
</tr>
<tr>
<td>Ile47</td>
<td>0.0578</td>
<td>0.0012</td>
<td>0.0505</td>
<td>0.0005</td>
<td>0.0440</td>
<td>0.0012</td>
</tr>
<tr>
<td>Gly40</td>
<td>0.0936</td>
<td>0.0012</td>
<td>0.0797</td>
<td>0.0027</td>
<td>0.0658</td>
<td>0.0008</td>
</tr>
<tr>
<td>Gly27</td>
<td>0.0527</td>
<td>0.0008</td>
<td>0.0534</td>
<td>0.0012</td>
<td>0.0585</td>
<td>0.0011</td>
</tr>
<tr>
<td>Asp25</td>
<td>0.0578</td>
<td>0.0012</td>
<td>0.0558</td>
<td>0.0016</td>
<td>0.0547</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

$^{1}H-^{15}N$ NOE values

<table>
<thead>
<tr>
<th>Residue</th>
<th>$Xaa = Gly^*$</th>
<th>$Xaa = L$-Ala</th>
<th>$Xaa = D$-Ala</th>
<th>$Xaa = Aib$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val56</td>
<td>0.6389</td>
<td>0.7808</td>
<td>0.6943</td>
<td>n/a</td>
</tr>
<tr>
<td>Ile54</td>
<td>0.4481</td>
<td>0.7206</td>
<td>0.7086</td>
<td>0.7718</td>
</tr>
<tr>
<td>Gly52</td>
<td>0.3087</td>
<td>0.3716</td>
<td>0.5432</td>
<td>0.7607</td>
</tr>
<tr>
<td>Xaa51</td>
<td>0.1906</td>
<td>0.2622</td>
<td>0.7388</td>
<td>n/a</td>
</tr>
<tr>
<td>Ile50</td>
<td>0.2743</td>
<td>0.2723</td>
<td>0.5748</td>
<td>0.4530</td>
</tr>
<tr>
<td>Gly49</td>
<td>0.4569</td>
<td>0.4603</td>
<td>0.5633</td>
<td>0.5276</td>
</tr>
<tr>
<td>Gly48</td>
<td>0.7287</td>
<td>0.6748</td>
<td>0.6964</td>
<td>0.7598</td>
</tr>
<tr>
<td>Ile47</td>
<td>0.6617</td>
<td>0.6877</td>
<td>0.7677</td>
<td>0.8199</td>
</tr>
<tr>
<td>Gly40</td>
<td>0.3569</td>
<td>0.3785</td>
<td>0.4225</td>
<td>0.3741</td>
</tr>
<tr>
<td>Gly27</td>
<td>0.8096</td>
<td>0.7788</td>
<td>0.7241</td>
<td>0.8205</td>
</tr>
<tr>
<td>Asp25</td>
<td>0.7584</td>
<td>0.8690</td>
<td>0.8841</td>
<td>0.7759</td>
</tr>
</tbody>
</table>

*previously reported by Freedberg et al. 

6
**Table E.5.** Generalized $S^2$ order parameters derived with model-free analysis using anisotropic diffusion tensor and $T_1$, $T_2$ and $^1$H-$^{15}$N NOE parameters. Residues located at the tips of the flaps are highlighted in magenta.

<table>
<thead>
<tr>
<th></th>
<th>Xaa = Gly*</th>
<th>Xaa = L-Ala</th>
<th>Xaa = D-Ala</th>
<th>Xaa = Aib</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val56</td>
<td>0.93</td>
<td>0.85</td>
<td>0.91</td>
<td>n/a</td>
</tr>
<tr>
<td>Ile54</td>
<td>0.77</td>
<td>0.89</td>
<td>0.83</td>
<td>0.9</td>
</tr>
<tr>
<td>Gly52</td>
<td>0.77</td>
<td>0.54</td>
<td>0.86</td>
<td>0.78</td>
</tr>
<tr>
<td>Xaa51</td>
<td>0.80</td>
<td>0.67</td>
<td>0.93</td>
<td>n/a</td>
</tr>
<tr>
<td>Ile50</td>
<td>0.56</td>
<td>0.68</td>
<td>0.84</td>
<td>0.81</td>
</tr>
<tr>
<td>Gly49</td>
<td>0.75</td>
<td>0.85</td>
<td>0.82</td>
<td>0.86</td>
</tr>
<tr>
<td>Gly48</td>
<td>0.83</td>
<td>0.85</td>
<td>0.86</td>
<td>0.9</td>
</tr>
<tr>
<td>Ile47</td>
<td>0.89</td>
<td>0.77</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>Gly40</td>
<td>0.59</td>
<td>0.65</td>
<td>0.73</td>
<td>0.69</td>
</tr>
<tr>
<td>Gly27</td>
<td>0.96</td>
<td>0.94</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>Asp25</td>
<td>0.83</td>
<td>0.87</td>
<td>0.79</td>
<td>0.84</td>
</tr>
</tbody>
</table>

*previously reported by Freedberg et al.\(^6\)
Figure E.7. $^1$H-detected CPMG $^{15}$N relaxation dispersion data for $^{15}$N-labelled residues in $[L$-Ala51/51']HIV-1 protease (left column), $[D$-Ala51/51']HIV-1 protease (middle column), and $[Aib51/51']$HIV-1 protease (right column). Fits shown are done according to 3-site exchange model globally for all flap residues except for $[Aib51/51']$HIV-1 protease (right column), which was fitted residue-by-residue with 2-site exchange model.
Figure E.8 (continuation of Figure E.7). $^1$H-detected CPMG $^{15}$N relaxation dispersion data for $^{15}$N-labelled residues in $[L$-Ala51/51']HIV-1 protease (left column), $[D$-Ala51/51']HIV-1 protease (middle column), and $[Aib51/51']$HIV-1 protease (right column). Fits shown are done according to 3-site exchange model globally for all flap residues except for [Aib51/51']HIV-1 protease (right column), which was fitted residue-by-residue with 2-site exchange model. In the case of $[L$-Ala51/51'] and $[D$-Ala51/51'] chemical analogues residues 25 and 27 were fitted together according to 2-site exchange model.
Table E.6. Initially all residues were fitted with 2-site exchange model without constrains on populations (only $k_{ex}$, s$^{-1}$ are listed, population of minor component varied in the range 1–3%). Clearly, in the case of $[L$-Ala51,51'] and $[D$-Ala51,51'] chemical analogues, $k_{ex}$ values were matching for residues in the flaps, whereas for $[Aib51/51']$ residues strongly varied exchange rates were found.

<table>
<thead>
<tr>
<th>Residue no.</th>
<th>$[L$-Ala51/51'] H1V-1 PR</th>
<th>$[D$-Ala51/51'] H1V-1 PR</th>
<th>$[Aib51/51']$ H1V-1 PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>2013 ± 75.30</td>
<td>1329 ± 46.88</td>
<td>n/a</td>
</tr>
<tr>
<td>54</td>
<td>2133 ± 80.87</td>
<td>1348 ± 49.72</td>
<td>2037 ± 175.5</td>
</tr>
<tr>
<td>52</td>
<td>2191 ± 90.86</td>
<td>1370 ± 52.04</td>
<td>poor dispersion</td>
</tr>
<tr>
<td>51</td>
<td>2180 ± 97.16</td>
<td>1340 ± 52.63</td>
<td>n/a</td>
</tr>
<tr>
<td>50</td>
<td>2246 ± 108.2</td>
<td>1342 ± 55.83</td>
<td>578.9 ± 134.7</td>
</tr>
<tr>
<td>49</td>
<td>2274 ± 120.3</td>
<td>1402 ± 62.40</td>
<td>2541 ± 109.9</td>
</tr>
<tr>
<td>48</td>
<td>2154 ± 166.9</td>
<td>1267 ± 71.35</td>
<td>poor dispersion</td>
</tr>
<tr>
<td>47</td>
<td>2142 ± 157.4</td>
<td>1114 ± 79.63</td>
<td>2200 ± 174.6</td>
</tr>
<tr>
<td>40</td>
<td>poor dispersion</td>
<td>poor dispersion</td>
<td>450.1 ± 251.6</td>
</tr>
<tr>
<td>27</td>
<td>2325 ± 192.3</td>
<td>848.3 ± 153.5</td>
<td>poor dispersion</td>
</tr>
<tr>
<td>25</td>
<td>1980 ± 192.2</td>
<td>536.6 ± 131.8</td>
<td>poor dispersion</td>
</tr>
</tbody>
</table>

Scheme E.2. Final fitting for flap residues in $[L$-Ala51/51'] and $[D$-Ala51/51'] chemical analogues was performed according to 3-site exchange model. 3-site exchange model was invoked based on interflap distance measurements by pulse-EPR$^7$ and previous MD simulations$^8$, where three principal conformers were identified. Populations of different conformers were taken directly from distance distribution profiles obtained by analysis of pulse-EPR data. In scheme, $p_i$ are populations of corresponding conformers. The 3-site exchange model improved quality of fitting (lower $\chi^2$) in comparison to 2-site exchange model.
Figure E.9. Pulse-EPR measurements data. Background subtracted dipolar echo evolution curves (left), dipolar spectra after Fourier transformation (center) and \( L \)-curves for Tikhonov regularization with optimal regularization parameter \( \lambda = 10 \) in all cases (right) for a, \([L]\text{-Ala51/51’;Asn25}]\text{HIV-1 protease}, b, \([D]\text{-Ala51/51’;Asn25}]\text{HIV-1 protease}, and c, \([\text{Aib51/51’;Asn25}]\text{HIV-1 protease}. Black line depicts experimental data, red line depicts fitting the data. For results for wild-type \([\text{Gly51/51’;Asp25}]\text{ and } \text{[Gly51/51’;Asn25]}\text{ enzymes, see previous communication.}\)
Figure E.10. a, ‘Nitroxide-to-nitroxide’ distance distribution profile for nitroxide-labelled homodimeric L-Ala51/51’ (top) and D-Ala51/51’ (bottom) unliganded HIV-1 protease molecules obtained in pulse-EPR measurements. b, Lys55(C')-to-Lys55’(C’') distance distribution profiles constructed based on 300 ns trajectories in all-atom MD simulations for corresponding homodimeric L-Ala51/51’ (top) and D-Ala51/51’ (middle) HIV-1 proteases and heterodimeric [L-Ala51;D-Ala51’] HIV-1 protease (bottom). c, Overlay of data shown in b.
**Surface Plasmon Resonance Assays.** Measurements were performed on Biacore 3000 instrument in HBS-EP buffer (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% surfactant P20, pH 5.6). Biotinylated substrate-derived inhibitor was immobilized onto streptavidin-coated Sensor Chip SA (from a Biocore) to get response difference of about 100 RU. All enzymes were then flown through microfluidic cell over immobilized inhibitor at multiple concentrations (at least 6 concentrations below and above calculated $K_d$ except analogues 7 and 8, where $K_d$ was too high) at flow rate 20 µL/min with association and dissociation kinetics studied for 120 seconds respectively. Curves were then fitted with BIAevaluation (v. 4.1) software with ‘two state reaction (conformational change)’ model. Only faster values of $k_{on}$ and $k_{off}$ are listed in the Table E.7.

![Figure E.11. Active-site titration of chemically synthesized enzymes by surface-plasmon resonance.](image)

Chemically synthesized enzymes (Table 5.1; entries 1-8) at 100 nM concentration were pumped through a microfluidic chip over a surface containing the substrate-based reduced isostere inhibitor MVT101. The asymmetric $[L$-Ala51; D-Ala51$']$ enzyme analogue 2 was determined to be the best binder to the inhibitor ($K_d = 100$ nM versus $K_d = 420$ nM for wild-type enzyme 1), which mimicks the earlier transition state of the proteolytic reaction. This is attributed to the higher concentration of catalytically-preorganized conformers present in the $[L$-Ala51; D-Ala51$']$ enzyme analogue.
Table E.7. Active-site titration of chemically synthesized enzymes by surface-plasmon resonance.

<table>
<thead>
<tr>
<th></th>
<th>Flap</th>
<th>Flap’</th>
<th>$k_{on,}$ $(\times 10^5 \text{ s}^{-1} \text{ M}^{-1})$</th>
<th>$k_{off,}$ s$^{-1}$</th>
<th>$K_d$, µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gly51</td>
<td>Gly51</td>
<td>1.01</td>
<td>0.042</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>L-Ala51</td>
<td>D-Ala51</td>
<td>2.1</td>
<td>0.02</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>L-Ala51</td>
<td>Gly51</td>
<td>1.62</td>
<td>0.038</td>
<td>0.24</td>
</tr>
<tr>
<td>4</td>
<td>Gly51</td>
<td>D-Ala51</td>
<td>1.08</td>
<td>0.066</td>
<td>0.61</td>
</tr>
<tr>
<td>5</td>
<td>L-Ala51</td>
<td>L-Ala51</td>
<td>0.74</td>
<td>0.075</td>
<td>1.01</td>
</tr>
<tr>
<td>6</td>
<td>Gly51</td>
<td>Aib51</td>
<td>0.6</td>
<td>0.094</td>
<td>1.57</td>
</tr>
<tr>
<td>7</td>
<td>D-Ala51</td>
<td>D-Ala51</td>
<td>0.24</td>
<td>0.197</td>
<td>8.2</td>
</tr>
<tr>
<td>8</td>
<td>Aib51</td>
<td>Aib51</td>
<td>0.1</td>
<td>0.174</td>
<td>17.5</td>
</tr>
</tbody>
</table>

AlaAsnThrAlaThrLeuLeu(ψ-CH$_2$NH)LecGlnArgGlyAsnPheArgAsnGlnCys

Figure E.12. Sequence of substrate derived reduced isostere inhibitor with hexapeptide binding site for HIV-1 protease and analogues depicted in red and polyethylene glycol and biotin moieties in blue. Molecule was synthesized with Boc-SPPS method.
References:


Appendix F. Supporting information for chapter 6.

Figure F.1. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of $^{13}$C-labeled (a) wild-type [1-99]HIV-1 protease, and (b) its L-Ala51 analogue.
Figure F.2. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of $^{13}$C-labeled (a) $D$-Ala51 [1-99]HIV-1 protease, and (b) Aib51 [1-99]HIV-1 protease.
Appendix G. Supporting information for chapter 7.

Table G.1. Data collection and refinement statistics for chemical analogues of HIV-1 protease (highest resolution is in parenthesis).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Inhibitor</th>
<th>Data collection</th>
<th>Refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[Gly51/51']</td>
<td>[L-Ala51/51']</td>
<td>[D-Ala51/51']</td>
</tr>
<tr>
<td>MVT-101</td>
<td>MVT-101</td>
<td>MVT-101</td>
<td>MVT-101</td>
</tr>
<tr>
<td>Beamline</td>
<td>APS 23ID-D</td>
<td>APS 23ID-D</td>
<td>APS 23ID-D</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>0.97934</td>
<td>0.97934</td>
<td>0.97934</td>
</tr>
<tr>
<td>Space group</td>
<td>$P2_12_12_1$</td>
<td>$P2_12_12_1$</td>
<td>$P2_12_12_1$</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$ (Å)</td>
<td>51.013</td>
<td>51.269</td>
<td>51.258</td>
</tr>
<tr>
<td>$b$ (Å)</td>
<td>58.401</td>
<td>58.783</td>
<td>58.626</td>
</tr>
<tr>
<td>$c$ (Å)</td>
<td>61.466</td>
<td>61.336</td>
<td>61.472</td>
</tr>
<tr>
<td>$\alpha = \beta = \gamma$ (º)</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 – 1.3</td>
<td>50 – 1.3</td>
<td>50 – 1.7</td>
</tr>
<tr>
<td>$R_{merge}$</td>
<td>5.6 (48.6)</td>
<td>7.5 (88.0)</td>
<td>9.5 (73.2)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>25.4 (1.8)</td>
<td>21.9 (2.3)</td>
<td>15.1 (2.2)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.1 (3.6)</td>
<td>5.2 (5.0)</td>
<td>5.1 (4.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Resolution (Å)</th>
<th>Completeness (%)</th>
<th>No. reflections</th>
<th>Work/free set</th>
<th>$R_{work}/R_{free}$</th>
<th>No. atoms</th>
<th>Protein</th>
<th>Inhibitor</th>
<th>Water</th>
<th>Mean $B$-value (Å²)</th>
<th>R.m.s. deviations</th>
<th>Bond lengths (Å)</th>
<th>Bond angles (º)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 – 1.3</td>
<td>97.8</td>
<td>42557 / 2264</td>
<td>17.9 / 21.5</td>
<td>12.66</td>
<td>1512</td>
<td>54</td>
<td>90</td>
<td>12.66</td>
<td>0.017</td>
<td>1.792</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 – 1.3</td>
<td>97.7</td>
<td>43001 / 2282</td>
<td>18.5 / 22.3</td>
<td>15.96</td>
<td>1514</td>
<td>54</td>
<td>117</td>
<td>15.96</td>
<td>0.016</td>
<td>1.672</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 – 1.7</td>
<td>98.4</td>
<td>19302 / 1233</td>
<td>19.7 / 24.7</td>
<td>23.98</td>
<td>1514</td>
<td>54</td>
<td>89</td>
<td>23.98</td>
<td>0.015</td>
<td>1.643</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 – 1.8</td>
<td>99.5</td>
<td>16792 / 1709</td>
<td>18.8 / 22.1</td>
<td>23.99</td>
<td>1516</td>
<td>54</td>
<td>103</td>
<td>23.99</td>
<td>0.016</td>
<td>1.618</td>
<td>0.016</td>
<td></td>
</tr>
</tbody>
</table>

**PDB ID** 3HAU 3HAW 3HBO 3HDK
**Table G.1, continued.** Data collection and refinement statistics for chemical analogues of HIV-1 protease (highest resolution is in parenthesis).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>[L-Ala51/Gly51']</th>
<th>[Gly51/D-Ala51']</th>
<th>[Gly51/Aib51']</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inhibitor</strong></td>
<td>MVT-101</td>
<td>MVT-101</td>
<td>MVT-101</td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td>to be reported elsewhere</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Beamline**                |                  |                  |                |
| **Wavelength, Å**           |                  |                  |                |
| **Space group**             |                  |                  |                |
| **Cell dimensions**         |                  |                  |                |
|   a (Å)                     |                  |                  |                |
|   b (Å)                     |                  |                  |                |
|   c (Å)                     |                  |                  |                |
|   α = β = γ (°)             |                  |                  |                |
| **Resolution (Å)**          |                  |                  |                |
| **R_merge**                 |                  |                  |                |
| **I/σI**                    |                  |                  |                |
| **Redundancy**              |                  |                  |                |

**Refinement** to be reported elsewhere

| **Resolution (Å)**          |                  |                  |                |
| **Completeness (%)**        |                  |                  |                |
| **No. reflections**         |                  |                  |                |
| **Work/free set**           |                  |                  |                |
| **R_work/R_free**           |                  |                  |                |
| **No. atoms**               |                  |                  |                |
| **Protein**                 |                  |                  |                |
| **Water**                   |                  |                  |                |
| **B-factors (Å²)**          |                  |                  |                |
| **Protein**                 |                  |                  |                |
| **Inhibitor**               |                  |                  |                |
| **Water**                   |                  |                  |                |
| **R.m.s. deviations**       |                  |                  |                |
| **Bond lengths (Å)**        |                  |                  |                |
| **Bond angles (°)**         |                  |                  |                |

**PDB ID** to be deposited to be deposited to be deposited
Table G.1, continued. Data collection and refinement statistics for chemical analogues of HIV-1 protease (highest resolution is in parenthesis).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Inhibitor</th>
<th>Data collection</th>
<th>Beamline</th>
<th>Wavelength, Å</th>
<th>Space group</th>
<th>Cell dimensions</th>
<th>a (Å)</th>
<th>b (Å)</th>
<th>c (Å)</th>
<th>α = β = γ (°)</th>
<th>Resolution (Å)</th>
<th>R_merge</th>
<th>I/σI</th>
<th>Redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MVT-101</td>
<td>to be reported elsewhere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MVT-101</td>
<td>to be reported elsewhere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MVT-101</td>
<td>to be reported elsewhere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MVT-101</td>
<td>to be reported elsewhere</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th>to be reported elsewhere</th>
<th>Resolution (Å)</th>
<th>Completeness (%)</th>
<th>No. reflections</th>
<th>Work/free set</th>
<th>R_work/R_free</th>
<th>No. atoms</th>
<th>Protein</th>
<th>Water</th>
<th>B-factors (Å²)</th>
<th>Protein</th>
<th>Inhibitor</th>
<th>Water</th>
<th>R.m.s. deviations</th>
<th>Bond lengths (Å)</th>
<th>Bond angles (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>to be deposited</th>
<th>to be deposited</th>
<th>to be deposited</th>
<th>to be deposited</th>
</tr>
</thead>
</table>
### Table G.4, continued. Data collection and refinement statistics for chemical analogues of HIV-1 protease (highest resolution is in parenthesis).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>[Asn25/25']</th>
<th>[D-Ala51/51']</th>
<th>[alloIle50/50']</th>
<th>[O-Ile50/50']</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibitor</td>
<td>MVT-101</td>
<td>KVS-1</td>
<td>KVS-1</td>
<td>KVS-1</td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>to be reported elsewhere</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha = \beta = \gamma) (°)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R_{merge})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I/\sigma I)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Redundancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Refinement**  |             |               |                 |              |
| Resolution (Å)  |             |               |                 |              |
| Completeness (%)|             |               |                 |              |
| No. reflections |             |               |                 |              |
| Work/free set   |             |               |                 |              |
| \(R_{work}/R_{free}\) |           |               |                 |              |
| No. atoms       |             |               |                 |              |
| Protein         |             |               |                 |              |
| Water           |             |               |                 |              |
| B-factors (Å\(^2\)) |           |               |                 |              |
| Protein         |             |               |                 |              |
| Inhibitor       |             |               |                 |              |
| Water           |             |               |                 |              |
| R.m.s. deviations |           |               |                 |              |
| Bond lengths (Å) |             |               |                 |              |
| Bond angles (°) |             |               |                 |              |
| **PDB ID**      | to be deposited | to be deposited | to be deposited | to be deposited |
Table G.5, continued. Data collection and refinement statistics for chemical analogues of HIV-1 protease (highest resolution is in parenthesis).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Data collection</th>
<th>Inhibitor</th>
<th>Enzyme</th>
<th>Data collection</th>
<th>Inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile50/O-Ile50′</td>
<td>to be reported elsewhere</td>
<td>KVS-1</td>
<td>O-Gly51/51′</td>
<td>JG-365</td>
<td>JG-365</td>
</tr>
<tr>
<td>[Ile50/O-Ile50′]</td>
<td>[O-Gly51/51′]</td>
<td>[Ile50/O-Ile50′]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data collection**

- **Beamline**
  - to be reported elsewhere

- **Wavelength, Å**
  - P2₁,2,2₁
  - P2₁,2,2₁
  - P2₁,2,2₁

- **Space group**
  - P2₁,2,2₁
  - P2₁,2,2₁
  - P2₁,2,2₁

- **Cell dimensions**
  - a (Å): 90.00
  - b (Å): 90.00
  - c (Å): 90.00
  - α = β = γ (°): 90.00

- **Resolution (Å)**
  - Rmerge
  - I/σI

**Refinement**

- to be reported elsewhere

- **Resolution (Å)**
  - Completeness (%)
  - No. reflections

- **Work/free set**
  - Rwork/Rfree
  - No. atoms

- **Protein**
  - Water

- **B-factors (Å²)**
  - Protein
  - Inhibitor

- **Water**

- **R.m.s. deviations**

- **Bond lengths (Å)**

- **Bond angles (°)**

| PDB ID | to be deposited | to be deposited | to be deposited |
Figure G.1. Active-site titration of chemically synthesized enzymes by surface-plasmon resonance. Chemically synthesized enzymes (Table 7.2; entries 1-7) at 100 nM concentration were pumped through a microfluidic chip over a surface containing the substrate-based reduced isostere inhibitor MVT101. WT is wild-type enzyme used as a reference.

**Surface Plasmon Resonance Assays.** Measurements were performed on a Biacore 3000 instrument in HBS-EP buffer (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% surfactant P20, pH 5.6). Biotinylated substrate-derived inhibitor was immobilized onto streptavidin-coated Sensor Chip SA (from Biacore) to obtain a response difference of about 100 RU. Each enzyme samples was then pumped through a microfluidic cell over immobilized inhibitor at multiple concentrations (at least 6 concentrations below and above calculated $K_d$) at a flow rate of 20 µL/min; association and dissociation kinetics were each studied for 120 seconds. Response curves were then fitted with BIAevaluation (v. 4.1) software with ‘two state reaction (conformational change)’ model. Only the faster values of $k_{on}$ and $k_{off}$ are listed in the Table H.6.
Table G.2. Active-site titration of chemically synthesized enzymes by surface-plasmon resonance.

<table>
<thead>
<tr>
<th>Flap</th>
<th>Flap'</th>
<th>$k_{\text{on}}$ ($\times10^5$ M$^{-1}$ s$^{-1}$)</th>
<th>$k_{\text{off}}$, s$^{-1}$</th>
<th>$K_d$, µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>OGly51</td>
<td>OGly51</td>
<td>1.62</td>
<td>0.085</td>
<td>0.52</td>
</tr>
<tr>
<td>OGly52</td>
<td>OGly52</td>
<td>0.92</td>
<td>0.067</td>
<td>0.72</td>
</tr>
<tr>
<td>OIle50</td>
<td>OIle50</td>
<td>1.46</td>
<td>0.122</td>
<td>0.84</td>
</tr>
<tr>
<td>Ile50</td>
<td>OIle50</td>
<td>1.28</td>
<td>0.093</td>
<td>0.73</td>
</tr>
<tr>
<td>Truncated</td>
<td>Gly51</td>
<td>0.019</td>
<td>0.017</td>
<td>8.9</td>
</tr>
<tr>
<td>Asn25</td>
<td>Asn25</td>
<td>1.44</td>
<td>0.122</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Figure G.2. Analytical HPLC ($\lambda = 214$ nm) and ESI mass-spectra of (a) alloIle50 [1-99]HIV-1 protease, and (b) O-Ile50 [1-99]HIV-1 protease.
Figure G.3. Analytical HPLC (λ = 214 nm) and ESI mass-spectra of (a) O-Gly51 [1-99]HIV-1 protease, and (b) O-Gly52 [1-99]HIV-1 protease.
obs $21870.5 \pm 0.6$ Da

calcd $21870.8$ Da

**Figure G.4.** (a) Analytical HPLC ($\lambda = 214$ nm) and (b) FT-ICR ESI mass-spectra of $[^{1}l]l$e50;$[^{1}l]l$e50′$]HIV-1 protease. In isotope patterns for charge state $9H^+$ (in c), red circles depict maxima for experimental measurements; blue rhombs correspond to maxima in simulated spectra.
Appendix H. Supporting information for chapter 8.

Figure H.1. Temperature dependence of 16.35 ppm and 15.45 ppm resonances. Spectra were processed with 30 Hz line broadening and were zero filled to 16 K before Fourier transformation. Lorentzian curve fitting of the peaks was used to determine their line widths (in blue).
Figure H.2. Arrhenius plots of the effect of temperature on the observed transverse relaxation rates ($1/T_{2\text{obs}} = \pi \cdot \text{line width}$) of the 16.35 ppm and 15.45 ppm peaks.

The data were fitted to the equation (1), where $T_{2\text{obs}}$ has two contributions $T_{2\text{dipolar}}$ and $k_{\text{exchange}}$. $1/T_{2\text{dipolar}}$ decreases with increasing temperature, while $k_{\text{ex}}$ increases:

$$\ln(1/T_{2\text{obs}}) = \ln[\exp(-E_d/RT + C_d) + \exp(-E_{ex}/RT + C_{ex})]$$

(1)

Table H.1. Parameters extracted from fitting to equation (1).

<table>
<thead>
<tr>
<th></th>
<th>$E_d$, kcal/mol</th>
<th>$C_d$, kcal/mol</th>
<th>$E_{ex}$, kcal/mol</th>
<th>$C_{ex}$</th>
<th>$k_{ex}$ at 273.9 K</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.35 ppm</td>
<td>-22.6</td>
<td>-37.2</td>
<td>7.17</td>
<td>17.5</td>
<td>$11.6 \cdot 10^6$ s$^{-1}$</td>
</tr>
<tr>
<td>15.45 ppm</td>
<td>-14.7</td>
<td>-22.1</td>
<td>11.6</td>
<td>25.3</td>
<td>3400 s$^{-1}$</td>
</tr>
</tbody>
</table>
Figure H.3. 900 MHz $^1$H-$^1$H NOESY with 1–1 water suppression was performed for $[L$-Ala51/51$']$HIV-1 protease with MVT-101 inhibitor in an attempt to correlate low-field resonances. Pattern of observed cross-peaks suggests similar environment for protons resonating at 16.3 ppm and 15.4 ppm, but different environment for signals at 13.0 ppm and 11.4 ppm.
Appendix I: Single-molecule fluorescence assays of proteolytic activity of HIV-1 protease.

Spectroscopic studies of enzyme catalysis\(^1\) on a single-molecule level provide an additional level of spatiotemporal resolution enabling the observation of transient intermediate states in catalysis, which may be lost in ensemble measurements. For an enzymatic reaction, where multiple turnovers occur, molecules are not synchronized in their dynamic behavior, that is, at a given point in time each enzyme molecule is at a different stage of the reaction mechanism\(^2\). When single molecules are observed, synchronization is not an issue – the reaction sequence can be monitored for each individual molecule. In such a way, the dynamic coupling between protein motions and catalysis within a single molecule can be obtained\(^3,4\). Elegant work has been done on protein enzymes acting on DNA\(^5,6\), folding and catalysis of ribozymes\(^7-9\), DNA replication machinery\(^10,11\), and protein folding\(^12-14\). In the pioneering works of Xie and co-workers\(^15,3\), it became apparent that enzymes display fluctuation of catalytic rates caused by conformational heterogeneity of enzyme molecules\(^2,16\).

Proteases are another important class of enzymes, which are, moreover, implicated in many diseases. Study of catalytic mechanisms of proteases with a single-molecule spectroscopy would also allow us to gain additional insights into substrate recognition\(^17\), and the autoprocessing of the protease out of larger polyprotein (zymogen)\(^18\). In addition, the methodology could be applied to ultra-sensitive and fast diagnosis / detection of a disease (e.g. cancer\(^19\)) or viral infection, where some work has already been done in this direction\(^20-22\). Herein we report attempts of proof-of-principle
experiments by monitoring the proteolytic cleavage of FRET (fluorescence resonance energy transfer)-labelled peptides by HIV-1 protease on a single-molecule level.

Two principally different schemes for performing single-molecule experiments were tested (Figure I.1). In the first approach, fluorophore-labelled peptide corresponding to p2/NC cleavage site in Gag/Pol polyprotein was equipped with a biotin-tag (Figure I.2). This strategy has allowed the immobilization of the substrate on a neutravidin-coated surface and the monitoring of the proteolytic cleavage reaction upon addition of HIV-1 protease with the help of total internal reflection (TIR) microscopy (Figure I.1 a). In the second set of experiments, ‘covalent dimer’ HIV-1 protease was labelled with a biotin-tag (Figure I.2) and further immobilized on the surface. A short peptide substrate labelled with a FRET-pair was then added to the flow cell and the reaction was again monitored by TIR microscopy (Figure I.1 b).

In the first experimental design, we have chosen a 17-residue polypeptide with only one NH$_2$-functionality at its N-terminus and one SH-group of the C-terminal cysteine residue (see Experimental Section). These two reactive groups were further used for fluorophore-labelling via succinimide active ester and maleimide chemistries, respectively (see Scheme I.1). A random coil conformation for such polypeptide was revealed by circular dichroism measurements in aqueous 50 mM phosphate buffer (pH 5.5), and, in the case of FRET Cy3 / Cy5 dye donor / acceptor pair, donor to acceptor distance was modeled to be $\sim$6.5 nm assuming $\beta$-strand peptide conformation.

The polypeptide was tethered to a polymer-passivated quartz surface coated with neutravidin and HIV-1 protease was added at 0.5-18 $\mu$M concentrations. The activity was monitored with time resolution of 30-500 ms. Initially high FRET-transfer was observed
resulting in higher fluorescence intensity of the acceptor dye (see Figure I.3a). In the case of substrate labelled with Cy-5 dye at its N-terminus, proteolytic cleavage resulted in the rise of the donor fluorescence signal (Figure I.4), combined with the disappearance of the acceptor signal due to dissociation of Cy5-containing product peptide. Interestingly, increase of the donor fluorescence and concomitant decrease of acceptor signal was not abrupt but step-wise (Figure I.4). Such behavior suggests either multi-step binding of the HIV-1 protease to the substrate and ‘search’ for the location of cleavage or conformational isomerisations of substrate-polypeptide and HIV-1 protease, or both.

To confirm that this is indeed a multistep binding / cleavage process and not the multi-step photobleaching, we performed the same experiments with a polypeptide where the native amide-bond \([\text{C(O)NH}]\) at cleavage p2/NC location was replaced by a reduced isostere fragment \([\psi\text{CH}_2\text{NH}]\), effectively transforming the substrate into an inhibitor of the HIV-1 protease. In the experiments with fluorophore-labelled inhibitor, we again observed increase of the donor fluorescence and corresponding decrease of acceptor fluorescence; however, the intermediate FRET-level state was much longer living. This result was encouraging, however, in some traces for substrate or inhibitor only experiments (no HIV-1 protease was added) we have also observed multi-step decrease of FRET, clearly indicating that photobleaching is a complicating factor with these fluorophore-labelled peptides. The experiments which support that proteolysis was indeed observed in single molecule traces are: 1) concentration dependence of apparent rate of decrease of Cy5-fluorescence and concomitant increase of Cy3-fluorescence (10-fold increase of the concentration of HIV-1 protease from 560 nM to 5.6 µM led to visibly faster FRET-decrease rate, see Figure I.4); and 2) high population of intermediate-
FRET state in experiments with corresponding inhibitor (compare Figure I.3 d and Figure I.6 d).

An alternative substrate was tested, where the N-terminus was labelled with the Cy3-donor and the Cy-5 acceptor was now attached closer to the surface (see Scheme I.2 c,d). In such a configuration, one would expect a decrease of Cy5(acceptor)-fluorescence and concomitant decrease of Cy3(donor)-fluorescence due to cleavage of Cy3-peptide product and its dissociation and diffusion into bulk solution. In many single-molecule trajectories we have, however, noticed that fluorescence of the Cy3-fluorophore does not diminish upon adding of HIV-1 protease, whereas fluorescence of the Cy5-fluorophore decreases. This may indicate that Cy3-dye does not evade volume of detection and remains attached either to the protein molecule or a surface nearby, perhaps due to Cy3-dye hydrophobicity or again the photobleaching of Cy5-dye might be complicating factor in the interpretation of the data.

In order to improve characteristics of the substrate we tried alternative dye configurations, such as Alexa555 / Alexa 647 and Cy3D / Cy5 (see Scheme I.3). Alexa dyes are known for their hydrophilic properties and high photostability – this could potentially improve the assay. Cy3B is a more photo-stable analogue of Cy3-dye (see Scheme I.4 a).

With the new substrates we observed noticeably higher photostability, however, in many single-molecule traces photobleaching was still taking place. Strikingly, in some traces even without enzyme we again found a multi-step decrease of acceptor signal and concomitant multi-step increase of donor signal. This strongly suggests that photobleaching events interfere with real proteolytic cleavage. We have performed assays
in which the initial timing of optical detection was set to be coincidental with the addition of HIV-1 protease. In such cases, we found Alexa647-fluorescence was more stable than without protein, meaning that the protein stabilizes acceptor dye fluorescence against photobleaching. One can speculate therefore that in the peptide which has a random coil structure, close interactions of donor dye and acceptor dye may take place inducing quenching and photobleaching.

In the second approach, we tethered ‘covalent dimer’ HIV-1 protease to the surface and monitored proteolytic cleavage reaction after addition of FRET-labelled substrate. Although in many cases we have seen single-molecule traces which could be interpreted as proteolytic cleavage, they can also be interpreted as photobleaching of dye molecules. Given that it is very difficult to design control experiments we abandoned this scheme. Because of the ambiguity in interpretation of the data we abandoned this project.
Figure I.1. Two approaches for single-molecule detection of proteolytic activity of HIV-1 protease: (a) peptidic substrate labeled with FRET-pair (Cy3 and Cy5 dyes) and equipped with biotin-tag for immobilization on neutravidin-coated quartz slide. Protease is then added and progress of proteolytic reaction is then monitored by fluorescence changes with the help of total internal reflection (TIR) fluorescence microscope; (b) HIV-1 protease chemically synthesized as covalent dimer (see Figure I.2) bearing biotin-tag immobilized on neutravidin-coated quartz surface. Substrate labeled with Cy3 and Cy5 dyes is then added and reaction is monitored with TIR microscope.
Figure I.2. In the total chemical synthesis of 203-amino acid ‘covalent dimer’ HIV-1 protease molecule, last alkylation step is replaced by the reaction with maleimide-containing PEGylated biotin-tag, which is further used to tether such HIV-1 protease construct to the quartz surface employing biotin-neutravidin method.
Figure I.3. Monitoring the cleavage reaction of Cy5-substrate-Cy3-biotin with TIR microscopy. In (a), panel combining ‘donor’ (left) and ‘acceptor’ channels for substrate immobilized on surface, where FRET efficiency reaches ~85% (in b). In (c) HIV-1 protease is added leading to cleavage of the substrate, where Cy5-dye diffuses from detection volume leading to reduction of high-FRET peak (in d). Note that the peak at 20% is leakage of ‘donor’ light due to imperfection of optical filtering.
Figure I.4. Average donor (green) and acceptor (red) signal intensities plotted versus time. For proteolytic reaction monitored after addition (signal spike in the beginning of the plot) of (a) 560 nM of HIV-1 protease in (note that total time is equal to 5.5 minutes) and (b) 5.6 µM of HIV-1 protease (total time in graph is equal to 2.8 minutes).
Figure I.5. Examples of single-molecule traces for proteolysis of Cy5-substrate-Cy3-biotin. (a) Fluorescence intensity (Cy5 in red and Cy3 in green) plotted versus time. (b) FRET efficiency calculated for (a) and displayed separately. (c) FRET efficiency for other single-molecules. Multistep decrease of FRET-level may be explained by conformational isomerisations or searching for cleavage site steps (indicated by arrows).
Figure I.6. Monitoring the binding of HIV-1 protease to Cy5-inhibitor-Cy3-biotin with TIR microscopy. In (a), panel combining ‘donor’ (left) and ‘acceptor’ channels for substrate immobilized onto surface, where FRET efficiency reaches ~85% (in b). In (c) HIV-1 protease is added leading to decease near equilibration of fluorescence intensities in donor and acceptor channels. Histogram in (d) shows high abundance of high FRET-peak (~85%) as well as medium FRET-peak (~55%) not observed in the case of substrate (see Figure I.3 d). Note that the peak at 20 % is leakage of ‘donor’ light due to imperfections in optical filtering.
Figure I.7. Single-molecule fluorescence traces for HIV-1 protease binding to FRET-labelled inhibitor. Many molecules displayed behavior where FRET-level of medium intensity was highly abundant.
Figure I.8. TIR-monitoring of proteolytic reaction of HIV-1 protease tethered to quartz slide (see Figure I.1 b and Figure I.2). (a) Donor and acceptor channels are shown, where bright spots are fluorescent single-molecules right after the Cy3-substrate-Cy5 is added, and in (b) the same region of the surface after 2 hours. (c) Single-molecule fluorescence traces, where the molecule first binds (high-FRET) then undergoes cleavage, leading to dissociation of one of the dyes.
Experimental section.

Peptides were synthesized with Boc-SPPS method\textsuperscript{23} and fluorophore-labelled as outlined in Scheme I.1 (see Schemes I.2 – I.4 and Figures I.9 – I.12 for structures of labeled peptides and analytical characterization). ‘Covalent dimer’ HIV-1 protease was synthesized analogously as described in Chapter 2, where the last alkylation step was replaced by reaction with biotin-containing tether unit trough maleimide-chemistry (see Figure I.2). Peptides or ‘covalent dimer’ HIV-1 protease were immobilized at 20-100 pM concentration in their stock solutions in 50 mM NaOAc, 0.4 % glucose, 1% BME (v/v), 10% glycerol (v/v), 1% Gloxy (catalase, glucose oxidase). Images were obtained in a wide-field total-internal-reflection microscope with 30-500 ms time resolution using an electron multiplying charge-coupled device (CCD) camera and homemade C++ program written by S. A. McKinney (Taekjip Ha laboratory at University of Illinois at Urbana-Champaign). All measurements were performed at 22 °C in 50 mM NaOAc, 0.4 % glucose, 1% BME (v/v) and 1% Gloxy (catalase, glucose oxidase) added as an oxygen scavenging system\textsuperscript{24} to slow photobleaching. HIV-1 protease was added at 0.5-18 µM concentrations. FRET values were calculated as the ratio between the acceptor intensity and the total intensity. More details of the experimental setup can be found elsewhere\textsuperscript{10,11,24}. 
Scheme I.1. In the synthesis of substrates and inhibitors for single-molecule spectroscopic observation of proteolytic activity of HIV-1 protease, the polypeptide synthesized with the help of Boc-SPPS chemistry contained one amino-group functionality (either peptide N-terminus or amino-group of lysine) and one sulfhydryl-group of cysteine. Such scheme has allowed stepwise site-specific labelling of peptide with two different dyes (‘donor’ and ‘acceptor’ fluorophores). Biotin-group was introduced for surface tethering purposes either during Boc-SPPS or after dye-labelling in a separate step.
Scheme I.2. Structures of fluorophore-labeled peptide substrates and reduced isostere inhibitors (p2/NC cleavage site in red).
(a) Alexa647-substrate-Alexa555-PEG-Btn

(b) Alexa555-substrate-Alexa647-PEG-Btn

**Scheme I.3.** Alexa647 / Alexa555 dyes were tested to improve photostability and make substrates more hydrophilic.
(a) Cy3B-substrate-Cy5-PEG-Btn

(b) Cy3-substrate-Cy5

Scheme I.4. Cy3B dye was introduced as a more photostable variant of Cy3-dye (in a). In (b) structure of Cy3 / Cy5 labelled substrate tested in experiment with HIV-1 protease construct tethered to the surface (see Figure I.1 b).
Figure I.9. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides. (a) Cy5-substrate-Cy3-Btn (Scheme I.2 a); (b) Cy5-inhibitor-Cy3-Btn (Scheme I.2 b).
Figure I.10. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides. (a) Cy3-substrate-Cy5-Btn (Scheme I.2 c); (b) Cy3-inhibitor-Cy5-Btn (Scheme I.2 d).
Figure I.11. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides. (a) Alexa647-substrate-Alexa555-Btn (Scheme I.3 a); (b) Alexa555-substrate-Alexa647-Btn (Scheme I.3 b).
Figure I.12. Analytical HPLC ($\lambda = 214$ nm) and ESI-MS characterization of chemically synthesized fluorophore labelled peptides. (a) Cy3B-substrate-Cy5-Btn (Scheme I.4 a); (b) Cy3-substrate-Cy5 (Scheme I.4 b).
References:


Appendix J. Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme.
Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme

Thomas Durek, Vladimir Yu. Torbeev, and Stephen B. H. Kent

PNAS published online Mar 8, 2007;
doi:10.1073/pnas.0610630104

This information is current as of March 2007.

<table>
<thead>
<tr>
<th>Supplementary Material</th>
<th>Supplementary material can be found at: <a href="http://www.pnas.org/cgi/content/full/0610630104/DC1">www.pnas.org/cgi/content/full/0610630104/DC1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This article has been cited by other articles: <a href="http://www.pnas.org#otherarticles">www.pnas.org#otherarticles</a></td>
</tr>
<tr>
<td>E-mail Alerts</td>
<td>Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.</td>
</tr>
<tr>
<td>Rights &amp; Permissions</td>
<td>To reproduce this article in part (figures, tables) or in entirety, see: <a href="http://www.pnas.org/mise/rightperm.shtml">www.pnas.org/mise/rightperm.shtml</a></td>
</tr>
<tr>
<td>Reprints</td>
<td>To order reprints, see: <a href="http://www.pnas.org/mise/reprints.shtml">www.pnas.org/mise/reprints.shtml</a></td>
</tr>
</tbody>
</table>

Notes:
Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme

Thomas Durek, Vladimir Yu. Torbee, and Stephen B. H. Kent*

Institute for Biophysical Dynamics, Department of Biochemistry and Molecular Biology, Department of Chemistry, University of Chicago, 929 East 57th Street, Chicago, Ill. 60637

Edited by Brian W. Matthews, University of Oregon, Eugene, OR, and approved January 9, 2007 (received for review November 30, 2006)

In this article, we report the total chemical synthesis of human lysozyme. Lysozyme serves as a widespread model system in various fields of biochemical research, including protein folding, enzyme catalysis, and amyloidogenesis. The 130-aa wild-type polypeptide chain of the human enzyme was assembled from four peptidyl segments by using native chemical ligation in a fully convergent fashion. Key to the assembly strategy is the application of the recently developed kinetically controlled ligation methodology, which affords an efficient control over the ligation of two peptide “thioesters” to yield a unique product. This result enables the facile preparation of a 64-residue peptide “thioester;” this segment is then linked by native chemical ligation to a 66-aa Cys peptide, to yield the target 130-aa polypeptide chain. The synthetic polypeptide chain was folded in vitro into a defined tertiary structure with concomitant formation of four disulfides, as shown by 2D TOSCY NMR spectroscopy. The structure of the synthetic human lysozyme was confirmed by high-resolution x-ray diffraction, giving the highest-resolution structure (1.04 Å) observed to date for this enzyme. Synthetic lysozyme was obtained in good yield and excellent purity and had full enzymatic activity. This facile and efficient convergent synthesis scheme will enable preparation of unique chemical analogs of the lysozyme molecule and will prove useful in numerous areas of lysozyme research in the future.

chemical protein synthesis | kinetically controlled ligation | native chemical ligation | peptide thioester | protein folding

Lysozyme is possibly one of the best-studied enzymes. The x-ray structure of hen egg-white lysozyme, initially reported in 1965, was the first high-resolution 3D structure of an enzyme molecule (1). Since then the protein has served as a model system for the study of protein folding and misfolding, enzyme catalysis and mechanism, x-ray crystallography, enzyme evolution, and protein engineering (2–9). Moreover, human lysozyme recently has attracted considerable interest because certain mutations in the enzyme have been shown to render the protein amyloidogenic (10, 11).

Despite extensive genetic, structural, and physico-chemical studies carried out over the last 50 years, many questions regarding lysozyme folding, catalysis, and amyloid fibril formation remain unsolved, unsatisfactorily explained, or controversial. This deficit is at least in part attributable to the limited means that could be used to modify the chemical structure of the lysozyme molecule. More powerful and versatile control over the structure of the enzyme is required for a detailed understanding of the properties of the protein on the molecular or atomic scale. Chemical protein synthesis has emerged as a powerful tool in this respect, especially because it grants nearly absolute control over the covalent structure of an enzyme molecule (12–14). Given the widespread and long-lasting interest in lysozyme research, it is not surprising that for exactly these reasons chemical synthesis of the full-length protein has been envisioned (and attempted) as early as the 1970s (15, 16).

However, these early experiments were unsuccessful, and despite significant progress made in the field, robust chemical access to the 130-aa family of lysozyme proteins has not been established so far.

Modern approaches to the chemical synthesis of proteins involve the synthesis by stepwise solid-phase peptide synthesis (SPPS) of relatively small (~50-residue) peptide segments, which span the entire sequence of the polypeptide chain of the protein. These unprotected segments then are linked by chemical ligation reactions to give the full-length polypeptide, which then is folded in vitro to form the active protein. Critical to this synthetic strategy was the development of chemical ligation approaches, which make it possible the chemoselective linking of unprotected peptide segments in good yield (17). Native chemical ligation (NCL) (18) is the most successful chemoselective ligation chemistry to date and has enabled the synthesis of a number of proteins, which often were equipped with nonnative features (such as biological probes, backbone modifications, D-amino acid residues, or glycans mimetics) to address specific experimental questions (19–22). NCL involves the reaction of an unprotected peptide “thioester with another unprotected peptide carrying an N-terminal cysteine. Initial reversible transthioesterification between the sulphydryl group of the N-terminal cysteine and the peptide “thioester gives a thioester-linked intermediate, which spontaneously rearranges in a rapid second step to form a native peptide bond (18).

Most proteins synthesized so far by NCL have been constructed from merely two peptide segments, thus limiting chemical access to target proteins of ~100 or fewer amino acids (13). To gain synthetic access to longer polypeptide chains, ligation of a larger number of peptide segment building blocks must be used. To date, essentially all three-segment syntheses have been performed in a rather inflexible fashion by sequential ligations starting from the C-terminal peptide segment with extension toward the N terminus. Multiple rounds of ligation and intermediate product purification typically result in substantial losses. This problem has been minimized by carrying out several ligations in a one-pot manner (23), but the rapid build up of impurities effectively limits such one-pot syntheses to only three segments. For these reasons, a more efficient convergent synthetic strategy is needed.

We recently introduced the concept of kinetically controlled ligation (KCL) (24), which enables the reaction of a peptide thioester and a Cys-peptide thioester to yield a single product. This process enables the synthesis of a protein in a fully convergent fashion (24). In a convergent synthesis (Scheme 1), each starting peptide segment is approximately the same number of chemical transformations away from the final product (25).
fact becomes particularly significant when multiple analogs of a
given target have to be prepared and the sites of modification are
scattered across the entire sequence. Convergent synthesis, in
principle, will increase final yields when compared with se-
quential assembly tactics (25).

In this article, we report the reproducible convergent chemical
synthesis of human lysozyme, a 130-residue protein molecule
containing four disulfides. The structure of the synthetic protein
was confirmed by mass spectrometry, 2D TOCSY NMR, and
high-resolution (1.9 Å) x-ray diffraction. The synthetic enzyme
had full catalytic activity.

Results and Discussion

Designing a Synthesis of Human Lysozyme. Our convergent strategy
for the total chemical synthesis of the human lysozyme molecule is
shown in Scheme 1. Human lysozyme contains eight cysteines,
which are distributed evenly across the polypeptide chain and form
four disulfide bonds in the folded enzyme molecule (Scheme 1). All
cysteines could serve as potential ligation sites. We envisioned a
final ligation of two large peptide segments to give the 130-residue
polypeptide chain. To optimize the synthesis, we used four peptide
segments of approximately equal length: peptides I–29, 30–64,
65–94, and 95–130. These peptides are of convenient length
(29–35 aa), and their preparation by optimized stepwise Boc
chemistry SPPS (26) was expected to be straightforward. In terms
of reaction rates, the chosen ligation sites (Met1–Cys10, Trp9–Cys15,
and Ala20–Cys25) were selected to be straightforward, i.e., product
formation was expected to occur rapidly (27) (unlike with Ile/Val/
Thr/Pro-Cys ligation sites).

According to Scheme 1, generation of the N-terminal half of the
enzyme (corresponding to residues 1–64) can be achieved by KCL
of [Lys1-Met10]-thiol-ester with [Cys10-Trp15]-thioester to form [Lys1-Trp15]-thioester. The peptide
thioester was generated by transthioesterification from the
thioester by using an excess of 4-mercaptophenyl-
acetic acid (MPA). The latter compound was shown to be a
superior thiol catalyst to the established and widely used
thiobenzyl, and it is particularly useful for the facile conversion of
a peptide thioester to the corresponding peptide thio-

Scheme 1. Convergent synthesis of human lyso-
zyme. The 130-residue polypeptide is assembled from four
segments of comparable length in a symmetrical fash-
ion. Key to the synthetic strategy used is the KCL of
[Lys1-Trp15]-thioester with [Cys10-Trp15]-thioester and the
temporary protection of Cys10. The 130-residue target build-up sequence
of human lysozyme with cysteines underlined. Steps:
(1) transthioesterification; (2) KCL (3) transthioester-
ification; (4) NCL, (5) Cys deprotection; (6) NCL, and
(7) AcCHO removal, disulfide formation, and fold-
ing. R = alkyl.

Synthesis of Peptide Segment Building Blocks. Peptide segments
were prepared by using in situ neutralization Boc chemistry
SPPS protocols as described in ref. 26. Segments I–29, 30–64,
and 65–94 were prepared on modified TAMPAL resins gen-
erating C-terminal thioesters upon HF cleavage (27).
Segment 95–130 carrying a free carboxyl group was synthe-
sized on β-CD-L carrier resin. All five tryptophans were incor-
porated as Trp(CHO), and His18 was incorporated as
His(Dnp). As expected, both the Trp(CHO) and His(Dnp)
side-chain protecting groups were unaffected by the
HF/pe-

superior reactivity of "thioesterlesters versus "thioalkylesters," Therefore, we decided to use the S-Acm group to protect the thiol-containing side chains of all five nonligation site cysteines.

Assembly of the N-Terminal Half. For KCl of [Lys-Met-Cys-Thr-Pro-Trp(CHO)]-thioester and [Cys-Thr-Pro-Trp(CHO)]-thioester, the segment 1–29 first was transthioesterified with MPAA in a separate reaction (Scheme 1a) (24, 25). The resulting purified [Lys-Met-Cys-Thr-Pro-Trp(CHO)]-thioester 1 was reacted with [Cys-Thr-Pro-Trp(CHO)]-thioalkylester 2 in 6 M guanidine hydrochloride/0.2 M sodium phosphate under KCL conditions, i.e., in the absence of added thiol catalyst (Scheme 1a). Liquid chromatography (LC) analysis (Fig. 1) indicated that most starting material was consumed within 5 h, and a heterogeneous compound mixture consisting mainly of [Lys-Met-Cys-Thr-Pro-Trp(CHO)]-Cy3-thiolactone 3 and the branched thioester [Lys-Met-Cys-Thr-Pro-Trp(CHO)]-thioester 4 had been formed (Fig. 1B). The product mixture was resolved by adding sodium 2-mercaptoethanesulfonate (MESNA), which effectively reversed all undesired side products by transthioesterification to give the N-terminal half [Lys-Met-Cys-Thr-Pro-Trp(CHO)]-MESNA thioester 5 in good yield (51%) after purification. LC-MS observed (obsd.) mass, 7,799.6 ± 0.8 Da; calculated (calcd.) mass, 7,798.7 Da (average isotopes) (Fig. 1C).

The formation of such unwanted thioester species is typical of KCL and represents an undesirable side reaction. Especially formation of the branched thioester 4 can be expected to be the principal reason for nonquantitative KCL (aside from oligomerization and cyclization) because it dissipates precious activated peptide "thioesterlester.

Assembly of the C-Terminal Half. The C-terminal half of lyszyme representing residues 65–130 was obtained by ligating [Thr-Ala-Ala]-thioester and [Cys-Ala-Ala-Ala]-thioester [25]. Segment 65–94 contains a C-terminal "thioester and an N-terminal cysteine. To prevent undesired side reactions (cyclization, oligomerization, etc.), during the envisaged first ligation reaction with segment 95–130 we protected Cys68 as the Thr. Conversion of Thr to Cys is achieved readily by treatment with alkylamines at pH 4, generating the desired free N-terminal cysteine, which can be used in another round of ligation (23, 29).

Before ligation, [Thr-Ala-Ala]-thioesterlester was transthioesterified with 120 mM 2-mercaptoethanol-lysine in 6 M guanidine hydrochloride and 0.2 M sodium phosphate at pH 6.8. This procedure not only generated the more reactive [Thr-Ala-Ala]-thioester[4-cysteinyl] ester 6 but also led to rapid and efficient dithioflavine (DNP) removal from His69, thus simplifying the ligation reaction product mixture (Scheme 1b). Ligation of purified [Thr-Ala]-thioester[4-cysteinyl] ester 6 and [Cys-Ala-Ala-Ala]7 was carried out in 6 M guanidinium hydrochloride, 0.2 M sodium phosphate, 30 mM MPAA, and 20 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) at pH 6.8 (Scheme 1b). The reaction was complete within 2 h (as determined by LC analysis) (Fig. 2) after which methionine HCl was added to 0.4 M and the pH was adjusted to 4.0 to effect the conversion of Thr68 to Cys68 (Scheme 1b). Deprotection was quantitative within 6 h, and the resulting 66-residue peptide 8 was purified by reversed-phase HPLC (RP-HPLC) (obsd., electrospray ionization (ESI)-MS mass, 7,539.8 ± 0.7 Da; calcd., 7,592.9 Da).
comparing AgOAc and Hg(OAc)$_2$ deprotection protocols under otherwise similar conditions. In our hands, AgOAc-mediated Acm removal was significantly slower than Acm deprotection by Hg(OAc)$_2$. However, MALDI-MS analysis indicated that decomposition of the formed Hg-thiolate compounds turned out to be more challenging and required longer reaction times and a larger excess of DTT when compared with the corresponding Ag-thiolates (data not shown). For this reason, we chose to stick to the AgOAc protocol, which also has been reported to be milder in the presence of sensitive amino acid residues (51, 32).

Trypsin and formyl protecting groups (five in total) were then removed by treating the crude peptide with 20% (v/vol) piperidine and 30% (v/vol) 2-mercaptoethanol in aqueous guanidine hydrochloride solution for 40 min at 0°C. The HPLC-purified fully deprotected and reduced 130-aa polypeptide [supporting information (SI) Fig. 6A] was obtained in reasonable yield (28% overall yield for [Ig4] + removal of Ac and formyl protecting groups + HPLC purification), based on the starting reagents 8 and 8: ESI-MS obsd. average mass, 14,700.2 ± 1.5 Da; calcld., 14,700.7 Da.

Folding, Disulfide Formation, and Product Characterization. The purified 130-residue polypeptide chain was folded by dialysis into buffer containing a redox system consisting of 5 mM oxidized glutathione and 2 mM DTT at pH 8 (final peptide concentration after dialysis was 0.175 mg/ml). After 14 h, formation of the correctly folded structure was revealed on LC analysis by the appearance of a sharp peak eluting at earlier retention time (SI Fig. 6). Such a shift in retention time has been observed in numerous cases upon (re)formation of disulfide cross-linked globular proteins and reflects the burial of hydrophobic residues within the protein core. The sharp peak was preceded by a poorly resolved uniden-

![Diagram](image_url)
Fig. 4. Characterization of synthetic human lysozyme. (A) LC analysis of purified and folded synthetic lysozyme. Chromatographic separations were performed as described in the Fig. 1 legend. (Inset) ES-MS spectrum. The calculated mass is 14,692.7 Da. Deconvolution of the ES-MS spectrum yields an observed mass of 14,693.4 ± 0.3 Da. (B) 2D TOCSY 1H-15N NMR spectrum showing aliphatic spin systems. (C) Clearance of a bacterial cell-wall suspension by synthetic human lysozyme (Curve I). The arrowhead indicates the time when enzyme or buffer was added to the cuvette. Negative control (buffer blank), curve II. The calculated specific enzymatic activity was 72,250 ± 1,000 units/mg of sample, comparable with recombinant human lysozyme preparations.

Fig. 5. X-ray structure of synthetic human lysozyme. (A) Ribbon representation of the x-ray structure of synthetic human lysozyme. (B) Final 2Fo – Fc electron density map around the active site at 1.64 Å resolution. (C) Superposition of the synthetic lysozyme structure (red) with lysozyme structures obtained from biosynthetic sources (PDB ID codes: 1SF5, green; 1W7T, blue).

A suspension of Micrococcus lysodeikticus cells led to a rapid decrease in turbidity, revealing that enzymatically active protein indeed had been obtained (Fig. 4C). The specific activity of the synthetic material was determined as 72,250 ± 1,000 units/mg of sample, which is comparable to reported values for commercially available recombinant material of high purity. Based on these biochemical and structural data, we conclude that chemically synthesized human lysozyme is a homogeneous preparation identical to human lysozyme isolated from biological sources. A typical preparation of synthetic lysozyme yielded ~2.5 mg of native enzyme. This amount of enzyme was more than sufficient to perform all biophysical and biochemical experiments described above.

Conclusions

In this article, we report on efficient and robust total chemical synthesis of human lysozyme. The 130-aa residue protein molecule

4850 | www.pnas.org/cgi/doi/10.1073/pnas.0610638104

Duruk et al.
was assembled from four peptide segments in a fully convergent manner by using modern chemical ligase methods. The size (≈30 an) of the peptide segments used makes them easily accessible by SPPS combined with standard HPLC purifications, even in the hands of only moderately experienced peptide chemists. Our convergent synthetic design is especially well suited for the preparation of various lysosome analogs because any desired chemical modifications or biosynthetic labels can be introduced at any position of the molecule without the need for a strategy change. Because of a high degree of sequence conservation within the c-type lysosome family and a highly conserved location of cysteine residues in particular (7), we also believe that the synthetic approach will be applicable to other c-type lysosome family members, for example, liver, kidney, and spleen.

Convergent synthesis is made possible by KCL, which is known to be a practical way of controlling the reactivity of peptide thioesters (24). The 13-oxoisole lysosome protein molecule is the second and thus far the largest example prepared in a convergent fashion by using a combination of NCL and KCL. The convergent synthesis of lysosome proteins such as this, which has been shown in this study, represents an important achievement in the field.

Materials and Methods

Peptide Synthesis and Purification. Peptides were synthesized as described in refs. 26 and 27 (see SI Text). The identity of the peptides was confirmed by liquid chromatography–mass spectrometry for each peptide was determined for a particular (7), we also believe that the synthetic approach will be applicable to other c-type lysosome family members, for example, liver, kidney, and spleen.

Convergent synthesis is made possible by KCL, which is known to be a practical way of controlling the reactivity of peptide thioesters (24). The 13-oxoisole lysosome protein molecule is the second and thus far the largest example prepared in a convergent fashion by using a combination of NCL and KCL. The convergent synthesis of lysosome proteins such as this, which has been shown in this study, represents an important achievement in the field.

Materials and Methods

Peptide Synthesis and Purification. Peptides were synthesized as described in ref. 26 and 27 (see SI Text). The identity of the peptides was confirmed by liquid chromatography–mass spectrometry for each peptide was determined for a particular (7), we also believe that the synthetic approach will be applicable to other c-type lysosome family members, for example, liver, kidney, and spleen.

Convergent synthesis is made possible by KCL, which is known to be a practical way of controlling the reactivity of peptide thioesters (24). The 13-oxoisole lysosome protein molecule is the second and thus far the largest example prepared in a convergent fashion by using a combination of NCL and KCL. The convergent synthesis of lysosome proteins such as this, which has been shown in this study, represents an important achievement in the field.

Materials and Methods

Peptide Synthesis and Purification. Peptides were synthesized as described in ref. 26 and 27 (see SI Text). The identity of the peptides was confirmed by liquid chromatography–mass spectrometry for each peptide was determined for a particular (7), we also believe that the synthetic approach will be applicable to other c-type lysosome family members, for example, liver, kidney, and spleen.

Convergent synthesis is made possible by KCL, which is known to be a practical way of controlling the reactivity of peptide thioesters (24). The 13-oxoisole lysosome protein molecule is the second and thus far the largest example prepared in a convergent fashion by using a combination of NCL and KCL. The convergent synthesis of lysosome proteins such as this, which has been shown in this study, represents an important achievement in the field.