

# HIV Protease Inhibition by Dimerization

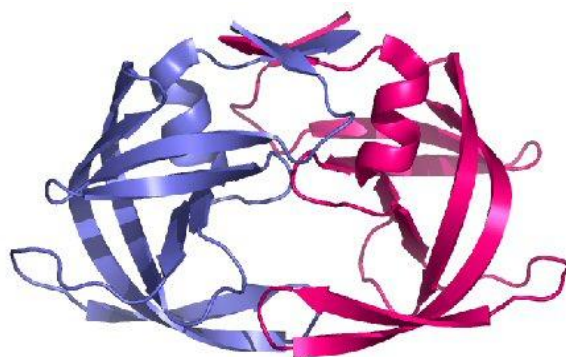
---

## Introduction

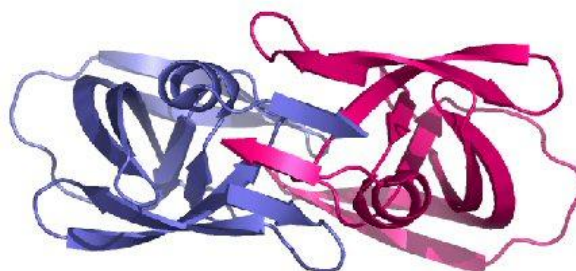
Human immunodeficiency virus was discovered in 1983 and ever since, efforts have been made to find an effective antiviral drug. There are a few enzymes, crucial for HIV replication, recognized in the HIV lifecycle. First, when the HIV enters a carrier cell, the HIV **reverse transcriptase** converts the viral RNA genome into double-stranded DNA. Subsequently, the viral DNA is integrated into the carrier cell DNA by HIV **integrase**. After integration follows transcription, translation and post-translational proteolytic processing by HIV **protease** (PR, see figure 1). The very first HIV drugs targeted inhibition of the reverse transcriptase. Along with the findings and characterization of HIV integrase and protease, more and more research has focused on inhibition of those enzymes as well. All three enzymes are viable targets and it is desirable to develop a number of diverse antiviral drugs to use against different strains and mutations of the HIV, possibly as multidrug treatments.<sup>1</sup>

Most HIV drugs in today's market are still reverse transcriptase inhibitors. HIV protease is the second most abundant target of the FDA approved anti-HIV drugs, and there are also a small number of approved drugs targeting HIV integrase and other targets.<sup>a</sup> The majority of the drugs target active and/or allosteric sites of the enzymes although some of the more recent commercially available drugs also inhibit protein-protein interactions within the enzymes, inhibiting polymerization of the monomers constituting the active enzymes.<sup>1</sup>

The function of the HIV protease is to cleave gag (group-specific antigen) and gag-polymerase precursor proteins into capsid proteins. Inhibiting this cleavage leaves the un-cleaved polypeptide substrates unable to infect new cells.<sup>1</sup>



*Figure 1. The three dimensional structure of the HIV protease homodimer. The two monomers are shown in blue and red.<sup>b</sup>*



*Figure 2. HIV protease with the anti-parallel four-strand  $\beta$ -sheet shown in the middle.<sup>b</sup>*

---

<sup>a</sup> <http://www.fda.gov/oashi/aids/virals.html>

<sup>b</sup> <http://chemistry.umeche.maine.edu/MAT500/Proteins12.html>

Rather much is known about the HIV protease. It belongs to the family of aspartyl proteases and it is a homodimer, made up of two 99 amino acid residue subunits. The active site of the protease is located at the interface between the two subunits, utilizing the Asp25 residues of each subunit as catalytic residues.

Traditional HIV protease inhibitors target the active site, acting as peptidomimetic substrate analogues. One of the major drawbacks of this approach is drug resistance from viral mutations. Another approach is to inhibit the dimerization of the protease. In order to make a large impact, large molecules have been used when it comes to such large targets as subunit interfaces. Small molecules have traditionally been used to target active sites and regions associated with conformational change of the target enzyme, but not extensively for subunit interface targeting. Although in a modern approach, considering small 'hot spots' being responsible for the vast majority of the dimerization energy, it would be rational to try to use smaller molecules to target those 'hot spots'.

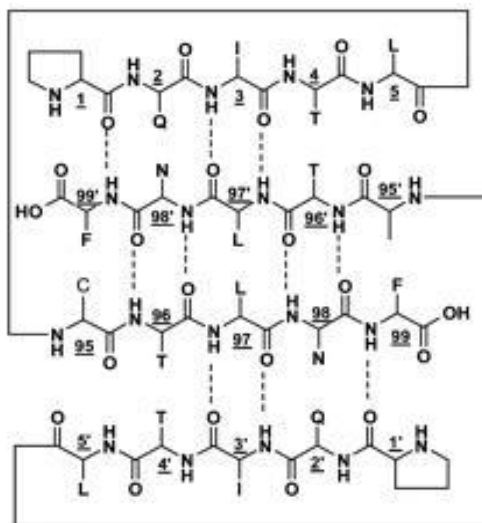


Figure 3. The amino acid sequence in the  $\beta$ -sheet.<sup>2</sup>

Investigations of the structural stability of HIV protease have identified a few 'hot spots' which contribute to the majority of the interaction. The most energetically favorable monomer-monomer interaction seems to be an anti-parallel four-strand  $\beta$ -sheet (see figures 2 and 3), formed by interdigitation of the C-terminal strands (Pro1-Gln2-Ile3-Thr4) and N-terminal strands (Thr96-Leu97-Asn98-Phe99) of both monomers<sup>2</sup>. The residues Cys95, Thr96, Leu97, Asn98, Phe99 and Pro1, Ile3, Leu5 are reported to supply almost 75% of the Gibbs free energy of dimerization, according to structure-based thermodynamic analysis.<sup>3</sup> Residues [Thr26, Gly27, Asp29] and [Gly49, Ile50, Gly51] have also been identified as two, less significant, 'hot spots'.<sup>3</sup> It has also been shown that mutations into alanine at positions 3, 5, 96, 97, 99 or at position 29 also interferes with dimerization.<sup>4</sup> This strengthens the hypothesis of the above mentioned 'hot spots' as crucial to dimerization. It has also been shown that the  $\beta$ -sheet sequence is highly conserved in HIV proteases.<sup>5</sup>

Because of these reasons (high importance in the dimer's stability and high conservation) we decided to create DOS/FOS-libraries which will target this  $\beta$ -sheet in particular to inhibit the dimerization of the protease.

## Diverse library

To achieve our goal we first decided to focus on making a diverse library which is supposed to target the  $\beta$ -sheet described above. These molecules mainly have a "tong shape", which means to synthesize two short peptide chains (between three and six each) linked by a linker. This "fork" encapsulates the monomer's  $\beta$ -sheet and inhibits the dimerization process of HIV protease (see figure 4).<sup>2, 6, 7</sup>

In order to make a library with novel compounds, one could in principle change either the linker or the peptide chain. The linker needs to be of a certain size to be able to position the peptides on the  $\beta$ -sheets of the monomer. Linkers of different flexibilities have been used in research, e.g. a simple carbon chain<sup>6, 7</sup> (very flexible) or a naphthalene based<sup>2</sup> linker (rigid).

A linker (shown in figure 5) with a naphthyridine core was chosen in this project. The nitrogen atoms in the aromatic rings introduce the possibility of hydrogen bonding to the C-terminal  $\text{COO}^-$  of the protease monomer. The two arms on the core have a *trans* double bond to increase rigidity and the functional carboxylic acid groups at the end.

Previous studies of HIV protease dimerization inhibitors indicate that three amino acids residues are sufficient to obtain an inhibiting effect, although larger peptides also have been successful.<sup>6</sup> Since it is better to keep the peptide small (in order to synthesize it with a good yield), we chose to limit the peptide to three residues. Both peptide chains will have the same amino acid sequence, in order to make a C2-symmetric compound. This is not required to obtain an inhibiting function, but it contributes to simplifying the molecule. Five different amino acids were chosen to build up the diverse library, namely glycine, isoleucine, serine, asparagine and tryptophan. The basis for this choice was as follows: 1) to use amino acids which have not been tested extensively in PR dimerization inhibitors, 2) to use amino acids with similar properties as the ones present in the protease, 3) to make a library with very diverse peptides, by choosing amino acids with varied properties.

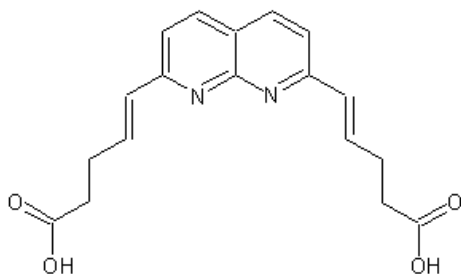


Figure 5. The molecular structure of the proposed linker.

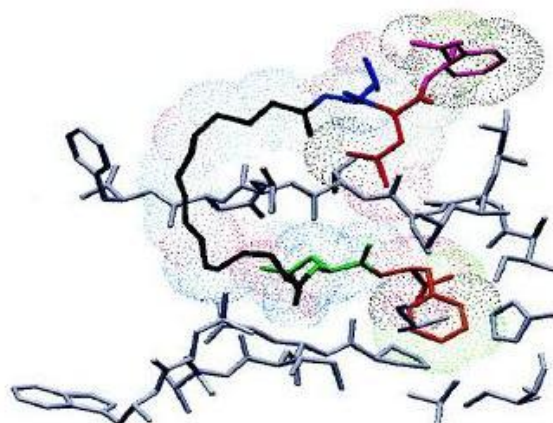
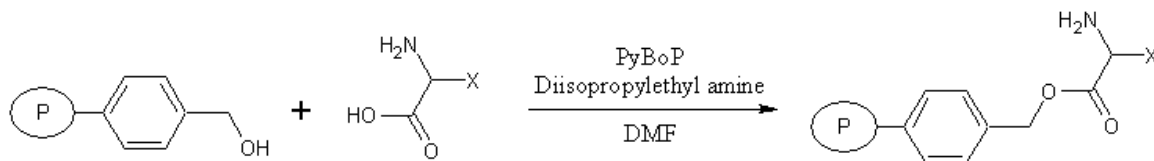


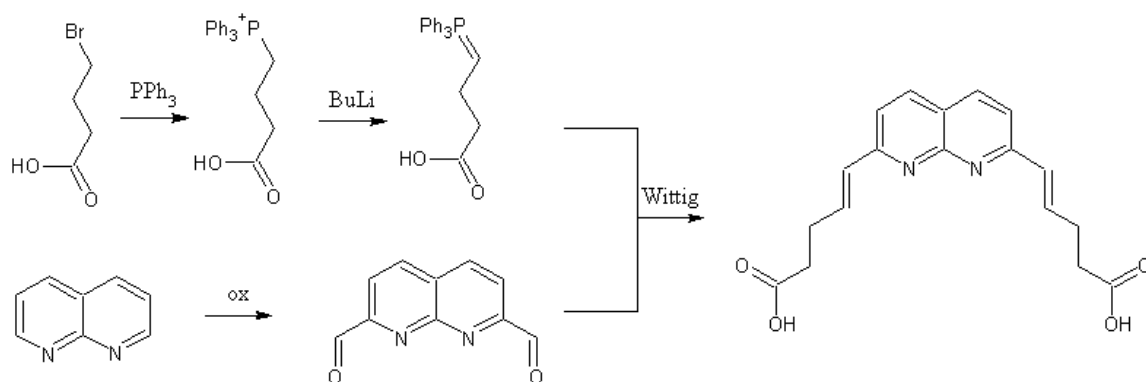
Figure 4. The shape of a molecular tong binding to the  $\beta$ -sheet.<sup>6</sup>

## Synthesis

We first decided to grow these two amino acids chains with solid state chemistry by binding it to a solid support (most often polystyrene). This is well known chemistry (see scheme 1): we need to have a hydroxyphenyl function on the polymer and perform an esterification on it with the first amino acid. Coupling is usually performed with



*Scheme 1. Coupling of the first amino acid to the solid support. P = polymer.*



*Scheme 2. Synthesis of the linker.*

PyBOP/diisopropylethyl amine or diisopropylcarbodiimide/HOBt in DMF. The amino acids need to be protected on its amine side, e.g. with Fmoc. When the reaction is done, we have to deprotect the amine function (with piperidine in DMF) and make a peptide bond with the next amino acid, and so on until we have three-amino acids chains.

The linker can be synthesized via oxidation of naphthyridine to the dialdehyde<sup>8</sup>, and subsequent coupling with the side chains (see scheme 2). A Wittig reaction is used to make the double bond linkage between the aromatic core and the “arms” of the linker. This is the most uncertain step, as there is a possibility that the nitrogens might disturb the Wittig reaction.

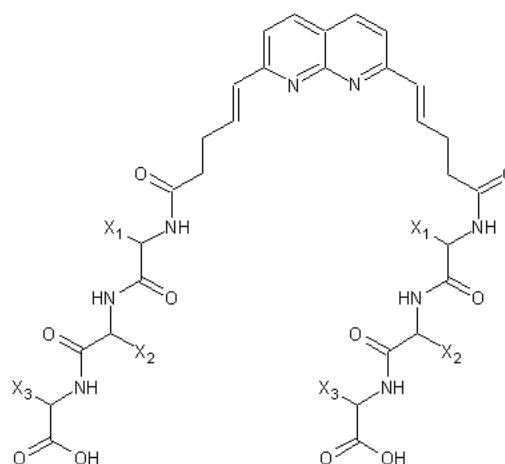
Finally we just need to cyclise two peptide chains with the linker by peptidic bonds and cleave the amino acid chain from the polymer by hydrolysis (using TFA). The resulting compound is the “tong molecule” which is a candidate for HIV’s protease inhibition. Figure 6 shows the general structure.

Since we want to create a whole library we decided to choose five different amino acids and to create all symmetric inhibitors possible in this condition (125). In order to do this we choose a multi-parallel method: automatically it is possible to do these 125 reactions simultaneously in 125 different tubes and it allows taking back pure compounds, which allows saving a lot of time and money.

Another synthesis route would be to attach the linker to the solid support and grow the peptides from the linker. This puts limitations on the linker as it needs more functionality, and it also requires protection of the nitrogens in the linker. Furthermore this route requires the peptide synthesis to be carried out “backwards”, from N-terminal to the C-terminal.

An alternative to performing multiple parallel syntheses is the “split-and-mix” method, which affords a mixture of all compounds in the end. This makes binding assays difficult, since one would either need to isolate every compound, or do deconvolution.

Yet another way to synthesize the compound is to perform the reactions in solution instead of on solid support. This is probably preferable on large scale, when an active inhibitor has actually been found. It should be possible to make this small peptide - and bind it to the linker - in solution, in larger quantities than what is possible on solid support.



*Figure 6. The molecular structure of the compounds in the diverse library.*

## Focused library

When the diverse library has been synthesized and all compounds tested for inhibiting effect, a focused library can be constructed with the most promising compound from the diverse library as a starting point. Small changes can be made to this compound to obtain the different molecules in the focused library. This library should naturally be much less diverse than the first.

Since it was not possible to actually perform the synthesis and binding assay of the diverse library in this project, we more or less arbitrarily choose one of the compounds in that library on which to base a focused library. This starting point was the compound with the peptide chain S, I, W (counted from the linker). The choice of this particular sequence was loosely based on the amino acids present in the monomer, which the inhibitor will have to interact with.

To make a focused library with this as a starting point, a few parts of the molecule needs to be changed. We decided to keep the general structure with the linker and identical peptide chains, and to make single changes at the amino acid positions. These changes are shown in table 1.

<i>Position</i>	1	2	3
<i>Amino acid in starting compound</i>	Serine (S)	Isoleucine (I)	Tryptophan (W)
<i>Characteristics</i>	Small, OH-group	Small, hydrophobic	Large, aromatic
<i>Amino acid in focused library</i>	Threonine (T)	Leucine (L)	N-methylated tryptophan (N-Me W)
	Cysteine (C)	Valine (V)	
	β-Serine (β-S)		

*Table 1. Changes to amino acid sequence in the focused library.*

The variations are small and include mostly natural amino acids. The exceptions are  $\beta$ -serine, i.e. serine with an extra carbon in the peptide backbone, and N-methylated tryptophan, i.e. tryptophan with a methyl group on the nitrogen in the peptide backbone. Making these changes, one at a time, will produce a library that consists of 24 compounds (four choices in position 1, three in position 2 and two in position 3).

All the amino acids in table 1 are commercially available. The synthesis and assay of these compounds is analogous to the diverse library.

To further develop the compound which shows the highest inhibition of the protease, one could decrease the peptidic character of the molecule. Many drugs used as dimerization inhibitors today are peptidomimetics. Changing the amide bonds into some other link can increase the stability of the molecule *in vivo*, since there are many proteases in the body that cleaves amide bonds.

## Binding tests

When the library is synthesized we need to know the efficiency of each compound, which can be expressed in binding affinity between the target and the protease and/or in inhibition. We propose three methods to test binding.

### Fluorimetric or Colorimetric

We need to use a protease's substrate which releases a chromophore or fluorescent group in solution. Then we compare how fast the solution's color properties change between solutions with and without inhibitor. The better the drug, the lesser the solution becomes fluorescent. We can even work with different concentrations and determine the dissociation constant between the drug and the enzyme.

### NMR Spectroscopy

If the molecule we synthesized binds to the monomer, as expected, then its surrounding magnetic field will change due to this enzyme presence. Performing NMR spectrometry can allow us to determine these changes and the new surrounding and calculate how strong the binding is (if there is any).

### Surface Plasmon Resonance

This is a method developed by the Swedish company Biacore and based on Surface Plasmon Resonance. On one side of a very thin sheet of gold the enzyme's monomer is fixed and there is a constant flow of drug solution at its contact. On the other side the reflection angle of a light ray is registered. It appears that this angle changes depending on the weight attached to the golden surface, which means that we can determine to what extent the drug is binding to the target and even study kinetic constants with this method.

## Bibliography

[1] = E. Clrecq, *Nat. Rev. Drug Disc.*, **2007**, *6*, 1001-1018

[2] = L. Bannwarth, A. Kessler, S. Pèthe, B. Collinet, N. Merabet, N. Boggetto, S. Sicsic, M. Reboud-Ravaux, S. Ongeri, *J. Med. Chem.*, **2006**, *49*, 4657-4664

[3] = M. J. Todd, N. Semo, E. Freire, *J. Mol. Biol.*, **1998**, *283*, 475-488

[4] = Y. Koh, S. Matsumi, D. Das, M. Amano, D. A. Davis, J. Li, S. Leschenko, A. Baldrige, T. Shioda, R. Yarchoan, A. K. Ghosh, H. Mitsuya, *J. Biol. Chem.*, **2007**, *282*, 28709-28720

[5] = M.-J. Camarasa S. Velazquez, A. San-Felix, M.-J. Perez-Perez, F. Gago, *Antiv. Res.*, **2006**, *71*, 260-267

[6] = M. Shultz, Y. Ham, S. Lee, D. Davis, C. Brown, J. Chmielewski, *J. Am. Chem. Soc.*, **2004**, *126*, 9886-9887

[7] = Y. Hwang, J. Chmielewski, *J. Med. Chem.*, **2005**, *48*, 2239-2242

[8] = C. Chandler, L. Deady, J. Reiss, V. Tzimos, *J. Heterocycl. Chem.*, **1982**, *19*, 1017-1019